

RESEARCH

Open Access



Aminoacyl tRNA synthetases as malarial drug targets: a comparative bioinformatics study

Dorothy Wavinya Nyamai and Özlem Tasthan Bishop*

Abstract

Background: Treatment of parasitic diseases has been challenging due to evolution of drug resistant parasites, and thus there is need to identify new class of drugs and drug targets. Protein translation is important for survival of malarial parasite, *Plasmodium*, and the pathway is present in all of its life cycle stages. Aminoacyl tRNA synthetases are primary enzymes in protein translation as they catalyse amino acid addition to the cognate tRNA. This study sought to understand differences between *Plasmodium* and human aminoacyl tRNA synthetases through bioinformatics analysis.

Methods: *Plasmodium berghei*, *Plasmodium falciparum*, *Plasmodium fragile*, *Plasmodium knowlesi*, *Plasmodium malariae*, *Plasmodium ovale*, *Plasmodium vivax*, *Plasmodium yoelii* and human aminoacyl tRNA synthetase sequences were retrieved from UniProt database and grouped into 20 families based on amino acid specificity. These families were further divided into two classes. Both families and classes were analysed. Motif discovery was carried out using the MEME software, sequence identity calculation was done using an in-house Python script, multiple sequence alignments were performed using PROMALS3D and TCOFFEE tools, and phylogenetic tree calculations were performed using MEGA vs 7.0 tool. Possible alternative binding sites were predicted using FTMap webserver and SiteMap tool.

Results: Motif discovery revealed *Plasmodium*-specific motifs while phylogenetic tree calculations showed that *Plasmodium* proteins have different evolutionary history to the human homologues. Human aaRSs sequences showed low sequence identity (below 40%) compared to *Plasmodium* sequences. Prediction of alternative binding sites revealed potential druggable sites in PfArgRS, PfMetRS and PfProRS at regions that are weakly conserved when compared to the human homologues. Multiple sequence analysis, motif discovery, pairwise sequence identity calculations and phylogenetic tree analysis showed significant differences between parasite and human aaRSs proteins despite functional and structural conservation. These differences may provide a basis for further exploration of *Plasmodium* aminoacyl tRNA synthetases as potential drug targets.

Conclusion: This study showed that, despite, functional and structural conservation, *Plasmodium* aaRSs have key differences from the human homologues. These differences in *Plasmodium* aaRSs can be targeted to develop anti-malarial drugs with less toxicity to the host.

Keywords: Aminoacyl tRNA synthetases, Motif analysis, Phylogenetic tree calculations, Homology modelling, Allosteric site

*Correspondence: o.tasthanbishop@ru.ac.za

Research Unit in Bioinformatics (RUBi), Department of Biochemistry and Microbiology, Rhodes University, Grahamstown 6140, South Africa



Background

Parasitic diseases like trypanosomiasis, malaria, leishmaniasis and filariasis affect millions of people in the world yearly [1–4]. These diseases cause a remarkable burden in economic development and health of affected countries and thus the need to devise control and prevention strategies. Currently, the main mode of prevention and treatment of these parasitic diseases is by use of drugs as there are no approved vaccines in the market [5]. However, most parasites have developed resistance against conventional drugs leading to the drugs being ineffective [6–10]. Thus, there is need to develop new classes of drugs and to identify drug targets to solve the shortcoming of drug resistance. Targeting housekeeping pathways such as protein translation may help deal with drug resistance as they are important for the survival of most parasites [11–13].

Plasmodium parasites cause malaria, which is a major public concern due to its high mortality and morbidity rates [13, 14]. There are five *Plasmodium* species that cause malaria in human, namely *Plasmodium falciparum*, *Plasmodium knowlesi*, *Plasmodium malariae*, *Plasmodium ovale* and *Plasmodium vivax* [15]. *Plasmodium* has three genomes; cytoplasm, mitochondrial and apicoplast, and each of them needs a functional protein translation mechanism for growth and survival [13, 16, 17]. *Plasmodium* proteins involved in protein translation machinery are generally encoded by the nuclear genome and exported to target organelles to carry out various functions in protein synthesis [16, 18–20].

