Published in final edited form as:

J Mol Biol. 2011 June 3; 409(2): 159–176. doi:10.1016/j.jmb.2011.03.026.

The double-length tyrosyl-tRNA synthetase from the eukaryote *Leishmania major* forms an intrinsically asymmetric pseudo-dimer

Eric T Larson^{a,b}, Jessica E Kim^{a,b}, Lisa J Castaneda^{a,c}, Alberto J Napuli^{a,c}, Zhongsheng Zhang^{a,b}, Erkang Fan^{a,b}, Frank H Zucker^{a,b}, Christophe LMJ Verlinde^{a,b}, Frederick S Buckner^{a,c}, Wesley C Van Voorhis^{a,c}, Wim G J Hol^{a,b}, and Ethan A Merritt^{a,b,*}

^a Medical Structural Genomics of Pathogenic Protozoa <http://www.msgpp.org>

^b Department of Biochemistry, University of Washington, Seattle, WA 98195 USA

^c Department of Medicine, University of Washington, Seattle, WA 98195 USA

Abstract

The single tyrosyl tRNA-synthetase (TyrRS) gene in trypanosomatid genomes codes for a protein that is twice the length of TyrRS from virtually all other organisms. Each half of the double-length TyrRS contains a catalytic domain and an anticodon-binding domain, however the two halves retain only 17% sequence identity to each other. The structural and functional consequences of this duplication and divergence are unclear. TyrRS normally forms a homodimer in which the active site of one monomer pairs with the anticodon-binding domain from the other. However, crystal structures of *Leishmania major* TyrRS show that instead the two halves of a single molecule form a pseudo-dimer resembling the canonical TyrRS dimer. Curiously, the C-terminal copy of the catalytic domain has lost the catalytically important HIGH and KMSKS motifs characteristic of Class I aminoacyl-tRNA synthetases. Thus the pseudo-dimer contains only one functional active site, contributed by the N-terminal half, and only one functional anticodon recognition site, contributed by the C-terminal half. Despite biochemical evidence for negative cooperativity between the two active sites of the usual TyrRS homodimer, previous structures have captured a crystallographically-imposed symmetric state. As the *L. major* TyrRS pseudo-dimer is inherently asymmetric, conformational variations observed near the active site may be relevant to understanding how the state of a single active site is communicated across the dimer interface. Furthermore, substantial differences between trypanosomal TyrRS and human homologs are promising for the design of inhibitors that selectively target the parasite enzyme.

Keywords

aminoacyl-tRNA synthetase; protozoa; drug target; tropical disease

© 2011 Elsevier Ltd. All rights reserved.

*Corresponding author: Phone: 206-543-1421 Fax: 206-685-7002, merritt@u.washington.edu.

Accession Numbers

Coordinates and structure factors have been deposited in the Protein Data Bank with accession numbers 3p0h, 3p0i, and 3p0j for the fisetin complex, the cubic crystal form of the tyrosinol complex, and the triclinic crystal form of the fisetin complex, respectively.

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.

Introduction

The aminoacyl-tRNA synthetases (aaRS) constitute an ancient family of essential enzymes that have retained conserved structural features across all branches of the tree of life. They group broadly into two classes defined by characteristic sequence motifs and by the consequent structural homology of their core domains^{1,2}. The Class I catalytic domain exhibits a Rossmann fold and is characterized by two hallmark sequence motifs whose canonical forms are HIGH and KMSKS. The Class II catalytic domain contains a core anti-parallel β -sheet surrounded by α -helices, and is uniquely identified by three conserved sequence motifs. The aaRS for each amino acid except lysine belongs consistently either to Class I or to Class II in all organisms. However, the evolutionary history of individual aaRS is complex, involving many horizontal gene transfers and leading to idiosyncratic phylogenetic trees for the 20+ individual aaRS sequence families^{3,4}. Models for the evolutionary history of eukaryotic TyrRS in particular have been modified multiple times as increasing numbers of relevant sequences and structures are reported^{5,6,7,8,9,10}.

Tyrosyl tRNA-synthetase (TyrRS) was the first aaRS to be examined crystallographically¹¹. It is a Class I aaRS, grouped together with tryptophanyl tRNA-synthetase (TrpRS) to form subclass Ic. Unlike other Class I aaRS, both TyrRS and TrpRS act as homodimers¹². With one exception, all Class Ic aaRS structurally studied to date exhibit the same core domain secondary structure and dimer geometry. Cognate tRNA molecules bind by bridging the anticodon-binding domain of one constituent monomer to the active site in the catalytic domain of the other constituent monomer¹³. The single known exceptional case is the structure of the TyrRS from the nucleocytoplasmic large DNA virus Mimivirus¹⁴, which retains the expected tertiary structure of each monomer but is observed crystallographically to assemble into a non-canonical homodimer¹⁵. The structural basis of tRNA recognition by TyrRS has been illuminated by crystal structures from bacterial¹³, archaeal^{16,17}, and eukaryotic¹⁸ TyrRS:tRNA^{Tyr} complexes, and the relative importance of specific residues has been further probed by mutational analysis¹⁹.

The mechanisms of tyrosyl activation by TyrRS to form tyrosyladenylate and the subsequent transfer of the tyrosyl moiety to tRNA^{Tyr} have both been extensively studied structurally and biochemically. The N-terminal domain active site, which carries out both of these reactions, is broadly similar in TyrRS homologs from archaea, eubacteria, and eukaryotes¹⁸ and similar also to the TrpRS active site¹². Although the TyrRS dimer and the two active sites it contains are symmetric as observed crystallographically, in solution only a single active site is competent to carry out the activation reaction at any one time; i.e. the dimer exhibits half-of-sites reactivity²⁰. This negative cooperativity has been further dissected in bacterial TyrRS through directed mutagenesis of the substrate binding site^{21,22}, but fewer details are known for eukaryotic TyrRS.

We chose *Leishmania major* tyrosyl-tRNA synthetase for study in the context of efforts by the Medical Structural Genomics of Pathogenic Protozoa collaboration to identify and characterize potential protein targets for the development of novel therapeutics for human diseases caused by parasitic infection²³. Among the target pathogens are *Trypanosoma brucei*, the causative agent of African sleeping sickness, *Trypanosoma cruzi*, the causative agent of Chagas disease, and various species of *Leishmania*, responsible for a spectrum of visceral and cutaneous diseases. The chronic forms of these diseases are fatal, and existing therapeutics are unsatisfactory due to poor effectiveness and to toxicity. All of these trypanosomatids are unicellular eukaryotes belonging to the class of kinetoplastids. As such, they are phylogenetically distant from their mammalian host species, a point that is dramatically illustrated by the phylogenetic tree for TyrRS (Figure 1). The trypanosomatid TyrRS sequences are more closely related to those of plants and certain other basal

eukaryotes than they are to either the cytosolic or mitochondrial TyrRS sequences of their mammalian hosts^{19,9,15,4}. Indeed, the closest TyrRS homolog for which a structure has previously been determined is that from Mimivirus, raising the possibility that TyrRS from trypanosomatids and other members of this clade similarly deviate from the canonical Class Ic dimeric assembly and its associated tRNA binding geometry. Thus it was not clear at the outset what quaternary structure to expect for the *L. major* TyrRS.

Furthermore, trypanosomatid TyrRS sequences have apparently doubled in length due to gene duplication or tandem insertion of an exogenous TyrRS sequence. That is, they contain a large C-terminal extension that is itself recognizably homologous to full-length TyrRS despite the absence of canonical Class I sequence motifs (Figure 2). This introduced further uncertainty as to their expected quaternary structure. It was possible that two molecules associate through their N-terminal domains to form either a canonical Class Ic dimer or a Mimivirus-like non-canonical dimer. The internal pseudo-symmetry of the gene structure also caused us to consider the possible formation of asymmetric assemblies; a pseudo-tetramer could form by head-to-tail association of two double-length monomers or a pseudo-dimer could form by self-association of the N-terminal and C-terminal halves from a single monomer. The first of these unusual possibilities would be analogous to the tetrameric assembly formed by two canonical Class Ic dimers of *Giardia lamblia* TrpRS²⁴; the second would essentially recreate a canonical (or possibly Mimivirus-like) TyrRS homodimer using the two copies of the core domains present within a single polypeptide chain.

This uncertainty is resolved by two crystal structures of *L. major* TyrRS in complex with the substrate analog L-tyrosinol, which we report here at 3.1 Å and 2.9 Å resolution. We also report a 3.0 Å resolution crystal structure of *L. major* TyrRS in complex with a flavonoid, fisetin, that was found through high-throughput screening of chemical libraries to identify potential inhibitors of the enzyme. Since trypanosomal TyrRS is substantially different from both homologs in the human host, these structures will aid in the rational design of selective inhibitors that may lead to novel anti-trypanosomal therapeutics.

