Crystal structure of an Fe-S cluster-containing fumarate hydratase enzyme from

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1073/pnas.1605031113">http://dx.doi.org/10.1073/pnas.1605031113</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>National Academy of Sciences (U.S.)</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Mon Dec 17 13:32:09 EST 2018</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/108745">http://hdl.handle.net/1721.1/108745</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>
Crystal structure of an Fe-S cluster-containing fumarate hydratase enzyme from *Leishmania major* reveals a unique protein fold

Patricia R. Feliciano\(^{a,b,c}\), Catherine L. Drennan\(^{b,c,d,1}\), and M. Cristina Nonato\(^{b,1}\)

*Laboratório de Cristalografia de Proteínas, Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo, São Paulo 14040-903, Brazil; \(^{d}\)Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139; \(^{\ast}\)Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139; and \(^{\dagger}\)Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139

Edited by Perry Allen Frey, University of Wisconsin–Madison, Madison, WI, and approved July 11, 2016 (received for review March 27, 2016)

Fumarate hydratases (FHs) are essential metabolic enzymes grouped into two classes. Here, we present the crystal structure of a class I FH, the cytosolic FH from *Leishmania major*, which reveals a previously undiscovered protein fold that coordinates a catalytically essential [4Fe-4S] cluster. Our 2.05 Å resolution data further reveal a dimeric architecture for this FH that resembles a heart, with each lobe comprised of two domains that are arranged around the active site. Besides the active site, where the substrate S-malate is bound bidentate to the unique iron of the [4Fe-4S] cluster, other binding pockets are found near the dimeric enzyme interface, some of which are occupied by malonate, shown here to be a weak inhibitor of this enzyme. Taken together, these data provide a framework both for investigations of the class I FH catalytic mechanism and for drug design aimed at fighting neglected tropical diseases.

**Results and Discussion**

**Overall Structure of LmFH-2**. The crystal structure of LmFH-2 was solved by single-wavelength anomalous dispersion (SAD) using iron as the anomalous scatterer and refined to 2.05 Å resolution (Table S1). The asymmetric unit contains one copy of the functional homodimeric enzyme, and the monomers are related by a noncrystallographic twofold axis (Fig. L4). The superposition of Ca atoms between chains A and B shows a high level of structural similarity with an rmsd of 0.22 Å. The quaternary structure of LmFH-2 resembles an upside-down heart (Fig. L4), with each monomer consisting of two structural domains arranged around the catalytic [4Fe-4S] cluster (Fig. L1B).

The LmFH-2 monomer contains 23 β-strands (β1 to β23) and 18 α-helices (α2 to α18), and can be described as being composed of two domains: an N-terminal domain (Asp-28 to Pro-375) and a C-terminal domain (Thr-385 to Ala-568), connected by a flexible linker (Asp-376 to Thr-384) (Fig. L1B and Fig. S1). The first 27 residues and the flexible linker were excluded from the structure because of the lack of interpretable electron density. The N-terminal domain can be divided in subdomains 1 (Asp-28 to Lys-107) and subdomains 2 (Asp-91 to Thr-384) (Fig. 1C).

**Significance**

Leishmaniases, Chagas disease, and sleeping sickness are parasitic diseases classified as neglected tropical diseases, affecting approximately one-sixth of the world’s population. Because of the absence of effective medicines to treat these diseases, there is a substantial interest in the identification of new targets for the development of therapeutic strategies to combat neglected tropical diseases. We have determined the crystal structure of a class I fumarate hydratase (FH) from *Leishmania major*, the parasite responsible for cutaneous leishmaniases, and find that the structure is distinct from class II human FH. Thus, *Leishmania* class I FH, which is an essential metabolic enzyme, offers a new perspective for the development of antileishmaniases therapies.

**Author contributions:** P.R.F. designed research; P.R.F. performed research; P.R.F., C.L.D., and M.C.N. analyzed data; P.R.F., C.L.D., and M.C.N. wrote the paper; and C.L.D. and M.C.N. supervised the research.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 5L2R).

