Published in final edited form as:

Mol Biochem Parasitol. 2011 May ; 177(1): 20–28. doi:10.1016/j.molbiopara.2011.01.003.

Crystal structures of three protozoan homologs of tryptophanyl-tRNA synthetase

Ethan A Merritt^{a,c}, Tracy L Arakaki^{a,c}, Robert Gillespie^{b,c}, Alberto J Napuli^{b,c}, Jessica E Kim^{a,c}, Frederick S Buckner^{b,c}, Wesley C Van Voorhis^{b,c}, Christophe L M J Verlinde^{a,c}, Erkang Fan^{a,c}, Frank Zucker^{a,c}, and Wim G J Hol^{a,c}

^aDepartment of Biochemistry, University of Washington, Seattle, WA 98195 USA

^bDepartment of Medicine, University of Washington, Seattle, WA 98195 USA

^cMedical Structural Genomics of Pathogenic Protozoa, <http://msgpp.org>

Abstract

Tryptophanyl-tRNA synthetase (TrpRS) is an essential enzyme that is recognizably conserved across all forms of life. It is responsible for activating and attaching tryptophan to a cognate tRNA^{Trp} molecule for use in protein synthesis. In some eukaryotes this original core function has been supplemented or modified through the addition of extra domains or the expression of variant TrpRS isoforms. The three TrpRS structures from pathogenic protozoa described here represent three illustrations of this malleability in eukaryotes. The *Cryptosporidium parvum* genome contains a single TrpRS gene, which codes for an N-terminal domain of uncertain function in addition to the conserved core TrpRS domains. Sequence analysis indicates that this extra domain, conserved among several apicomplexans, is related to the editing domain of some AlaRS and ThrRS. The *C. parvum* enzyme remains fully active in charging tRNA^{Trp} after truncation of this extra domain. The crystal structure of the active, truncated enzyme is presented here at 2.4 Å resolution. The *Trypanosoma brucei* genome contains separate cytosolic and mitochondrial isoforms of TrpRS that have diverged in their respective tRNA recognition domains. The crystal structure of the *T. brucei* cytosolic isoform is presented here at 2.8 Å resolution. The *Entamoeba histolytica* genome contains three sequences that appear to be TrpRS homologs. However one of these, whose structure is presented here at 3.0 Å resolution, has lost the active site motifs characteristic of the Class I aminoacyl-tRNA synthetase catalytic domain while retaining the conserved features of a fully formed tRNA^{Trp} recognition domain. The biological function of this variant *E. histolytica* TrpRS remains unknown, but, on the basis of a completely conserved tRNA recognition region and evidence for ATP but not tryptophan binding, it is tempting to speculate that it may perform an editing function. Together with a previously reported structure of an unusual TrpRS from *Giardia*, these protozoan structures broaden our perspective on the extent of structural variation found in eukaryotic TrpRS homologs.

Keywords

aminoacyl-tRNA synthetase

© 2011 Elsevier B.V. All rights reserved.

Correspondence to: Ethan A Merritt.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

The protozoans *Cryptosporidium parvum*, *Trypanosoma brucei*, and *Entamoeba histolytica* are human pathogens responsible for significant public health burden. These organisms are eukaryotes, but they are phylogenetically distant from higher eukaryotes such as their mammalian hosts. There can be substantial differences in the structure and detailed biological role of homologous proteins from protozoans and from humans. The three structures of protozoan tryptophanyl-tRNA synthetase (TrpRS) homologs we describe here were determined as part of a larger effort undertaken by the Medical Structural Genomics of Pathogenic Protozoa (MSGPP) collaboration [1]. We have previously reported the structure of the unique TrpRS from *Giardia lamblia*, which unexpectedly was observed to form a homotetramer [2]. The overall MSGPP goal is to identify and characterize proteins from these disease-causing organisms that may constitute targets for the design and development of new anti-parasitic drugs.

TrpRS is a class I aminoacyl-tRNA synthetase (aaRS). All members of this class are characterized by a Rossmann-fold catalytic domain whose active site is recognizable by the presence of two conserved sequence motifs with consensus sequences HIGH and KMSKS [3,4] (Figure 1, red boxes). The active site of TrpRS sequences exhibits a third, more weakly conserved, motif (AIDQ in the human sequence) involved in binding ATP [5]. The catalytic domain carries out the two half-reactions necessary to activate and attach L-tryptophan to its cognate tRNA:

Cognate tRNA recognition is carried out by a second, α -helical, domain whose three-dimensional fold is also found in TyrRS. Both TrpRS and TyrRS are homodimeric. Both employ a tRNA binding mode in which the cognate tRNA anticodon arm is recognized and bound by one monomer within the dimer, positioning the acceptor stem of that same tRNA molecule in the active site of the other monomer. Two cognate tRNA molecules can bind simultaneously to one dimer [6,7,8]. Crystal structures of human cytosolic TrpRS in complex with the cognate tRNA^{Trp} (PDB ID: 2ake, 2dr2; [7]) provide a model for TrpRS:tRNA^{Trp} interaction by eukaryotic homologs including the three protozoan TrpRS examined here. Residues involved in the recognition of tRNA^{Trp} are shown in Figure 1 (blue boxes), including a strictly conserved lysine residue that interacts directly with the central tRNA anticodon base. In some species TrpRS has additional idiosyncratic N-terminal domains whose biological function is not directly related to the primary catalytic activity [9,10].

The crystal structures described here correspond to the conserved catalytic and anticodon-binding domains of TrpRS homologs from the pathogens *C. parvum*, *T. brucei*, and *E. histolytica*, determined to 2.42 Å, 2.8 Å, and 3.0 Å resolution, respectively. The *C. parvum* and *T. brucei* enzymes are essential to their respective organisms. Unfortunately, structural comparison of their active sites to that of human cytosolic TrpRS does not reveal obvious opportunities for the design of selective inhibitors that might serve as lead compounds for drug design. While the *G. lamblia* TrpRS whose structure we reported previously differs from the human cytosolic TrpRS in the structural mechanism for tryptophan activation, we find that the stabilization of the *C. parvum* and *T. brucei* enzymes by substrates and substrate analogs is consistent with the structural mechanism used by the human enzyme. Interestingly, for each of the three protozoan structures that we report here there is direct or indirect evidence for the addition of a novel biological function that supplements, modifies,

or replaces the primary catalytic activity. Further investigation of these differences between the parasite and host enzymes may eventually illuminate other opportunities for the development of anti-parasitic drugs.

2. Methods

2.1. Target selection and expression

Following standard MSGPP procedures, we designed multiple expression constructs by trimming target genomic sequences at one or both ends, guided by consensus domain boundaries that we identified through multiple sequence alignment [11]. For each of the three protozoan TrpRS targets we obtained crystallizable protein from an N-terminal truncation of the full-length sequence (Figure 1).

Attempts to express soluble protein from the full-length genomic sequence cgd7_1490 (GenBank accession no. **EAK9023**) of the sole TrpRS gene in the *C. parvum* genome did not succeed. However soluble protein was expressed from a construct corresponding to residues 206–593 cloned from *C. parvum* (ATCC Iowa strain) into *E. coli* expression vector AVA421 [12]. The AVA421 vector introduces a cleavable N-terminal His-tag. The protein was purified by Ni-NTA chromatography and the bound protein was cleaved by protease 3C overnight at 4°C, leaving five tag residues (GPGSM) attached N-terminal to residue 206 of the genomic sequence. The released protein was further purified by gel filtration on a HiLoad Superdex 200 size exclusion column.