Results

Overall structure

The X-ray crystal structure of the double-length tyrosyl-tRNA synthetase from the parasitic protozoan *Leishmania major* has been determined using a combination of molecular replacement and single wavelength anomalous dispersion methods (MRSAD; Table 1, Supplemental Figure S1). The architecture of the enzyme is that of a pseudo-dimer formed by the N- and C-terminal halves of a single double-length polypeptide chain with only 17% sequence identity between them. The pseudo-dimer mimics the canonical TyrRS homodimer observed for all previous homologs except that from Mimivirus (Figure 3). A linker consisting of residues 355–373 connects the two halves of the pseudo-dimer, stretching between helix $\alpha 15$ in the N-terminal half and strand $\beta 2'$ in the C-terminal half. There are no structural equivalents in the C-terminal half for helix $\alpha 1$, helix $\alpha 2$, or strand $\beta 1$; rather these structural elements have been replaced by a sequence that adopts an extended conformation and is incorporated into the linker connecting the two halves of the pseudo-dimer (colored magenta in Figure 3). Electron density for the entire linker is visible but it is hard to model individual residues unambiguously. The linker does not assume any regular secondary structure and does not correspond in sequence or structure to any part of the canonical TyrRS dimer.

The inter-domain interface between the N-terminal and C-terminal halves mimics the canonical dimer interface. Disregarding interactions with the linker, the majority of this pseudo-dimer interface is made up of residues from the connective peptide (CP1) region, as

in other Class Ic aaRS structures²⁵. This corresponds to residues 120–166 in the N-terminal half and residues 458–503 in the C-terminal half. An additional small contribution to the dimer interface is made by interactions involving residues 74–84 (helix $\alpha 4$) and 416–429 (helix $\alpha 4'$). The buried surface area of the inter-domain pseudo-dimer interface averaged over the crystallographically independent instances in the current set of structures is 1280 Å², which is in the same range seen for the analogous interface in the canonical dimer. The corresponding free energy of dissociation $\Delta^i G$ is –22 kcal/mol as calculated by PISA²⁶, although this number is not physically meaningful as both sides of the interface are contributed by the same polypeptide chain. As a point of reference, PISA calculates $\Delta^i G$ = –23 kcal/mol for the 1317 Å² true dimer interface in the human tyrosinol:TyrRS complex (PDBID 1q11)⁷.

Very unexpectedly, there is also an extensive packing interaction between the N-terminal portions of paired crystallographically independent protein chains that buries 1200 Å² of accessible surface. This interaction greatly complicated the interpretation of our initial structure solutions (described in detail in the Supplemental Methods section) because it is present in both the cubic and triclinic crystal forms reported here, and also in a second triclinic crystal form (data not shown). However, the $\Delta^i G$ for this second interface is only –2 kcal/mol as calculated by PISA. Thus we believe it is only a crystallization artifact, despite its recurrence in multiple crystal forms.

The Rossmann fold and anticodon-binding domains in both the N-terminal and C-terminal halves of the enzyme conform overall to the expected conserved Class Ic core secondary structure and fold (Figures 3 and 4, Supplemental Figure S2). However, there is a difference in the relative orientation of the two constituent domains within each half. SSM superposition of the entire C-terminal half onto the entire N-terminal half yields an overall r.m.s.d. of 3.5 Å for 246 paired C α atoms. Superposition of the two Rossmann fold domains alone yields 2.3 Å r.m.s.d. for 156 C α atoms from 205 residues total in the smaller C-terminal domain. After this superposition, it requires an additional rotation of ~21° to superimpose instead the two anticodon-binding domains yielding an r.m.s.d. of 2.2 Å for 100 C α atoms from 120 residues total in the C-terminal anticodon-binding domain (Supplemental Figure S2).

Active site conformation in the N-terminal pseudo-monomer

The Class I aaRS active site is characterized by the presence of a signature motif KMSKS, whose first lysine is the most highly conserved residue of the Class I aaRS sequence family as a whole. It is strictly conserved in eubacterial TyrRS, and is also present in TyrRS sequences from plants and from various pro-tozoa including the trypanosomatids. This lysine stabilizes the transition state of the amino acid activation reaction by interacting with the pyrophosphate leaving group of the substrate ATP²². However, the second lysine of this sequence motif is not retained by all archaeal TyrRS, nor is it retained by cytosolic TyrRS from opisthokonts, including the previously studied human^{7,25} and yeast¹⁸ homologs. In these enzymes its role is functionally replaced by a metal ion, an obligate potassium ion in human cytosolic TyrRS²⁷. In both cases, interaction of the lysine or the metal with the β and γ phosphates of ATP is believed to stabilize an asymmetric state of the TyrRS homodimer, giving rise to the observed half-of-sites reactivity²⁷.

The present structures are the first to illuminate the structural state of the bacterial-like KMSKS motif, which includes the second lysine, in the context of a eukaryal active site. The residues making up the KMSKS motif are well ordered in the N-terminal pseudo-monomer of all of the *L. major* structures described here. Curiously, they are in the ‘closed’ conformation that in bacterial TyrRS has been observed only in complexes with ATP or with analogs of the tyrosyladenylate reaction product. Thus it seems that in *L. major* TyrRS,

the ATP-binding site is at least partially pre-formed. This is in distinct contrast to the induced-fit model for ATP binding by bacterial TyrRS²², which is supported by a series of structures that capture snapshots of ‘open’, ‘semi-open’, and ‘closed’ conformations [reviewed by Kobayashi *et al.*²⁸].

Two distinct conformational states of the active site are observed in the three *L. major* TyrRS structures discussed here (Figure 5). The first conformational state is observed in both molecules of the tyrosinol complex in the cubic crystal form and in two of the four crystallographically independent copies in the triclinic crystal form. In this conformation, the loop containing residues 146–154 is well-ordered and positioned so that the residues at the tip of the loop can interact with the α - and β -phosphates of ATP or with the single phosphate of tyrosyladenylate. At the same time, residues 73–89 on the opposite side of the active site pocket assume a conformation in which helix $\alpha 4$ and the loop connecting it to helix $\alpha 5$ curl over to interact with residues 40–43 immediately preceding the HIGH motif, partially uncoiling helix $\alpha 5$. This relatively closed conformation, consisting of a well-ordered loop containing residues 146–154 and a curled $\alpha 4$ – $\alpha 5$ loop, is similar to the conformation previously observed for the equivalent residues in archaeal (PDBID 1j1u, 2cya, 2cyb, 2cyc)^{16,17} and eukaryotic (PDBID 1q11, 2dlc)^{7,18} TyrRS structures. This conformation has no direct parallel in bacterial TyrRS or the related human mitochondrial enzyme. In bacterial homologs the loop between helices $\alpha 4$ and $\alpha 5$ is much longer than in the eukaryotic/archaeal homologs and extends over the binding pocket so that the residues at its tip can interact with the phosphates of the bound substrate or product. Although the tip of the bacterial $\alpha 4$ – $\alpha 5$ loop is functionally analogous to the tip of the eukaryotic/archaeal 146–154 loop, the corresponding loops originate from opposite sides of the active site pocket and thus are not structurally equivalent.

A second conformational state of the TyrRS active site is observed for the first time in both molecules of the fisetin complex and in the remaining two molecules in the triclinic tyrosinol complex (Figure 5). In these molecules helix $\alpha 5$ is longer, helices $\alpha 4$ and $\alpha 5$ run parallel to each other, and the intervening residues form a short cap across the top of the two helices. In this extended conformation, helix $\alpha 4$ shifts to interact more extensively with helix $\alpha 7'$ across the pseudo-dimer interface rather than interacting with the residues preceding the HIGH motif. In all of the molecules exhibiting this extended conformation, residues 146–154 are disordered. Thus the active site in these molecules is more open on both sides than has been observed in previous TyrRS structures, despite the presence of a ‘closed’ conformation of the KMSKS loop itself.