1To whom correspondence may be addressed. Email: cdrennan@mit.edu or cristy@fcrp.usp.br.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1605031113/-/DCSupplemental.
Subdomain 1 contains 11 β-strands (β1 to β9, β15, and β16) and six α-helices (α1 to α5 and α11). Subdomain 2 is located between the β9 and β15 of subdomain 1 and contains five β-strands (β10 to β14) and five α-helices (α6 to α10). The C-terminal domain contains seven α-helices (α12 to α18) and seven β-strands (β17 to β23) organized as independent β-barrel and α-helical motifs (Fig. 1D). The monomer has a [4Fe-4S] cluster coordinated by three cysteine residues (Cys-133, Cys-252, and Cys-346; the motif is C-X_{118}-C-X_{93}-C) from the N-terminal domain (Fig. 2A). The fourth iron of the cluster has no protein ligand.

The DALI server (13) was used to search the Protein Data Bank to identify known structures with similar folds to LmFH-2. Unsurprisingly, DALI finds the β-subunit of a putative class I FH from *Archaeoglobus fulgidus* (PDB ID code 2ISB) as the strongest match to the C-terminal domain of LmFH-2 (z-score of 22.9). The high structural similarity is evident in the structural superposition, which has an rmsd of 2.2 Å for 172 Ca atoms (of the 183 residues comprising the C-terminal domain) (Fig. S2A). In fact, this conserved region is classified by Structural Classification of Proteins database (SCOP) (14) as a swiveling β/β/α domain, known to be a mobile motif in multidomain proteins. Although the structure of the α-subunit of this putative class I FH from *A. fulgidus* may turn out to be similar to the N-terminal domain of LmFH-2, it has not been determined. Moreover, the DALI server finds no close matches to the full N-terminal domain of LmFH-2. The best matches are to the Ni-binding domain of HypA from *Thermococcus kodakaraensis* KOD1 (PDB ID code 3A43) (15) and the N-terminal β-domain of l-serine dehydratase from *Legionella pneumophila* (PDB ID code 4RQO) (16), but in both cases the z-scores are low: 6 and 5.7, respectively. Z-scores below 2 indicate structural dissimilarity. Structural comparisons to the identified proteins reveal that only a fragment of LmFH-2 subdomain 1, comprised of residues from one β-sheet (β4, β8, β9, β15, and β16) and two helices (α2 and α5), is structurally similar (Fig. S2B). The rmsds between this fragment of LmFH-2 and PDB ID codes 3A43 and 4RQO models are 2.7 and 2.7 Å for 74 and 100 aligned Ca atoms, respectively. Thus, we can interpret the relatively low

![Scheme 1.](image-url)

Subdomain 1 contains 11 β-strands (β1 to β9, β15, and β16) and six α-helices (α1 to α5 and α11). Subdomain 2 is located between the β9 and β15 of subdomain 1 and contains five β-strands (β10 to β14) and five α-helices (α6 to α10). The C-terminal domain contains seven α-helices (α12 to α18) and seven β-strands (β17 to β23) organized as independent β-barrel and α-helical motifs (Fig. 1D).

![Fig. 1.](image-url)

