

# Structure of the prolyl-tRNA synthetase from the eukaryotic pathogen *Giardia lamblia*

Eric T. Larson,<sup>a,b</sup> ‡ Jessica E. Kim,<sup>a,b</sup> Alberto J. Napuli,<sup>a,c</sup> Christophe L. M. J. Verlinde,<sup>a,b</sup> Erkang Fan,<sup>a,b</sup> Frank H. Zucker,<sup>a,b</sup> Wesley C. Van Voorhis,<sup>a,c</sup> Frederick S. Buckner,<sup>a,c</sup> Wim G. J. Hol<sup>a,b</sup> and Ethan A. Merritt<sup>a,b,\*</sup>

<sup>a</sup>Medical Structural Genomics of Pathogenic Protozoa, <http://msgpp.org>, USA, <sup>b</sup>Department of Biochemistry, University of Washington, Seattle, WA 98195, USA, and <sup>c</sup>Department of Medicine, University of Washington, Seattle, WA 98195, USA

‡ Current address: Department of Medicinal Chemistry, Boehringer Ingelheim, Ridgefield, Connecticut, USA.

Correspondence e-mail: [merritt@u.washington.edu](mailto:merritt@u.washington.edu)

The genome of the human intestinal parasite *Giardia lamblia* contains only a single aminoacyl-tRNA synthetase gene for each amino acid. The *Giardia* prolyl-tRNA synthetase gene product was originally misidentified as a dual-specificity Pro/Cys enzyme, in part owing to its unexpectedly high off-target activation of cysteine, but is now believed to be a normal representative of the class of archaeal/eukaryotic prolyl-tRNA synthetases. The 2.2 Å resolution crystal structure of the *G. lamblia* enzyme presented here is thus the first structure determination of a prolyl-tRNA synthetase from a eukaryote. The relative occupancies of substrate (proline) and product (prolyl-AMP) in the active site are consistent with half-of-the-sites reactivity, as is the observed biphasic thermal denaturation curve for the protein in the presence of proline and MgATP. However, no corresponding induced asymmetry is evident in the structure of the protein. No thermal stabilization is observed in the presence of cysteine and ATP. The implied low affinity for the off-target activation product cysteinyl-AMP suggests that translational fidelity in *Giardia* is aided by the rapid release of misactivated cysteine.

Received 29 March 2012

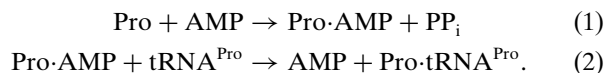
Accepted 30 May 2012

**PDB Reference:** prolyl-tRNA synthetase, 3ial.

## 1. Introduction

The intestinal parasite *Giardia lamblia* (alternatively *G. intestinalis*) is a eukaryote belonging to the class Diplomonadida. This water-borne pathogen is prevalent worldwide and can be transmitted *via* inadequately purified water or between infected humans. *G. lamblia* is one of several pathogenic protozoa whose genomes are being mined by the MSGPP structural genomics collaboration in order to identify potential targets for the development of new antiparasitic drugs (Van Voorhis *et al.*, 2009). Aminoacyl-tRNA synthetases (aaRSs) constitute one such class of potential drug targets, as they are key enzymes for protein synthesis in all organisms and hence, with rare exceptions, are essential for the growth or survival of the organism. Many eukaryotic genomes code for separate cytosolic and mitochondrial aaRS orthologs. However, *G. lamblia* is notable both for the small size of its genome and for its lack of mitochondria, so unsurprisingly it contains only a single set of aaRS genes.

Each aaRS carries out two sequential reactions. It must recognize and activate the corresponding amino acid by attaching AMP, and it must specifically recognize and bind the cognate tRNA in order to transfer the activated amino acid to the terminal adenosine residue of the tRNA. In the case of prolyl-tRNA synthetase (ProRS) this corresponds to the two reactions (1) and (2),



Subsequently, the anticodon loop of the charged tRNA is matched by the ribosome to a complementary codon on an mRNA, leading to the incorporation of the amino acid that it carries into a growing protein chain. If specificity is lost at any of these three steps, *i.e.* if the wrong amino acid is activated, if a noncognate tRNA is mistakenly charged or if the anticodon is paired with the wrong codon, then the result is the incorporation of an incorrect amino acid into a nascent protein.

Protein synthesis clearly depends on the existence of a well tuned pathway for the accurate activation and incorporation of each amino acid. Therefore, it is quite surprising that the genomes of some archaeal hyperthermophilic methanogens lack a recognizable gene coding for CysRS (Jacquin-Becker *et al.*, 2002; Ruan *et al.*, 2004). The missing essential functionality was at first attributed to the compensatory presence of a dual-specificity Pro/Cys-tRNA synthetase that was imputed to activate Cys and transfer it to tRNA<sup>Cys</sup> as well as to activate Pro and transfer it to tRNA<sup>Pro</sup> (Lipman *et al.*, 2000; Stathopoulos *et al.*, 2000). Furthermore, the ProRS from *G. lamblia* was initially reported to exhibit this same Pro/Cys dual specificity based on the observation that Cys was incorporated into bulk tRNA in the presence of *G. lamblia* ProRS (Bunjun *et al.*, 2000). However, it was later shown that misactivation of Cys is a general property of ProRS homologs from archaea, eukaryotes and some bacteria, but this activity is not accompanied by an ability to recognize or charge tRNA<sup>Cys</sup> (Ahel *et al.*, 2002; Ambrogelly *et al.*, 2002). The previously reported incorporation of Cys into unfractionated tRNA in the presence of ProRS is adequately explained by the formation of misacylated Cys-tRNA<sup>Pro</sup>. Thus, despite its initial annotation as a dual-specificity aaRS, the *G. lamblia* homolog for which the structure is reported here, functions biologically as a typical eukaryotic ProRS, albeit one with a surprisingly high off-target activity in activating Cys.