Binding modes of tyrosinol and fisetin

Notwithstanding the conformational differences of the KMSKS loop and the $\alpha 4$ – $\alpha 5$ helices that distinguish the active site of the current structures from those of previously reported complexes with tyrosine analogs, the position of tyrosinol within the *L. major* TyrRS active site is essentially identical to that previously observed for tyrosine itself and for the tyrosyl moiety of tyrosyladenylate (Figure 6). The plane of the tyrosine ring stacks against the sidechain of Gln167. The tyrosinol O ^{η} is oriented through hydrogen bonds with O ^{η} of Tyr36 and O ^{δ} of Asp170. These residues and interactions are the same as found for the tyrosyl moiety in structures of human, yeast, and archaeal TyrRS complexes (Figures 6c and 6d). The amide N is oriented through hydrogen bonding to O ^{η} of Tyr163, also a conserved interaction.

Fisetin (3,7,3',4'-tetrahydroxyflavone) was one of several flavonoids identified as binding tightly to TyrRS in the course of high-throughput screening of chemical libraries. Although fisetin itself is not effective as an inhibitor of parasite growth, characterization of its binding mode at the active site of trypanosomal TyrRS may serve to guide the identification or

design of more potent inhibitors. The 7-hydroxyl group of the fisetin benzopyran ring system binds in approximately the same manner and position as the tyrosinol O'' . The remainder of the fisetin molecule extends into the position occupied by the ribose of tyrosyladenylate as seen in the yeast complex. The pair of hydroxyls from the fisetin catechol moiety lie at approximately the positions of the ribose $O2'$ and $O3'$. The remaining hydroxyl substituents of fisetin project into the solvent region and make no interactions with the protein (Figure 6c and 6d).

Defective active site in C-terminal pseudo-monomer

In the C-terminal copy of the Rossmann fold domain, the region corresponding to the typical active site does not appear to be capable of carrying out amino acid activation. Neither of the highly conserved Class I aaRS consensus motifs that play a role in the catalytic mechanism are conserved (Figures 2 and 4). The loop containing residues 556–560 (sequence APAVL), which corresponds to the missing KMSKS motif in the C-terminal half, is not well ordered in any the *L. major* structures reported here. Diffraction quality crystals grew only when the protein was cocrystallized with either the substrate analog tyrosinol or the flavonoid fisetin. Electron density that may be attributed to either of these ligands is not observed in the C-terminal Rossmann fold domain in any of the twelve crystallographically independent chains for four refined structures (the three structures reported here and a second triclinic form of the tyrosinol complex [data not shown]). Ligand density is observed in the active site of each copy of the N-terminal Rossmann fold domain, however. This strongly suggests that the C-terminal half of the *L. major* enzyme is not capable of binding these ligands. The $\alpha4'$ – $\alpha5'$ loop is in the 'curled' conformation, one of two states described above for the equivalent region in the functional active site, but the significance of this observation is not clear.

Anticodon binding domains

In the known binding mode of tRNA to dimeric Class Ic aaRS, the anticodon arm of the tRNA is recognized by the anticodon-binding (C-terminal) domain of one monomer, which positions the acceptor stem of that same tRNA molecule adjacent to the active site in the Rossmann fold domain of the other monomer. The distance between the anticodon loop and the acceptor stem of the tRNA is too great for both to bind to the same monomer. Thus, in the current *L. major* pseudo-dimer, the identification of a single functional active site located in the N-terminal copy of the Rossmann fold domain necessarily implies a functional C-terminal copy anticodon-binding domain to facilitate the aminoacylation of tRNA^{Tyr} (see supplemental Figure S3 for a model of tRNA-binding by trypanosomal TyrRS).

Although both the N-terminal and C-terminal copies of the anticodon-binding domain retain the pair of sequence motifs called AC1 and AC2 (Figures 2 and 4) associated with recognition of the tRNA anticodon stem, they differ structurally. There is a 6-residue deletion between $\beta7$ and $\beta8$ in the N-terminal copy of the anticodon-binding domain, which has the effect of severely shortening the loop connecting them. This loop would normally form a binding pocket for anticodon base G34 (Figure 7). In contrast to this, the equivalent loop in the C-terminal anticodon-binding domain is of normal length and its conformation is very similar to that seen in the yeast TyrRS:tRNA complex (Figure 7) and in archaeal homologs. Note that while the human mitochondrial TyrRS, like bacterial homologs, does not contain the $\beta7$ -loop- $\beta8$ structural element at all, these proteins have instead an additional C-terminal domain that makes extensive interactions with tRNA¹³. Therefore despite the presence of the anticodon stem binding motifs on helices $\alpha11$ and $\alpha14$ of the N-terminal copy of this domain, there is doubt as to whether this copy is capable of recognizing or binding tRNA.

A second notable feature that differs between the two copies of the anticodon-binding domain is the hairpin formed by residues 256–271 following the AC1 motif in the N-terminal copy of the anticodon-binding domain (colored cyan in Figure 3). This hairpin is not present in TyrRS structures from yeast, human or Mimivirus. An insertion at this point is characteristic of TyrRS sequences from the plant/plastid clade, although it is typically only 8 residues long rather than the 15-residue sequence in *L. major*. Curiously, neither a sequence insertion nor a corresponding hairpin is present in the functional C-terminal copy of this domain in the *L. major* enzyme (Figure 4). This suggests a complex evolutionary origin for the observed double-length TyrRS gene in these and other organisms.

Thermal denaturation curves

The thermal denaturation curve for *L. major* TyrRS in the absence of ligands shows a transition with maximal slope at $T_M = 48.7^\circ\text{C}$ (Figure 8a). The analogous measurement from the closely related *T. cruzi* TyrRS yielded $T_M = 47.4^\circ\text{C}$ (Figure 8b). The protein is stabilized in the presence of either 2 mM L-tyrosine or 10 mM MgATP as indicated by a shift $\Delta T_M = +4^\circ\text{C}$. In the presence of both L-tyrosine and MgATP, the two substrates of the activation reaction, the protein is further stabilized to yield a net increase of $\Delta T_M = +13^\circ\text{C}$. This likely indicates the formation of the reaction product, tyrosyladenylate. A similar shift is observed in the presence of MgATP and the substrate analog L-tyrosinol, which only lacks the carbonyl oxygen, so it is possible that the protein is further stabilized upon binding of both substrates together (Figure 8b). However, reaction between tyrosinol and ATP is feasible and yields a tyrosyladenylate analog also lacking the carbonyl oxygen suggesting that this oxygen of the true adenylate product does not interact with the enzyme, as is indeed the case in the complex of yeast TyrRS with a tyrosyladenylate analog (PDBID 2dlc).

Thermal denaturation shift assays were also used for high-throughput screening of chemical libraries for potential inhibitors. The results of this screening program will be reported separately, but we note that one compound identified was the flavonoid fisetin (3,7,3',4'-tetrahydroxyflavone). Fisetin yielded an increase of $\Delta T_M = +7^\circ\text{C}$, which is larger than either of the natural substrates alone (Figure 8a). Importantly, fisetin was also used successfully in cocrystallization trials of the selenomethionyl protein, resulting in the TyrRS:fisetin complex crystals used for initial structure determination.

Discussion

Two observed conformations of the active site

The homodimeric assembly usually seen for TyrRS exhibits negative cooperativity between the two subunits, leading to an asymmetric dimer in which only one active site is competent to perform catalysis. The structural basis for this asymmetry has remained unclear because all crystal structures to date have contained symmetric dimers regardless of the presence or absence of ligands. As pointed out by Yaremchuk *et al.*¹³, this may be a consequence of the high ligand concentration used for crystallization, which favors a symmetric state in which both active sites are occupied by the highest affinity ligand present. Crystallographically imposed symmetry is not relevant to the current *L. major* TyrRS structures since the pseudodimer is intrinsically asymmetric and contains a functional active site only in the N-terminal “monomer”. It is therefore of interest to consider whether any of the structural features observed in the current asymmetric structures are relevant to the asymmetric state assumed for typical homodimeric homologs in solution. One obvious candidate is the conformational variation observed in the $\alpha 4$ – $\alpha 5$ loop. In the curled conformation seen in previous structures these residues do not interact with the second monomer. But in the extended conformation seen for the first time in the *L. major* structures, these same residues form part of the dimer interface, structurally linking the state of the active site to the second monomer. Intriguingly,

there is additional evidence for the conformation of this $\alpha 4$ – $\alpha 5$ loop correlating with substrate binding. There are two available structures of apo TyrRS from archaeal/eukaryotic homologs. In the apo structure of human cytosolic TyrRS (PDBID 1n3l)²⁵, the $\alpha 4$ – $\alpha 5$ loop is intermediate in conformation between the extended and curled forms seen here, although it is closer to the latter. The same is true for the *M. jannaschii* apo structure (PDBID 1u7d)²⁹. The conformational variability at this location is of interest because it has the obvious consequence of creating a dynamic dimer interface (pseudo-dimer interface in the case of *L. major*) that potentially represents a means of communicating the state of the active site of one monomer to the adjacent monomer. One may speculate that an equivalent conformational change in the active site of typical, homodimeric, TyrRS could underlie the mechanism of induced asymmetry that leads to negative cooperativity and half-of-sites reactivity. A similar argument has been advanced for the possible involvement of residues 82–86 in *Staphylococcus aureus* TyrRS in communication between the two monomers³⁰ but, as described above, these residues belong to a loop in bacterial TyrRS that has no direct structural equivalent in the archaeal/eukaryotic homologs.

