PROTEIN STRUCTURE REPORT

Crystal structures and proposed structural/functional classification of three protozoan proteins from the isochorismatase superfamily

JONATHAN CARUTHERS,¹ FRANK ZUCKER,¹ ELIZABETH WORTHEY,⁴
PETER J. MYLER,⁴ FRED BUCKNER,² WES VAN VOORHUIS,² CHRIS MEHLIN,¹
ERICA BONI,¹ TIFFANY FEIST,¹ JOSEPH LUFT,⁵ STACEY GULDE,⁵
ANGELA LAURICELLA,⁵ OLEKSANDR KALUZHNIY,¹ LORI ANDERSON,¹
ISOLDE LE TRONG,¹ MARGARET A. HOLMES,¹ THOMAS EARNEST,⁶
MICHAEL SOLTIS,⁷ KEITH O. HODGSON,⁷ WIM G.J. HOL,^{1,3} AND
ETHAN A. MERRITT¹

¹Biomolecular Structure Center, Department of Biochemistry, ²Division of Infectious Disease, and

³Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA

⁴Seattle Biomedical Research Institute, Seattle, Washington 98109, USA

⁵Hauptman-Woodward Medical Research Institute, Buffalo, New York 14203, USA

⁶Berkeley Center for Structural Biology, Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

⁷SSRL, Stanford University, Stanford, California 94305, USA

(RECEIVED August 15, 2005; FINAL REVISION August 15, 2005; ACCEPTED August 17, 2005)

Abstract

We have determined the crystal structures of three homologous proteins from the pathogenic protozoans *Leishmania donovani*, *Leishmania major*, and *Trypanosoma cruzi*. We propose that these proteins represent a new subfamily within the isochorismatase superfamily (CDD classification cd004310). Their overall fold and key active site residues are structurally homologous both to the biochemically well-characterized *N*-carbamoylsarcosine-amidohydrolase, a cysteine hydrolase, and to the phenazine biosynthesis protein PHZD (isochorismase), an aspartyl hydrolase. All three proteins are annotated as mitochondrial-associated ribonuclease Mar1, based on a previous characterization of the homologous protein from *L. tarentolae*. This would constitute a new enzymatic activity for this structural superfamily, but this is not strongly supported by the observed structures. In these protozoan proteins, the extended active site is formed by inter-subunit association within a tetramer, which implies a distinct evolutionary history and substrate specificity from the previously characterized members of the isochorismatase superfamily. The characterization of the active site is supported crystallographically by the presence of an unidentified ligand bound at the active site cysteine of the *T. cruzi* structure.

Keywords: structural genomics; *Leishmania*; *Trypanosoma*; functional annotation; protein families; evolutionary relationships; cysteine hydrolase

Reprint requests to: Ethan A. Merritt, Biomolecular Structure Center M/S 357742, University of Washington, Seattle, WA 98195, USA; e-mail: merritt@u.washington.edu; fax: (206) 685-7002.

Article published online ahead of print. Article and publication date are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.051783005>.

An interesting phenomenon revealed by structural biology is the diverse functionality of proteins with similar tertiary folds and even active site features. We have determined structures for a set of protozoan proteins, which show high structural homology to members of the isochorismatase

superfamily of proteins, despite low sequence identity and a disparate functional annotation. The set of protein structures presented here were determined as part of the consortium for Structural Genomics of Pathogenic Protozoa (SGPP). All three proteins are annotated as mitochondrial-associated endoribonuclease Mar1 on the basis of prior biochemical characterization of a close homolog from *Leishmania tarentolae* (GeneDB no. AF083881) (Alfonzo et al. 1998). The two *Leishmania* species homologs reported here share ~90% sequence identity with the *L. tarentolae* homolog. The *Trypanosoma cruzi* protein reported here is more distantly related (51% sequence identity) but is found to be essentially identical in structure and active site features. In this study we have used structural analysis and ligand binding experiments to show that the active site characteristics of these proteins closely resemble both the cysteine-mediated and aspartyl-mediated catalytic sites of previously characterized superfamily members.

SGPP target proteins are routinely cloned and expressed in parallel from multiple leishmania species in order to maximize the chance of eventual success in purification and crystallization. Parallel expression of this target was attempted from three species: *Leishmania major*, *Leishmania donovani*, and *Leishmania mexicana*. Crystals were obtained first for target construct Lmaj001686AAA, but these diffracted only to low resolution. Subsequently, better-diffracting native crystals were obtained from target Tcru003547AAA, and SeMet crystals were obtained from

target Ldon001686AAA. The structure was solved by Se MAD from the *L. donovani* crystals and used for molecular replacement to solve the *L. major* and *T. cruzi* structures (Table 1).