Crystal structure of LmFH-2. (A) Overall structure of the LmFH-2 functional dimer. The Upper and Lower panels represent two orthogonal views of the structure with two domains: N terminal (blue and green) and C terminal (yellow). The Left and Right panels show the cartoon and electrostatic surface potential representation of the LmFH-2 dimer, respectively. The [4Fe-4S] clusters are shown in magenta. (B) Ribbon diagram of LmFH-2 monomer. The N-terminal domain is divided into two nonsequential subdomains 1 (light blue and dark blue) and 2 (green), and is connected to C-terminal domain (yellow) by a linker (black arrow), as indicated in the linear schematic. (C) Ribbon diagram of LmFH-2 N-terminal subdomains 1 and 2. (D) Ribbon diagram of LmFH-2 C-terminal domain.
The dimer interface of LmFH-2 involves 32 residues of the N-terminal domain from chains A and B, and is stabilized by 42 hydrogen bonds (Table S2) and two cation–π interactions between Phe-63 and Lys-69. The great majority of residues at the interface are completely conserved in class I FHs (Fig. S3). Phe-63 and Lys-69 are not fully conserved, showing that the cation–π interaction is a unique feature of cytosolic FHs from *Leishmania* spp.

**Active Site of LmFH-2.** The substrate S-malate is clearly identified in both active sites of the LmFH-2 structure and it is found coordinated to the unique iron (Fe4) of the [4Fe-4S] cluster via C2 carboxyl and hydroxyl oxygen atoms, as observed by the final σA-weighted 2Fo–Fo electron density map (Fig. 2A). Notably, aconitase, a [4Fe-4S] cluster-containing enzyme that catalyzes the stereo-specific dehydration/rehydration of citrate to isocitrate via cis-aconitate, also coordinates its substrate isocitrate via a carboxyl and hydroxyl oxygen to the unique Fe atom of its [4Fe-4S] (17). The active site of LmFH-2 is located in a deep cleft formed between the N- and C-terminal domains, and comprises the [4Fe-4S] cluster, a water molecule, and 12 residues from chain A and 1 residue from chain B, suggesting that the dimerization can be important for enzyme activity (Fig. 2 A and B). The N-terminal domain of chain A provides seven residues (Cys-133, Gln-134, Asp-135, Arg-173, Gly-216, Cys-252, Cys-346), with only Gly-216 and Phe-63 located in a helix (α6), whereas all of the other residues are found in loops. The C-terminal domain of chain A provides five residues (Arg-421, Thr-467, Thr-468, Arg-471, Lys-491), with Arg-421 and Arg-471 located in α13 and α15, respectively, and other residues in loops. The N-terminal domain of chain B provides the His-334 that is located in a loop and is not directly involved in substrate binding. Sequence comparisons with 26 members of the class I FHs indicate that active site residues are fully conserved (Fig. S3).

The electrostatic surface potential of LmFH-2 dimer reveals a positively charged cavity located at the interface between N- and C-terminal domains from each monomer (Fig. 1A). This cavity, which leads to the active site, contains two S-malate molecules (Fig. S4), suggesting that access to the active site is favored by the charge distribution within this region. The residues Asn-219, Gln-225, and Tyr-222 coordinate to S-malate in this positive cavity (Fig. 3B), and sequence comparison indicates that only Tyr-222 is not conserved in class I FHs (Fig. S3). Movement of the “swiveling” C-terminal domain, which has higher B-factors than average for the rest of the structure (Fig. S5), may also regulate access to the active site.

Experimental evidence suggests that class I FHs catalyze the dehydration of S-malate to fumarate by a carbanion intermediate (E1cB) mechanism (Fig. 2C), with the [4Fe-4S] cluster acting as a Lewis acid (9). The first step is thought to be deprotonation at C3 to form the carbanion, and surprisingly we find that a Thr (Thr-467) is the closest residue to C3 (3.34 Å), which is near a water molecule and two Arg residues (Arg-421 and Arg-471), any of which could accept the proton. The residue Asp-135, and its hydrogen-bonding partner His-334/B, are ideally positioned to play a role as a catalytic acid to protonate the C2 hydroxyl group of S-malate for elimination as H2O and subsequent formation of fumarate.