## 2. Methods

### 2.1. Protein production

The nucleotide sequence of GiardiaDB accession No. GL50803\_15983 (Aurrecoechea *et al.*, 2009) corresponding to protein residues 34–542 of the 542-residue *G. lamblia* ProRS was PCR-amplified from genomic DNA of strain WB and cloned into *Escherichia coli* expression vector BG1861, a derivative of pET14b. The protein was purified by Ni-NTA affinity chromatography followed by size exclusion on an XK 26/60 Superdex 75 column (Amersham Pharmacia Biotech) in MSGPP standard buffer (20 mM HEPES, 0.5 M sodium chloride, 2 mM  $\beta$ -mercaptoethanol, 5% glycerol, 0.025% sodium azide pH 7.5; Mehlin *et al.*, 2006). The purified protein retained a noncleavable eight-residue expression tag at the N-terminus.

**Table 1**

Crystallographic data-processing and model statistics.

Data processing	
Space group	$P2_12_12_1$
Unit-cell parameters (Å)	$a = 83.34, b = 87.62, c = 142.41$
Wavelength (Å)	0.9795
Resolution (Å)	40–2.20 (2.28–2.20)
Unique reflections	52868 (4857)
$R_{\text{merge}}$	0.089 (0.369)
Mean $I/\sigma(I)$	8.3 (2.8)
Completeness (%)	99 (92)
Multiplicity	4.5 (3.5)
Wilson $B$ (Å <sup>2</sup> )	37
Model refinement	
Resolution (Å)	30–2.20
Reflections	52791
Reflections (test set)	2726
$R_{\text{work}}/R_{\text{free}}$	0.172/0.215
No. of atoms in asymmetric unit	
Protein	7924
Ligand	96
Water	215
TLS groups	8 (4 per chain)
Mean $B_{\text{eq}}$ (Å <sup>2</sup> )	35
Root-mean-square non-ideality	
Bond lengths (Å)	0.007
Bond angles (°)	1.070

### 2.2. Differential scanning fluorimetry

The effect of ligand binding on the thermal stability of the protein was assayed by adding substrate or other potential ligands to 0.5 mg ml<sup>-1</sup> protein in MSGPP standard buffer and monitoring fluorescence from the hydrophobic dye SYPRO Orange (Sigma–Aldrich) over a temperature range of 293–363 K. In the presence of a high-affinity ligand a protein will in general be more resistant to thermal denaturation, resulting in a positive shift  $\Delta T_m$  in the inflection point of the melting curve (Lo *et al.*, 2004; Niesen *et al.*, 2007). The differential scanning fluorimetry (DSF) assay was carried out in 96-well trays using a DNA Engine Opticon 2 RT-PCR machine (Bio-Rad). Each condition assayed was represented twice in the tray. Separate transitions contributing to biphasic denaturation curves were modeled by Levenberg–Marquardt curve fitting using the program *gnuplot*.

### 2.3. Crystallization

Purified protein was concentrated to 23 mg ml<sup>-1</sup> and was supplemented with 1 mM TCEP, 10 mM L-proline and 10 mM MgATP. Crystals were grown from sitting drops equilibrated at 277 K against a reservoir consisting of 28% (w/v) PEG 3350, 0.2 M magnesium chloride, 0.1 M Tris–HCl pH 7.8. The initial crystallization drops consisted of 0.2  $\mu$ l protein solution and 0.2  $\mu$ l reservoir solution. The crystals were cryoprotected by adding 1  $\mu$ l of 30% glycerol, 0.175 M magnesium chloride, 24.5% (w/v) PEG 3350, 0.7 mM TCEP, 0.105 M Tris–HCl pH 8.2 prior to cooling in liquid nitrogen. The space group was  $P2_12_12_1$ , with one dimer of ProRS in the asymmetric unit.