It was apparent upon inspection that these proteins were structurally homologous to the isochorismatase superfamily of enzymes, CDD classification cd00431 (Murzin et al. 1995), despite sequence identity of <20% to previously characterized family members. This superfamily has been subdivided by SCOP into five families, each represented by at least one structure (Murzin et al. 1995). These are (1) nicotinamidase E.C. 3.5.1.19, (2) nicotinamidase related, (3) *N*-carbamoylsarcosine amidohydrolase (CSHase) E.C. 3.5.1.59, (4) isochorismatase E.C. 3.3.2.1, and (5) a family of bacterial sequences of unknown function exemplified by the *Escherichia coli* yac gene product. Although the quaternary structures observed for these proteins are quite diverse, their tertiary fold is homologous, and key features of the active site are highly conserved. For example, all structures determined to date from this superfamily exhibit a rare nonproline *cis*-peptide bond at the active site, which helps to position the substrate-binding residues appropriately. There is not, however, universal conservation of catalytic residues across the multiple families. The first superfamily members to be characterized exhibited a conserved Cys residue at the active site, so the grouping was originally annotated as

Table 1. Summary of data collection and refinement statistics

	Target		
	Lmaj001686AAA	Ldon001686AAA	Tcru003547AAA
Space group	I4	I432	P432
Unit cell dimensions (Å)	a = b = 100.4, c = 44.7	a = b = c = 147.4	a = b = c = 121.1
Resolution range (Å)	50–3.8	50–2.4	120–2.0
Unique reflections	2294	10,935	21,028
Completeness ^a	100% (99%) ^a	100% (100%)	100% (100%)
Redundancy	3.7 (3.5)	20.8 (20.6)	13.1 (11.5)
R _{sym} ^b	0.126 (0.91)	0.075 (0.91)	0.164 (0.67)
Structure solution method	MR	MAD	MR
R _{cryst} ^c	0.303	0.225	0.157
R _{free} ^d	0.310	0.276	0.187
RMSD bond lengths (Å)	0.003	0.02	0.019
RMSD bond angles (°)	0.630	1.9	1.6
Residues fit	191	192	195
Average B factors (Å ²)	92.75	63.7	28.6
Protein atoms in model	1504	1517	1495
Water molecules in model	0	0	165

^aNumbers in parentheses refer to numbers for the outer shell.

^bR_{sym} = $\sum |I - \langle I \rangle| / \sum I$, where I is observed intensity; $\langle I \rangle$, average intensity from multiple measurements.

^cR_{cryst} = $\sum ||F_o| - |F_c|| / \sum |F_o|$, where $|F_o|$ is observed structure factor amplitude; $|F_c|$ is calculated structure factor amplitude.

^dR_{free} R_{cryst} based on 5% of the data excluded from refinement.

Table 2. Conserved active site structural features of the isochorismatase superfamily

Protein	Catalytic cysteine	Cis-peptide	Triad Asp	Triad Lys/Arg	Conserved polar/charged	Conserved Thr	Conserved Gln
Ldon001686AAA	Cys112	Ile107–Glu108	Asp19	Lys82	His59	Thr57	Gln21
Lmaj001686AAA	Cys112	Ile107–Glu108	Asp19	Lys82	His59	Thr57	Gln21
Tcr003547AAA	Cys115	Phe110–Glu111	Asp19	Lys83	Gln60	Thr58	Gln21
1YAC	Cys118	Val113–Val114	Asp19	Arg84	Ser59	Thr57	Gln21
1J2R	NA	Ile140–Ser141	Asp26	Lys112	NA	NA	Gln28
1ILW	Cys133	Val128–Ala129	Asp10	Lys94	Asp52	Thr50	Gln12
1NBA	Cys177	Ala172–Thr173	Asp51	Lys144	Asn94	Thr92	NA
1NF8	NA	Val150–Tyr151	Ala38 ^a	Lys122	Gln78	Thr76	Gln40

^a 1NF8 is a D38A mutant; the native enzyme has an Asp residue in this location.

the “cysteine hydrolase” superfamily. But this Cys residue is not conserved in isochorismatase itself (Table 2).