Aminoacyl tRNA synthetases (aaRSs) are a group of key enzymes in protein translation pathway; they catalyze the first reaction, where an amino acid is added to the cognate tRNA molecule in the presence of ATP and magnesium (Mg^{2+}) ions. This reaction takes place in two steps; first ATP activates the amino acid through formation of aminoacyl-adenylate intermediate, while the second step involves ligation of the adenylate intermediate to the cognate tRNA molecule through a covalent bond generating AMP [11, 12, 21]. Although the canonical function of these enzymes is to add amino acids to tRNA for translation and they are highly conserved in their catalytic domains, in general, aaRSs show sequence, structural and functional diversity across organisms [22]. Furthermore, in some organisms, aaRSs have evolved to perform non-canonical functions such as angiogenesis, RNA splicing, signaling events, transcription regulation, apoptosis and immune responses [23–25]. *Plasmodium falciparum* tyrosyl-tRNA synthetases (PfTyrRS), for instance, have cytokine-like functions, while eukaryotic methionyl-tRNA synthetases (MetRS) have glutathione-S-transferase domains that play a key role in protein-protein interactions [26, 27]. *Plasmodium falciparum*

lysyl-tRNA synthetase (PfLysRS) synthesizes diadenosine polyphosphate, a signaling molecule that plays a role in gene expression, DNA replication and regulation of ion channels of the parasite [28, 29].

Of the five human malaria parasites, *P. falciparum* causes the most severe form of malaria, and is responsible for most of the malaria mortality cases reported across the world [30]. *Plasmodium falciparum* has a total of 36 aaRSs that are asymmetrically distributed in either the cytoplasm, mitochondria or the apicoplast compartments. Of the 36 *P. falciparum* aaRSs, 15 reside in the apicoplast, 16 in the cytoplasm and four in mitochondria: AlaRS, GlyRS, ThrRS and CysRS are found both in the apicoplast and the cytoplasm and each of the four is encoded by a single gene and exported to the two compartments while only phenylalanine-tRNA synthetase (PheRS) is encoded in the mitochondria [20, 31, 32]. *Plasmodium falciparum* protein translation in the mitochondria relies on enzymes imported from the cytoplasm including aaRSs [31]. The apicoplast encodes AspRS, PheRS, ValRS, LysRS, HisRS, AsnRS, ProRS, SerRS, TrpRS, ArgRS, IleRS, GluRS, LeuRS, TyrRS and MetRS while AlaRS, CysRS ThrRS and GlyRS are reported to have a single gene encoding both the cytoplasm and apicoplast enzyme [18, 20, 32, 33]. A single transcript for each gene is spliced alternatively to generate the two isoforms for each protein which are then targeted to either the cytosol or the apicoplast [20, 32]. Each of these genes encodes a protein with a N-terminal extension that corresponds to a signal and transit peptide and is conserved in the apicomplexa phylum [20]. *Plasmodium falciparum* cytoplasm has genes that encode ProRS, AspRS, IleRS, LysRS, HisRS, PheRS, AsnRS, ArgRS, GlnRS, SerRS, TrpRS, ValRS, MetRS, LeuRS, GluRS and TyrRS [33–38].

In human, aaRSs carry out aminoacylation reactions in the cytoplasm, nucleus and the mitochondria. After tRNA is encoded in the nucleus, it is transported to the cytoplasm where protein translation takes place [11]. The human mitochondria acquires nuclear-encoded aaRSs with the aid of translation signals within the aaRSs proteins to carry out protein synthesis [39]. The cytoplasm is the only compartment where both aminoacylation and protein synthesis exclusively takes place in humans. Human aaRSs are, thus, classified as mitochondrial or cytoplasmic based on the compartment where they are localized [39]. In human, a total of 36 aaRSs have been reported with 17 of them in the mitochondrion and 16 aaRSs exclusively functioning in the cytoplasm while the other three catalyze aminoacylation reactions in both organelles [11, 39]. The three bifunctional aaRSs in human are GlnRS, GlyRS and LysRS. In the cytoplasm, aminoacylation of proline and glutamate is catalyzed by a single bifunctional enzyme (Glu/ProRS). Thus, both