Constructs based on *T. brucei* genomic sequence Tb927.3.5580 (GenBank accession no. **AAZ10589**) were PCR amplified from genomic DNA of *T. brucei* strain TREU927 GUTat 10.1. This gene encodes the cytosolic isoform of TrpRS [13]. A fragment consisting of a non-cleavable N-terminal tag MAHHHHHH followed by residues 3–388 of the full-length sequence was expressed using the BG1861 expression vector, a modified form of pET14b. The protein was purified by Ni-NTA chromatography followed by filtration on a Superdex 75 column.

A fragment of sequence 58.m00157 (GenBank accession no. **EAL48880**) corresponding to residues 4–381 was cloned from genomic DNA from *E. histolytica* strain HM-1:IMSS (courtesy of Bill Petri, University of Virginia) into the BG1861 expression vector. The expression construct included a non-cleavable N-terminal tag MAHHHHHH. Protein was expressed in *E. coli* strain BL21(DE3). The protein was purified by Ni-NTA chromatography followed by filtration on a Superdex 75 column.

2.2. Protein crystallization and data collection

Crystals for data collection were grown by vapor diffusion at room temperature from sitting drops placed in 96-well trays by a PHOENIX liquid handling robot (Art Robbins Instruments).

C. parvum TrpRS was dissolved to 32.1 mg/ml in standard buffer (0.5 M NaCl, 25 mM HEPES pH 7.5, 5% glycerol, 0.025% NaAzide). Cocrystallants were added to yield final concentrations of 10 mM ATP, 1 mM TCEP, 10 mM MgCl₂, 2 mM L-tryptophan, 24.7 mg/ml protein. The reservoir solution contained 0.2 M MgCl₂, 0.1 M bis-tris buffer at pH 6.2, 28% w/v PEG 3350. Crystals grew from drops consisting of 0.15 µL protein solution, 0.15 µL reservoir solution. Crystals were obtained in two different space groups, *P*2₁ and *P*6₅. Both forms contained one dimer of *C. parvum* TrpRS in the asymmetric unit, with clear electron density for a tryptophan molecule present in the active site. However, the hexagonal crystal form diffracted only to 3.2 Å resolution and its structure is not further described here. Diffraction data to 2.4 Å resolution were measured from a single crystal of the monoclinic

form at SSRL beamline 11-1 and processed using Mosflm [14], Scala [15], and Truncate [16].

T. brucei TrpRS was dissolved to 17.8mg/ml in standard buffer plus 1mMTCEP (tris(2-carboxyethyl)phosphine). The reservoir solution contained 0.2 M Na malonate at pH 7.0, 20% w/v PEG 3350. Triclinic crystals grew from drops consisting of 0.1 μ L protein solution and 0.2 μ L reservoir solution. Diffraction data to 2.8 Å were measured from a single crystal using a rotating anode source with a Saturn CCD detector (Rigaku America) and processed using D*trek [17]. The crystal asymmetric unit contained one dimer.

E. histolytica TrpRS was dissolved to 13.2 mg/ml in standard buffer. The reservoir solution contained 0.2 M Na citrate, 20% w/v PEG 3350. Crystals in space group $P2_1$ grew at room temperature from drops consisting of 0.1 μ L protein solution and 0.2 μ L reservoir solution. Diffraction data were measured from a single crystal on beamline 9-2 of the Stanford Synchrotron Radiation Lightsource, processed using Mosflm, and processed to 3.0 Å resolution using Scala and Truncate. The crystal asymmetric unit contained three dimers, corresponding to a total molecular weight of 258 kDa.

2.3. Structure determination

Initial phases for all three structures were generated by molecular replacement using the program Phaser [18]. The probe structure was based on one monomer of the human cytosolic TrpRS [19,7]. In the case of the *E. histolytica* TrpRS, only 5 of the 6 monomers in the asymmetric unit were found correctly by the program; the sixth monomer was then placed manually by superimposing a copy of one of the two complete dimers onto the remaining un-paired monomer. The models were iteratively improved by manual rebuilding in Coot [20] and automated refinement in Refmac [21]. Strong non-crystallographic symmetry restraints were applied to the *E. histolytica* and *T. brucei* TrpRS models. Atomic displacements in the *E. histolytica* model were described purely by TLS (no individual B_{iso} terms). Atomic displacements in the other two structures were modeled by a combination of TLS and B_{iso} contributions. TLS groups were chosen on the basis of TLSMD analysis [22]. The MolProbity server [23] was used to guide optimization of side chain conformations and to validate the refined $[\phi, \psi]$ backbone torsion angles. Data quality and refinement statistics are shown in Table 1.

2.4. Enzyme activity assays

Enzyme activity was quantified by measuring the attachment of ^3H -labeled tryptophan to tRNA in the presence of TrpRS. Reactions were carried out in a 96 well plate with a total reaction volume per well of 75 μ L. The initial buffer consisted of 50 mM Hepes pH 7.0, 10 mM MgCl_2 , 10 mM KCl, 2.5 mM DTT. Reactants were added to a final concentration of 1 mg/ml bulk *E. coli* tRNA (Sigma #R4251), 1 mM ATP, 0.5 μ M unlabeled L-tryptophan, 40 μ Ci/ml ^3H -labeled L-tryptophan, 30 μ g/ml TrpRS, 10 U/ml pyrophosphatase (Sigma #I1643). Paired control experiments used additional unlabeled L-tryptophan at 100 μ M as a competitive inhibitor. The reaction was carried out for 90 minutes at room temperature before quenching with cold stop buffer (100 μ L 10% Trichloroacetic acid). Reaction components were separated from tRNA by filtration through a Durapore® membrane with low protein binding (Millipore #MSHVN4B10) and washed 3 $\times$ with stop buffer. Scintillation fluid was added after drying overnight, and radiation was quantified using a scintillation counter.

2.5. Ligand binding assays

Ligand binding was assayed by measuring protein thermal denaturation in solution in the presence or absence of the putative ligand. Denaturation was tracked by measuring

fluorescence from a hydrophobic dye, Sypro Orange (SigmaAldrich) as a function of increasing temperature. The dye is weakly fluorescent in aqueous solution surrounding a well-folded protein but gives increased fluorescent signal as a consequence of binding to hydrophobic patches that are exposed as the protein releases a bound ligand and eventually denatures. In the presence of a high-affinity ligand, the protein will in general be more resistant to thermal denaturation, resulting in a shift ΔT_M in the inflection point of the denaturation curve [24]. The effect of ligand binding on the thermal denaturation of TrpRS homologs was assayed by adding substrate or other potential ligands to 0.5 mg/ml protein in standard buffer and monitoring fluorescence over a temperature range of 20–90° C. The assay was carried out in 96-well trays in a DNA Engine Opticon 2 RT-PCR machine (BioRad). Thermal denaturation curves were measured for the *C. parvum*, *T. brucei*, and *E. histolytica* TrpRS constructs used for structural studies, and also for the full-length TrpRS from *G. lamblia* (GenBank accession no. **EAA40921**) [2].