The best biochemically characterized structural representative of cysteine hydrolases within this structural superfamily is CSHase (Romao et al. 1992; Zajc et al. 1996). Because of the conserved nature of the active site of CSHase with the homologous region of Ldon001686AAA, we hypothesized that this protein may bind known inhibitors of CSHase, including glyoxylate and succinic semialdehyde (SSA). Ligand binding studies of these suicide inhibitors did, in fact, show this to be the case. Further, upon determination of the Tcr003547AAA structure, electron density corresponding to an unknown ligand molecule was present in the protein molecule in the pocket corresponding to the active site pockets of both CSHase (Romao et al. 1992; Zajc et al. 1996) and the phenazine biosynthesis protein PHZD (Parsons et al. 2003).

Results and Discussion

The Ldon001686AAA, Lmaj001686AAA, and Tcr003547AAA proteins appear to be biological tetramers made up of identical 21.6-kDa subunits. The monomer fold is a β -sandwich domain in which a central β sheet is flanked by α helices (Fig. 1A). This fold is characteristic of the isochorismatase/cysteine hydrolase superfamily (CDD classification cd00431). The three structures are in crystallographically distinct space groups, but in each case four of these domains oligomerize to form a pinwheel tetramer that is presumably the active conformation (Fig. 1B). Previously known structural representatives of this superfamily are variously monomers, tetramers, and octamers, but in each case the location and key structural features of the active site within the monomer fold are strongly conserved. In the present set of structures, there are four putative active sites in the tetrameric assembly, each formed at the interface between two monomers.

Three of the five previously assigned families within this superfamily are cysteine hydrolases (Table 2). Although most of these structural homologs have been poorly characterized biochemically and enzymatically, one homolog in particular, CSHase, has been studied extensively (Romao et al. 1992; Zajc et al. 1996). CSHase is involved in one of two creatine degradation pathways in microorganisms. Through this route, creatine is degraded into sarcosine, which is eventually degraded into glycine. CSHase itself degrades *N*-carbamoylsarcosine to sarcosine, ammonia, and carbon dioxide. CSHase is thought to work through the attack of the carbamoyl carbon by the thiol group from an active site cysteine residue. The resultant thiohemiacetal is stabilized by the main-chain backbone amide in a *cis*-peptide bond between Ala172 and Thr173.

An optimized structural alignment (Kleywegt and Jones 1997; Guda et al. 2004) of CSHase and Ldon001686AAA revealed remarkable structural homology to the active site of CSHase. The *L. donovani* protein contains an active site residue Cys 112, a *cis*-peptide linkage between Ile107 and Glu108, and a very similar arrangement of polar groups (Table 2). Further additional conserved, or semiconserved, residues can be seen in the *L. donovani* pocket such as Thr57 and Gln21 (Table 2). The role that these conserved residues play in the function of the protein is not clear. This led us to hypothesize that Ldon001686AAA might plausibly function by using the same fundamental chemistry as the amidohydrolase.

To test this theory, we have attempted to bind the CSHase inhibitor ligands glyoxylate and SSA to Ldon001686AAA. Both of these bind irreversibly to the active site cysteine of CSHase. When Ldon001686AAA/glyoxylate cocrystals were analyzed, the unit cell was found to have shrunk by 10 Å, and clear density for the glyoxylate ion was observed bound to the putative active site cysteine residue, Cys112, in the same orientation as was observed in the CSHase structure (data not shown). Difference maps resulting from soaking of SSA into Ldon001686AAA crystals revealed that

Figure 1. Overview of monomer fold and tetrameric quaternary structure shared by Ldon001686AAA, Lmaj001686AAA, and Tcr003547AAA. (A) Stereo view of Tcr003547AAA monomer, with the N terminus at the *bottom right* and C terminus at the *top center*, showing the characteristic fold of the isochorismatase superfamily. (B) The tetrameric quaternary assembly of Ldon001686AAA. The three proteins crystallize in different space groups, but in all three structures a crystallographic fourfold axis runs through the center of the tetramer. The resulting tetramers are essentially identical.

SSA bound in a similar fashion to glyoxylate (data not shown). While these structures do not by themselves indicate the true natural substrate[s] of the *L. donovani*, *L. major*, and *T. cruzi* proteins, they provide evidence that the active site geometry is consistent with the cysteine-mediated catalytic mechanism of CSHase (Romao et al. 1992; Zajc et al. 1996).