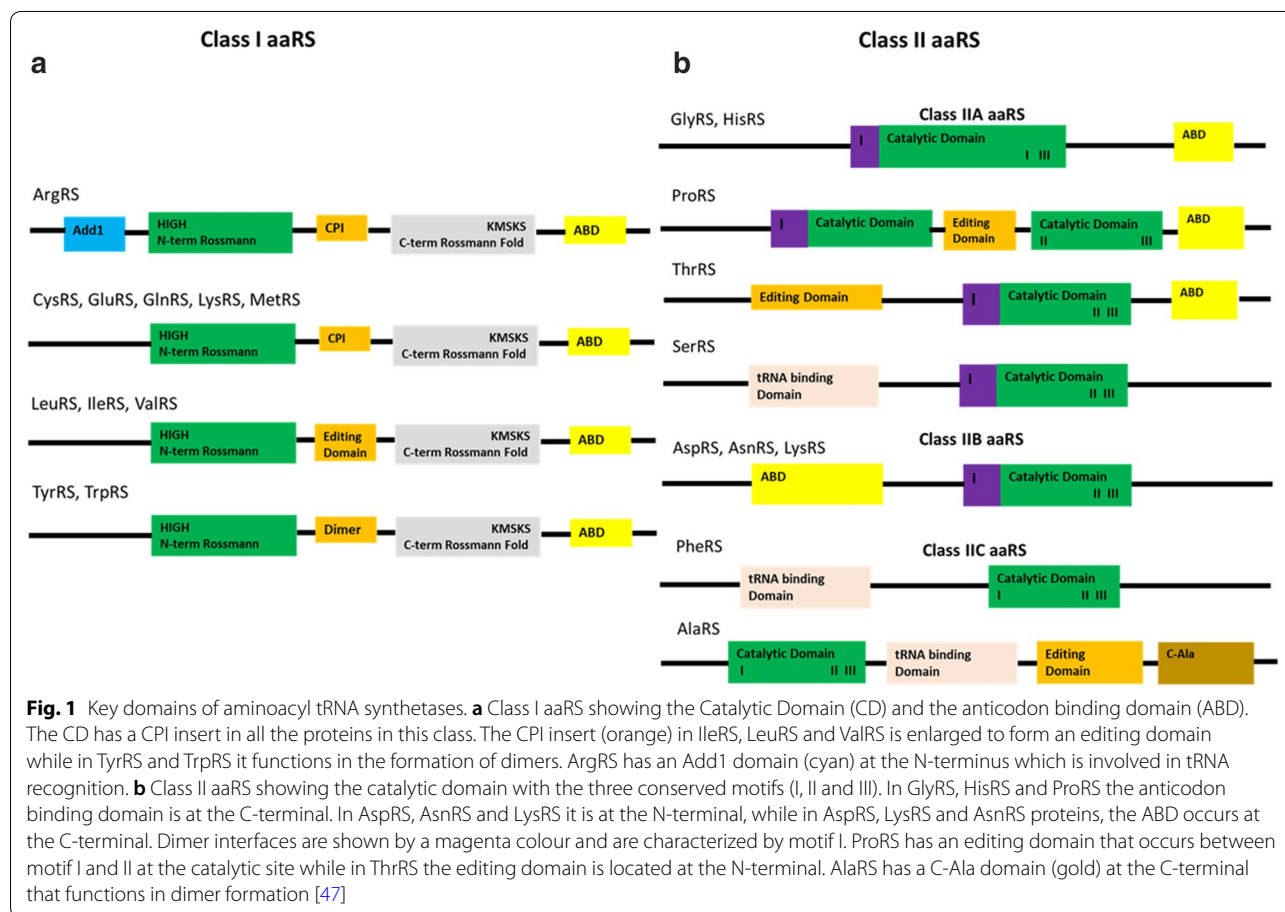
compartments have enzymes for charging all the 20 amino acids [39, 40].

