Title

Crystal structure of a member of a novel family of dioxygenases (PF10014) reveals a conserved cupin fold and active site

Journal

Proteins Structure Function and Bioinformatics, 82(1)

ISSN

0887-3585

Authors

Xu, Qingping
Grant, Joanna
Chiu, Hsiu-Ju
et al.

Publication Date

2014

DOI

10.1002/prot.24362

Peer reviewed

Published in final edited form as:
Proteins. 2014 January ; 82(1): 164–170. doi:10.1002/prot.24362.

Crystal structure of a member of a novel family of dioxygenases (PF10014) reveals a conserved cupin fold and active site

Qingping Xu[1,2], **Joanna Grant**[1,3], **Hsiu-Ju Chiu**[1,2], **Carol L. Farr**[1,4], **Lukasz Jaroszewski**[1,5,6], **Mark W. Knuth**[1,3], **Mitchell D. Miller**[1,2], **Scott A. Lesley**[1,3,4], **Adam Godzik**[1,5,6], **Marc-André Elsliger**[1,4], **Ashley M. Deacon**[1,2], and **Ian A. Wilson**[1,4,*]

[1]Joint Center for Structural Genomics (http://www.jcsg.org)

[2]Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, California 92045 USA

[3]Protein Sciences Department, Genomics Institute of the Novartis Research Foundation, San Diego, California 92121 USA

[4]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, California 92037 USA

[5]Center for Research in Biological Systems, University of California, San Diego, La Jolla, California 92093-0446 USA

[6]Program on Bioinformatics and Systems Biology, Sanford-Burnham Medical Research Institute, La Jolla, California 92037 USA

Abstract

PF10014 is a novel family of 2-oxyglutarate-Fe^{2+}-dependent dioxygenases that are involved in biosynthesis of antibiotics and regulation of biofilm formation, likely by catalyzing hydroxylation of free amino acids or other related ligands. The crystal structure of a PF10014 member from *Methylibium petroleiphilum* at 1.9 Å resolution shows strong structural similarity to cupin dioxygenases in overall fold and active site, despite very remote homology. However, one of the β-strands of the cupin catalytic core is replaced by a loop that displays conformational isomerism that likely regulates the active site.

Keywords

PF10014/BsmA; cupin dioxygenase; free amino acids; 2-oxyglutarate; ferrous iron

Introduction

The cupin fold consists of a double-stranded β-helix barrel. Cupin-like proteins form one of the largest and most functionally diverse protein superfamilies, consisting of at least 50 PFAM families [1]. The largest subset, 2-oxyglutarate-Fe^{2+}-dependent dioxygenases, catalyzes diverse reactions such as hydroxylation, desaturation, and ring expansion/closure,

*Correspondence to: Ian A. Wilson, Department of Integrative Structural and Computational Biology, BCC206, 10550 N. Torrey Pines Rd., La Jolla, California 92037, USA. Tel.: 1- 858-784-9706; Fax: 1-858-784-2980; wilson@scripps.edu.

Accession numbers

The structure factors and atomic coordinates of MpDO are deposited in the RCSB Protein Data Bank (http://www.rcsb.org/) with PDB code 3pl0. The plasmid for producing recombinant MpDO was deposited in PSI:Biology-Materials Repository with clone ID MpCD00325753 (http://dnasu.asu.edu).

involving a wide-range of substrates in many important biological processes, such as biosynthesis of antibiotics and plant products, post-translational modification of protein side chains, damage repair of alkylated DNA/RNA, lipid metabolism, and biodegradation of numerous compounds. These enzymes bind cofactor Fe^{2+}, through coordination by an Hx_1E/Dx_nH motif, and cosubstrate, 2-oxyglutarate (2OG), which is required for oxidation [2–4].

DUFs are protein families representing Domains of Unknown Function [5]. The PF10014 (or BsmA) family, previously known as DUF2257, consists of at least 320 unique proteins. PF10014 members are predominantly found in bacteria, but also in eukaryotes. The first characterized member, BsmA from *Serratia liquefaciens*, is involved in quorum sensing-controlled biofilm development [6]. Other known members are found in gene clusters responsible for biosynthesis of polyketides or other antibiotics, such as Azinomycin B in *Streptomyces sahachiroi* (Azi27) [7], polyoxin H in *Streptomyces cacaoi* (PolL) [8], hormaomycin in *Streptomyces griseoflavus* (HrmJ) [9], antifugal glidobactins (GlbB) [10], and tryptoquialanine in fungus (TqaL) [11]. A PF10014 member isolated from *Bacillus thuringiensis* was recently shown to convert L-isoleucine into (2S,3R,4S)-4-hydroxyisoleucine [12]. The enzyme was characterized as a 2OG-Fe^{2+}-dependent L-Ile dioxygenase (IDO), since its function requires 2OG, Fe^{2+}, and ascorbic acid [12]. Other PF10014 family members catalyze the hydroxylation of free amino acids [13] and, thus, likely have a conserved enzymatic function. These enzymes are potentially valuable biotechnological tools for the production of modified amino acids and other derivative compounds.