### 2.4. X-ray diffraction and structure determination

Diffraction images obtained using the apparatus on SSRL beamline 9-2 tuned to an X-ray energy of 12.658 keV were

integrated and scaled using *HKL-2000* (Otwinowski & Minor, 1997). The *Methanothermobacter thermautotrophicus* ProRS structure (PDB entry 1nj5; Kamtekar *et al.*, 2003) was truncated using the *CCP4* program *CHAINS*AW (Winn *et al.*, 2011; Stein, 2008) in order to serve as a molecular-replacement probe in *Phaser* (McCoy *et al.*, 2007). Unsatisfactory regions in the initial molecular-replacement solution were removed or rebuilt manually in *Coot* (Emsley & Cowtan, 2004) before submitting the model for automatic rebuilding by the *ARP/wARP* server (Cohen *et al.*, 2008). Alternating sessions of manual rebuilding and real-space refinement in *Coot* and automated refinement in *REFMAC* (Murshudov *et al.*, 2011) yielded a final model with crystallographic residuals  $R = 0.172$ ,  $R_{\text{free}} = 0.215$  for data to 2.2 Å resolution. The model for both chains *A* and *B* was complete for residues 38–542, except that the electron density for the  $\beta$ -hairpin formed by residues 426–435 in chain *A* was too indistinct to support a model. Flexibility of the protein chain was modeled using an eight-segment translation/libration/screw description generated by the *TLSMD* server (Painter & Merritt, 2006*a,b*). In the final model, 979 of 997 peptides had  $(\varphi, \psi)$  geometry in the most favored energy regions according to *MolProbity* (Chen *et al.*, 2010). There were no  $(\varphi, \psi)$  outliers. Data and model-quality statistics are given in Table 1. Model superpositions were performed in *Coot* using the *SSM* algorithm (Krissinel & Henrick, 2004). Figures were prepared using *PyMOL* (DeLano, 2002) and *Raster3D* (Merritt & Bacon, 1997). The structure factors and final model have been deposited in the PDB as entry 3ial.

The effective net occupancy for the ligands in the active site was estimated by comparing the ligand-atom *B* values that resulted from parallel refinement of models varying only in the discrete occupancies assigned to the prolyl and adenylyl moieties of the prolyladenylate molecule. Although the strong correlation between crystallographic *B* values and site occupancy makes it difficult to refine these parameters jointly as independent variables, it is possible to refine either quantity after applying an *a priori* constraint to the other. In the present case, we may invoke two *a priori* expectations. First, the *B* values of alternative well ordered ligands in a well ordered binding site should be approximately equal to each other. Second, the *B* values of the ligand atoms should be no lower than the mean *B* value of the surrounding protein atoms, which for the *G. lamblia* ProRS active site is 31 Å<sup>2</sup>. The first expectation is best satisfied for the current data by assigning the prolyl moiety twice the occupancy of the adenylyl moiety. If one assumes that no sites are empty, this corresponds to site occupancies of 1.0 and 0.5, respectively, which results in mean *B* factors of 43 Å<sup>2</sup> for the proline atoms and 43 Å<sup>2</sup> for the adenylyl atoms after refinement. The second expectation sets a lower occupancy limit of 0.70 for the prolyl moiety, as occupancies of <0.7 cause the refined *B* values to be lower than those of the surrounding protein. If we maintain the 2:1 ratio of proline:prolyl-AMP occupancy but allow some fraction of the sites to be empty, the limiting case corresponds to occupancies of 0.70 and 0.35, which results in a mean *B* of 32 Å<sup>2</sup> for the proline atoms and 33 Å<sup>2</sup> for the adenylyl atoms.

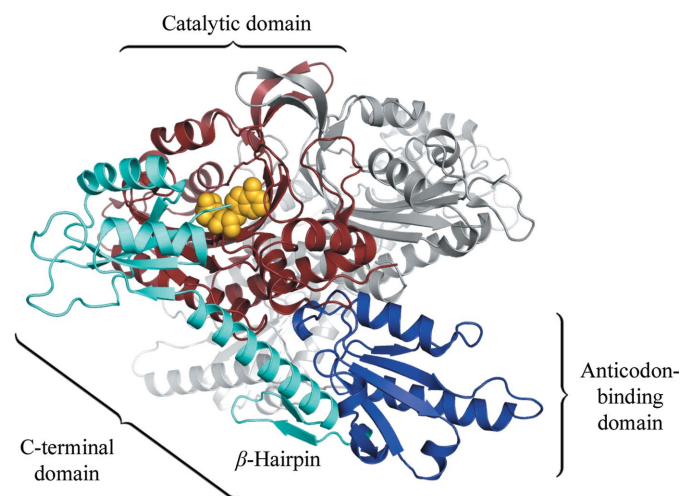
The crystallographic residuals for prolyl occupancy 1.0 are  $R = 0.165$  and  $R_{\text{free}} = 0.212$ ; the residuals for prolyl occupancy 0.7 are  $R = 0.166$  and  $R_{\text{free}} = 0.213$ . This difference is suggestive, but not in itself statistically significant; however, the proline-binding site still contains residual difference density after refinement of the model with a prolyl occupancy of 0.7. After taking everything into consideration, in the final model we assigned the prolyl moiety atoms with full occupancy; *i.e.* there are no empty binding sites. Depending on the exact occupancy assigned to the adenylyl moiety atoms, the crystallographic model is thus consistent with the product prolyl-AMP being present in ~50% of the active sites, with the remainder of the active sites being occupied by unreacted proline.

### 3. Results and discussion

#### 3.1. Overall structure

Class II aaRSs, including ProRS, are characterized by a catalytic domain that is uniquely identified by three conserved sequence motifs and a three-dimensional fold consisting of a core antiparallel  $\beta$ -sheet surrounded by  $\alpha$ -helices. All ProRS homologs contain a second conserved  $\alpha/\beta$  domain responsible for anticodon binding. Most bacterial ProRS homologs contain a third domain responsible for editing mischarged tRNA<sup>Pro</sup>. The editing domain comprises an insertion between motifs 2 and 3 of the canonical class II catalytic domain. Archaeal and eukaryotic ProRS homologs do not contain an editing domain at this position, but generally do contain a third domain of uncertain function at the C-terminus (Fig. 1).