Generally, aaRSs proteins are classified into two distinct classes based on key features of the catalytic site architecture and the manner of charging tRNA [21, 23]. Class I aaRSs include IleRS, LeuRS, MetRS, CysRS, GlnRS, GluRS, TrpRS, ValRS, ArgRS and TyrRS. Proteins in this class have a catalytic domain (Fig. 1a) characterized by a Rossmann fold (RF) located near the N-terminal [41]. The catalytic domain of this class comprises five parallel β -sheet strands flanked by α -helices. The RF possesses highly conserved HIGH and KMSKS motifs separated by a loop [42, 43] as shown in Fig. 1a. The HIGH motif is located in a region formed by a loop linking the first β -sheet strand and the adjacent α -helix while the KMSKS motif occurs after the fifth β -sheet strand [11]. The RF domain has an insert known as the connective peptide I (CPI) in all enzymes in this class whose structure is characteristic of mixed α and β folds. Proteins in this class have common domains that include an alpha-helical anticodon binding domain (ABD), connective peptide (CPI) and the tRNA stem contact fold [44]. The CPI insert is found towards the end of the first half of

the fourth β -strand of the RF joining the N-terminal and C-terminal sections of the catalytic domain [11].

With the exception of TyrRS, MetRS and TrpRS, all Class I enzymes are monomeric [11]. In monomeric enzymes, the CPI binds tRNA at the 3'-single stranded end while in TrpRS and TyrRS it forms the dimer interface of these dimeric enzymes [41, 45]. In ValRS, IleRS and LeuRS, the CPI insert is enlarged (250–275 amino acid residues as compared to CysRS and MetRS where it is 50 and 100 residues respectively) to include an editing domain for editing misacylated tRNA through hydrolysis [46]. The editing domain proofreads the aminoacylation process through pre-transfer or post-transfer editing [11]. Post-transfer editing involves hydrolyzing of misacylated tRNA to amino acid and tRNA while pre-transfer modification hydrolyzes the mis-activated aminoacyl adenylate to AMP and amino acid [11]. The ABD of proteins in Class I occurs at the C-terminal which binds the anticodons in the cognate tRNA [47].

Class I enzymes binds to the tRNA acceptor end through the minor groove and these enzymes aminoacylate the 2'-OH group of adenosine nucleotide [11, 47]. Proteins in this class can further be classified into