An effective approach to characterizing DUFs is through high-throughput structural biology, as implemented in structural genomics efforts [5,14]. At the time of targeting members of the PF10014 family for structure determination, very little was known as to the potential biochemical function(s) of this protein family or relationships to other known structures. Here, we report the first crystal structure of a member of PF10014, a *Methylibium petroleiphilum* PM1 dioxygenase (Mpe_A2762, referred to as MpDO hereafter). *M. petroleiphilum* is the only known bacterium capable of utilizing methyl tert-butyl ether (MTBE), a fuel additive that poses significant environmental risk, as its sole carbon source [15]. The MpDO structure reveals that PF10014 represents a novel family of 2OG-Fe^{2+} dioxygenases within the cupin superfamily, consistent with independent biochemical studies of other family members [12,13,16,17]. Flexibility in a conserved loop at the active site suggests an induced fit catalytic mechanism.

Material and Methods

Cloning, protein production, and crystallization

Protocols and reagents for cloning, purification and crystallization of MpDO (GenBank: YP_001021951, UniProt: A2SJH7) are as previously reported [18]. Primers (forward primer, 5′-ctgtacttccagggcATGCACGTCGATATCGAACTGCCCCTCG-3′ and reverse primer, 5′-aattaagtcgcgttaGGCCGGCGCCTGGAAGCCGCCGC-3′, target sequence in upper case) were used to amplify the gene encoding MpDO from genomic DNA. Purified protein was concentrated to 14.8 mg/ml for crystallization trials. Analytical size exclusion analyses were performed as reported previously [18]. Sitting drops composed of 100 nl protein solution mixed with 100 nl crystallization solution were equilibrated against a 50 μl mother liquor reservoir at 277 K for 54 days prior to harvest. The crystallization reagent consisted of 0.9 M LiCl, 6% PEG 6000, and 0.1M HEPES pH 6.5 with glycerol [10% (v/v)] added as cryoprotectant.

Data collection, structure solution, and refinement

Multiple-wavelength anomalous diffraction (MAD) data were collected at wavelengths corresponding to the inflection, high energy remote, and peak of a selenium MAD experiment at 100 K using an ADSC Quantum CCD Q315 detector at ALS beamline 8.2.2. Data processing and structure determination were carried out automatically using our distributed structure solution pipeline [19]. Model completion and refinement were performed manually with COOT [20] and REFMAC [21].

Results and Discussion

Structure determination and model quality

The cytoplasmic protein MpDO contains seven methionines within its 253 amino acids. The MpDO crystal structure was determined using the semi-automated, high-throughput JCSG pipeline [22]. The selenomethionine derivative of full-length MpDO was expressed in *Escherichia coli* with an N-terminal, TEV-cleavable, His-tag and purified by metal affinity chromatography. The purification tag was removed prior to crystallization. 179 crystals were harvested from several conditions and screened for diffraction. Although most crystals diffracted poorly (<3.0 Å), one condition yielded crystals suitable for structure determination. The crystal structure was determined in space group $P2_1$ using the MAD method and refined to 1.9 Å resolution with R_{cryst} of 16.9% and R_{free} of 21.1% with good geometry [23]. The asymmetric unit (asu) contains two monomers (A and B), consisting of residues 0 to 253 (Gly0 remains after cleavage of the N-terminal purification tag). Two disordered residues (229–230) in monomer B were not included in the model. One ethylene glycol, one chloride, and 413 water molecules were also modeled. Four surface side chains (A104, A112, A132 and B59) were only partially modeled due to lack of electron density. As no ligand or cofactor was found in the active site, the MpDO structure represents the enzyme apo-form. Data and refinement statistics are summarized in Table I.

Overall structure

The MpDO monomer (Fig. 1A) consists of 16 β-strands (β1–β16), five α-helices (α1– α5), and two short 3_{10} helices (η1– η2). The majority of the β-strands form two anti-parallel β-sheets: a large seven-stranded β-sheet (Sheet1; β1, β3, β5, β8, β9, β13, and β16) and smaller three-stranded β-sheet (Sheet2; β11, β12, and β14). β4, β10, and β15 extend strands β3, β11 and β14, respectively. These two β-sheets pack against each other, forming a cup-shaped β-sandwich with a topology characteristic of the double-stranded β-helix fold (Fig. 1B). The helices are all on one side of Sheet1 and contribute to its stabilization. The two monomers in the asu are very similar to each other, especially when the three loops around the active site entrance are excluded from the calculation (Fig. 1 C), which then results in an RMSD of 0.5 Å (221 equivalent C^{α} atoms). The flexible loops (RMSD 3.5 Å for 31 C^{α} atoms) are located between β2 and β3 (L1), β6 and β7 (L2), and β8 and β9 (L3) (Fig. 1B).