3. Results and Discussion

3.1. *Cryptosporidium parvum* TrpRS

Although the *C. parvum* TrpRS crystals used for data collection were grown in the presence of tryptophan and ATP, the active site of both monomers shows evidence for only the presence of tryptophan as judged by electron density maps of the refined structure (Figure 2). This was somewhat unexpected, as the *C. parvum* protein is able to activate tryptophan and charge tRNA^{Trp} in solution (Figure 3) and we had anticipated finding the tryptophanyl-AMP half-reaction product rather than unreacted tryptophan.

Bacterial TrpRS is known to bind ATP via an induced-fit mechanism that involves reordering of the KMSKS loop region [25,26]. On the other hand, human cytosolic TrpRS, a closer homolog to the protozoan enzymes, exhibits a more closed conformation of the active site in which the KMSAS loop can be ordered even in the absence of ATP [5]. In the current *C. parvum* crystal structure, residues 460–472, which include the ATP-binding motif KMSAS, are disordered. This is consistent with the absence of ATP or tryptophanyl-AMP in the binding site of both crystallographically distinct subunits. It is also consistent with thermal stability measurements that show stability is increased by the addition of tryptophan, but not by the addition of ATP (Figure 4a). A plausible explanation for these observations is that ATP binding and tryptophan activation are dependent on the presence of tRNA. As discussed below, this dependence is probably an artifact of the truncated expression product, which alters a functionally relevant β -hairpin at the N-terminus of the catalytic domain.

The biological function of the excised *C. parvum* TrpRS N-terminal domain is not known, but we find that it is not required for activity (Figure 3). Using the N-terminal 180 residues as a BLAST probe against the NCBI non-redundant protein sequence data set yields hits against the TrpRS N-terminus from other apicomplexans (*Toxoplasma gondii*, *Babesia bovis*, *Theileria annulata*) and from a related alveolate (*Perkinsus marinus*). Sequence similarity to plasmodial sequences is weaker. Intriguingly, the greatest similarity to a protein sequence other than another apicomplexan TrpRS is to the editing domain from AlaRS of the hyperthermophilic archaeon *Pyrococcus horikoshii*. Weak sequence similarity of the plasmodial TrpRS N-terminus and the editing domain of AlaRS has been noted previously [27]. This relationship is strongly supported by an evaluation of possible three-dimensional folds assumed by the N-terminal sequence domain. We threaded the sequence of the *C. parvum* TrpRS N-terminal domain into known protein structures using the Phyre algorithm [28]. This search assigns the sequence to the fold-family of the ThrRS/AlaRS editing domain at a significance level of $1.4e^{-19}$, which strongly suggests there is indeed a shared fold despite pairwise sequence identity of only 7%–20% against existing representative structures with this fold in the PDB.

3.2. *Trypanosoma brucei* TrpRS

Trypanosomatids, including *T. brucei*, express a single form of most aminoacyl-tRNA synthetases. However, they have independently evolved a separate isoform of TrpRS used in the mitochondrion, possibly by gene duplication of the two core domains of the cytosolic homolog followed by divergence of the tRNA recognition domain and the addition of an N-terminal mitochondrial targeting sequence [29]. The cytosolic and mitochondrial isoforms are 44% sequence identical. The structure presented here is that of the cytosolic member of the pair. These organisms require a separate mitochondrial TrpRS isoform because the majority of tRNA^{Trp} in the trypanosomatid mitochondrion is post-transcriptionally modified by thiolation of U33 and by RNA-editing of the anticodon (residues 34–36) from CCA to UCA [30,31,32]. These modifications allow recognition of the non-standard Trp codon UGA used in trypanosomatid mitochondrial genes. The cytosolic TrpRS isoform is not capable of charging this modified tRNA^{Trp} [29]. The essentiality of the separate mitochondrial form has been confirmed by RNA interference [33]. Neither isoform is capable of charging *E. coli* tRNA^{Trp} [29].

From the perspective of identifying possible drug targets in trypanosomes, it would be of interest to understand the structural basis for the altered tRNA recognition by the *T. brucei* TrpRS isoforms that makes them specific for the modified or unmodified trypanosomal tRNA^{Trp}. Unfortunately a key portion of the protein expected to interact extensively with the tRNA anticodon arm (first blue box in Figure 1) is disordered in both monomers of the current structure, although the helical region containing residues directly involved in binding the anticodon bases (second blue box in Figure 1) is well ordered and superimposes closely on the corresponding helical region of the human and yeast TrpRS:tRNA^{Trp} complexes [7,34]. In the human complex, TrpRS sidechains Lys 431, Ser 378, and Thr 427 make direct interactions with the bases of the anticodon. The lysine residue, whose N^ε interacts with both O2 and O2' of the middle anticodon base C35, is strictly conserved in TrpRS including both trypanosomal isoforms. The serine residue equivalent to human Ser 378 is also conserved in all the homologs considered here. The residue equivalent to human Thr 427, whose O^γ accepts a hydrogen bond from N4 of anticodon base C34, is a point of difference. This residue is a serine in the mitochondrial *T. brucei* isoform, whose cognate tRNA contains a UCA anticodon. But this residue is also a serine in other TrpRS sequences whose cognate tRNA contains the usual CCA anticodon, including that of *C. parvum*, suggesting that this Ser/Thr difference is not a key factor in recognizing the variant *T. brucei* mitochondrial tRNA^{Trp}.

It is also difficult to explain the differential specificity of the two trypanosomal isoforms as depending on direct recognition of the modified or unmodified tRNA residue U33 adjacent to the anticodon. As observed in the human TrpRS:tRNA^{Trp} complex, U33 extends from the tRNA in a direction away from the protein rather than interacting directly with it. Since thiolation of this residue nevertheless abrogates recognition by the cytosolic TrpRS isoform in *T. brucei* [29], it may be that it induces a conformational change in the trypanosomal tRNA^{Trp} that allosterically alters its interaction with TrpRS, perhaps mediated by a change in the uridine sugar pucker [31]. For all of these reasons, our structural understanding of TrpRS:tRNA^{Trp} recognition in trypanosomes remains incomplete.