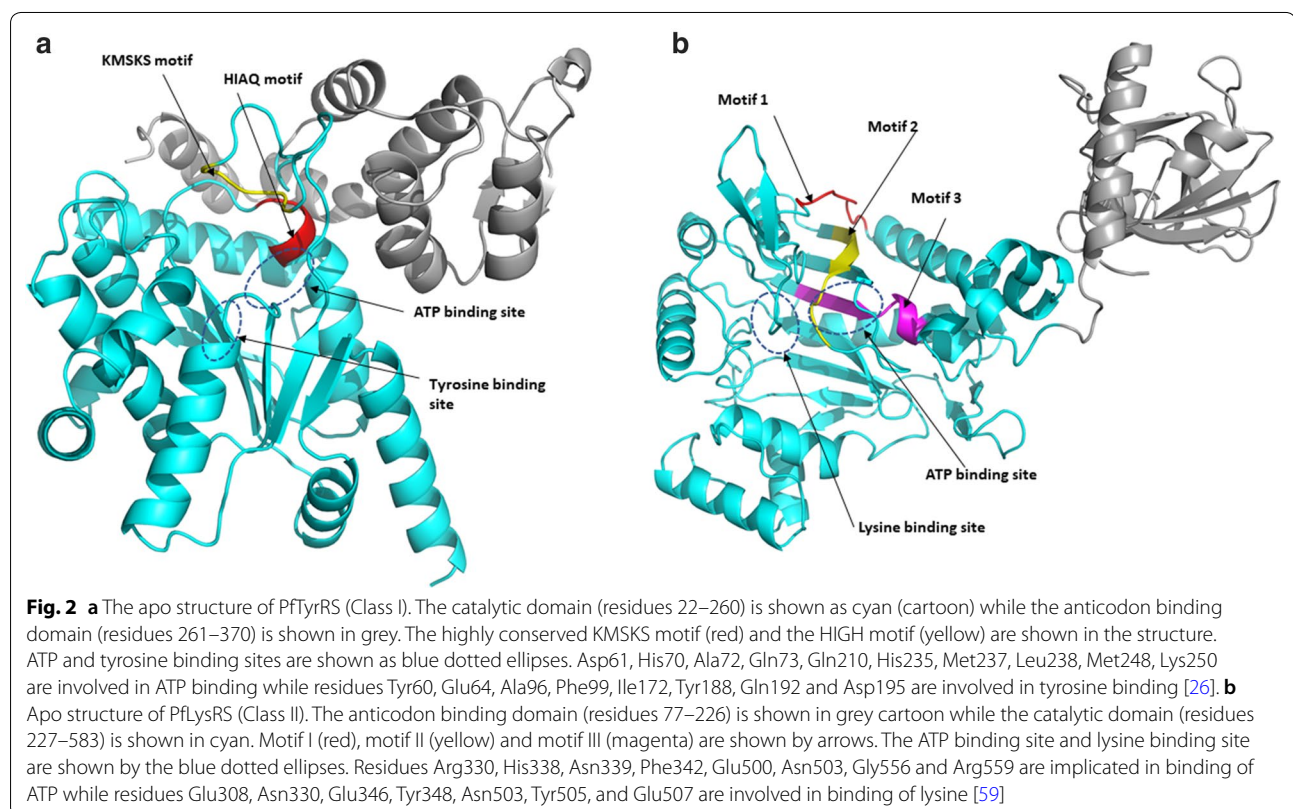


five subclasses based on sequence similarity and physico-chemical properties of their substrates [48, 49]. Subclass Ia members charge hydrophobic amino acids that have aliphatic side chains and include ValRS, MetRS, IleRS and LeuRS. Subclass Ib proteins have charged amino acids as their substrates and include GlnRS, CysRS and GluRS. Members of subclass Iib bind to the cognate tRNA before carrying out the aminoacylation process [11, 50]. TrpRS and TyrRS belong to subclass Ic and their substrates are aromatic amino acids. ArgRS is the only member of subclass Id and it possesses an Add1 domain at the N-terminal whose function is to recognize the D-loop in the tRNA core (Fig. 1a) [11, 47]. Class I LysRS found in some bacteria and archaea shares structural similarity with subclass Ib, but it has a unique alpha helix cage and is thus grouped in subclass Ie [51].

Class II aaRSs include HisRS, ProRS, LysRS, SerRS, AspRS, ThrRS, AlaRS, GlyRS, PheRS and AsnRS. Proteins in this class are further grouped in three subclasses whose members are more closely related than other subclasses [52, 53]. Class IIa proteins exist as dimers and includes ProRS, SerRS, GlyRS, ThrRS, HisRS and all have the aminoacylation domain at the N-terminal [11]. Members of this subclass have an ABD at the C-terminal (Fig. 1b). The anticodon binding domain is absent in SerRS as this protein does not require an anticodon

to discriminate its cognate tRNA [54, 55]. ProRS has editing domains located between motifs I and II at the catalytic domain while in ThrRS the editing domain is at the N-terminus (Fig. 1b) [47, 49]. Members of Class IIb are dimers and have a C-terminal catalytic domain that is structurally similar and include AspRS, LysRS and AsnRS. The ABD in this subclass is located at the N-terminal (Fig. 1b). Class IIc includes PheRS, AlaRS and GlyRS and all exist in tetrameric conformation [11, 53]. AlaRS possesses a C-Ala domain at the C-terminal which is absent in other members of Class IIc. The editing domain in AlaRS occurs between the tRNA binding domain and the C-Ala domain (Fig. 1b) [47].

Class II enzymes possess a catalytic site domain characterized by seven β -sheet strands connected by α -helices [56]. This domain, just like the Class I catalytic domain couples amino acid, ATP and tRNA 3'-terminus during catalytic reactions [47, 57]. Class II catalytic domain has three weakly conserved motifs (Figs. 1b, 2b); Motif I found at the N-terminal of the catalytic region is characterized by a long α -helix linked to a short β -strand with a proline residue at the end which is highly conserved and is involved in homo dimerization [41]. Motif II juxtaposes amino acid, ATP and tRNA and comprise β -sheet strands. Motif III is located at the C-terminal of the catalytic domain and binds ATP and comprise alternate



β -strands and α -helices [43]. LysRS can be classified in both classes based on the structure and mode of charging tRNA, with Class I LysRS occurring in some bacteria and most archaea [58] while Class II LysRS occurs in most bacteria and all eukaryotes [11].