Size-exclusion chromatography coupled with static light scattering indicates that MpDO is a dimer in solution (data not shown). The two monomers are related by a non-crystallographic two-fold axis (Fig. S1), presumably corresponding to the dimer observed in solution. However, each monomer has a buried interface of only 875 $Å^2$ (~7% of the overall solvent accessible surface), which is considered too small for a stable dimer by PISA [24]. The monomers associate through the exposed side of the Sheet2 surface, but the interface residues are not well conserved except for those in the L3 loop. Thus, additional evidence is needed to confirm the physiological relevance of the dimer observed in the crystal lattice.

Structural and sequence comparisons

Structural similarity searches [25] revealed that MpDO is similar to other cupin superfamily proteins. The top hits include a putative dioxygenase SO_2589 from *Shewanella oneidensis* (PDB code 3on7; PF14226; JCSG, unpublished), hypoxia inducible factor (HIF) prolyl 4-hydroxylase PHD2 (PDB code 3ouj) [26], and human AlkB-like DNA/RNA demethylase FTO (PDB code 3lfm) [27]. Most secondary structural elements are highly conserved (Fig. 2A), despite very low sequence identities (~10%, Fig. 2B). Both PHD2 and FTO belong to the cupin superfamily of 2OG-Fe^{2+}-dependent dioxygenases with well-conserved active site features. Cupin superfamily members generally share a double-stranded β-helix fold containing two 4-stranded β-sheets, which correspond to β8, β9, β13, and β16 (Sheet1), and β11, β12, and β14 (Sheet2) in MpDO. Thus, the MpDO catalytic core resides on the C-terminal half (residues 141–247), while the N-terminal half stabilizes the cupin core and helps shape the active site pocket (near L1 and L2). Interestingly, Sheet2 of MpDO contains only three β-strands, since the first β-strand of the canonical cupin fold is absent. The equivalent region of MpDO (L3 of Fig. 1B) contains a ^{159}PxPE^{162} motif, which forms a bulge that adopts different conformations in the two monomers (Fig. 1C).

Residues involved in binding Fe^{2+} and 2OG are strictly conserved between MpDO and its closest structural homologs (Fig. 2C), suggesting His165, Asp167 and His218 of MpDO are involved in coordinating Fe^{2+} and Arg237 in binding 2OG. In the structure of apo-MpDO, a water molecule occupies the equivalent position as Fe^{2+} with similar coordination interactions. The RMSD between these four conserved residues of MpDO and FTO is only 1.2 Å (39 atoms). The His218 imidazole in MpDO is flipped by ~180° and stabilized by a hydrogen bond between His218Nε and Asp167Oδ1, compared to the FTO Fe^{2+} and N-oxalylglycine (OGA) bound structure. The conservation of overall structure and active site clearly supports assignment of MpDO as a cupin superfamily dioxygenase.

To assess the relationship between PF10014 and other PFAM families, we carried out a profile-profile based similarity search using HHpred [28]. All top matches belong to the Cupin CLAN (CL0029), including the Tet_JBP family (PF12851, E-value=0.0028 and Probability=96.7%), 2OG-FeII_Oxy_3 family (PF13640, E=0.0043 and Probability=96.3%), and DUF2086 family (PF09859, E=1.2 and Probability=85%). The most significant sequence similarity mapped to the C-terminal region of MpDO (residues 165–253), encompassing the entire conserved catalytic core (seq id ~20%). The Tet_JBP family contains proteins involved in modification of nucleic acids, while the 2OG-FeII_Oxy_3 family includes Egl-9, an enzyme that regulates HIF by prolyl hydroxylation. These results are in good agreement with the structural results, suggesting PF10014 is phylogenetically most closely related to the cupin dioxygenase families that hydroxylate proteins or nucleic acids.

Active site and reaction mechanism

The MpDO active site is formed mainly by Sheet1 and the L1–L3 loops (Figs. 1B and S2). Residues involved in binding Fe^{2+} and 2OG are strictly conserved in most PF10014 members (Fig. S3), suggesting they likely share the same enzymatic function. Notably, Arg237 involved in binding 2OG is not present in BsmA. While the same role could be fulfilled by another functionally equivalent residue, homology modeling indicates no plausible candidates. Thus, BsmA may lack the capacity to bind 2OG but can still chelate a metal, which likely indicates a different enzymatic function.