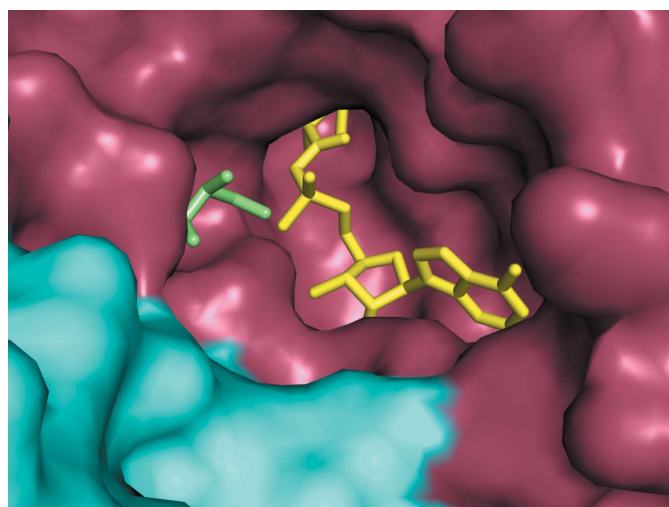
The overall fold of the *G. lamblia* ProRS is very similar to those of the two archaeal homologs *Methanocaldococcus jannaschii* and *M. thermautotrophicus*, for which structures



**Figure 1**  
The *G. lamblia* ProRS dimer. The conserved class II catalytic core domain is shown in rust, the anticodon-binding domain is shown in blue and the C-terminal domain is shown in cyan, including the *Giardia*-specific  $\beta$ -hairpin formed by residues 424–439 where this domain adjoins the anticodon-binding domain. The active site is occupied by a prolyl-AMP molecule (yellow) formed as the product of the activation reaction L-proline + MgATP. The entire second monomer of the dimer is colored gray.

have been determined (Kamtekar *et al.*, 2003). The superposition of individual domains onto the *M. jannaschii* structure yielded an r.m.s.d. of 1.0 Å for 257 C $\alpha$  atoms in the catalytic domain and an r.m.s.d. of 1.1 Å for 97 C $\alpha$  atoms in the anticodon-binding domain. One notable point of divergence is a 16-residue  $\beta$ -hairpin (residues 424–439) at the very start of the C-terminal domain in the *Giardia* ProRS that has no counterpart in previously observed structures. The tip of this hairpin extends outward to reach the anticodon-binding domain of the other monomer in the dimer, contacting the surface of the domain furthest from the anticodon-recognition site.

The secondary structure of the remainder of the C-terminal domain (residues 440–542) is topologically similar to that of the two known archaeal structures and to the archaeal-type homolog from the bacterium *Thermus thermophilus* (Yaremchuk *et al.*, 2000). Of these, only the *M. thermautotrophicus* homolog superimposes well structurally onto the *Giardia* ProRS (r.m.s.d. of 1.4 Å for 95 C $\alpha$  atoms; 19% sequence identity). Furthermore, the C-terminal domain of *Giardia* ProRS retains only one of the four Cys residues that form a zinc-binding site that is conserved across many archaeal and eukaryotic ProRS sequences, including those of *M. thermautotrophicus* ProRS and human cytosolic ProRS. The C-terminal domain wraps around one entire face of the catalytic domain, positioning the C-terminus of the protein so that the terminal Tyr residue reaches into the active site (Figs. 1 and 2). Although the biological function of this domain is not known precisely, deletion of the C-terminal 80 residues from *M. jannaschii* ProRS causes a 25-fold loss in proline-activation activity and a sixfold loss in aminoacylation efficiency (Hati *et al.*, 2006). The observed involvement of the C-terminus in forming the active-site surface makes it plausible that it mediates positioning the acceptor CCA-3' end of the tRNA



**Figure 2**

Active site of the *G. lamblia* ProRS. The surface of the catalytic domain is shown in rust and the surface of the C-terminal domain is shown in cyan. The product of the activation reaction, prolyl-AMP, is shown in yellow. A well ordered glycerol molecule (green) occupies the site expected to accommodate pyrophosphate, which is the other product of the activation reaction.