Protein translation has been explored as a target in the development of antimalarial drugs with most compounds interfering with the ribosome [60]. DDD107498 compound has been reported to target the blood stages of *Plasmodium* and the mechanism of action is believed to be inhibition of translation elongation factor 2 (eEF2) which is responsible for translocation of ribosome across mRNA [61]. Recently, there has been increased interest in exploring *P. falciparum* aaRSs as potential drug targets [18, 28, 33, 36, 60, 62, 63]. *Plasmodium* aaRSs inhibitors have been identified that target either the ATP pocket, the amino acid or tRNA binding site or the editing domains of some of these enzymes. Some of the compounds reported to target *P. falciparum* aaRSs are halofuginone, cladosporin, 3-aminomethyl benzoxaborole AN6426, bicyclic azetidine BRD3444, glyburide and TCMDC-124506 [35, 36, 63–65]. Halofuginone, a derivative of febrifugine, targets ProRS tRNA and proline binding site mimicking tRNA 3'-Adenine 76 and L-pro in an ATP dependent manner [62, 66, 67]. Halofuginone binding to human and *Plasmodium* ProRS involves identical residues and in both the compound mimics proline and adenine substrates binding pose thus leading to toxicity in human cells [64, 68, 69]. Cladosporin, a secondary metabolite from fungi, is reported to have activity against blood and liver stage *P. falciparum* and its activity is selective to only the parasite LysRS protein [28, 35]. Cladosporin, an adenosine analogue binds at the ATP binding site of PfLysRS [28, 35]. Cladosporin can, thus, be used as a basis for development of other scaffolds with improved drug-like properties. The compound 3-aminomethyl benzoxaborole AN6426 was reported to be active against LeuRS in drug resistant *P. falciparum* but did not impair growth of the wild type [36]. This compound binds to the editing domain of PfLeuRS and inhibits it inactivating the 3' Adenine 76 nucleotide of the cognate tRNA covalently and the catalytic turnover of *P. falciparum* resistant strains [36].

Glyburide and TCMDC-124506 are reported to bind to a site adjacent to the ATP binding site of PfProRS and displace key residues involved in ATP binding thus inhibiting the enzyme activity [65]. Glyburide and TCMDC are selective to PfProRS and do not cause toxicity to human cells and thus can be used as a basis for development of drugs targeting PfProRS [65]. BRD444, a bicyclic azetidine inhibits *P. falciparum* blood stages. In vitro studies on resistant *P. falciparum* showed non-synonymous single-nucleotide variant at the locus that encodes the alpha

subunit of cytosolic PfPheRS [63]. Assays on recombinant PfPheRS showed that bicyclic azetidines inhibits aminoacylation of PfPheRS in a concentration-dependent manner confirming that this protein is the molecular target for bicyclic azetidines [63].

Due to these shortcomings of the current compounds that target aaRSs and the ever-increasing antimalarial drug resistance [6, 18, 19, 22, 30, 70], there is need to develop novel drugs and identify more targets to counter this resistance. In addition, the development of drugs that are active against the liver, blood stage parasites [70] and the sexual stages of the parasites thus terminating the infection cycle would help in malaria eradication [71]. With aaRSs proteins being present in all stages of the parasite life cycle, identification of subtle differences between the *Plasmodium* and human proteins would help in achieving this goal.