The plasticity of the active site is believed to be important for catalysis in cupin dioxygenases[2]. Flexible conformations of the L1–L3 loops result in differences in size and accessibility of the active site (Fig. 3A–B). The L3 loop is well defined in both monomers

and not involved in crystal contacts other than at the putative dimer interface. Thus, the L3 conformations observed in the crystal likely represent two stable conformations in solution. In monomer A, the conformation is "open" with the entire active site fully accessible (Fig. 3A). In contrast, in monomer B, the conserved ^{159}PxPE^{162} region moves into the active site pocket, covering the cosubstrate 2OG binding site, even though the cavity for binding 2OG and Fe^{2+} remains intact (Fig. 3B). Thus, monomer B may mimic an intermediate "closed" conformation, presumably similar to the state when Fe^{2+} and 2OG are bound. The MpDO substrate binding site is expected to be at the same location as in other cupin dioxygenases. Structural comparisons and modeling studies suggest the substrate occupies a location near Pro161 of monomer B.

The physiological function and substrate specificity of MpDO is currently unknown. Recent biochemical studies suggest that PF10014 members can hydroxylate free L-amino acids and can utilize methionine as a substrate to produce L-methionine sulfoxide [13]. Due to the highly conserved active site, MpDO serves as a good framework for understanding substrate specificity within the family. To probe the potential roles of conserved residues in recognizing free amino acids, we modeled 2OG and Fe^{2+} into the conserved location based on the FTO structure (Fig. 3) and then manually methionine docked into the substrate-binding site (Fig. 3C). A cluster of three conserved arginines (Arg65, Arg67 and Arg104) may be important for substrate specificity by recognizing its carboxyl group. The methionine amine is likely to be stabilized by the conserved Gln91 or Glu162. These interactions help position the site of hydroxylation close to the catalytic center.

Based on the features observed in the crystal structure, and earlier studies of other dioxygenases, we expect that MpDO will exhibit a conserved mechanism for substrate binding and catalysis. Loops at the gate of the active site sample different conformations such that Fe^{2+} and 2OG can bind into the active site during an "open" state (monomer A). Binding of cofactor Fe^{2+} and cosubstrate 2OG induces a conformational change in L3 that protects 2OG from solvent (similar to the "monomer B" structure). The reorganization of the ^{159}PxPE^{162} region may enable the active site to become competent for binding the primary substrate and oxygen. The hydroxylation reaction, which results in the transfer of one oxygen atom to the succinate byproduct and the other oxygen to the primary substrate, likely proceeds through a radical mechanism involving an iron-oxo intermediate [2–4].

The JCSG and other structural genomics efforts have significantly contributed to our understanding of DUFs through structure determination [5,14]. Since structural space is significantly smaller than sequence space, it is expected that many of these novel DUF sequences will likely exhibit structural similarity to known folds, which should give some clues as to function. Here, we determined the crystal structure of a member of PF10014, which allowed classification of this family to the cupin superfamily of 2OG-Fe^{2+}-dependent dioxygenases. In parallel, the function of PF10014 was also illustrated through independent biochemical studies [12,13,16,17]. Both types of studies are consistent and complementary to each other, highlighting the advantages of studying these DUFs using multiple approaches.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the members of the JCSG high-throughput structural biology pipeline for their contribution to this work. Portions of this research were carried out at the Stanford Synchrotron Radiation Lightsource (SSRL) and the Advanced Light Source (ALS). The SSRL is a Directorate of SLAC National Accelerator Laboratory and an Office of Science User Facility operated for the U.S. Department of Energy Office of Science by Stanford University. The

ALS is supported by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences Division of the US Department of Energy under Contract No. DE-AC02-05CH11231 at Lawrence Berkeley National Laboratory. The Berkeley Center for Structural Biology is supported in part by NIH, NIGMS and the Howard Hughes Medical Institute. Genomic DNA from *Methylibium petroleiphilum* PM1 was a gift from Dr. Staci Kane, Lawrence Livermore National Laboratory, and Drs. Krassimira Hristova and Kate Scow, UC-Davis, USA. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIGMS or NIH.

Grant sponsor: NIH, National Institute of General Medical Sciences (NIGMS), Protein Structure Initiative, Grant Number U54 GM094586; The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the NIH, National Institute of General Medical Sciences (including P41GM103393) and the National Center for Research Resources (P41RR001209).