3.3. *Entamoeba histolytica* TrpRS homolog

The *E. histolytica* genome contains three sequences that appear to be TrpRS homologs: GenBank accessions **EAL51632**, **EAL43803**, and **EAL48880**. The first two differ from each other in amino acid sequence at only two positions, and are 55% sequence identical to the human cytosolic TrpRS. The gene product of the third sequence, whose structure is reported here, is anomalous. It is only 30% sequence identical to the other two *E. histolytica*

TrpRS isoforms, and 33% sequence identical to human cytosolic TrpRS. It notably lacks not only the HIGH and KMSKS motifs characteristic of the ATP-binding site in all Class I aaRS, but also the AIDQ motif common to the Class Ic ATP-binding site (Figure 1). Despite this, the nominal active site region is well ordered in the crystal structure and the backbone superimposes well onto the substrate-bound active sites of the *C. parvum* and human homologs. The canonical role of the KMSKS motif is to promote formation of the transition state intermediate of the activation reaction by direct interaction with the β and γ phosphates of ATP [26,35]. However, replacement of this canonical motif by KFDIP (residues 264–268) in the *E. histolytica* sequence means that the surface where the ATP phosphates would bind is more hydrophobic. The surface formed by Ile 267 and Pro 268 in particular is unfavorable for phosphate binding. Nevertheless, the adenosine binding pocket itself is well-formed. The thermal stability of the protein is increased in the presence of either MgADP or MgATP (Figure 4c), implying that the nucleotide can bind even though the protein is not competent to carry out the activation of tryptophan. By contrast, the thermal stability of the protein is not significantly altered by the presence of tryptophan (Figure 4c). These observations are in distinction to equivalent characterization of the *C. parvum*, *T. brucei*, and *G. lamblia* TrpRS homologs, whose thermal stability is increased in the presence of tryptophan or tryptophanol (Figures 4a, 4b, 4c). The implied lack of affinity for tryptophan by the *E. histolytica* homolog is consistent with the observed occlusion of the canonical tryptophan binding site caused by the substitution of a glutamine residue at position 83 of the *E. histolytica* protein where other, active, TrpRS sequences contain a glycine (e.g. Gly 280 in *C. parvum* as seen in Figures 1 and 2). The sidechain of Gln 83 in the *E. histolytica* structure protrudes directly into the volume of space occupied by the indole ring of tryptophan in previously characterized tryptophan:TrpRS complexes including the *C. parvum* TrpRS (Figure 2).

Given these alterations in sequence at residue positions involved binding both the ATP phosphates and tryptophan itself, it is not surprising that this *E. histolytica* homolog is consequently unable to charge tRNA^{Trp} (Figure 3). We also note that the N-terminal sequence of the *E. histolytica* protein lacks the characteristic VxxWxV motif of the β -hairpin that is formed transiently to mediate the first half-reaction carried out by human TrpRS [36,37,2]. This portion of the *E. histolytica* protein is well-ordered in the crystal structure, unlike the hairpin seen for the corresponding region of the human homolog only in the presence of ATP [19]. It resembles the human structure through one strand of the hairpin, but does not fold back on itself to form the second strand.

Nevertheless, this third *E. histolytica* TrpRS homolog does retain completely the sequence motifs and structural framework involved in anticodon recognition (Figures 1, 5). Furthermore, the entire protein starting at residue 8 is well-ordered in the crystal structure, including the nominal active site region of the catalytic domain. It may be that this gene product is from a 'dead' gene that has lost all biological function, but alternatively it may be that it has diverged to perform a separate function not requiring the original catalytic activity. This would be analogous to the mini splice variant of human TrpRS, which independent of its normal role as a synthetase also acts to promote apoptosis in endothelial cells, a non-canonical activity that is dependent on the presence of the tRNA^{Trp} anticodon arm binding motif [9]. More speculatively, we note that the combined retention of tRNA^{Trp} anticodon arm binding motifs, retention of an adenosine binding pocket, but loss of the KMSAS motif and occlusion of the binding site area occupied by the indole ring of tryptophan would all be features consistent with acquisition of an editing function. Normal recognition of the tRNA^{Trp} anticodon arm would position the terminal acceptor arm adenosine of a mis-charged tRNA^{Trp} into the retained adenosine binding pocket, but only if the mis-charged amino acid were small enough to fit in the reduced volume of the sidechain binding pocket.

3.4. Overall similarity of the protozoan homologs to human TrpRS

The sequences of both the catalytic and tRNA-binding domains in these protozoan proteins are similar to those of cytosolic TrpRS from other eukaryotes. Sequence identity between the human cytosolic TrpRS and the *C. parvum*, *T. brucei*, and *E. histolytica* homologs studied here is 57%, 56%, and 33%, respectively. Sequence identity between human cytosolic TrpRS and the separately reported *Giardia lamblia* homolog is 44% [2]. As is the case for TrpRS homologs from higher eukaryotes, but unlike those from bacteria, the second lysine residue of the usual class I motif KMSKS is replaced in the protozoan sequences by serine or alanine (KMSAS).

The monomers within the dimeric protozoan TrpRS structures are each superimposable onto a monomer of the human cytosolic TrpRS starting at human residue Phe 107, with a root-mean-square deviation of the C $^{\alpha}$ atoms less than 1.5 Å (Figure 5). The region of maximal difference between monomers is the loop containing the KMSAS motif of the ATP-binding site. This loop is disordered in the *C. parvum* and *T. brucei* structures. A second region of difference is the loop between helix α 12 and strand β 6 (Figure 1). The human and *C. parvum* structures are nearly identical here, but the number of residues in this loop varies considerably in other species. For example, the *G. lamblia* TrpRS contains a 49-residue insertion at this point.

In addition to expressing a cytosolic homolog of the proteins studied here, many higher eukaryotes also express a second, mitochondrial, TrpRS that is similar to bacterial forms of TrpRS rather than to cytosolic TrpRS [38]. For example, the sequence identity between the human cytosolic and mitochondrial isoforms is only 22% even in the core of the catalytic domain. In contrast to this, many parasitic protozoa have a reduced genome containing only a single set of aminoacyl-tRNA synthetases that act in both the cytoplasm and in the mitochondrion. This is the case for both *Giardia* and *Cryptosporidia* [39,40]. The single TrpRS gene sequence that is present in these organisms groups with the cytosolic form of the enzyme from higher eukaryotes rather than with the mitochondrial form. The genomes of trypanosomatids and of *Entamoeba* also contain, in general, only a single set of aminoacyl-tRNA synthetases [41,42]; TrpRS is one of several exceptions to this generalization. Nevertheless, the substantial sequence similarity of the homologs presented here to each other and to the human cytosolic TrpRS is reflected in the overall similarity of their respective structures (Figure 5).

3.5. Differences among protozoan and human homologs at the N-terminus of the catalytic domain

Archaeal and eukaryotic cytosolic TrpRS sequences contain N-terminal extensions of varying length that precede the conserved catalytic core domain. In some cases the extension comprises an entire additional domain with secondary biological functions. For example, the human cytosolic TrpRS contains an N-terminal WHEP domain that mediates association of the enzyme with other aaRS to form a multi-protein complex [43,44,45]. The protozoan TrpRS studied here do not contain a region homologous to the human WHEP domain, although as discussed above the *C. parvum* enzyme does contain a larger, unrelated N-terminal domain. However, independent of these additional nonconserved domains, the sequence of the residues extending immediately N-terminal to the universally conserved catalytic core domain is weakly conserved among eukaryotes. This region has been called the eukaryote-specific extension (ESE) by Yang et al. [37], and corresponds to residues 82–154 of the human sequence [37]. In human cytosolic TrpRS, the ESE plays a role in the primary enzymatic activity [46,5]. Beginning with human residue 107, the three-dimensional conformation of these residues is shared by the human and protozoan enzymes. This region

of conserved structure contains structural elements $\beta 1$, $\alpha 2$, $\alpha 3$, $\beta 2$, and $\alpha 4$ following the notation in Figure 1.