**Fig. 2.** LmFH-2 active site. (A) The residues of chains A and B are shown in white and light blue, respectively. The substrate S-malate, [4Fe-4S] cluster and water molecule are shown in green, yellow (S) and orange (Fe), and cyan, respectively. Mesh represents the final 2Fo–Fo electron density map contoured at 1.5σ level (blue) for S-malate and the [4Fe-4S] cluster. A stereoview is shown in Fig. S4. (B) Interactions between S-malate and the active site residues in LmFH-2. The water molecule, C, N, O, Fe, and S atoms are shown in cyan, black, blue, red, orange, and yellow, respectively. The hydrogen bonds are shown as green dashed lines. The distance between the OH group of Thr-467 to the S-malate C3 carbon is shown as a black dashed line. Image created with LigPlot (30). (C) Proposed mechanism for class I FHs to catalyze the dehydration of S-malate to fumarate. The first step is thought to be deprotonation of S-malate C3 by a catalytic base (B). The closest residue to C3 is Thr-467 (3.34 Å), which is near a water molecule and two Arg residues (Arg-421 and Arg-471), any of which could accept the proton. The residue Asp-135, and its hydrogen-bonding partner His-334/B, are ideally positioned to play a role as a catalytic acid to protonate the C2 hydroxyl group of S-malate for elimination as H2O and subsequent formation of fumarate.
H$_2$O and subsequent formation of fumarate. Here, Asp-135 and its hydrogen-bonding partner, His-334/B, are ideally positioned to play this role.

The structure of class II FHs is well known to be a homotramer, with each monomer consisting of three domains, and its active site is formed by residues of three monomers (7), which is different from that found in LmFH-2. In addition, class II FH performs the reaction without a [4Fe-4S] cluster as a cofactor. Because of these differences in structure and cofactor use, along with their essential metabolic roles, class I FHs appear to represent an excellent target for structure-based drug design against NTD.

**LmFH-2 Tunnel as a Ligand-Binding Site.** Interestingly, the dimerization of LmFH-2 reveals a deep cavity on the top of the protein formed between N-terminal domains from both chains (Fig. 4 and Movie S1). This cavity is $\sim$15 Å from the [4Fe-4S] cluster (Fig. 4B), and does not connect to the active site cavity described above, which is at the bottom of the structure. This “top” cavity generates a tunnel that goes through the entire protein and exhibits a volume of 1,168.5 Å$^3$ (Fig. 4A). Although tunnels have been observed in a number of enzyme structures, they commonly provide passage from the protein surface to a buried active site (18) or provide a route from one active site to another to protect a highly reactive (19) or very valuable (20) reaction intermediate. None of these functions would seem to apply here. Interestingly, analysis of the difference electron density map in this tunnel reveals the presence of ligands from the crystallization condition, such as polyethylene glycol and malonate, as well as glycerol.

Malonate is a known inhibitor of porcine class II FH, although not a strong one, with a $K_i$ of 40 mM (21), and has been observed bound in the active site of class II FH from Rickettsia prowazekii (22). Given that inhibition of class I FHs by malonate has not been examined previously, we investigated its ability to inhibit the *L. major* class I FH. Our results demonstrate that malonate is
a weak inhibitor of LmFH-2, with an IC50 of 9.8 ± 0.3 mM against S-malate and 5.6 ± 0.3 mM against fumarate (Fig. S6). Consistent with this finding, we observe S-malate and not malonate bound in the active site when both are present in the crystallization buffer. Instead of being bound in the active site, malonate occupies two different types of pockets within the “top” cavity (Fig. 4B), one near the tunnel entrance (Fig. 4C) and the other at the tunnel center near the dihedral interface (Fig. 4D). The residues His41, Lys-144, Glu-207, Ala-354, and His-355 coordinate to malonate at a pocket near the tunnel entrance, and sequence comparison indicates that those residues are not conserved within class I FHs. However, the malonate bound near the dipter interface is coordinated by residues Gln-195/A, Asp-197/A, and Glu-267/B, where Gln-195 and Glu-267 are conserved and Asp-197 is either Asp or Glu. Although it is unclear if this interface binding site is the source of the weak inhibition by malonate, the interface site does seem to be a conserved small molecule binding site at an interesting position in the structure.