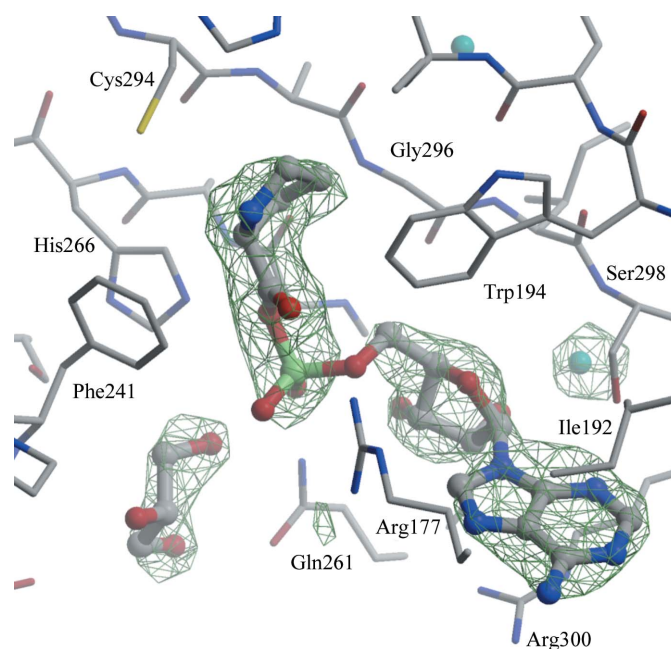
for transfer of the activated proline in the second half reaction (2).

The functional form of ProRS is a homodimer (Fig. 1). In the present crystal structure the two protein chains forming the dimer are crystallographically independent, but there are no significant conformational differences between the two monomers (r.m.s.d. of 0.5 Å for 497 C $\alpha$  atoms after superposition). The dimer interface buries 2290 Å<sup>2</sup> of solvent-accessible surface from each chain as calculated by PISA (Krissinel & Henrick, 2007), which is comparable to the interface seen in the archaeal homologs. The dimer association is particularly intimate in the immediate region of the active site of each monomer, making allosteric interaction between the two active sites structurally plausible.

### 3.2. Active site

The complete active site is well ordered in both of the monomers present in the *Giardia* ProRS crystal structure. There is well defined electron density in both copies of the active site for the reaction product prolyl-AMP, which is formed enzymatically from the L-proline and MgATP present in the crystallization drop. However, the electron density for the adenylyl moiety is weaker than that for the prolyl moiety. We interpret this as evidence of incomplete reaction, such that 30–50% of the active sites in the crystal are occupied by proline alone.

The hydrophobic surface of the proline-binding pocket is formed by the conserved residues Trp194, Glu196, His198, Phe241, Cys294 and Gly296. The proline-ring N atom is suitably positioned to donate a hydrogen bond to Thr146 O $\gamma$  or Glu148 O $\epsilon^2$ , which are also highly conserved residues. As



**Figure 3**

Difference electron density in the active site of chain A. The isodensity contours are drawn at +3 $\sigma$  in an ( $mF_{\text{obs}} - F_{\text{calc}}$ ) difference Fourier map after refinement of the protein model without inclusion of ligand atoms. The refined models for prolyl-AMP and glycerol are superimposed.

noted for the *M. thermautotrophicus* ProRS by Kamtekar *et al.* (2003), the proline-binding pocket can sterically accommodate cysteine with essentially no rearrangement of the protein. Note that this would orient the S atom of the free cysteine to point away from the S atom of Cys294. However, despite the known ability of the *Giardia* ProRS to activate cysteine, our attempts at cocrystallization failed to yield any crystals that showed evidence of bound cysteine, either with or without MgATP.

The phosphate group of prolyl-AMP is coordinated by water-mediated hydrogen bonding to the side chain of Arg177, which is strongly conserved in ProRS sequences. The likely position of the pyrophosphate leaving group immediately after reaction (1) is also evident from consistent electron density in both copies of the active site (Figs. 2 and 3). However, in the present crystal structure this density is better fitted by a glycerol molecule, which presumably displaced pyrophosphate during cryoprotection of the crystal. The protein side chains coordinating to the glycerol are Arg235 and Gln261. These residues are not part of the conserved class II aaRS core catalytic domain motifs, but they are conserved among archeal-type ProRS sequences and may be inferred to coordinate the  $\beta$ -phosphate and  $\gamma$ -phosphate of the ATP prior to reaction. The imidazole ring of His266 is also suitably placed to coordinate the  $\beta$ -phosphate.

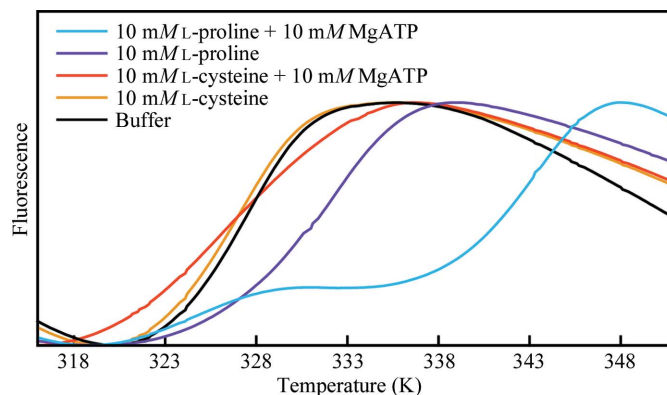
One might expect that the highly conserved functions carried out at the active site of individual aaRSs would lead to strong structural similarity across species. However, the cross-species variation between homologs can be substantial, offering an opportunity for the identification of selective inhibitors that may be developed into drugs targeting specific pathogens (Hurdle *et al.*, 2005). For example, Yu and coworkers reported the synthesis of a series of ATP-competitive quinoline derivatives that selectively inhibit *Candida albicans* ProRS relative to human ProRS (Yu *et al.*, 2001). From this perspective, it is encouraging that *Giardia* ProRS has less than 30% sequence identity overall to either of the two human ProRS orthologs (one cytosolic and one mitochondrial). Furthermore, of the 23 residues that line the active site in the *G. lamblia* ProRS structure, 19 differ from the homologous residue in at least one of the human enzymes, while 11 differ from the equivalent residue in both human enzymes. One notable point of divergence is residue Ile192, the side chain of which forms the hydrophobic surface accommodating one face of the ATP adenine (Fig. 3). Both human ProRS orthologs have a phenylalanine residue at this position, as does *T. thermophilus* ProRS. Comparison of the *G. lamblia* and *T. thermophilus* ProRS crystal structures (Yaremchuk *et al.*, 2001) confirms that the adenine-binding pocket in the *Giardia* homolog therefore provides additional ligand-accessible volume that might be exploitable to introduce inhibitor specificity.