Structural clues to phylogeny and relationship to other double-length TyrRS sequences

The sequence families of TyrRS and TrpRS are recognizably related. However the low sequence identity among distantly related TyrRS and TrpRS homologs makes it difficult to determine with confidence the depth and history of their evolutionary split. Phylogenetic analyses of the two sequence families jointly or of TyrRS alone have variously lead to non-congruent phylogenetic trees^{5,6,7,8,31,9}. One approach to improving the accuracy of comprehensive multiple sequence alignment is to use the close structural similarity of the two aaRS to provide additional structure-based constraints on alignment^{8,7,10}. Most recently Dong *et al.*¹⁰ were able to draw on crystal structures of both TyrRS and TrpRS from eubacteria, archaea, and eukarya to reexamine the phylogeny of both families. This work produced a structure-guided phylogenetic tree in which the root of the TrpRS sequences is located in the archaeal branch of TyrRS. However, the analysis considered only the portion of the sequences corresponding to the Rossmann fold domain, and the authors explicitly noted that the power of the analysis was limited by the absence of any structures from the clade of plant/plastid TyrRS. The availability of the *L. major* TyrRS structure may offer an opportunity to address both of these limitations. We note, for example, that the first part of the anticodon-binding domain of plant/plastid sequences is longer than in other eukaryotic sequences. The *L. major* TyrRS structure indicates that this extra length is accounted for by the presence of a distinct structural element, a hairpin immediately preceding helix $\alpha 12$, whose location may help to anchor the alignment of other plant/plastid sequences.

From the complexity of the phylogenetic tree for eukaryotic TyrRS sequences alone, it is evident that contemporary homologs stem from multiple points of origin^{9,4,10}. Even within narrow groups of eukaryotes there is variation in TyrRS gene and domain structure, often involving additional domains C-terminal to the core TyrRS catalytic and anticodon-binding domains (Figure 2). For example, the human cytosolic TyrRS gene (YARS) has a C-terminal domain that acts as an EMAP-II-like cytokine^{32,33}. Many plants have an additional C-terminal WD40-like domain of unknown function. In both of these cases the additional C-terminal domain is a structural appendage to the canonical homodimer formed by the conserved N-terminal TyrRS core domains. Trypanosomatid TyrRS sequences are unusual in containing four domains that appear to represent two sets of the usual core TyrRS catalytic and anticodon-binding domains arranged head-to-tail (Figure 2). The possibility existed that these double-length homologs might also form the typical homodimer from two fully functional N-terminal halves, leaving the C-terminal halves as TyrRS-like appendages. The structures reported here show that this is not the case. Instead the N-terminal and C-terminal halves of a single molecule form a functional pseudo-dimer.

Although both the N-terminal and C-terminal halves of the *L. major* sequence clearly belong to the clade of plant/plastid TyrRS sequences when considered separately, the two halves are only 17% sequence identical to each other. The pairwise identity of the full double-length sequences between the trypanosomatids *L. major*, *T. brucei*, and *T. cruzi* is 60%–70%. The pairwise sequence identity between the N-terminal TyrRS core domains from trypanosomatids and the normal-length TyrRS from plasmodia is ~50%. The relative degree of divergence between the two halves of the trypanosomatid TyrRS sequences suggest that they originated either from an ancient event of gene duplication followed by substantial divergence, or through introduction of the tandem copy through horizontal gene transfer. One point of difference between the two halves is that only the N-terminal half contains an insertion following sequence motif AC1 that is characteristic of sequences from the plant/plastid clade. This suggests that if the duplication arose through horizontal transfer, the event involved import of what became the C-terminal half of the current double-length sequence from a source outside the plant/plastid clade.

We were thus curious whether this double-length TyrRS gene structure is limited to trypanosomatids or has its origin in an event deeper in the plant/plastid lineage³⁴. Indeed there are a small number of TyrRS sequences in other genomes that, like the trypanosomal TyrRS, contain two head-to-tail homologous copies of the TyrRS core domains. Some of these preserve recognizable instances of the characteristic Class Ic aaRS motif structure in both copies; others do not (Figure 2). We did not, however, find evidence for the double-length trypanosomatid sequences being representative of a monomeric TyrRS variant distributed within a larger clade. One possible exception is the TyrRS from the diatom *Phaeodactylum tricornutum*. Like the *L. major* TyrRS, the two halves of this sequence are only 17% sequence identical to each other, and exhibit the same pattern of preserved motifs as the trypanosomatid sequences. The *P. tricornutum* sequence is 64% identical to that of *L. major* TyrRS in the N-terminal half, and 45% identical overall, suggesting a common origin.

Whether or not they are directly related to trypanosomatid sequences, the remaining double-length TyrRS homologs that have been identified as lacking two full sets of characteristic motifs are likely also to behave as pseudo-dimers containing only one fully active set of core domains. The *Schistosoma mansoni* TyrRS is an extreme case where the two copies of the core domains are even more divergent than those in the *L. major* TyrRS. The *S. mansoni* sequence has not only lost the canonical motifs from the C-terminal copy of the catalytic domain, it also no longer has recognizable tRNA-binding motifs in the N-terminal copy of the nominal anticodon-binding domain (Figure 2).

In contrast to this, the double-length TyrRS from *Arabidopsis thaliana* is the result of a recent gene duplication³⁵ that has retained two full complements of the characteristic motifs from both catalytic and anticodon-binding domains. The N- and C-terminal halves of this sequence are 76% sequence identical to each other, which is greater than their individual sequence identity to the nearest currently known homologs from other organisms. The functional state of the active *A. thaliana* enzyme is unknown. Based on the sequence and in light of the current known TyrRS structures, its quaternary structure might range from a monomeric pseudo-dimer to a true dimer to a hypothetical, but unlikely, repeating chain of molecules in which each set of N-terminal core domains forms a canonical TyrRS dimer interaction with the C-terminal set of domains from the next molecule in the chain.

Potential of TyrRS as an anti-trypanosomal drug target

Aminoacyl-tRNA synthetases are central to the key biological function of translating an RNA message into a protein with the corresponding amino acid sequence. The structural elements mediating this function are remarkably maintained across all forms of life. Nevertheless individual aaRS from specific organisms can exhibit idiosyncratic differences.

Many higher eukaryotes, including mammals, have separate cytosolic and mitochondrial forms for each aminoacyl-tRNA synthetase. In contrast, the genomes of many lower eukaryotes, including trypanosomatids, code for only a single aaRS sequence, which is therefore essential. This combination of essential biological function and individual variation suggests that the aaRS from these pathogens may be useful targets for drugs targeting infectious disease [review: Hurdle *et al.*³⁶]. Several anti-aaRS compounds have already shown promise as drugs against eukaryotic pathogens. For example, borrelidin is a highly specific anti-ThrRS macrolide produced by *Streptomyces* that shows anti-malarial activity³⁷. Icofungipen is a competitive inhibitor of IleRS that has been shown to be effective against fluconazole-resistant *Candida albicans*³⁸. Ding *et al.*³⁹ have characterized a compound targeting LeuRS that inhibits growth of bloodstream-form *T. brucei* but is non-toxic to mammalian cell lines. These examples demonstrate that it is possible exploit differences between homologous aaRS in eukaryotic pathogens and their human host as a basis for developing pathogen-specific drugs.