However, neither the sequence nor the structure of residues preceding $\beta 1$ is conserved across all eukaryotic homologs. This became evident from examination of the structure of the *G. lamblia* TrpRS, in which key features of the first part of the ESE diverge from the human cytosolic enzyme [2]. In the human enzyme, residues 82–93 form a β -hairpin upon ATP binding that mediates the first half reaction, activation of tryptophan to form tryptophanyl-AMP [19,5,47]. The β -hairpin is stabilized by interaction of hairpin residues Val 85, Trp 88, and Val 90, which constitute a recognizable VxxWxV motif, with active site residues Trp 203 and Phe 339 [37]. Point mutations that perturb these interactions either seriously compromise the ability of the enzyme to activate tryptophan or cause activation to become tRNA dependent [36]. In the *Giardia* TrpRS sequence the first part of the ESE does not contain residues homologous to the β -hairpin motif, nor does the *Giardia* sequence contain an analog to the paired active site tryptophan residue. The crystal structure of *G. lamblia* TrpRS shows that instead its N-terminus forms a much longer and reoriented helix $\alpha 1$ that provides a novel interaction surface stabilizing a homotetrameric assembly [2]. Thermal denaturation curves (Figure 4d) strongly suggest that ATP binding by the *G. lamblia* enzyme requires the presence of tryptophan, but the consequent activation reaction is not tRNA-dependent (Figure 3). The characteristic β -hairpin motif is also absent from the sequences of TrpRS homologs from the protozoans *Trichomonas vaginalis* and *Toxoplasma gondii* [2]. Thus it is clear that for at least some protozoan TrpRS the specific structural mechanism for the activation reaction must differ in detail from that of the human cytosolic enzyme.

The sequences of both the *T. brucei* and the *C. parvum* TrpRS homologs reported here do contain the VxxWxV β -hairpin motif found in the human enzyme, while the *E. histolytica* sequence does not (Figure 1). The presumptive β -hairpin containing this motif constitutes the extreme N-terminus of the *T. brucei* cytosolic homolog (residues 3–14) but is not ordered in the crystal structure, consistent with the absence of ATP. The structural equivalent of the human cytosolic TrpRS active site residue Trp 203 that interacts with the hairpin is a phenylalanine in the *T. brucei* sequence (Phe 120), as it is in the homolog from *C. parvum* (Phe 353). Presumably the phenylalanine sidechain in these protozoan homologs plays the same role as that of the human homolog's tryptophan in stacking against the central W of the VxxWxxV motif upon ATP binding. The thermal denaturation curves for the *T. brucei* enzyme (Figure 4b) are consistent with hairpin-mediated binding of ATP similar to the human enzyme and tRNA-independent tryptophan activation.

In contrast to this, tryptophan activation by the *C. parvum* construct is observed to be tRNA-dependent, which we interpret as evidence for the importance of residue Val 204. The characteristic β -hairpin motif comprises residues 204–209 of the full-length *C. parvum* TrpRS sequence, with *C. parvum* residue Val 204 corresponding to human residue Val 85 (Figure 1). However, in the *C. parvum* TrpRS construct expressed for crystallization the N-terminal 205 residues were truncated, leaving only the conserved catalytic and anticodon binding domains. Although this truncated construct is still competent to carry out both half-reactions in the presence of tRNA (Figure 3), it is apparently unable to carry out the activation of tryptophan to tryptophanyl-AMP in the absence of tRNA (Figure 4a). In retrospect it is entirely plausible that this truncation could introduce tRNA-dependence. It has previously been reported that single site mutations of residue Val 85 in the human TrpRS reduce hydrophobic interaction of the surrounding β -hairpin with the conserved AIDQ motif, similarly causing tryptophan activation to become tRNA-dependent [36]. In the truncated *C. parvum* construct all residues N-terminal to Pro 206 have been replaced by the short sequence GPGSM remaining after cleavage of an expression tag (Figure 1). It is

likely that this substitution perturbs the same β -hairpin-mediated interaction as the single site mutations characterized previously in human TrpRS. As is the case for the mutated human enzyme, the truncated *C. parvum* enzyme becomes unable to bind ATP in the absence of tRNA (Figure 4a) and the activation reaction therefore becomes tRNA-dependent. Thus although the observed tRNA-dependence of activation by the truncated *C. parvum* enzyme is probably an artifact, it is entirely consistent with the full-length *C. parvum* enzyme employing a β -hairpin to stabilize ATP-binding during the activation reaction as previously characterized for the human enzyme.

4. Conclusions

The three structures described here represent different states in the evolution of TrpRS gene function. The *C. parvum* enzyme may be considered an example of a typical eukaryotic TrpRS whose catalytic and anticodon binding domains have not diverged very far from those of higher eukaryotes in either sequence or structure. Nevertheless it has acquired an additional N-terminal domain that is shared with several other protozoans, including *Plasmodium*. The biological function of this extra domain has yet to be determined experimentally, but the fold similarity to other aaRS editing domains would be consistent with an editing function fused to the catalytically active enzyme. The *T. brucei* enzyme is one member of a pair of trypanosomatid TrpRS isoforms that have evolved to fill complementary roles, one cytoplasmic and the other mitochondrial. The anticodon recognition domains of both isoforms has diverged from other TrpRS homologs to the point where they do not recognize, for example, tRNA^{Trp} from other species such as *E. coli*. From the perspective of identifying pathogen-specific binding surfaces, it would be interesting to obtain the structure of the mitochondrial isoform as well, particularly if it could be captured in complex with the unusual modified mitochondrial tRNA^{Trp} specific to trypanosomatids. Finally, the *E. histolytica* protein represents a TrpRS homolog that has diverged from typical members of the family to the extent that it is no longer able to perform the normal catalytic function of a TrpRS. Nevertheless it has retained an anticodon recognition site that is nearly identical to those of active homologs and apparently has also retained a functional adenosine binding pocket. These features invite speculation that the protein may have evolved a new, as yet unproven, biological role, possibly involving the editing of mis-charged tRNA^{Trp}.

Acknowledgments

We thank Bill Petri for providing *E. histolytica* genomic DNA, and Angela Criswell of Rigaku Americas Corporation for collecting diffraction data from crystals of *T. brucei* TrpRS. Financial support from NIAID award AI067921 is gratefully acknowledged. Portions of this research were carried out at the Stanford Synchrotron Radiation Lightsource, 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 by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program, and the National Institute of General Medical Sciences.

References

1. Van Voorhis WC, Hol WGJ, Myler PJ, Stewart LJ. The role of medical structural genomics in discovering new drugs for infectious diseases. PLoS Comput Biol. 2009
2. Arakaki TL, Carter M, Napuli AJ, Verlinde CLMJ, Fan E, Zucker F, Buckner FS, Van Voorhis WC, Hol WGJ, Merritt EA. The structure of tryptophanyl-tRNA synthetase from *Giardia lamblia* reveals divergence from eukaryotic homologs. J Struct Biol. 2010; 171:238–243. [PubMed: 20438846]
3. Eriani G, Delarue M, Poch O, Gangloff J, Moras D. Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. Nature. 1990; 347:203–206. [PubMed: 2203971]