### 3.3. Biological implications of $\Delta T_m$ assays

The relative affinity of chemically similar compounds for a binding site can be conveniently determined by comparing

their respective effects on the thermal denaturation curve of the protein. Compounds with higher affinity cause a greater shift  $\Delta T_m$  in the inflection point of the denaturation curve ( $T_m$ ) relative to that of protein with no ligand present (Niesen *et al.*, 2007). We assayed the effect of L-cysteine, L-proline, MgAMP, MgADP and MgATP alone or in combination on the thermal denaturation of *G. lamblia* ProRS. The addition of 10 mM L-proline alone produced a shift  $\Delta T_m = +5$  K, the addition of 10 mM MgATP alone produced a shift  $\Delta T_m = +3$  K and the addition of both 10 mM L-proline and 10 mM MgATP produced a shift  $\Delta T_m = +16$  K (Fig. 4). The very large shift in the presence of both substrates is a strong indication that the activation reaction has occurred and that the product prolyl-AMP has significantly greater affinity than either substrate alone. Note, however, that the curve observed in the presence of L-proline and MgATP is biphasic (Fig. 4). If the curve is modeled as arising from two transitions, the first corresponds to a  $T_m$  at 327 K ( $\Delta T_m = 0$  K) and the second to a  $T_m$  at 344 K ( $\Delta T_m = +17$  K). The partial contributions of the two transitions to the observed net denaturation curve are 40 and 60%, respectively. This implies that 40% of the monomers in solution either contain an empty binding site or have been destabilized by some effect that counteracts the stabilization of a bound ligand. A possible explanation for this is allosteric anticooperativity (half-of-the-sites reactivity) between the two monomers that make up each dimer, as discussed below.

ProRS in general, and *G. lamblia* ProRS in particular, has been shown to activate L-cysteine at low levels even in the absence of tRNA. This led to the initial misannotation of the *Giardia* enzyme as a dual-function Pro/Cys aaRS. Based on measured  $k_{cat}/K_m$  values, the relative Cys:Pro activation ratio *in vitro* ranges from 1:1400 for yeast to 1:30 for the archaea *M. jannaschii* and *M. thermautotrophicus* and the eukaryote



**Figure 4** Stabilization of *G. lamblia* ProRS as measured by differential scanning fluorimetry. Increasing fluorescence indicates increasing exposure of hydrophobic surface area as the protein unfolds. Each curve is the mean of two replicate experiments. Addition of L-cysteine alone does not stabilize the protein. The denaturation curve after the addition of L-cysteine and MgATP is essentially identical to the curve for MgATP alone (not shown). We interpret the two-phase curve for L-proline and MgATP as arising from incomplete reaction of the two added substrates; the smaller transition near 328 K arises from a subset of molecules where the reaction has not occurred and thus no product is bound, while the larger transition at 343 K arises from stabilization by the reaction product prolyl-AMP.

*G. lamblia* (Ahel *et al.*, 2002). Cys activation increases slightly in the presence of unfractionated tRNA, but occurs even in the absence of tRNA (Lipman *et al.*, 2002; Jacquin-Becker *et al.*, 2002). However, we see no evidence for thermal stabilization of the protein in the presence of either cysteine alone or cysteine plus MgATP (Fig. 4). The implied very low affinity of the enzyme for either cysteine or cysteinyl-AMP is consistent with our lack of success in growing crystals with either species present in the active site. It is also consistent with a passive biological mechanism for reducing the misincorporation of cysteine instead of proline during protein synthesis, as any cysteinyl-AMP that is formed may be released quickly rather than remaining in the ProRS active site to undergo a second reaction that would misacylate tRNA<sup>Pro</sup> and would thus eventually lead to the misincorporation of Cys residues into a protein.