The similarity in character between the active site in the present structures and the active sites of CSHase and PHZD does not extend beyond the immediate vicinity of the catalytic residues. To gain further insight into the

possible evolutionary and functional relationships to other members of the isochorismatase superfamily, we constructed structure-based sequence alignments using the program CEMC (Guda et al. 2004). These are shown in Figures 2 and 3.

This alignment reveals that in four of the five currently assigned subfamilies, one side of the active site cleft is formed by a ~25-residue loop that is not found in our present structures. In the 1NBA, 1J2R, 1IM5, and 1NF8 structures, this allows formation of a complete active site from the monomeric protein. Residues from this loop are involved in substrate recognition, even in the case of the tetrameric proteins 1NBA and 1J2R, whose tetrameric assemblies are observed to be in different configurations than the regular fourfold (C4) symmetry seen in the present structures. By contrast, this portion of the active site surface in the present structures is contributed by helix $\alpha 8$ (residues 180–190) from a neighboring monomer.

Figure 2. Structural superposition of representative isochorismatase superfamily members onto Tcr003547AAA. Each structure is representative of one subfamily in the current SCOP classification. A color-coded backbone trace is shown for each protein: red, Tcr003547AAA (reported here; 1YZV) reference monomer; brown, Tcr003547AAA adjacent monomer; magenta, *E. coli* yac gene product (1YAC); yellow, *Arthrobacter* sp. *N*-carbamoylsarcosine amidohydrolase (1NBA); green, *P. aeruginosa* Phenazine biosynthesis protein PHZD (isochorismatase) (1NF8); cyan, *Pyrococcus horikoshii* pyrazinamidase (nicotinamidase) (1IM5); and blue, *E. coli* yecD gene product (nicotinamidase-related) (1J2R). Difference electron density, corresponding to the unknown ligand found at the active site of the Tcr003547AAA structure, is shown contoured at 3σ in a ($mF_o - F_c$) map. The key feature of this superposition is that one surface of the extended active site in the isochorismatase subfamilies represented by 1NBA, 1NF8, 1IM5, and 1J2R is formed by a characteristic loop of residues visible at the *top* of this figure. The residues making up this loop are not present in the protozoan sequences or in the *E. coli* yac gene product (Fig. 3). In the present protozoan structures, the analogous binding surface is formed instead by residues from a neighboring subunit in the monomer (brown trace in this figure).

Figure 3. Structure-based primary sequence alignment of representative isochorismatase superfamily members. PDB codes 1YZV, 1NF8, 1NBA, 1IM5, 1J2R, and 1YAC. Only the region of structural homology is shown; the individual sequences extend both N-terminal and C-terminal to this region. In particular, the C-terminal residues 175–196 (*T. cruzi* numbering) of the protozoan proteins, including helix $\alpha 8$, are not shown. Conserved regions are colored as follows: green, active site putative catalytic triad Cys, Asp, and Lys/Arg residues; yellow, active site *cis*-peptide linkage; red background, conserved Asp; magenta/black background, conserved Thr; brown, 25-residue loop present in only some families within the superfamily; and magenta, residues interacting with active sites of other monomers in the tetrameric assembly.

The one previously assigned isochorismatase subfamily that lacks this same 25-residue loop is structurally represented by the *E. coli* yac gene product (Protein Data Bank [PDB] entry 1YAC), of unknown function (Colovos et al. 1998). This protein was observed to form an octamer consisting of two back-to-back tetramers, each exhibiting regular C4 symmetry as in the present protozoan structures. While the conserved residues of the active site of 1YAC are homologous to their counterparts of the protozoan structures, the primary sequence as a whole shows <20% identity to the protozoan sequences. Furthermore, homology between the yac gene sequence and the protozoan sequences ends at approximately residue 174 of the *T. cruzi* sequence (Fig. 3) and thus does not encompass that part of the active site binding surface formed by helix $\alpha 8$ in the present

structures. The longer (208-residue) yac sequence instead continues with an additional C-terminal 15-residue α helix that has no counterpart in the protozoan structures.