4. Moras D. Structural and functional relationships between aminoacyl-tRNA synthetases. *Trends Biochem Sci.* 1992; 17:159–164. [PubMed: 1585461]
5. Yu Y, Liu Y, Shen N, Xu X, Xu F, Jia J, Jin Y, Arnold E, Ding J. Crystal structure of human tryptophanyl-tRNA synthetase catalytic fragment - Insights into substrate recognition, tRNA binding, and angiogenesis activity. *J Biol Chem.* 2004; 279:8378–8388. [PubMed: 14660560]
6. Yaremchuk A, Kriklivyi I, Tukalo M, Cusack S. Class I tyrosyl-tRNA synthetase has a class II mode of cognate tRNA recognition. *EMBO J.* 2002; 21:3829–3840. [PubMed: 12110594]
7. Shen N, Guo L, Yang B, Jin Y, Ding J. Structure of human tryptophanyl-tRNA synthetase in complex with tRNA(Trp) reveals the molecular basis of tRNA recognition and specificity. *Nucleic Acids Res.* 2006; 34:3246–3258. [PubMed: 16798914]
8. Tsunoda M, Kusakabe Y, Tanaka N, Ohno S, Nakamura M, Senda T, Moriguchi T, Asai N, Sekine M, Yokogawa T, Nishikawa K, Nakamura KT. Structural basis for recognition of cognate tRNA by tyrosyl-tRNA synthetase from three kingdoms. *Nucleic Acids Res.* 2007; 35:4289–4300. [PubMed: 17576676]
9. Kise Y, Lee SW, Park SG, Fukai S, Sengoku T, Ishii R, Yokoyama S, Kim S, Nureki O. A short peptide insertion crucial for angiostatic activity of human tryptophanyl-tRNA synthetase. *Nat Struct Mol Biol.* 2004; 11:149–156. [PubMed: 14730354]
10. Jia J, Arif A, Ray PS, Fox PL. WHEP domains direct noncanonical function of glutamyl-Prolyl tRNA synthetase in translational control of gene expression. *Mol Cell.* 2008; 29:679–690. [PubMed: 18374644]
11. Arakaki T, Trong IL, Phizicky E, Quartley E, DeTitta G, Luft J, Lauricella A, Anderson L, Kalyuzhnyi O, Worthey E, Myler PJ, Kim D, Baker D, Hol WGJ, Merritt EA. Structure of Lmaj006129AAA, a hypothetical protein from *Leishmania major*. *Acta Cryst.* 2006; F62:175–179.
12. Alexandrov A, Vignali M, LaCount DJ, Quartley E, deVries C, Rosa DD, Babulski J, Mitchell SF, Schoenfeld LW, Fields S, Hol WG, Dumont ME, Phizicky EM, Grayhack EJ. A facile method for high-throughput co-expression of protein pairs. *Mol Cell Proteomics.* 2004; 3:934–938. [PubMed: 15240823]
13. Charriere F, Helgadottir S, Horn EK, Soll D, Schneider A. Dual targeting of a single tRNA(Trp) requires two different tryptophanyl-tRNA synthetases in *Trypanosoma brucei*. *Proc Natl Acad Sci U S A.* 2006; 103:6847–6852. [PubMed: 16636268]
14. Leslie AGW. Recent changes to the mosflm package for processing film and image plate data. *Joint CCP4 and ESF-EACBM Newsletters on Protein Crystallography.* 1992; 26
15. Evans P. Scaling and assessment of data quality. *Acta Cryst.* 2006; D62:72–82.
16. Collaborative Computational Project No. 4. The ccp4 suite: programs for protein crystallography. *Acta Cryst.* 1994; D50:760–763.
17. Pflugrath JW. The finer things in X-ray diffraction data collection. *Acta Cryst.* 1999; D55:1718–1725.
18. Storoni LC, McCoy AJ, Read RJ. Likelihood-enhanced fast rotation functions. *Acta Cryst.* 2004; D60:432–438.
19. Yang X, Otero F, Skene R, McRee D, Schimmel P, dePouplana L. Crystal structures that suggest late development of genetic code components for differentiating aromatic side chains. *Proc Natl Acad Sci U S A.* 2003; 100:15376–15380. [PubMed: 14671330]
20. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Cryst.* 2004; D60:2126–2132.
21. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst.* 1997; D53:240–255.
22. Painter J, Merritt EA. Optimal description of a protein structure in terms of multiple groups undergoing tls motion. *Acta Cryst.* 2006; D62:439–450.
23. Lovell S, Davis I, Arendall WB III, deBakker P, Word J, Prisant M, Richardson J, Richardson D. Structure validation by $\text{C}\alpha$ geometry: ϕ , ψ and $\text{C}\beta$ deviation. *Proteins: Structure, Function, and Genetics.* 2003; 50:437–450.
24. Lo M-C, Aulabaugh A, Jin G, Cowling R, Bard J, Malamas M, Ellestad G. Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Anal Biochem.* 2004; 332:153–159. [PubMed: 15301960]