The design of new anti-trypanosomal drugs targeting TyrRS therefore seems promising, as the current structures show that the active site of *L. major* TyrRS exhibits extensive differences to both the cytosolic and mitochondrial isoforms of human TyrRS. In making this comparison, the crystal structure of the yeast TyrRS:tyrosyladenylate:tRNA^{Tyr} complex (PDBID 2dlc)¹⁸ provides a model for the active site of the human enzyme. There are twenty-six residues within 4 Å of the intermediate reaction product tyrosyladenylate in the yeast complex; all but one of these are identical in the human cytosolic isoform. Thirteen of these twenty-six residues are different in the *L. major* enzyme (Figures 4 and 6d). An equivalent amount of divergence is present at these twenty-six positions between the *L. major* TyrRS and the human mitochondrial TyrRS isoform (Table 2). One key point of difference between the *L. major* and human cytosolic TyrRS is the absence of a metal ion in the active site of the *L. major* TyrRS structure. In this regard the active site of the trypanosomal enzyme is more like that of bacterial homologs than that of the yeast or human cytosolic enzymes. In the present case, none of the four residues whose sidechains coordinate K⁺ in the human TyrRS:tyrosinol complex⁷ are conserved in the trypanosomatid sequence (Figure 4). Two of these differences, *L. major* Phe39 rather than the threonine in human TyrRS, and *L. major* Leu96 rather than a tyrosine, result in the introduction of an additional area of hydrophobic surface relative to the active site of the human cytosolic enzyme. In this light, it is particularly interesting that the binding mode observed for fisetin in the *L. major* TyrRS active site is reminiscent of that observed for a series of potent inhibitors with 4×10⁴-fold greater affinity for *S. aureus* TyrRS relative to yeast TyrRS (PDBIDs 1jii, 1jij, 1jik, and 1jil)^{40,30}. This is strong evidence that high selectivity for the trypanosomatid TyrRS over the host TyrRS should likewise be achievable.

Summary

Crystal structures of double-length TyrRS from the parasitic protozoan *L. major* reveal it to be a pseudo-dimer that resembles the canonical dimer observed for the majority of TyrRS enzymes. Each half of the *L. major* enzyme folds into a unit that is typical for the Class I aaRS family, with a Rossmann fold domain and a predominantly helical tRNA-binding domain. Interestingly, only the N-terminal half of the pseudo-dimer contains the canonical Class I aaRS motifs HIGH and KMSMS. This combined with the observation that the C-terminal half of the enzyme is not able to bind substrate analogs strongly suggests that only the N-terminal half of the enzyme possesses a catalytically functional active site. A functional active site in the N-terminal half necessarily implies, and sequence and structural information supports, that the tRNA-binding domain of the C-terminal half is functional since the active site is only accessible to the terminal adenosine of the tRNA^{Tyr} if the anticodon binds in the other “monomer”.

TyrRS homologs from the sequence clade that includes plants and protozoa have been sparsely represented in structural work until now. The two structures determined so far from this clade both show intriguing differences from other TyrRS structures: the Mimivirus enzyme forms a non-canonical homodimer; the *L. major* enzyme forms a pseudo-dimeric monomer. A short insertion following the AC1 motif that is characteristic of plant/plastid TyrRS sequences is observed to form a well-ordered hairpin structure. We speculate that additional structural variation may be found among both the typical and the double-length TyrRS homologs from plants. The availability of the *L. major* structures presented here may also allow extension of recent structure-based analyses of the phylogenetic relationship between TrpRS and TyrRS¹⁰, helping to resolve the ambiguous evidence for the evolutionary origins of this pair of aminoacyl-tRNA synthetases.

Finally, the active site of the sole TyrRS present in *L. major* and related trypanosomatids is sufficiently different from the homologous human enzymes to offer opportunities for the design of selective anti-trypanosomal drugs. In this regard, the structures presented here will serve as a platform for a structure-based drug design program targeting this essential enzyme. These efforts are currently underway.

Materials and Methods

Protein production

The sequence corresponding to GenBank accession no. **AAZ11768.1**, coding for the 682 residue *L. major* TyrRS, was PCR amplified from genomic DNA of *L. major* strain Friedlin and cloned into *E. coli* expression vector BG1861, derived from pET14b⁴¹. Protein was purified by Ni-NTA affinity chromatography followed by size exclusion over a XK 26/60 Superdex 75 column (Amersham Pharmacia Biotech) in MSGPP standard buffer (20mM HEPES, 0.5 M sodium chloride, 2 mM β -mercaptoethanol, 5% glycerol, 0.025% sodium azide at pH 7.5)⁴². Purified protein retained a non-cleavable eight-residue expression tag.

Crystallization

The protein solution was concentrated to 14 mg/ml and supplemented with 1 mM TCEP and 10 mM L-tyrosinol. Crystals were grown by vapor diffusion from sitting drops at 4° C. Crystals of native protein in complex with tyrosinol grew from drops equilibrated against a reservoir containing 16% w/v PEG 3350 and 0.2 M potassium formate at pH 7.5. The initial crystallization drops consisted of 1 μ L of protein solution, 1 μ L of the reservoir solution, and 0.5 μ L suspension of seed microcrystals in MSGPP buffer. Crystals were cryoprotected by soaking in a mixture of 15% ethylene glycol, 10 mM L-tyrosinol, 19% PEG 3350, and 188 mM potassium formate at pH 7.5 prior to being frozen in liquid nitrogen.

Selenomethionyl-substituted protein was grown using the protocols of Studier and SGPP^{43,42}. Crystals of the selenomethionyl protein in complex with fisetin grew from drops equilibrated against a reservoir containing 25% w/v PEG 3350, 0.1 M tri-sodium citrate pH 5.5, and 4% w/v acetone. Initial drops consisted of 0.15 μ L protein at 24 mg/ml in MSGPP buffer containing 5 mM DTT and 5 mM fisetin, plus 0.15 μ L reservoir solution. A mixture of 30% PEG 3350, 10% MSGPP buffer, 10% ethylene glycol, and 0.1 M tri-sodium citrate at pH 5.5 was added to the crystallization drop for cryoprotection prior to mounting and freezing crystals in liquid nitrogen. Crystals of the selenomethionyl protein in complex with tyrosinol grew from drops equilibrated against a reservoir containing 25% w/v PEG 3350, 0.1 M tri-sodium citrate pH 5.5, and 0.01 M Ferric(III) chloride. The initial crystallization drops consisted of 0.15 μ L protein at 24 mg/ml in MSGPP buffer containing 5 mM DTT and 10 mM L-tyrosinol, plus 0.15 μ L reservoir solution. A mixture of 26% PEG 3350, 8% MSGPP buffer, 8% ethylene glycol, and 0.09 M tri-sodium citrate at pH 5.5 was added to

the crystallization drop for cryoprotection prior to mounting and freezing crystals in liquid nitrogen.

Crystals could also be grown in the presence of MgATP alone or in combination with tyrosinol or tyrosine, but these crystals did not diffract.

Crystallographic diffraction images were collected at beamline 9-2 of the Stanford Synchrotron Radiation Lightsource by remote access using an automated crystal handling system⁴⁴. Images were processed using Mosflm⁴⁵ and Scala⁴⁶ or HKL2000⁴⁷.

Structure determination

The X-ray crystal structure of *L. major* TyrRS was solved using a combination of molecular replacement and single wavelength anomalous dispersion methods. Selenomethionyl-substituted protein cocrystallized with the flavonol fisetin, identified through a high throughput chemical library screen, in the cubic space group I23 and diffracted to a resolution of ~3 Å (Table 1). Diffraction data were measured at three X-ray energies spanning the selenium K-absorption edge but the crystal suffered from radiation damage so the data collected first (the inflection point, 0.97931 Å) proved to be the best and was thus used for all steps from structure determination through final refinement.

Processed data was submitted to the BALBES automated MR server⁴⁸. This yielded a partial MR solution composed of two copies of the Mimivirus TyrRS monomer (PDBID 2j5b)¹⁵, both with sequence assigned from the N-terminal half of the *L. major* protein. Residual density in the resulting electron density map was present outside of the partial MR solution and formed clear secondary structural elements that likely belonged to the unmodeled C-terminal halves of both proteins in the asymmetric unit. However, this MR electron density map was still not of sufficient quality to model the remainder of the protein. To improve the electron density map, phase information from the partial MR model was combined with experimental phase information from the anomalous scattering contribution of selenium atoms using the experimental phasing module of Phaser (PhaserEP)⁴⁹. The program identified 36 selenium sites, consistent with an asymmetric unit containing two protein chains and a solvent content of approximately 68%. Indeed, the experimental phases yielded significantly improved electron density maps and many of the selenium sites were situated in readily interpretable electron density outside of the partial MR model. Resolve⁵⁰ was then used for density modification and automated chain tracing starting with only the experimental phase information and the 36 selenium sites. Much of both full N-terminal halves, corresponding essentially to the MR solution, were traced and additionally much of both Rossmann-fold domains from the C-terminal halves. This model was then manually developed in increments using successive cycles of building in Coot⁵¹ and refinement with Refmac⁵². The structure of TyrRS from *A. permix* (PDBID 2cya¹⁷) was used to guide the building of the C-terminal anticodon-binding domains. For a more detailed description of the initial structure determination, please see the Supplemental Information. The improvement of the electron density from partial molecular replacement solution to the incorporation of experimental phasing information to final refinement is shown in Supplemental Figure S1.