References

1. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD. The Pfam protein families database. Nucleic Acids Res. 2012; 40:D290–301. [PubMed: 22127870]
2. Hewitson KS, Granatino N, Welford RW, McDonough MA, Schofield CJ. Oxidation by 2-oxoglutarate oxygenases: non-haem iron systems in catalysis and signalling. Philos Transact A Math Phys Eng Sci. 2005; 363:807–828. discussion 1035–1040.
3. Purpero V, Moran GR. The diverse and pervasive chemistries of the alpha-keto acid dependent enzymes. J Biol Inorg Chem. 2007; 12:587–601. [PubMed: 17431691]
4. Hausinger RP. FeII/alpha-ketoglutarate-dependent hydroxylases and related enzymes. Crit Rev Biochem Mol Biol. 2004; 39:21–68. [PubMed: 15121720]
5. Bateman A, Coggill P, Finn RD. DUFs: families in search of function. Acta Crystallogr Sect F. 2010; 66:1148–1152.
6. Labbate M, Queck SY, Koh KS, Rice SA, Givskov M, Kjelleberg S. Quorum sensing-controlled biofilm development in *Serratia liquefaciens* MG1. J Bacteriol. 2004; 186:692–698. [PubMed: 14729694]
7. Zhao Q, He Q, Ding W, Tang M, Kang Q, Yu Y, Deng W, Zhang Q, Fang J, Tang G, Liu W. Characterization of the azinomycin B biosynthetic gene cluster revealing a different iterative type I polyketide synthase for naphthoate biosynthesis. Chem Biol. 2008; 15:693–705. [PubMed: 18635006]
8. Chen W, Huang T, He X, Meng Q, You D, Bai L, Li J, Wu M, Li R, Xie Z, Zhou H, Zhou X, Tan H, Deng Z. Characterization of the polyoxin biosynthetic gene cluster from *Streptomyces cacaoi* and engineered production of polyoxin H. J Biol Chem. 2009; 284:10627–10638. [PubMed: 19233844]
9. Hofer I, Crusemann M, Radzom M, Geers B, Flachshaar D, Cai X, Zeeck A, Piel J. Insights into the biosynthesis of hormaomycin, an exceptionally complex bacterial signaling metabolite. Chem Biol. 2011; 18:381–391. [PubMed: 21439483]
10. Schellenberg B, Bigler L, Dudler R. Identification of genes involved in the biosynthesis of the cytotoxic compound glidobactin from a soil bacterium. Environ Microbiol. 2007; 9:1640–1650. [PubMed: 17564599]
11. Gao X, Chooi YH, Ames BD, Wang P, Walsh CT, Tang Y. Fungal indole alkaloid biosynthesis: genetic and biochemical investigation of the tryptoquialanine pathway in *Penicillium aethiopicum*. J Am Chem Soc. 2011; 133:2729–2741. [PubMed: 21299212]
12. Kodera T, Smirnov SV, Samsonova NN, Kozlov YI, Koyama R, Hibi M, Ogawa J, Yokozeki K, Shimizu S. A novel l-isoleucine hydroxylating enzyme, l-isoleucine dioxygenase from *Bacillus thuringiensis*, produces (2S,3R,4S)-4-hydroxyisoleucine. Biochem Biophys Res Commun. 2009; 390:506–510. [PubMed: 19850012]
13. Smirnov SV, Sokolov PM, Kodera T, Sugiyama M, Hibi M, Shimizu S, Yokozeki K, Ogawa J. A novel family of bacterial dioxygenases that catalyse the hydroxylation of free L-amino acids. FEMS Microbiol Lett. 2012; 331:97–104. [PubMed: 22448874]

14. Jaroszewski L, Li Z, Krishna SS, Bakolitsa C, Wooley J, Deacon AM, Wilson IA, Godzik A. Exploration of uncharted regions of the protein universe. PLoS Biol. 2009; 7:e1000205. [PubMed: 19787035]
15. Bruns MA, Hanson JR, Mefford J, Scow KM. Isolate PM1 populations are dominant and novel methyl tert-butyl ether-degrading bacterial in compost biofilter enrichments. Environ Microbiol. 2001; 3:220–225. [PubMed: 11321538]
16. Hibi M, Kawashima T, Kodera T, Smirnov SV, Sokolov PM, Sugiyama M, Shimizu S, Yokozeki K, Ogawa J. Characterization of *Bacillus thuringiensis* L-isoleucine dioxygenase for production of useful amino acids. Appl Environ Microbiol. 2011; 77:6926–6930. [PubMed: 21821743]
17. Smirnov SV, Kodera T, Samsonova NN, Kotlyarova VA, Rushkevich NY, Kivero AD, Sokolov PM, Hibi M, Ogawa J, Shimizu S. Metabolic engineering of *Escherichia coli* to produce (2S, 3R, 4S)-4-hydroxyisoleucine. Appl Microbiol Biotechnol. 2010; 88:719–726. [PubMed: 20665018]
18. Xu Q, van Wezel GP, Chiu HJ, Jaroszewski L, Klock HE, Knuth MW, Miller MD, Lesley SA, Godzik A, Elsliger MA, Deacon AM, Wilson IA. Structure of an MmyB-like regulator from *C.aurantiacus*, member of a new transcription factor family linked to antibiotic metabolism in actinomycetes. PLoS One. 2012; 7:e41359. [PubMed: 22844465]
19. van den Bedem H, Wolf G, Xu Q, Deacon AM. Distributed structure determination at the JCSG. Acta Crystallogr Sect D. 2011; 67:368–375. [PubMed: 21460455]
20. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr Sect D. 2004; 60:2126–2132. [PubMed: 15572765]
21. Winn MD, Murshudov GN, Papiz MZ. Macromolecular TLS refinement in REFMAC at moderate resolutions. Methods Enzymol. 2003; 374:300–321. [PubMed: 14696379]
22. Elsliger MA, Deacon AM, Godzik A, Lesley SA, Wooley J, Wuthrich K, Wilson IA. The JCSG high-throughput structural biology pipeline. Acta Crystallogr Sect F. 2010; 66:1137–1142.
23. Davis IW, Murray LW, Richardson JS, Richardson DC. MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. Nucleic Acids Res. 2004; 32:W615–619. [PubMed: 15215462]
24. Krissinel E. Crystal contacts as nature's docking solutions. J Comput Chem. 2010; 31:133–143. [PubMed: 19421996]
25. Holm L, Rosenstrom P. Dali server: conservation mapping in 3D. Nucleic Acids Res. 2010; 38:W545–549. [PubMed: 20457744]
26. Rosen MD, Venkatesan H, Peltier HM, Bembenek SD, Kanelakis KC, Zhao LX, Leonard BE, Hocutt FM, Wu X, Palomino HL, Brondstetter TI, Haugh PV, Cagnon L, Yan W, Liotta LA, Young A, Mirzadegan T, Shankley NP, Barrett TD, Rabinowitz MH. Benzimidazole-2-pyrazole HIF prolyl 4-hydroxylase inhibitors as oral erythropoietin secretagogues. ACS Med Chem Lett. 2010; 1:4. [PubMed: 20730035]
27. Han Z, Niu T, Chang J, Lei X, Zhao M, Wang Q, Cheng W, Wang J, Feng Y, Chai J. Crystal structure of the FTO protein reveals basis for its substrate specificity. Nature. 2010; 464:1205–1209. [PubMed: 20376003]
28. Soding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. 2005; 33:W244–248. [PubMed: 15980461]