25. Ilyin V, Temple B, Hu M, Li G, Yin Y, Vachette P, Carter C. 2.9 angstrom Crystal structure ligand-free tryptophanyl-tRNA synthetase: Domain movements fragment the adenine nucleotide binding site. *Protein Science*. 2000; 9:218–231. [PubMed: 10716174]
26. Retailleau P, Huang X, Yin Y, Hu M, Weinreb V, Vachette P, Vonrhein C, Bricogne G, Roversi P, Ilyin V, Carter CWJ. Interconversion of ATP binding and conformational free energies by tryptophanyl-tRNA synthetase: structures of ATP bound to open and closed, pre-transition-state conformations. *J Mol Biol*. 2003; 325:39–63. [PubMed: 12473451]
27. Ahel I, Korencic D, Ibba M, Soll D. Trans-editing of mischarged tRNAs. *Proc Natl Acad Sci U S A*. 2003; 100:15422–15427. [PubMed: 14663147]
28. Kelley LA, Sternberg MJE. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc*. 2009; 4:363–371. [PubMed: 19247286]
29. Charriere F, Helgadottir S, Horn E, Soll D, Schneider A. Dual targeting of a single tRNA(Trp) requires two different tryptophanyl-tRNA synthetases in *Trypanosoma brucei*. *Proc Natl Acad Sci U S A*. 2006; 103:6847–6852. [PubMed: 16636268]
30. Alfonzo JD, Blanc V, Estevez AM, Rubio MA, Simpson L. C to U editing of the anticodon of imported mitochondrial tRNA(Trp) allows decoding of the UGA stop codon in *Leishmania tarentolae*. *EMBO J*. 1999; 18:7056–7062. [PubMed: 10601027]
31. Crain PF, Alfonzo JD, Rozenski J, Kapushoc ST, McCloskey JA, Simpson L. Modification of the universally unmodified uridine-33 in a mitochondria-imported edited tRNA and the role of the anticodon arm structure on editing efficiency. *RNA*. 2002; 8:752–761. [PubMed: 12088148]
32. Wohlgamuth-Benedum JM, Rubio MAT, Paris Z, Long S, Poliak P, Lukes J, Alfonzo JD. Thiolation controls cytoplasmic tRNA stability and acts as a negative determinant for tRNA editing in mitochondria. *J Biol Chem*. 2009; 284:23947–23953. [PubMed: 19574216]
33. Garcia LT, Leite NR, Alfonzo JD, Thiemann OH. Effects of *Trypanosoma brucei* tryptophanyl-tRNA synthetases silencing by RNA interference. *Mem Inst Oswaldo Cruz*. 2007; 102:757–762. [PubMed: 17924007]
34. Yang X-L, Otero FJ, Ewalt KL, Liu J, Swairjo MA, Kohrer C, RajBhandary UL, Skene RJ, McRee DE, Schimmel P. Two conformations of a crystalline human tRNA synthetase-tRNA complex: implications for protein synthesis. *EMBO J*. 2006; 25:2919–2929. [PubMed: 16724112]
35. Retailleau P, Weinreb V, Hu M, Carter CW Jr. Crystal structure of tryptophanyl-tRNA synthetase complexed with adenosine-5' tetraphosphate: Evidence for distributed use of catalytic binding energy in amino acid activation by class I aminoacyl-tRNA synthetases. *J Mol Biol*. 2007; 369:108–128. [PubMed: 17428498]
36. Guo L-T, Chen X-L, Zhao B-T, Shi Y, Li W, Xue H, Jin Y-X. Human tryptophanyl-tRNA synthetase is switched to a tRNA-dependent mode for tryptophan activation by mutations at V85 and I311. *Nucleic Acids Res*. 2007; 35:5934–5943. [PubMed: 17726052]
37. Yang X-L, Guo M, Kapoor M, Ewalt KL, Otero FJ, Skene RJ, McRee DE, Schimmel P. Functional and crystal structure analysis of active site adaptations of a potent anti-angiogenic human tRNA synthetase. *Structure*. 2007; 15:793–805. [PubMed: 17637340]
38. Brindefalk B, Viklund J, Larsson D, Thollessen M, Andersson SGE. Origin and evolution of the mitochondrial aminoacyl-tRNA synthetases. *Mol Biol Evol*. 2007; 24:743–756. [PubMed: 17182897]
39. Morrison HG, McArthur AG, Gillin FD, Aley SB, Adam RD, Olsen GJ, Best AA, Cande WZ, Chen F, Cipriano MJ, Davids BJ, Dawson SC, Elmendorf HG, Hehl AB, Holder ME, Huse SM, Kim UU, Lasek-Nesselquist E, Manning G, Nigam A, Nixon JEJ, Palm D, Passamaneck NE, Prabhu A, Reich CI, Reiner DS, Samuelson J, Svard SG, Sogin ML. Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. *Science*. 2007; 317:1921–1926. [PubMed: 17901334]
40. Abrahamsen MS, Templeton TJ, Enomoto S, Abrahante JE, Zhu G, Lancto CA, Deng M, Liu C, Widmer G, Tzipori S, Buck GA, Xu P, Bankier AT, Dear PH, Konfortov BA, Spriggs HF, Iyer L, Anantharaman V, Aravind L, Kapur V. Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science*. 2004; 304:441–445. [PubMed: 15044751]
41. Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, Lennard NJ, Caler E, Hamlin NE, Haas B, Bohme U, Hannick L, Aslett MA, Shallom J, Marcello L, Hou L,

- Wickstead B, Alsmark UCM, Arrowsmith C, Atkin RJ, Barron AJ, Bringaud F, Brooks K, Carrington M, Cherevach I, Chillingworth T-J, Churcher C, Clark LN, Corton CH, Cronin A, Davies RM, Doggett J, Djikeng A, Feldblyum T, Field MC, Fraser A, Goodhead I, Hance Z, Harper D, Harris BR, Hauser H, Hostetler J, Ivens A, Jagels K, Johnson D, Johnson J, Jones K, Kerhornou AX, Koo H, Larke N, Landfear S, Larkin C, Leech V, Line A, Lord A, Macleod A, Mooney PJ, Moule S, Martin DMA, Morgan GW, Mungall K, Norbertczak H, Ormond D, Pai G, Peacock CS, Peterson J, Quail MA, Rabbinowitsch E, Rajandream M-A, Reitter C, Salzberg SL, Sanders M, Schobel S, Sharp S, Simmonds M, Simpson AJ, Tallon L, Turner CMR, Tait A, Tivey AR, Aken SV, Walker D, Wanless D, Wang S, White B, White O, Whitehead S, Woodward J, Wortman J, Adams MD, Embley TM, Gull K, Ullu E, Barry JD, Fairlamb AH, Opperdoes F, Barrell BG, Donelson JE, Hall N, Fraser CM, Melville SE, El-Sayed NM. The genome of the African trypanosome *Trypanosoma brucei*. *Science*. 2005; 309:416–422. [PubMed: 16020726]
42. El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, Aggarwal G, Caler E, Renauld H, Worthey EA, Hertz-Fowler C, Ghedin E, Peacock C, Bartholomeu DC, Haas BJ, Tran A-N, Wortman JR, Alsmark UCM, Angiuoli S, Anupama A, Badger J, Bringaud F, Cadag E, Carlton JM, Cerqueira GC, Creasy T, Delcher AL, Djikeng A, Embley TM, Hauser C, Ivens AC, Kummerfeld SK, Pereira-Leal JB, Nilsson D, Peterson J, Salzberg SL, Shallom J, Silva JC, Sundaram J, Westenberger S, White O, Melville SE, Donelson JE, Andersson B, Stuart KD, Hall N. Comparative genomics of trypanosomatid parasitic protozoa. *Science*. 2005; 309:404–409. [PubMed: 16020724]
 43. Wolf YI, Aravind L, Grishin NV, Koonin EV. Evolution of aminoacyl-tRNA synthetases—analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events. *Genome Res*. 1999; 9:689–710. [PubMed: 10447505]
 44. Cahuzac B, Berthonneau E, Birlirakis N, Guittet E, Mirande M. A recurrent RNA-binding domain is appended to eukaryotic aminoacyl-tRNA synthetases. *EMBO J*. 2000; 19:445–452. [PubMed: 10654942]
 45. Rho SB, Lee KH, Kim JW, Shiba K, Jo YJ, Kim S. Interaction between human tRNA synthetases involves repeated sequence elements. *Proc Natl Acad Sci U S A*. 1996; 93:10128–10133. [PubMed: 8816763]
 46. Otani A, Slike BM, Dorrell MI, Hood J, Kinder K, Ewalt KL, Cheresch D, Schimmel P, Friedlander M. A fragment of human TrpRS as a potent antagonist of ocular angiogenesis. *Proc Natl Acad Sci U S A*. 2002; 99:178–183. [PubMed: 11773625]
 47. Shen N, Zhou M, Yang B, Yu Y, Dong X, Ding J. Catalytic mechanism of the tryptophan activation reaction revealed by crystal structures of human tryptophanyl-tRNA synthetase in different enzymatic states. *Nucleic Acids Res*. 2008; 36:1288–1299. [PubMed: 18180246]
 48. Beitz E. TeXshade: shading and labeling multiple sequence alignments using LaTeX2e. *Bioinformatics*. 2000; 16:135–139. [PubMed: 10842735]
 49. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Cryst*. 2004; D60:2256–2268.