The structure from the fisetin complex was then transferred to the isomorphous crystal form of the ty-rosinol complex and placed by MR into the triclinic crystal forms of the tyrosinol complex. Refinement of all three reported structural models proceeded via alternating cycles of refinement in Refmac⁵² and model building in Coot⁵¹. Crystallographic data handling and project management used the CCP4 program suite^{53,54}. Model validation was performed using Molprobity⁵⁵. The refined crystallographic models included a two-segment TLS description for the atomic displacements in each protein chain⁵⁶. Structural superpositions

were performed using Coot and SSM⁵⁷. Structural figures were created and rendered using Molscript⁵⁸ and Raster3D⁵⁹ or with PyMOL⁶⁰. Crystallographic statistics are shown in Table 1.

Thermal melt assays

Protein stability was measured in solution as a function of temperature by adding the hydrophobic dye Sypro Orange (SigmaAldrich). The dye emits low fluorescence in aqueous solution surrounding a well-folded protein, but in the presence of increasing temperature the dye gives increased fluorescent signal as a consequence of binding to hydrophobic patches that are exposed as the protein denatures. In the presence of a high-affinity ligand, the protein will in general be more resistant to thermal denaturation, resulting in a positive shift ΔT_M in the inflection point of the melting curve⁶¹. The effect of substrate binding on trypanosomal tyrosyl-tRNA synthetase 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), permitting high-throughput screening of chemical libraries. Melting curve shifts were measured for the *T. cruzi* TyrRS (GenBank [EAN95560.1](#)) as well as for the *L. major* TyrRS used in structural studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful for the contributions of other members of the MSGPP collaboration, and offer special thanks to Tracy Arakaki, Megan Carter, Liren Xiao, and Li Zhang. This work was funded by NIAID award P01AI067921 (Medical Structural Genomics of Pathogenic Protozoa). 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.

Abbreviations used

TyrRS	tyrosyl tRNA-synthetase
aaRS	aminoacyl tRNA-synthetase
TLS	translation/libration/screw

References

1. 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; 347:203–6. [PubMed: 2203971]
2. Cusack S, Berthet-Colominas C, Hartlein M, Nassar N, Leberman R. A second class of synthetase structure revealed by X-ray analysis of Escherichia coli seryl-tRNA synthetase at 2.5 Å. *Nature*. 1990; 347:249–255. [PubMed: 2205803]
3. 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]
4. Brindelfalk B, Viklund J, Larsson D, Thollesson M, Andersson SGE. Origin and evolution of the mitochondrial aminoacyl-tRNA synthetases. *Mol Biol Evol*. 2007; 24:743–56. [PubMed: 17182897]
5. Brown J, Robb F, Weiss R, Doolittle W. Evidence for the early divergence of tryptophanyl- and tyrosyl-tRNA synthetases. *J Molec Evol*. 1997; 45:9–16. [PubMed: 9211729]

6. Diaz-Lazcoz Y, Aude J, Nitschke P, Chiapello H, Landes-Devauchelle C, Risler J. Evolution of genes, evolution of species: The case of aminoacyl-tRNA synthetases. *Mol Biol Evol.* 1998; 15:1548–1561. [PubMed: 12572618]
7. Yang XL, Otero FJ, Skene RJ, McRee DE, Schimmel P, dePouplana LR. Crystal structures that suggest late development of genetic code components for differentiating aromatic side chains. *Proc Natl Acad Sci U S A.* 2003; 100:15376–80. [PubMed: 14671330]
8. O'Donoghue P, Luthey-Schulten Z. On the evolution of structure in aminoacyl-tRNA synthetases. *Microbiol Mol Biol Rev.* 2003; 67:550–573. [PubMed: 14665676]
9. Huang J, Xu Y, Gogarten J. The presence of a haloarchaeal type tyrosyl-tRNA synthetase marks the opisthokonts as monophyletic. *Mol Biol Evol.* 2005; 22:2142–2146. [PubMed: 16049196]
10. Dong X, Zhou M, Zhong C, Yang B, Shen N, Ding J. Crystal structure of Pyro-coccus horikoshii tryptophanyl-tRNA synthetase and structure-based phylogenetic analysis suggest an archaeal origin of tryptophanyl-tRNA synthetase. *Nucleic Acids Res.* 2010; 38:1401–12. [PubMed: 19942682]
11. Irwin MJ, Nyborg J, Reid BR, Blow DM. The crystal structure of tyrosyl-transfer RNA synthetase at 2–7 Å resolution. *J Mol Biol.* 1976; 105:577–86. [PubMed: 972395]
12. Doublié S, Bricogne G, Gilmore C, Carter C. Tryptophanyl-transfer-RNA synthetase crystal-structure reveals an unexpected homology to Tyrosyl-transfer-RNA synthetase. *Structure.* 1995; 3:17–31. [PubMed: 7743129]
13. 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; 21:3829–40. [PubMed: 12110594]
14. Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, Scola BL, Suzan M, Claverie JM. The 1.2-megabase genome sequence of Mimivirus. *Science.* 2004; 306:1344–50. [PubMed: 15486256]
15. Abergel C, Rudinger-Thirion J, Giege R, Claverie JM. Virus-encoded aminoacyl-tRNA synthetases: structural and functional characterization of mimivirus TyrRS and MetRS. *J Virol.* 2007; 81:12406–17. [PubMed: 17855524]
16. Kobayashi T, Nureki O, Ishitani R, Yaremchuk A, Tukalo M, Cusack S, Sakamoto K, Yokoyama S. Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion. *Nat Struct Biol.* 2003; 10:425–32. [PubMed: 12754495]
17. Kuratani M, Sakai H, Takahashi M, Yanagisawa T, Kobayashi T, Murayama K, Chen L, Liu ZJ, Wang BC, Kuroishi C, Kuramitsu S, Terada T, Bessho Y, Shirouzu M, ichiSekine S, Yokoyama S. Crystal structures of tyrosyl-tRNA synthetases from Archaea. *J Mol Biol.* 2006; 355:395–408. [PubMed: 16325203]
18. 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; 35:4289–300. [PubMed: 17576676]
19. Bonnefond L, Giege R, Rudinger-Thirion J. Evolution of the tRNA(Tyr)/TyrRS aminoacylation systems. *Biochimie.* 2005; 87:873–83. [PubMed: 16164994]
20. Jakes R, Fersht AR. Tyrosyl-tRNA synthetase from Escherichia coli. Stoichiometry of ligand binding and half-of-the-sites reactivity in aminoacylation. *Biochemistry.* 1975; 14:3344–50. [PubMed: 1096941]
21. Fersht AR. Dissection of the structure and activity of the tyrosyl-tRNA synthetase by site-directed mutagenesis. *Biochemistry.* 1987; 26:8031–7. [PubMed: 3442641]
22. Fersht AR, Knill-Jones JW, Bedouelle H, Winter G. Reconstruction by site-directed mutagenesis of the transition state for the activation of tyrosine by the tyrosyl-tRNA synthetase: a mobile loop envelopes the transition state in an induced-fit mechanism. *Biochemistry.* 1988; 27:1581–7. [PubMed: 3284584]
23. Fan E, Baker D, Gelb MH, Buckner FS, Van Voorhis WC, Phizicky E, Dumont M, Mehlin C, Grayhack EJ, Sullivan M, Verlinde CL, DeTitta G, Meldrum D, Merritt EA, Earnest TN, Soltis M, Zucker F, Myler P, Schoenfeld L, Kim D, Worthey EA, LaCount D, Vignali M, Li J, Mondal S, Massey A, Carroll B, Gulde S, Luft JR, DeSoto L, Holl M, Caruthers JM, Bosch J, Robien MA,