### 3.4. Half-of-the-sites reactivity

Although the class II aaRSs are symmetric homodimers with two equivalent active sites per dimer, at least some of them can exhibit a negative cooperativity known as half-of-the-sites reactivity in which only one active site of the dimer is functionally active at a time. This phenomenon was first characterized for the class Ic TyrRS (Jakes & Fersht, 1975), which together with TrpRS is a homodimeric exception to the generally monomeric class I aaRSs. Cooperative or anti-cooperative intermonomer allostery has subsequently been reported for the class IIb enzymes AspRS (Kern *et al.*, 1985) and LysRS (Hughes *et al.*, 2003) and for the class IIa enzymes GlyRS (Freist *et al.*, 1996) and ProRS (Ambrogelly *et al.*, 2005). The implication in each case is that activation of the amino-acid substrate in the active site of one monomer in the dimer somehow perturbs the state of the active site in the other monomer so that it becomes either more competent or less competent to carry out the same reaction. The structural mechanism of this allostery remains a great mystery, as virtually all crystal structures to date of aaRSs with half-of-the-sites reactivity exhibit a structurally symmetric dimer in the crystal in spite of the presence of bound activation reaction substrates, products or analogs. The sole exception is a crystal structure of the unusual double-length TyrRS from *Leishmania major*, which is intrinsically asymmetric (Larson *et al.*, 2011).

ProRS from archaea and eukaryotes is inferred to exhibit half-of-the-sites reactivity on the basis of experiments showing that in the presence of saturating amounts of ATP and <sup>14</sup>N-labeled proline the specific activity retained by *M. jannaschii* ProRS after washing corresponds to only one labeled atom per dimer (Ambrogelly *et al.*, 2005). This negative cooperativity provides a qualitative explanation for the biphasic denaturation curve we observe by DSF for *G. lamblia* ProRS in the presence of ATP and proline (Fig. 4), although the relative ratio of inactive:active sites inferred from fitting a two-transition model to the DSF curve in the present case is 40%/60% rather than the 50%/50% that would correspond to perfect half-of-the-sites binding. The high-temperature tran-

sition component of the biphasic curve corresponds to denaturation of those sites that are stabilized by the presence of the high-affinity reaction product prolyl-adenylate. The  $T_m$  of the low-temperature transition component in the biphasic curve cannot be distinguished from the  $T_m$  of the apo protein, but this does not necessarily imply that it corresponds to the denaturation of sites that are identical to the apo state. It is also possible that the  $\Delta T_m = 0$  K observed for the low-temperature transition is the result of compensatory stabilization from bound proline in one monomer and allosteric destabilization from the presence of the reaction product in the other monomer. That is, the observed denaturation curve is compatible with two possible explanations. The second active site in each dimer could be entirely empty or it could be occupied by an unreacted proline molecule whose binding affinity has been reduced by an allosterically induced change to the local environment.

Evidence for incomplete formation of the reaction product is also provided by the crystal structure, as the electron density corresponding to the prolyl moiety of prolyl-AMP is clearly greater than the electron density corresponding to the adenylyl moiety. The crystallographic experiment yields only an imprecise value for the fraction of monomers with the reaction product in the active site, but gives a strong indication that either proline or prolyl-AMP is present in no less than 70% of the monomers in the crystal. Under the assumption that the mean  $B$  values refined for the two possible ligands proline or prolyl-AMP should be roughly equal, the crystallographic data are consistent with a 1:1 ratio of unreacted proline and prolyl-AMP. However, any structural asymmetry induced by activation of proline in one monomer is apparently insufficient to favor one orientation of the dimer in preference to the other within the crystal lattice. Thus, the two crystallographically independent monomers are each observed as a mixture of the two states: reacted and unreacted.

### 3.5. Correcting for misactivated cysteine

Because ProRS misactivates amino acids other than proline at a significant rate, some mechanism of error correction is necessary to assure translational fidelity. Eubacterial ProRS contains a separate editing domain that acts in *cis* to correct misrecognition of glycine and alanine. Many prokaryotes also contain a separate Cys-tRNA<sup>Pro</sup> deacylase that is capable of post-transfer editing in *trans* to remove the amino acid from a mischarged tRNA resulting from improperly recognized cysteine (An & Musier-Forsyth, 2004). Eukaryotes in general lack both of these mechanisms, relying at least in part on a three-way kinetic balance at the ProRS active site between the forward and reverse activation reactions and the nonproductive release of misactivated amino acids. The relative importance of these various editing mechanisms is species-specific and indeed may constitute one path of approach to the development of selective inhibitors of correct proline incorporation during protein syntheses (Ahel *et al.*, 2002; Splan *et al.*, 2008). In *M. jannaschii* the correction of misactivated Ala-AMP by ProRS arises primarily from catalysis of the reverse

reaction and secondarily from nonproductive release of Ala-AMP (Splan *et al.*, 2008). An equivalent quantitative analysis of misactivated cysteine error correction has not been reported, but the low affinity for Cys-AMP implied by the DSF results presented here suggests that nonproductive release of misactivated cysteine is important for translational fidelity in *Giardia*.

This work was supported by National Institutes of Health award P01AI067921 (MSGPP). Portions of this work were carried out at the Stanford Synchrotron Radiation Light-source, a national user facility operated by Stanford University on behalf of the US Department of Energy, Office of Basic Energy Sciences.