Figure 1. Sequence alignment of five protozoan TrpRS with human cytosolic TrpRS
Residue numbering is given for the *E. histolytica* sequence. The full-length *T. brucei* mitochondrial, human cytosolic, and *C. parvum* TrpRS contain non-homologous N-terminal extensions whose sequences are not shown here. *C. parvum* residues N-terminal to Pro 206 (thin yellow box) have been replaced by the sequence GPGSM remaining after cleavage of an expression tag from the construct studied here. Sequence motifs associated with the canonical TrpRS active site are boxed in red. These are notably absent in the *E. histolytica* homolog. Regions involved in binding the tRNA anticodon arm as seen in the human TrpRS:tRNA^{Trp} complex (PDB ID: 2dr2; [7]) are boxed in blue. Residues whose sidechains are directly involved in anticodon base recognition in the human TrpRS:tRNA^{Trp} complex

are further highlighted in blue. One of these is a strictly conserved lysine. The β -hairpin motif VxxWxV and the paired aromatic residue at the active site present in the human TrpRS and three of the protozoan homologs are boxed in purple. Secondary structure elements are shown for *C. parvum*; they are essentially the same in the human homolog and in the other two protozoan structures described here. The first and last well-ordered residue in each of the respective crystal structures is indicated by blue square brackets. Figure composed using texshade [48].

Figure 2. Difference electron density at the active site of *C. parvum* TrpRS

Electron density contours enclosing the bound tryptophan are drawn at 3.5σ in a Sigma A weighted (mFo-Fc) map with all non-protein atoms omitted from the calculation of Fc. Glycine 280 is strongly conserved in TrpRS sequences, but the *E. histolytica* **EAL48880** sequence has a glutamine at this position (residue 83), consistent with its inability to bind tryptophan.

Figure 3. Activity of three protozoan TrpRS homologs

The N-terminally truncated *C. parvum* TrpRS is able to charge *E. coli* tRNA^{Trp} with ^3H -labeled tryptophan. The active full-length TrpRS homolog from *G. lamblia* serves as a positive control. As expected, the defective active site of the *E. histolytica* homolog is not capable of charging tRNA. The *T. brucei* TrpRS was omitted from the assay because it is known to not recognize *E. coli* tRNA^{Trp} [29]. Unshaded bars indicate incorporation in the presence of competition from additional 100 μM unlabelled tryptophan.

Figure 4. Thermal denaturation curves in the presence or absence of substrates

Each curve is normalized to the maximum fluorescence recorded. **(a)** Thermal denaturation curves for the truncated *C. parvum* TrpRS in the absence of tRNA^{Trp} indicate binding of tryptophan, but not of ATP alone. Also the presence of both substrates does not lead to a formation of a tighter binding reaction product. However in the presence of tRNA^{Trp} the enzyme is active (Figure 3), suggesting that ATP binding and subsequent tryptophan activation are both tRNA-dependent. **(b)** Thermal denaturation curves for the *T. brucei* TrpRS. In the absence of ligands the protein shows a multi-component denaturation curve that can be modelled as comprising two transitions, one at about 45° and the other at about 62°. In the presence of either tryptophan or ATP the low temperature transition is suppressed; the high temperature transition is stabilized to about 68° in the presence of ATP, and to about 72° in the presence of tryptophan. If both substrates are present, the transition is further shifted to about 78° by the formation of the tighter-binding product of the activation reaction, Trp-AMP. If the non-reactive substrate analog L-tryptophanol is used instead of L-tryptophan itself, the ternary complex does not show this further shift relative to the binary complex (curves not shown). **(c)** Thermal denaturation curves for the *E. histolytica* TrpRS homolog show evidence that it binds ATP, but addition of tryptophan does not significantly stabilize the protein (shift ≤ 1°) regardless of the presence or absence of ATP. The curves shown are averages of two replicate experiments. **(d)** Thermal denaturation curves for the *G. lamblia* TrpRS indicate binding of tryptophan, but not of ATP alone. However, in the presence of both substrates the enzyme carries out the activation reaction to produce tighter-binding reaction product tryptophanyl-AMP, which shifts the denaturation curve by 12° relative to the unliganded protein.

Figure 5. Structural superposition of three protozoan TrpRS homologs onto the human TrpRS:tRNA^{Trp} complex

Each protozoan TrpRS dimer was superimposed onto the human TrpRS structure by SSM optimization of the C^α atoms in one monomer [49]. The superimposed monomers are shown at the right of the figure as a backbone trace: *C. parvum* (green), *T. brucei* (cyan), *E. histolytica* (purple), human (gold). A semi-transparent surface for the second monomer in each protozoan dimer, but not the human, is shown with the same coloring. A small displacement in the relative positions of the second monomer after superposition of the first monomer is visible; i.e. the angle between the two *T. brucei* monomers (cyan) is smallest, while the angle between the two *E. histolytica* monomers (purple) is largest. The position of the second human monomer is intermediate (not shown in the figure). Also shown is the position of one bound tRNA^{Trp} molecule as seen in the human TrpRS:tRNA^{Trp} complex (PDB ID: 2dr2; [7]). The anticodon is at the far left of the figure. The r.m.s.d. of C^α atoms after pairwise superposition of one monomer from each protozoan structure onto the 2.1 Å resolution human TrpRS structure (PDB ID: 1r6t; [19]) is: *C. parvum*: 1.38 Å for 336 C^α; *T. brucei*: 0.96 Å for 316 C^α; *E. histolytica*: 1.42 Å for 351 C^α; *G. lamblia*: 1.39 Å for 341 C^α[2].

Table 1

Data collection and refinement statistics

	<i>C. parvum</i>	<i>T. brucei</i>	<i>E. histolytica</i>
PDB entry	3hv0	3i05	3hxr
Space Group	<i>P</i> 2 ₁	<i>P</i> 1	<i>P</i> 2 ₁
Unit Cell (a b c Å)	58.32 85.22 73.11	51.29 51.43 89.97	93.20 158.88 100.28
($\alpha\beta\gamma$ °)	90 101.9 90	95.3 100.1 106.1	90 107.3 90
Wavelength (Å)	0.9787	1.54	0.9795
Resolution (Å)^a	57–2.42 (2.55–2.42)	31–2.80 (2.90–2.80)	49–3.00 (3.16–3.00)
Total number of unique reflections	25491 (3517)	20576 (2145)	54415(7976)
<i>R</i> _{sym}	0.096 (0.930)	0.077 (0.324)	0.115 (1.05)
Completeness (%)	96 (91)	97 (96)	98 (99)
<i>I</i> / σ (<i>I</i>)	9.7 (1.7)	9.2 (2.6)	9.7 (1.2)
Redundancy	4.4 (4.3)	3.1 (3.1)	3.2 (3.1)
Wilson B (Å²)	54	76	86
Contents of asymmetric unit	1 dimer	1 dimer	3 dimers
Refinement Resolution (Å)	40–2.42	30–2.80	48–3.0
<i>R</i>	0.227	0.244	0.228
<i>R</i> _{free}	0.273	0.301	0.250
RMSD bonds (Å)	0.010	0.010	0.012
RMSD angles (°)	1.119	1.238	1.361
Protein atoms	5586	5215	17738
Non-protein atoms	74	0	0
Residues in favored regions (%)	98	98	94
Residues in disallowed regions (%)	0	0	2
TLS groups	4 per chain	6 per chain	1 per chain
Mean <i>B</i>_{iso}+<i>B</i>_{TLS} protein atoms (Å²)	66.5	62.1	77.8
Mean <i>B</i>_{iso} non-protein atoms (Å²)	30.5	-	-

^a Numbers in parentheses refer to the highest resolution shell of data.