Although aaRS are desirable drug targets, selectivity of drugs to only parasitic aaRS and not human proteins is a challenge as human aaRS have bacterial and eukaryotic origin [17, 72, 73]. High conservation of aaRS across *Plasmodium* and the human host may hinder development of parasite specific inhibitors [13, 74, 75]. Comparative studies between host and parasite sequences and structures are important in identifying differences that can be exploited for drug development [74, 76–78]. The aim of this study was to discern sequence and structural differences of aaRS between human and *Plasmodium* proteins despite the functional conservation of these proteins. The differences that occur at the active pockets and the predicted druggable sites can thus be exploited for development of drugs with good selectivity [79]. Targeting of cytosolic protein machinery in *Plasmodium* shows immediate death while inhibition of apicoplast protein translation machinery is reported to show delayed death where parasites die only during the next replication process after treatment [80]; thus, in this study the cytosolic aaRSs were used. The sequences were classified into two groups based on differences in structure of their catalytic domain and further into the different aaRS families based on their amino acid substrates [21, 23, 81]. The study was divided into two parts. First, sequence-based analysis which involved motif search, multiple sequence alignment and phylogenetic tree calculations was carried out. Secondly, structure-based analysis was carried out which involved modeling of 3D structures of proteins, mapping of identified motifs to these structures and identification of probable allosteric drug targeting sites on the 3D models. The results showed striking differences in motifs and at residue level between parasite and human proteins. The results from this study thus form a basis for further research on aaRS as potential antimalarial drug targets and other parasitic diseases.

Methods

Sequence retrieval

Plasmodium falciparum aminoacyl tRNA synthetases (PfaaRS) were retrieved from NCBI-Protein database [82]. Protein sequences of other *Plasmodium* species and human ones were searched by BLAST in UniProt using each PfaaRS as the query sequence for the specific family [83]. The BLASTp algorithm with the default BLOSUM62 matrix was used for the search of homologous sequences. (Additional file 1). The data set consisted of the five *Plasmodium* species that infect human, *P. berghei*, *P. yoelii*, *P. fragile* and human homologues. For phylogenetic tree calculations, other apicomplexan (Cryptosporidium and Toxoplasma) sequences and prokaryote sequences were also retrieved (Additional file 1). The sequences were then grouped into 20 groups based on the different aaRS families. Retrieved sequences were also grouped into two classes (Class I and Class II), each consisting of ten protein families [48, 84]. Crystal structures for human and *P. falciparum* ArgRS, TrpRS, MetRS, TyrRS, LysRS and ProRS proteins (Additional file 2) were retrieved from Protein Data Bank (PDB) [85].

Motif discovery

Motif discovery was done using Multiple Expectation Maximisation for Motif Elicitation (MEME) vs 4.11 to identify highly conserved motifs in each aaRS class [86]. A total of 90 motifs with a motif width of 6–50 residues were run for each of the non-homologous classes. The MAST tool was used to identify overlapping motifs [87]. A Python script was used to analyse MAST files and MEME log files. Motif conservation was represented as a number of sites per a total number of class sequences, and the results were displayed as heatmaps. Further, motif discovery was performed for each aaRS family and the results also displayed as heatmaps. For each aaRS family, the default parameters were used with motif width of 6–50 residues and the number of motifs run for each family varied (Additional file 3).

Homology modelling and model quality assessment

3D structures of *Homo sapiens*, *P. falciparum*, *P. knowlesi*, *P. malariae*, *P. ovale*, *P. vivax*, *Plasmodium fragile*, *Plasmodium berghei* and *Plasmodium yoelii*, and proteins were built by homology modelling using MODELLER v9.15 [88]. Templates were identified using HHpred and PRotein Interactive MOdeling (PRIMO) webserver [89, 90] for the six ArgRS, TyrRS, TrpRS, MetRS, LysRS and ProRS families (Additional file 2). The other families had no good quality templates hence models were not built. For ArgRS-5JLD [91]; for TrpRS-4J75 [92]; for TyrRS-5USF [93]; for ProRS-4NCX [65] and for LysRS-4DPG

[94] was used; for MetRS-4DLP, crystal structure of MetRS from *Brucella melitensis* was used [95]. For each protein, 100 models were calculated and the top three models with the lowest z-DOPE (Discrete Optimized Protein Energy) score were selected for validation. Structure quality assessment was done using Protein Structure Analysis (PROSA) webserver [96], Verify3D [97] and Qualitative Model Energy Analysis (QMEAN) [98] and the model with the best scores was selected for allosteric site prediction and motif mapping.