Another close homolog of the structures presented here is the structure of the isochorismatase PhzD gene product from *Pseudomonas aeruginosa* (PDB entry 1NF8). Although this structure lacks the catalytic cysteine residue present in our structures and CSHase, it contains the other catalytic triad residues thought to be required for reaction in the CSHase structure. In addition, similar to CSHase and unlike either our structures or the 1YAC structure, the 1NF8 structure has the 25-residue loop present in other family members. As seen in the CSHase structure, this loop has the effect of cloistering the active site and presumably sequestering the substrate during reaction. The structure of 1NF8 was

co-crystallized with a substrate isochorismate molecule, after alanine mutation of the triad Asp38 residue. It was determined that the active site Gln78 residue was in contact with the carboxylate of the substrate isochorismate. Of the three structures presented here, only Tcr003547AAA possesses a glutamic acid residue in the corresponding position (Gln60) (Table 2). The homologs from *Leishmania* instead have a histidine residue at this site. It is interesting to ponder what role this particular position may play in substrate specificity.

When the structure of the *T. cruzi* homolog was solved, additional electron density was found in the active site of this protein, presumably from an exogenous ligand molecule introduced during protein expression in *E. coli* (Fig. 4). The density in the immediate region of Cys115 is consistent with a ligand conformation similar to the glyoxylate molecule seen in the Ldon-001686AAA structure; however, the *T. cruzi* ligand is clearly much larger in size. Although the overall electron density is similar in volume and spatial location to the isochorismate ligand observed in the *P. aeruginosa* PHZD structure, its specific shape is not compatible. It is clearly not a nucleotide. Attempts to model other potential ligands into this density have so far been unsuccessful.

We have determined the structures of three homologous protozoan proteins. Although currently annotated as an endoribonuclease, these proteins are found to be structurally homologous to the isochorismatase superfamily. This superfamily contains no members annotated as ribonuclease. Although the observed active site in the present structures is large enough to accommodate a nucleic acid chain, the active site features do not seem compatible with oligonucleotide binding or endonuclease activity. Therefore the current annotation is called into question. The active site cysteine would suggest a

cysteine hydrolase activity consistent with the activity characterized for certain other superfamily members such as CSHase. However, these other branches of the superfamily share an extended binding site formed by a loop of residues encoded by a region of sequence which is missing from the present family of protozoan sequences. Key interactions provided by this loop in other superfamily proteins but missing in ours are instead provided by the neighboring monomer of the tetrameric quaternary assembly. We therefore propose that these protozoan proteins represent a new, evolutionarily distinct subfamily within the isochorismatase superfamily.

Materials and methods

Target selection

Following accepted guidelines for structural genomics, these targets were chosen for structural determination primarily because of low sequence homology to known structures. PSI-BLAST searches revealed them to belong to a family of sequences from various protozoans with sequence identities up to 90% but low sequence homology <40% (nearest sequence homolog is LOC495931 protein from *Xenopus laevis* [accession no. AAH87293-NCBI]) to sequences from other organisms. The protozoan sequences were annotated as Mar1 endonuclease on the basis of biochemical characterization of the *L. tarentolae* homolog. This nuclease was reported to have weak, nonspecific ribonucleolytic activity within the mitochondria (Alfonso et al. 1998).

Cloning, expression, and purification

DNA from *L. major* Friedlin and *L. donovani* Ld1S was extracted from parasites grown in culture with a simple phenol extraction. RNA was not removed prior to using this as a template. Polymerase chain reaction (PCR) was employed to amplify the target using the following primers: forward primer, CTCACCACCACCACCACCATATGTCTCGCTTGATGCGCATTA; reverse primer, ATCCTATCTTACTCACTTAGAGCGGGATCGGAGGCTCCT. (Underlined regions are specific for the gene.) The PCR protocol was as follows: 120 sec at 95°C; 30 repetitions of 30 sec at 94°C, 60 sec at 60°C, and 270 sec at 72°C; and 600 sec at 72°C.

For the *L. major* target, PCR was done by using the PfuTurbo kit (Stratagene) according to the manufacturer's instructions on a PTC 200 (MJ Research) thermal cycler. The *L. donovani* target was amplified in a similar manner using the same (unoptimized) primers and standard Taq polymerase. The amplified PCR product was purified by agarose gel electrophoresis, extracted by using a QiaQuick 96 kit (Qiagen), and spliced into BG1861, a modified pET14b vector that appends MAHHHHHH onto the N terminus of the protein, by ligase-independent cloning (LIC) using T4 polymerase (Novagen). HT-96 *E. coli* (Novagen) were transformed, and plasmid was extracted with a QiaPrep 96 Turbo kit (Qiagen) following overnight growth of a single colony inoculated into 600 μ l of Terrific Broth. This plasmid was used to transform BL-21 Star cells (Invitrogen), and a single colony was expanded into a liter