- Arakaki T, Holmes MA, LeTrong I, Hol WG. Structural genomics of pathogenic protozoa: An overview. *Methods in Molecular Biology*. 2008; 426:497–513. [PubMed: 18542886]
24. 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]
  25. Yang XL, Skene RJ, McRee DE, Schimmel P. Crystal structure of a human aminoacyl-tRNA synthetase cytokine. *Proc Natl Acad Sci U S A*. 2002; 99:15369–74. [PubMed: 12427973]
  26. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *J Molec Biol*. 2007; 372:774–797. [PubMed: 17681537]
  27. Austin J, First EA. Potassium functionally replaces the second lysine of the KMSKS signature sequence in human tyrosyl-tRNA synthetase. *J Biol Chem*. 2002; 277:20243–8. [PubMed: 11927599]
  28. Kobayashi T, Takimura T, Sekine R, Kelly VP, Kamata K, Sakamoto K, Nishimura S, Yokoyama S. Structural snapshots of the KMSKS loop rearrangement for amino acid activation by bacterial tyrosyl-tRNA synthetase. *J Mol Biol*. 2005; 346:105–17. [PubMed: 15663931]
  29. Zhang Y, Wang L, Schultz PG, Wilson IA. Crystal structures of apo wild-type *M. jannaschii* tyrosyl-tRNA synthetase (TyrRS) and an engineered TyrRS specific for O-methyl-L-tyrosine. *Protein Sci*. 2005; 14:1340–9. [PubMed: 15840835]
  30. Qiu X, Janson CA, Smith WW, Green SM, McDevitt P, Johanson K, Carter P, Hibbs M, Lewis C, Chalker A, Fosberry A, Lalonde J, Berge J, Brown P, Houge-Frydrych CS, Jarvest RL. Crystal structure of *Staphylococcus aureus* tyrosyl-tRNA synthetase in complex with a class of potent and specific inhibitors. *Protein Sci*. 2001; 10:2008–16. [PubMed: 11567092]
  31. Moreira D, Lopez-Garcia P. Comment on “The 1.2-megabase genome sequence of Mimivirus”. *Science*. 2005; 308:1114. [PubMed: 15905382]
  32. Wakasugi K, Schimmel P. Two distinct cytokines released from a human aminoacyl-tRNA synthetase. *Science*. 1999; 284:147–51. [PubMed: 10102815]
  33. Yang XL, Liu J, Skene R, McRee DE, Schimmel P. Crystal structure of an EMAP-II-like cytokine released from a human tRNA synthetase. *Helvetica Chimica Acta*. 2003; 86:1246–1257.
  34. Burki F, Shalchian-Tabrizi K, Pawlowski J. Phylogenomics reveals a new ‘megagroup’ including most photosynthetic eukaryotes. *Biol Lett*. 2008; 4:366–9. [PubMed: 18522922]
  35. Duchene AM, Giritch A, Hoffmann B, Cognat V, Lancelin D, Peeters NM, Zaepfel M, Marechal-Drouard L, Small ID. Dual targeting is the rule for organellar aminoacyl-tRNA synthetases in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A*. 2005; 102:16484–9. [PubMed: 16251277]
  36. Hurdle JG, O’Neill AJ, Chopra I. Prospects for aminoacyl-tRNA synthetase inhibitors as new antimicrobial agents. *Antimicrob Agents Chemother*. 2005; 49:4821–33. [PubMed: 16304142]
  37. Otoguro K, Ui H, Ishiyama A, Arai N, Kobayashi M, Takahashi Y, Masuma R, Shiomi K, Yamada H, Omura S. In vitro antimalarial activities of the microbial metabolites. *J Antibiot (Tokyo)*. 2003; 56:322–4. [PubMed: 12760691]
  38. Hasenoehrl A, Galic T, Ergovic G, Marsic N, Skerlev M, Mittendorf J, Geschke U, Schmidt A, Schoenfeld W. In vitro activity and in vivo efficacy of icofungipen (PLD-118), a novel oral antifungal agent, against the pathogenic yeast *Candida albicans*. *Antimicrob Agents Chemother*. 2006; 50:3011–8. [PubMed: 16940096]
  39. Ding D, Meng Q, Gao G, Zhao Y, Wang Q, Nare B, Jacobs R, Rock F, Alley MRK, Plattner JJ, Chen G, Li D, Zhou H. Design, Synthesis, and Structure-Activity Relationship of *Trypanosoma brucei* Leucyl-tRNA Synthetase Inhibitors as Antitrypanosomal Agents. *J Med Chem*. 2011; 54:1276–1287. [PubMed: 21322634]
  40. Stefanska AL, Coates NJ, Mensah LM, Pope AJ, Ready SJ, Warr SR. SB-219383, a novel tyrosyl tRNA synthetase inhibitor from a *Micromonospora* sp. I. Fermentation, isolation and properties. *J Antibiot (Tokyo)*. 2000; 53:345–50. [PubMed: 10866215]
  41. 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, Gray-hack EJ. A facile method for high-throughput co-expression of protein pairs. *Mol Cell Proteomics*. 2004; 3:934–938. [PubMed: 15240823]

42. Mehlin C, Boni E, Buckner FS, Engel L, Feist T, Gelb M, Haji L, Kim D, Liu C, Mueller N, Myler PJ, Reddy JT, Sampson JN, Subramanian E, Van Voorhis WC, Worthey E, Zucker F, Hol WGJ. Heterologous expression of proteins from plasmodium falciparum: results from 1000 genes. *Molecular and Biochemical Parasitology*. 2006; 148:144–160. [PubMed: 16644028]
43. Studier FW. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif*. 2005; 41:207–34. [PubMed: 15915565]
44. McPhillips TM, McPhillips SE, Chiu HJ, Cohen AE, Deacon AM, Ellis PJ, Garman E, Gonzalez A, Sauter NK, Phizackerley RP, Soltis SM, Kuhn P. Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines. *J Synchrotron Radiat*. 2002; 9:401–6. [PubMed: 12409628]
45. 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.
46. Evans P. Scaling and assessment of data quality. *Acta Cryst*. 2006; D62:72–82.
47. Otwinowski, Z.; Minor, W. Processing of X-ray diffraction data collected in oscillation mode. In: Carter, CW., Jr; Sweet, RM., editors. *Methods Enzymol*. Vol. 276. Academic Press; New York: 1997. p. 307–326. *Macromolecular Crystallography, Part A* edition
48. Long F, Vagin AA, Young P, Murshudov GN. BALBES: a molecular-replacement pipeline. *Acta Cryst*. 2008; D64:125–132.
49. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. *Phaser* crystallographic software. *Journal of Applied Crystallography*. 2007; 40:658–674. [PubMed: 19461840]
50. Terwilliger TC. Solve and resolve: automated structure solution and density modification. *Methods Enzymol*. 2003; 374:22–37. [PubMed: 14696367]
51. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Cryst*. 2010; D66:486–501.
52. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst*. 1997; D53:240–255.
53. Collaborative Computational Project No. 4. The ccp4 suite: programs for protein crystallography. *Acta Cryst*. 1994; D50:760–763.
54. Potterton L, McNicholas S, Krissinel E, Gruber J, Cowtan K, Emsley P, Murshudov GN, Cohen S, Perrakis A, Noble M. Developments in the ccp4 molecular-graphics project. *Acta Cryst*. 2004; D60:2288–94.
55. Lovell S, Davis I, Arendall WB III, deBakker P, Word J, Prisant M, Richardson J, Richardson D. Structure validation by α geometry: ϕ , ψ and $C\beta$ deviation. *Proteins: Structure, Function, and Genetics*. 2003; 50:437–450.
56. Painter J, Merritt EA. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Cryst*. 2006; D62:439–450.
57. 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.
58. Kraulis P. Molscript: a program to produce both detailed and schematic plots of protein structures. *J Appl Cryst*. 1991; 24:946–95.
59. Merritt EA, Bacon DJ. Raster3D - photorealistic molecular graphics. *Meth Enzymol*. 1997; 277:505–524. [PubMed: 18488322]
60. DeLano, W. The PyMOL Molecular Graphics System. 2002. <http://www.pymol.org>
61. Lo MC, 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; 332:153–9. [PubMed: 15301960]
62. Beitz E. TeXshade: shading and labeling multiple sequence alignments using LaTeX2e. *Bioinformatics*. 2000; 16:135–139. [PubMed: 10842735]

Figure 1. Phylogenetic relationship of TyrRS sequences

Cartoon representation of the phylogenetic relationship of TyrRS sequences from trypanosomatids, Mimivirus, plants, opisthokont cytosolic homologs, mitochondrial homologs, archaea, and bacteria. The cartoon combines information from previously published work^{4,15,19}.

Figure 2. Gene structure and motifs in double-length TyrRS homologs, other eukaryotes and archaea

The coloring of the catalytic (brown) and anticodon-binding (blue) domains corresponds to that of the *L. major* TyrRS structure shown in Figure 3. The HIGH and KMSKS active site motifs are common to the Class I catalytic domain. The AIDQ motif is characteristic of the ATP-binding site in TyrRS and TrpRS. The AC1 motif (KIKKAYS in *L. major* TyrRS) corresponds to the residues in helix $\alpha 11$ of the eukaryotic/archaeal TyrRS anticodon-binding domain that interact with the anticodon stem of tRNA^{Tyr}. The AC2 motif (LHPADLK in *L. major* TyrRS) contains the residues in helix $\alpha 14$ that specifically recognize the anticodon bases G34 and U/ ψ 35.