## References

- Ahel, I., Stathopoulos, C., Ambrogelly, A., Sauerwald, A., Toogood, H., Hartsch, T. & Söll, D. (2002). *J. Biol. Chem.* **277**, 34743–34748.
- Ambrogelly, A., Ahel, I., Polycarpo, C., Bunjun-Srihari, S., Krett, B., Jacquin-Becker, C., Ruan, B., Köhrer, C., Stathopoulos, C., RajBhandary, U. L. & Söll, D. (2002). *J. Biol. Chem.* **277**, 34749–34754.
- Ambrogelly, A., Kamtekar, S., Stathopoulos, C., Kennedy, D. & Söll, D. (2005). *FEBS Lett.* **579**, 6017–6022.
- An, S. & Musier-Forsyth, K. (2004). *J. Biol. Chem.* **279**, 42359–42362.
- Aurrecochea, C. *et al.* (2009). *Nucleic Acids Res.* **37**, D526–D530.
- Bunjun, S., Stathopoulos, C., Graham, D., Min, B., Kitabatake, M., Wang, A. L., Wang, C. C., Vivarès, C. P., Weiss, L. M. & Söll, D. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 12997–13002.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* **D66**, 12–21.
- Cohen, S. X., Ben Jelloul, M., Long, F., Vagin, A., Knipscheer, P., Lebbink, J., Sixma, T. K., Lamzin, V. S., Murshudov, G. N. & Perrakis, A. (2008). *Acta Cryst.* **D64**, 49–60.
- DeLano, W. (2002). *PyMOL*. <http://www.pymol.org>.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Freist, W., Logan, D. T. & Gauss, D. H. (1996). *Biol. Chem. Hoppe Seyler*, **377**, 343–356.
- Hati, S., Ziervogel, B., Sternjohn, J., Wong, F.-C., Nagan, M. C., Rosen, A. E., Siliciano, P. G., Chihade, J. W. & Musier-Forsyth, K. (2006). *J. Biol. Chem.* **281**, 27862–27872.
- Hughes, S. J., Tanner, J. A., Hindley, A. D., Miller, A. D. & Gould, I. R. (2003). *BMC Struct. Biol.* **3**, 5.
- Hurdle, J. G., O'Neill, A. J. & Chopra, I. (2005). *Antimicrob. Agents Chemother.* **49**, 4821–4833.
- Jacquin-Becker, C., Ahel, I., Ambrogelly, A., Ruan, B., Söll, D. & Stathopoulos, C. (2002). *FEBS Lett.* **514**, 34–36.
- Jakes, R. & Fersht, A. R. (1975). *Biochemistry*, **14**, 3344–3350.
- Kamtekar, S., Kennedy, W. D., Wang, J., Stathopoulos, C., Söll, D. & Steitz, T. A. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 1673–1678.
- Kern, D., Lorber, B., Boulanger, Y. & Giegé, R. (1985). *Biochemistry*, **24**, 1321–1332.
- Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* **D60**, 2256–2268.
- Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
- Larson, E. T., Kim, J. E., Castaneda, L. J., Napuli, A. J., Zhang, Z., Fan, E., Zucker, F. H., Verlinde, C. L. M. J., Buckner, F. S., Van Voorhis, W. C., Hol, W. G. J. & Merritt, E. A. (2011). *J. Mol. Biol.* **409**, 159–176.
- Lipman, R. S. A., Beuning, P. J., Musier-Forsyth, K. & Hou, Y.-M. (2002). *J. Mol. Biol.* **316**, 421–427.
- Lipman, R. S. A., Sowers, K. R. & Hou, Y.-M. (2000). *Biochemistry*, **39**, 7792–7798.
- Lo, M.-C., Aulabaugh, A., Jin, G., Cowling, R., Bard, J., Malamas, M. & Ellestad, G. (2004). *Anal. Biochem.* **332**, 153–159.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Mehlin, C. *et al.* (2006). *Mol. Biochem. Parasitol.* **148**, 144–160.
- Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.
- Niesen, F. H., Berglund, H. & Vedadi, M. (2007). *Nature Protoc.* **2**, 2212–2221.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Painter, J. & Merritt, E. A. (2006a). *Acta Cryst.* **D62**, 439–450.
- Painter, J. & Merritt, E. A. (2006b). *J. Appl. Cryst.* **39**, 109–111.
- Ruan, B., Nakano, H., Tanaka, M., Mills, J. A., DeVito, J. A., Min, B., Low, K. B., Battista, J. R. & Söll, D. (2004). *J. Bacteriol.* **186**, 8–14.
- Splan, K. E., Ignatov, M. E. & Musier-Forsyth, K. (2008). *J. Biol. Chem.* **283**, 7128–7134.
- Stathopoulos, C., Li, T., Longman, R., Vothknecht, U. C., Becker, H. D., Ibba, M. & Söll, D. (2000). *Science*, **287**, 479–482.
- Stein, N. (2008). *J. Appl. Cryst.* **41**, 641–643.
- Van Voorhis, W. C., Hol, W. G. J., Myler, P. J. & Stewart, L. J. (2009). *PLoS Comput. Biol.* **5**, e1000530.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- Yaremchuk, A., Cusack, S. & Tukalo, M. (2000). *EMBO J.* **19**, 4745–4758.
- Yaremchuk, A., Tukalo, M., Grøtli, M. & Cusack, S. (2001). *J. Mol. Biol.* **309**, 989–1002.
- Yu, X. Y., Hill, J. M., Yu, G., Yang, Y., Kluge, A. F., Keith, D., Finn, J., Gallant, P., Silverman, J. & Lim, A. (2001). *Bioorg. Med. Chem. Lett.* **11**, 541–544.