Figure 4. Active site of Tcr003547AAA showing observed difference electron density corresponding to an unidentified ligand. Density cage is contoured at 3σ in a $(mF_o - F_c)$ difference map. Relevant active site residues are labeled.

of ZYP-5052 autoinduction media and grown overnight at 37°C and then overnight at 18°C in order to generate unlabeled protein. For selenomethionine-labeled protein, a single colony inoculated a liter of PA-0.5G, a phosphate-buffered, defined media, and grown overnight at 37°C with constant shaking. The resulting *E. coli* were collected by centrifugation and resuspended in 2 L of PASM-5052 selenomethionine media; these were allowed to grow for 6 h at 37°C and then overnight at 18°C. *E. coli* from either standard or selenomethionine media were harvested by centrifugation and frozen at -80°C. The pellet was then resuspended in standard buffer—25 mM HEPES (pH 7.25), 120 mM NaCl, 5% glycerol, and 0.025% sodium azide—to which was added 0.2% cholate, 1 mg/mL lysozyme (Sigma), 1 mM 2ME, and protease inhibitors (Roche Complete, EDTA-free) and then was sonicated on ice to disrupt *E. coli*. The selenomethionine derivatives of the *L. donovani* protein were isolated by using the same standard buffer with a NaCl concentration raised to 500 mM. Cellular debris was removed by 20 min of centrifugation at 18,000g, and the supernatant was tumbled with 10 mL of nickel-NTA resin (Superflow NTA, Qiagen) for 30 min at 4°C. The resin was allowed to settle, the supernatant was discarded, and the resin then was rinsed once with 10 mM imidazole and once with 20 mM imidazole in standard buffer. The resin was then recovered and added to a disposable column; the resin was then rinsed with 20 mM imidazole in standard buffer, the protein was eluted with 15 mL of 500 mM imidazole in standard buffer, and the eluent was dialyzed against 4 L of standard buffer overnight. The dialyzed material was concentrated to 10 mL by centrifugal ultrafiltration (Amicon Ultra, Millipore), DTT was added to 1 mM, and the dialyzed product was then applied to a prepacked Superdex 75 26/60 gel chromatography column (Amersham Biosciences) at 4°C in standard buffer. After running at 1 mL/min, peak fractions were collected and pooled, protease inhibitors (Roche Complete, EDTA free) were added, and the solution was concentrated in standard buffer with 2 mM DTT. Protein was aliquotted into a PCR plate, flash-frozen in liquid nitrogen, and stored at -80°C (Deng et al. 2004; Mehlin et al. 2004).

Use of common primers to clone protein from multiple species

The *L. major* primers were used to amplify DNA for the homologous targets from *L. donovani* and *L. mexicana* strains as described above. These were cloned and expressed as SGPP targets Lmaj001686AAA, Ldon001686AAA, and Lmex001686AAA. It was not possible to purify soluble protein from the *L. mexicana* expression, but purified protein from the other two species was sent for crystallization trials.

Crystal screening/optimization

Initial crystal conditions were found by large-scale screening at the Hauptman-Wood Medical Research Institute (Luft et al. 2003). Crystallization experiments were set up by using the microbatch-under-oil technique (Chayen et al. 1992) with a Robbins Scientific Tango liquid handling system. Each of the 1536 experiments contained 200 nL of crystallization cocktail solution combined with 200 nL of protein solution under paraffin oil (Fluka catalog no. 76235) contained in a 1536-well plate (Greiner BioOne catalog no. 79101). The experiment plate was stored for

1 wk at 4°C and imaged at 23°C. Images were manually reviewed. Of the 1536 crystallization conditions, nine produced outcomes for Ldon001686AAA considered suitable for further optimization trials, as did 15 for Tcr003547AAA.