Figure 3. Pseudo-dimeric *L. major* TyrRS and true dimeric human TyrRS

The inherently asymmetric *L. major* pseudo-dimer shown is that of the tyrosinol:TyrRS complex (PDBID 3p0i). The human dimer is perfectly symmetric as seen in the crystal structure of the two core TyrRS domains in complex with tyrosinol (PDBID 1q11)⁷. The catalytic domain is colored brown in both structures, dark brown for the N-terminal (functional) copy in *L. major* TyrRS, light brown for the C-terminal (presumed nonfunctional) copy in *L. major* TyrRS, intermediate brown in both monomers of the human TyrRS. The anticodon-binding domain is colored blue in the *L. major* TyrRS, dark blue for the N-terminal copy and light blue for the C-terminal (functional) copy. The anticodon-binding domain is colored yellow in both monomers of human TyrRS. The linker residues connecting the N- and C-terminal halves of the *L. major* TyrRS are shown in magenta; the analogous residues in human TyrRS are also shown in magenta in the second monomer. Crystallographically observed bound tyrosinol molecules in both structures are shown in green. Figure generated using PyMOL⁶⁰.

Figure 4. Sequence alignment and secondary structure of *L. major* and human cytosolic TyrRS
 Structure-based alignment of the N- and C-terminal halves of the *L. major* TyrRS sequence together with the sequence of the core domains of human cytosolic TyrRS. Secondary structure elements for the N-terminal half of *L. major* TyrRS from the cubic form of the tyrosinol complex are shown above the sequences, except that the extended part of the longer $\alpha 5$ helix seen in the fisetin complex is shown in green (residues 84–88). Secondary structure elements for the C-terminal half of *L. major* TyrRS are shown underneath the sequences. Blue boxes indicate the canonical Class Ic catalytic domain motifs. Red boxes labeled AC1 and AC2 indicate the regions which are observed to interact with the tRNA anticodon stem in the yeast TyrRS:tyrosyladenylate:tRNA^{Tyr} complex (PDBID 2dlc)¹⁸. Residues of the human sequence that are within 4 Å of the superimposed tyrosyladenylate molecule from the yeast complex are highlighted in cyan. Residues coordinating a K⁺ ion in the human structure are highlighted in magenta. Figure generated using T_EXshade⁶².

Figure 5. Two conformational states of the active site are observed in *L. major* TyrRS structures
 The closed state of active site in the N-terminal half of *L. major* TyrRS as seen in the tyrosinol complex is shown in green. This state is characterized by a well-ordered 148–149 loop and a ‘curled’ conformation of the $\alpha 4$ – $\alpha 5$ loop. The open state of active site as seen in the fisetin complex is shown in yellow. In this state, the 148–149 loop is disordered and the $\alpha 4$ – $\alpha 5$ loop is present in an ‘extended’ conformation that is in contact with helix $\alpha 7'$ across the “pseudo-dimer” interface (shown in cyan). This is analogous to the active site of one monomer from a canonical TyrRS dimer establishing contact with the second monomer across the dimer interface.

Figure 6. Active site of *L. major* TyrRS; with electron density for ligands, and highlighting substantial sequence divergence from human TyrRSs

(a) Active site of *L. major* TyrRS with bound tyrosinol. Electron density is contoured at 4σ in an ($mF_{obs} - F_{calc}$) difference map with tyrosinol omitted from F_{calc} . (b) Active site of *L. major* TyrRS with bound fisetin. Electron density is contoured at 4σ in an ($mF_{obs} - F_{calc}$) difference map with fisetin omitted from F_{calc} . (c) Superposition of *L. major* TyrRS complexes with tyrosinol (yellow; PDBID 3p0h), fisetin (cyan; PDBID 3p0i), and yeast TyrRS with tyrosyladenylate (green; PDBID 2dlc). (d) Stereo view of the same superimposed ligands from panel c in the *L. major* TyrRS active site pocket, shown as a transparent surface so the rather buried ligands are not mostly concealed. Residues within 8 Å of the superimposed tyrosyladenylate are colored to highlight the substantial differences between the *L. major* enzyme and the two human homologs. Identical residues are light green, those that differ from the mitochondrial enzyme are yellow, those that differ from the cytoplasmic enzyme are red, and those that differ from both human enzymes are orange. The view into the pocket is approximately 50° counterclockwise around the Y-axis with respect to the orientation in Figure 3.

Figure 7. Comparison of anticodon recognition regions of the N- and C-terminal halves of *L. major* TyrRS and yeast TyrRS

The N-terminal (dark blue) and C-terminal (light blue) copies of the *L. major* anticodon-binding domain are shown superimposed onto the yeast TyrRS:tRNA^{Tyr} complex (PDBID 2dlc) (yellow). The three anticodon bases are labeled. The C-terminal region is more structurally similar that of the yeast enzyme than either are to that of the N-terminal region.

Figure 8. Thermal denaturation curves

(a) Thermal denaturation curves for *L. major* TyrRS alone and in the presence of tyrosine, ATP, tyrosine and ATP both, or fisetin. **(b)** Thermal denaturation curves for *T. cruzi* TyrRS alone and in the presence of tyrosinol, ATP, or tyrosinol and ATP both.

Table 1

Crystallographic data collection and refinement statistics

PDBID	TyrRS:fisetin 1c46b2	TyrRS:tyrosinol 1652b1	TyrRS:tyrosinol 1541b3
Data collection	SeMet	SeMet	native
Wavelength (Å)	0.97931	0.97908	0.97946
Resolution (Å)	47.1-3.0 (3.16-3.0)	85.3-3.13 (3.3-3.13)	47.6-2.89 (3.0-2.89)
Space group	I23	I23	P1
Unit cell (Å)	a=b=c=240.2	a=b=c=241.4	a=57.5 b=111.9 c=141.0 $\alpha=74.47^\circ \beta=80.83^\circ \gamma=77.13^\circ$
Monomers per asymmetric unit	2	2	4
Unique reflections	46076 (6696)	41253 (5946)	72452 (7220)
R_{merge}	0.15 (1.58)	0.182 (1.42)	0.084
Completeness (%)	99.9 (99.8)	99.9 (99.8)	98.6 (97.9)
Mean $I/\sigma(I)$	14.0 (1.7)	7.3 (1.3)	11.3 (1.1)
Redundancy	13.8 (10.5)	8.0 (7.0)	3.6 (3.5)
Wilson B factor (Å ²)	86.8	99.1	84.6
Refinement			
Resolution (Å)	47.1-3.0	85.3-3.13	47.6-2.89
R/R_{free}	0.23/0.27	0.24/0.27	0.24/0.29
Number of protein atoms	9723	9287	19650
Number of other atoms	42	24	50
Mean B_{eq} protein (Å ²)	80.8	122	143
rmsd deviation from ideal bond lengths (Å)	0.006	0.010	0.008
rmsd deviation from ideal bond angles (°)	0.90	1.16	1.05
TLS groups per chain	2	2	2
Residues with ϕ, ψ in favored regions (%)	96.9	94.0	95.1
Residues with ϕ, ψ in allowed regions (%)	99.8	99.3	99.8
Number of ϕ, ψ outliers (residues)	2	8	6

Table 2
Residue differences at the the active site of trypanosomal and human TyrRS

At each of these positions the side chain of the corresponding residue in yeast TyrRS is within 4 Å of the reaction intermediate tyrosyladenylate as seen in the crystal structure of the complex (PDBID 2dlc). The human mitochondrial enzyme is not structurally homologous in the region of *L. major* residue 149.

<i>L. major</i>	human cytosolic	human mitochondrial
Asp 37	Trp 40	Cys 78
Phe 39	Thr 42	Phe 80
Glu 40	Ala 43	Asp 81
Pro 41	Thr 44	Pro 82
Gln 49	Tyr 52	His 91
Lys 53	Pro 55	Ala 94
Trp 70	Leu 72	Leu 111
Phe 75	His 77	Thr 116
Met 149	Val 153	—
Cys 179	Gln 182	Gln 241
Leu 181	Gly 184	Gly 243
Leu 183	Ile 186	Ser 245
His 210	Asn 212	Val 271
His 211	Pro 213	Pro 272
Met 212	Met 214	Leu 273
Leu 213	Val 215	Ile 274