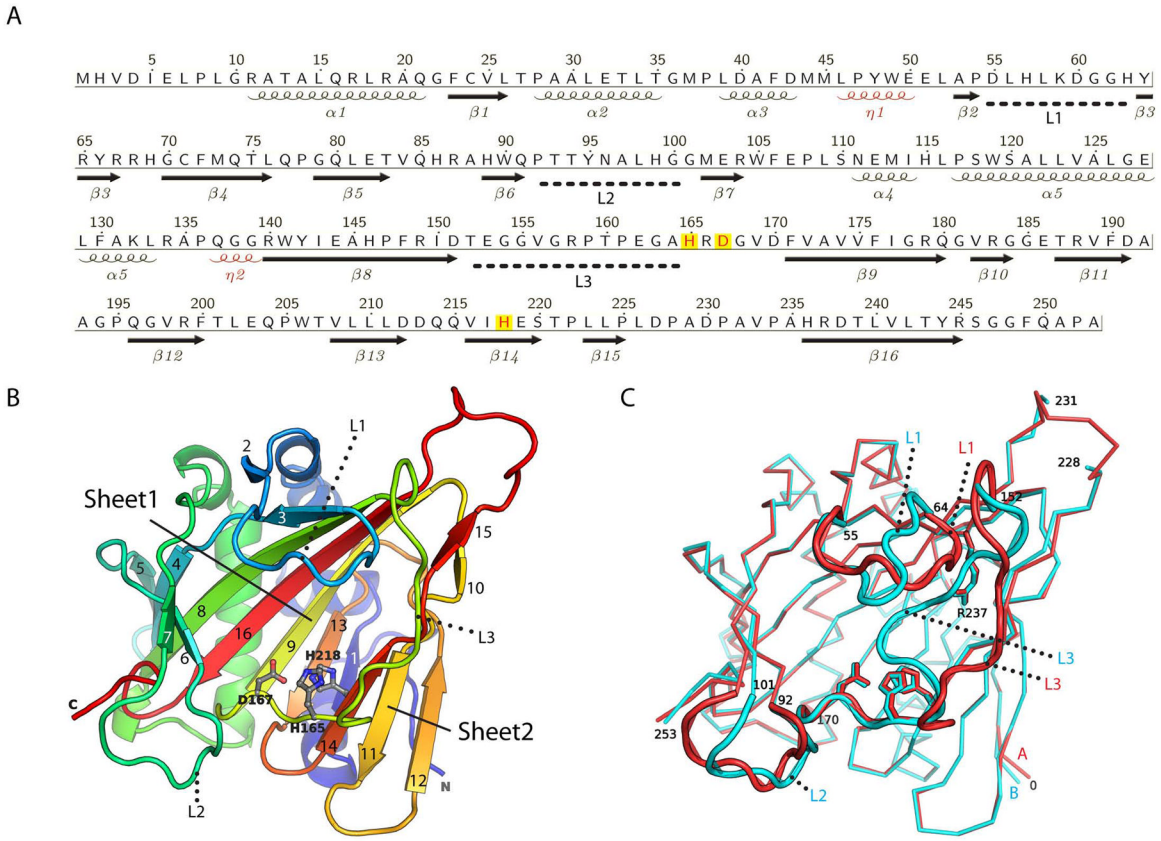

Fig. 1.