Optimization of crystal growth conditions was performed in Seattle. All liquid handling and data tracking tasks were accomplished by using a modified CrystalMation platform (RoboDesign). Solution matrices were generated by the Alchemist I screen making system as programmed automatically from the CrystalTrak database; 400 nL and 400 nL sitting drop vapor-diffusion experiments were performed in 96-well Intelli-plates (Hampton Research catalog no. HR3-299) prepared by the automated high-throughput dispenser Hydra-Plus-One as described elsewhere (Krupka et al. 2002). Crystallization plates were imaged at regular intervals by using the RoboMicroScope II and scored manually with CMView software provided by the manufacturer. Final optimized conditions for each target are as follows: (1) Lmaj001686AAA at 100 mM MOPS (pH 6.5), 15% PEG 8K, and 100 mM potassium phosphate; (2) Ldon001686AAA at 100 mM MES (pH 5.6), 18% PEG 1K, and 50 mM potassium phosphate; and (3) Tcr003547AAA at 100 mM MES (pH 6) and 1.4 M ammonium sulfate

Data collection/phasing/refinement

Data were collected at the ALS and SSRL synchrotron radiation laboratories, using Quantum Q4 and Q315 CCD detectors. All crystals were flash frozen in liquid nitrogen to 100K. Native data for Lmaj001686AAA and Ldon001686AAA were processed by using the program HKL2000 (Otwinowski and Minor 1997). Selenium derivative data for Ldon001686AAA were initially autoindexed by the program MOSFLM (Leslie 1992), collected by using the collection software Blu-Ice, and processed by using MOSFLM via the automated scripting procedure Elves (Leslie et al. 1986; Leslie 1992; CCP4 1994; Steller et al. 1998; Holton and Alber 2004). Data for Tcr003547AAA and for the Ldon001686AAA + glyoxylate cocrystal were autoindexed and processed by Elves, and collected by using the program DCS at ALS beamline 8.2.1 (Table 1).

Experimental phases were obtained by three-wavelength MAD phasing of the SeMet Ldon001686AAA. Five initial Se sites were found by using the program SOLVE (Terwilliger and Berendzen 1999). These sites were then phased in the CCP4 program MLPHARE (CCP4 1994). A synthesis map was generated by using the CCP4 program FFT, and this map was then used to fit the Ldon1686AAA model. Initial refinement of Ldon001686AAA was carried out initially by using the program CNS (Brunger et al. 1998). Final refinement of Ldon001686AAA, using TLS refinement, was carried out by using the program Refmac5 (Table 1). The structures of Lmaj001686AAA and Tcr003547AAA were solved by placement of the *L. donovani* structure into the corresponding unit cells by using the program EPMR (Kissinger et al. 1999). This model was initially refined by using CNS, but final refinement was completed by using the program Refmac5 (Murshudov et al. 1997; Table 1).

Model building and analysis of electron density

Model building was accomplished through manual fitting of electron density by using the program O. Idealized helices and

strands were placed into density, and loops were added to link the poly-alanine chain together. Anomalous difference maps of the Selenium positions were used as guideposts during the fitting process.

Fitting of the glyoxylate ligand was carried out through automated real-space refinement routines in the program Xfit (McRee 1999). The identity of the ligand corresponding to the unexpected electron density found in the structure of the *T. cruzi* homolog remains unknown. Attempts have been made to model various molecules such as isochorismate and adenosine monophosphate into the density cage but have not yielded plausible structural models. We are currently attempting the characterization of this unidentified ligand through biochemical and biophysical means.

Structural superpositions were carried out by using the CE-MC server (Guda et al. 2004). Figure 3 was prepared by using TeXShade (Beitz 2000).

Accession numbers

Atomic coordinates and structure factors for the three structures reported here have been deposited with the PDB (accession codes 1X9G, 1XN4, 1YZV).

Acknowledgments

We wish to acknowledge the essential contributions from other members of the SGPP consortium, including Gholam Fazelinia, Christy Vogt, Tim Louie, Grace Huang, Chris Fong, and Ellen Sisk at the Seattle Biomedical Research Institute; Christina Veatch and Jennifer Smith at the Hauptman Woodward Institute; and Laurie Bachrach, Martin Crimiale, J.T. Reddy, Larry DeSoto, Tracy Arakaki, Christophe Verlinde, and Erkang Fan at the University of Washington. This work was supported by NIH Protein Structure Initiative award GM64655. Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory, a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, and the NIH, National Center for Research Resources, Biomedical Technology Program, and the National Institute of General Medical Sciences. Other portions were carried out at the Advanced Light Source, which is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under contract no. DE-AC02-05CH11231.