Conclusion

The crystal structure of LmFH-2 represents an important step toward the validation of the metabolically essential class I FH enzymes as targets against leishmaniasis, Chagas disease, and sleeping sickness. We are excited to find that this class I FH has a fold that shares no resemblance to mammalian class II FHs or to any other protein of known structure, thus representing a structurally unique drug target. With a structure of class I FH finally in hand, the catalytic mechanism can be interrogated and designing a structure-based drug to combat these NTDs can begin.

Materials and Methods

Crystallization and Data Collection. Recombinant LmFH-2 was expressed in E. coli T7 express and purified by nickel affinity chromatography, as described previously (5). The purification of LmFH-2 was performed with 1 mM DTT in all buffers at 4 °C in an MBraun anaerobic glovebox. Initial LmFH-2 crystallization conditions were identified using a Mosquito robot (TPP Labtech) in a room-temperature MBraun anaerobic glovebox, and optimized using a hanging-drop vapor-diffusion method at room temperature in a Coya anaerobic chamber. Drops were prepared by mixing 1 μL of protein solution (8.7–10 mg/mL in 50 mM Tris, pH 8.5, 150 mM NaCl, 1 mM DTT) and 1 μL of reservoir solution (2–4% (v/v) vol/vol) tacsimate, pH 5 (Hampton Research), 12–14% (vol/vol) polyethylene glycol (PEG) 3,350 (Hampton Research), equilibrated against 400 μL of reservoir solution. Tacsimate is composed of a mixture of titrated organic salts (Table S1), containing the substrate S-malate and the inhibitor malonate, which resulted in a structure with both molecules bound without further addition of these molecules to the crystallization buffer. After 1 d, brownish needle cluster-like crystals were obtained. The optimization of the crystals was performed using microseeding techniques (24) and ethanol (2.7% (vol/vol) as an additive. The crystals were transferred to a cryoprotectant solution (8% (vol/vol) tacsimate, pH 5, 18% (vol/vol) PEG 3,350, 25% (vol/vol) glycerol), and flash-cooled in liquid nitrogen in the cryo anaerobic chamber. Data collection was performed by an inverse-beam method (Friedel mates were measured rotating the crystal 180° every 15 frames with 1° oscillation and exposure time of 1 s) at the 24-ID-C beamline of the Advanced Photon Source. Diffraction data were processed and scaled using HKL2000 (25). The statistics are summarized in Table S1.

Structure Determination and Refinement. The crystal structure of LmFH-2 was solved by iron-SAD. The positions of the one sulfur and eight iron sites per functional dimer were determined and refined using phenix.autosol (26). A partial model was built in Coot (27) using a 4 Å resolution experimental map with a figure of merit of 0.488, followed by model building using phenix. autosol to 2.7 Å resolution and figure of merit of 0.368. Native data from 200 to 2.05 Å were used for structure refinement and iterative rounds of model building and addition of water molecules in Coot (27). Refinement in phenix.refine (26) used noncrystallographic symmetry restraints, TLS (translation, libration, and screw), and positional and individual B-factor restraints with S-malate and malonate geometry restraints generated by phenix.elbow (26). The Ramachandran statistics are 97.4% in the most favored region and 2.6% in the allowed region. The final model contains 2 poly peptide chains, 2 [4Fe-4S] clusters, 4-S-malate molecules, 4 malonate molecules, 8 glyceroles molecules, 1 PEG molecule, and 772 water molecules (Table S1). Figures were created with PyMol Software (28). The electrostatic surface potentials were calculated using the Adaptive Poisson-Boltzman Solver (29) plugin implemented in PyMol, using default parameters.

ACKNOWLEDGMENTS. This work is based upon research conducted at the Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences from the National Institutes of Health through the National Institutes of Health and Howard Hughes Medical Institute Investigator.