Crystal structure of MpDO. (A) The sequence of MpDO annotated with secondary structure elements. β-strands are labeled β1 to β16, α-helices α1 to α5, and 3_{10} helices η1 to η2. The three conserved residues most likely involved in binding Fe^{2+} are highlighted in yellow boxes. (B) Ribbon representation of MpDO color-coded from blue (N-terminus) to red (C-terminus). The β-strands as defined in (A) are labeled as 1 through 16, and the three loops around the active site as L1 to L3. (C) Structural superposition of the two monomers in the asu (A: red; B: cyan). The regions with large structural differences are highlighted as thicker tubes.

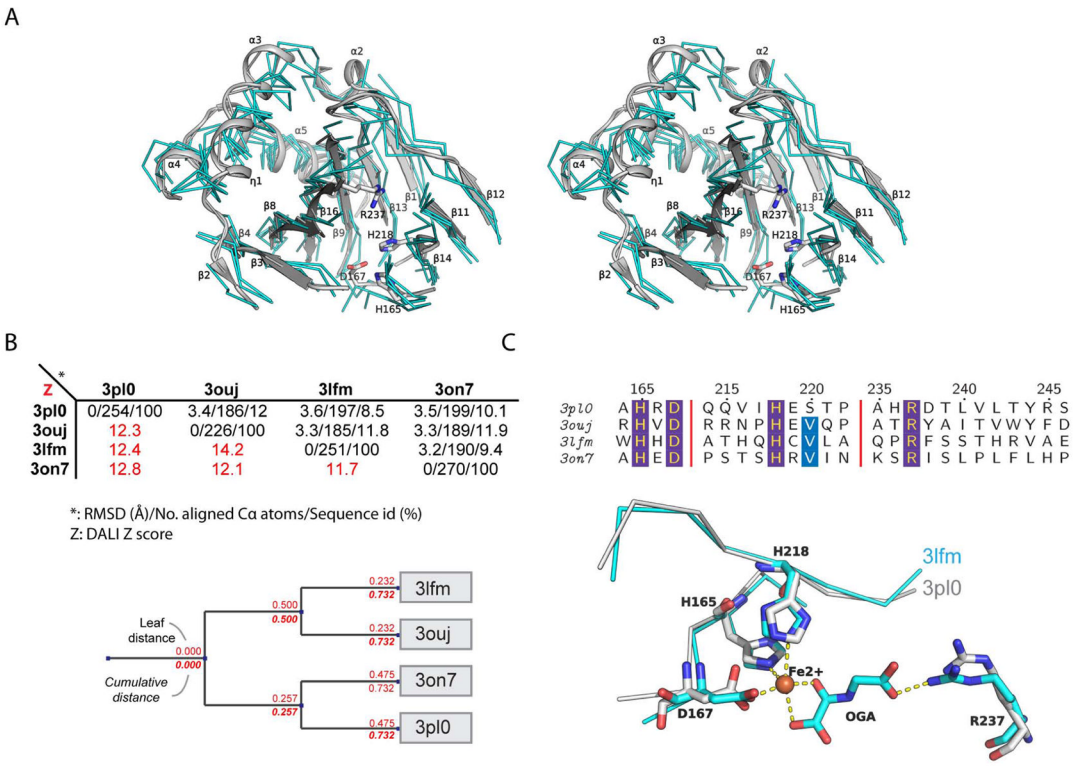

Fig. 2.

Structural comparisons between MpDO (gray) and the top structural homologs (cyan). (A) Common secondary structural elements within MpDO, FTO (PDB code 3lfm), SO_2589 (PDB code 3on7) and PHD2 (PDB code 3ouj). (B) Pair-wise structural comparisons between the four structures (top) and a tree representation of their relationship based on Dali Z-score (bottom). (C) Conserved residues involved in binding Fe^{2+} and 2OG in both sequence (top) and structure (bottom).