References

- Alfonzo, J.D., Thiemann, O.H., and Simpson, L. 1998. Purification and characterization of MAR1: A mitochondrial associated ribonuclease from *Leishmania tarentolae*. *J. Biol. Chem.* **273**: 30003–30011.
- Beitz, E. 2000. TeXshade: Shading and labeling of multiple sequence alignments using LaTeX2e. *Bioinformatics* **16**: 135–139.
- Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. 1998. Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**: 905–921.
- Chayen, N.E., Shaw Stewart, P.D., and Blow, D.M. 1992. Microbatch crystallization under oil: A new technique allowing many small-volume crystallization trials. *J. Crystal Growth* **122**: 176–180.
- Collaborative Computational Project, Number 4 (CCP4). 1994. The CCP4 Suite: Programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**: 760–763.
- Colovos, C., Cascio, D., and Yeates, T.O. 1998. The 1.8 Å crystal structure of the ycaC gene product from *Escherichia coli* reveals an octameric hydrolase of unknown specificity. *Structure* **10**: 1329–1337.
- Deng, J., Davies, D.R., Wisedchaisri, G., Wu, M., Hol, W.G., and Mehlin, C. 2004. An improved protocol for rapid freezing of protein samples for long-term storage. *Acta Crystallogr. D Biol. Crystallogr.* **60**: 203–204.
- Guda, C., Lu, S., Scheeff, E.D., Bourne, P.E., and Shindyalov, I.N. 2004. CE-MC: A multiple protein structure alignment server. *Nucleic Acids Res.* **32**: W100–W103.
- Holton, J.M. and Alber, T. 2004. Automated protein crystal structure determination using ELVES. *Proc. Natl. Acad. Sci.* **101**: 1537–1542.
- Kissinger, C.R., Gehlhaar, D.K., and Fogel, D.B. 1999. Rapid automated molecular replacement by evolutionary search. *Acta Crystallogr. D Biol. Crystallogr.* **55**: 484–491.
- Kleywegt, G.J. and Jones, T.A. 1997. Detecting folding motifs and similarities in protein structures. *Methods Enzymol.* **277**: 525–545.
- Krupka, H.I., Rupp, B., Segelke, W., Lekan, T., Wright, D., Wu, H.C., Todd, P., and Azarani, A. 2002. The high-speed Hydra-Plus-One system for automated high-throughput protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **58**: 1523–1566.
- Leslie, A.G.W. 1992. Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography*, No. 26.
- Leslie, A.G.W., Brick, P., and Wonacott, A. 1986. Recent changes to the MOSFLM package for processing film and image plate data. *Daresbury Lab. Inf. Quart. Protein Crystallogr.* **18**: 33–39.
- Luft, J.R., Collins, R.J., Fehrman, N.A., Lauricella, A.M., Veatch, C.K., and DeTitta, G.T. 2003. A deliberate approach to screening for initial crystallization conditions of biological macromolecules. *J. Struct. Biol.* **142**: 170–179.
- McRee, D.E. 1999. XtalView/Xfit: A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**: 156–165.
- Mehlin, C., Boni, E.E., Andreyka, J., and Terry, R.W. 2004. Cloning grills: High throughput cloning for structural genomics. *J. Struct. Funct. Genomics* **5**: 59–61.
- Murshudov, G.N., Vagin, A.A., and Dodson, E.J. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**: 240–255.
- 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.
- Otwinowski, Z. and Minor, W. 1997. Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**: 307–326.
- Parsons, J.F., Calabrese, K., Eisenstein, E., and Ladner, J.E. 2003. Structure and mechanism of *Pseudomonas aeruginosa* PhzD: An isochorismatase from the phenazine biosynthetic pathway. *Biochemistry* **42**: 5684–5693.
- Romao, M.J., Turk, D., Gomis-Ruth, F.X., Huber, R., Schumacher, G., Mollering, H., and Russmann, L. 1992. Crystal structure analysis, refinement and enzymatic reaction mechanism of *N*-carbamoylsarcosine amidohydrolase from *Arthrobacter* sp. at 2.0 Å resolution. *J. Mol. Biol.* **226**: 1111–1130.
- Steller, I., Bolotovskiy, R., and Rossmann, M., 1998. The use of partial reflections for scaling and averaging x-ray area-detector data. *J. Appl. Crystallogr.* **30**: 1036–1040.
- Terwilliger, T.C. and Berendzen, J. 1999. Automated MAD and MIR structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **55**: 849–861.
- Zajc, A., Romao, M.J., Turk, B., and Huber, R. 1996. Crystallographic and fluorescence studies of ligand binding to *N*-carbamoylsarcosine amidohydrolase from *Arthrobacter* sp. *J. Mol. Biol.* **263**: 269–283.