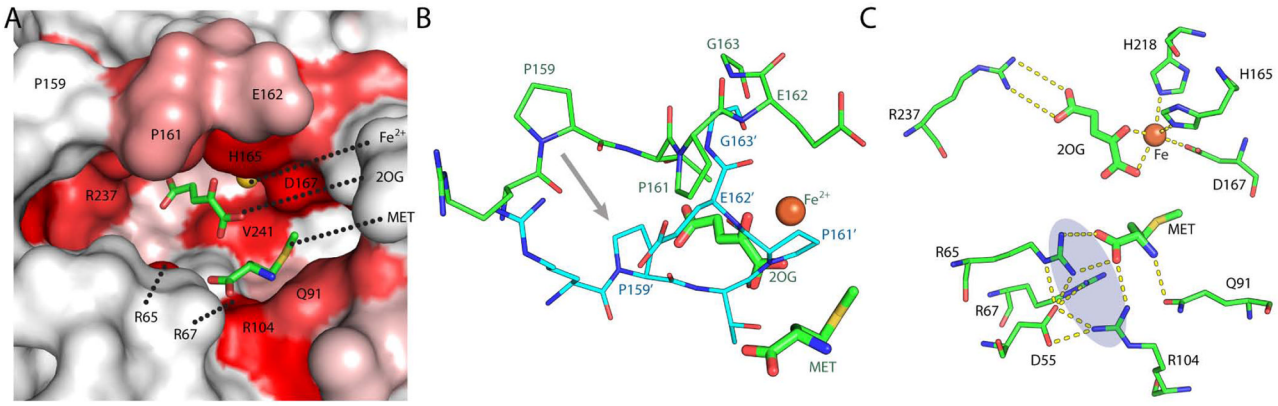

Fig. 3.

The active site of MpDO. (A) Surface representation of the active site MpDO of monomer A, colored by a scale of sequence conservation ranging from not conserved (white) to highly conserved (red). Modeled 2OG, Fe^{2+}, and substrate (methionine) are shown as sticks/sphere. (B) The conformational change of the L3 flexible region (monomer A, green; monomer B, cyan) alters accessibility of the active site. (C) Predicted interactions between protein, substrate, 2OG and Fe^{2+} (monomer A). Potential hydrogen bonds are shown as dashed lines.

Table I

Data collection and refinement statistics

Data collection			
Space group	$P2_1$		
Unit cell	*a*=43.5, *b*=63.2, *c*=104.5 Å, β=97.8°		
Data	λ_1 MADSe-infl	λ_2 MADSe-remo	λ_3 MADSe-peak
Wavelength (Å)	0.9796	0.9537	0.9793
Resolution range (Å)	28.9–1.91	28.9–1.91	28.9–2.03
Highest resolution shell	2.01–1.91	2.01–1.91	2.14–2.03
No. observations	132,709	132,999	131,366
No. unique reflections	43,626	43,643	35,835
Completeness (%)[a]	99.1 (98.4)	99.2 (98.6)	98.9 (97.9)
Mean I/σ (I)[a]	14.5 (3.0)	13.9 (2.8)	14.4 (2.8)
R_{merge} on I[a]	0.049 (0.455)	0.053 (0.488)	0.067 (0.599)
R_{meas} on I[a]	0.059 (0.549)	0.064 (0.588)	0.079 (0.701)
R_{pim} on I[a]	0.033 (0.304)	0.036 (0.326)	0.041 (0.361)
Model and refinement statistics			
Data used in refinement		λ_2 MADSe	
No. reflections (total)		43,627	
No. reflections (test)		2199	
Cutoff criteria		|F|>0	
R_{cryst} (%)		16.9	
R_{free} (%)		21.1	
Stereochemical parameters			
Restraints (RMSD observed)			
Bond lengths (Å)		0.016	
Bond angles (°)		1.51	
MolProbity score			
All atom clash score		5.6	
Ramachandran plot (%)[b]		99.0 (0)	
Rotamer outliers (%)		1.0	
Average isotropic B-value (Å^2)[c]		35.8 (32.4)	
ESU based on R_{free} (Å)		0.14	
No. protein residues / chains		492 / 2	
Non-protein entities		413 H_2O, 1 GOL and 1 CL	

[a] Highest resolution shell in parentheses.

[b] Percentage of residues in favored regions of Ramachandran plot (outliers in parenthesis).

[c] This value represents the total B that includes TLS and residual B components (Wilson B-value in parenthesis).

ESU = Estimated Standard Uncertainty in coordinates.

$R_{merge}=\Sigma_{hkl}\Sigma_i|I_i(hkl)-<I(hkl)>|/\Sigma_{hkl}\Sigma_i I_i(hkl)$, R_{meas} (redundancy-independent R_{merge}) $=\Sigma_{hkl}[N_{hkl}/(N_{hkl}-1)]^{1/2}\Sigma_i|I_i(hkl)-<I(hkl)>|/\Sigma_{hkl}\Sigma_i I_i(hkl)$, and R_{pim} (precision-indicating R_{merge}) $=\Sigma_{hkl}[1/(N_{hkl}-1)]^{1/2}\Sigma_i|I_i(hkl)-<I(hkl)>|/\Sigma_{hkl}\Sigma_i I_i(hkl)$.

$R_{cryst}=\Sigma||F_{obs}|-|F_{calc}||/\Sigma|F_{obs}|$, where F_{calc} and F_{obs} are the calculated and observed structure factor amplitudes, respectively.

R_{free} = as for R_{cryst}, but for 5.0% of the total reflections chosen at random and omitted from refinement.