Sequence alignment

For each family of sequences, multiple sequence alignment was carried out using Profile Multiple Alignment with Local Structures and 3D constraints (PROMALS3D) and Tree-based Consistency Objective Function Evaluation (TCOFFEE) alignment tools [99, 100]. Visualization and editing of the alignments were done using the Jalview vs. 2.10 software [101]. The alignment results from the two alignment tools were compared, and, in both, it was observed that the sequences were aligned identically except for the less conserved C-terminal and N-terminal regions. TCOFFEE sequence alignments were used for the phylogenetic tree calculations as well as for all versus all pairwise sequence identity calculations via a Python script. The sequence identity results were translated into heatmaps using a Matlab script.

Molecular phylogenetic analysis

Phylogenetic tree calculations were carried out for each family of aaRSs to study evolutionary relationships within the protein families using Molecular Evolutionary Genetic Analysis (MEGA) vs7.0 tool [102]. For sequence alignment of each family, three gap deletion options—90%, 95% and 100%—were used to calculate the models, and the best three models for each deletion option were selected based on the lowest Bayesian information criterion (BIC) scores. Maximum Likelihood (ML) statistical method was used to infer evolutionary relationship while calculating trees for the top three models for each gap deletion option for each protein families [103]. Total of 180 ($3 \times 3 \times 20$) trees were calculated. Nearest-Neighbour-Interchange search was performed for all the constructed trees. BioNJ and Neighbour Join algorithms were used for a matrix of pairwise distances calculated using JTT model to obtain the initial trees for the heuristic search and the topology with the highest log likelihood selected [104]. A strong branch swap filter and 1000 bootstrap replicates were used for each tree calculation. The trees were then compared to the bootstrap consensus trees to ensure that branching patterns were accurate and the best model and gap deletion for each case was, then, chosen.

Prediction of alternate druggable sites

Structure-based drug design and development requires understanding of the structure and function of the binding sites of the target protein. Identification of new drug targeting sites different from the validated active sites is key in development of new classes of drugs. In this study, probable druggable sites of our protein models were determined using FTMap webserver [105] and SiteMap [106, 107]. Homology models were used as input for the prediction of probable druggable sites. The FTMap webserver identifies probable binding sites by screening of small compounds that vary in shape, polarity and size using an empirical energy function and the CHARMM force field [105]. The webserver docks isopropanol, acetaldehyde, phenol, benzaldehyde, urea, dimethyl ether, acetonitrile, ethane, acetamide, benzene, methylamine, cyclohexane, ethanol, N,N-dimethylformamide, isobutanol and acetone at the surface of the protein [108]. Clusters of low energy conformations are calculated and ranking of the probes is done based on the average energy [105]. The site that binds most of the compounds is considered the active binding site while other regions that bind several compounds are the predicted binding sites.

SiteMap, a tool in Schrödinger suites assigns site points in cavities that are likely to contribute to protein–protein or protein–ligand interactions based on energetic and geometric properties [106, 107]. The tool uses an algorithm that depends on how well sheltered the sites from solvents are and how close they are from the protein surface to determine the likeness of a site point. The sites are classified based on different properties which include; how enclosed is the site by the protein, the size of the site as measured by the number of points, the degree by which a ligand can accept or donate hydrogen bonds, how tight the site interacts with the protein, how exposed the site is to the solvent and the hydrophilic and hydrophobic nature of the site [106]. The predicted binding sites are then ranked based on a SiteScore calculated using a linear combination of these factors [106].

Results and discussion

In this study, 92 Class I and 89 Class II proteins were analysed for the eight *Plasmodium* species and their human homologues. More mammalian sequences were included for MSA and motif search within each aaRS family to avoid bias. A protein from each aaRS family was represented for each organism except for PbAspRS which was reported as a putative protein thus we did not include it in the study. Overall, the study is divided into two parts. In the first part, sequence related analyses such as MSA, phylogenetic tree calculations and motif identification were performed with the aim of understanding the general differences between plasmodial and human proteins.

The second part included homology modelling, mapping of motif information into 3D structures and identification of alternative drug targeting sites, as the active site within a family of proteins is generally highly conserved, hence identification of plasmodial protein specific inhibitors might be challenging.

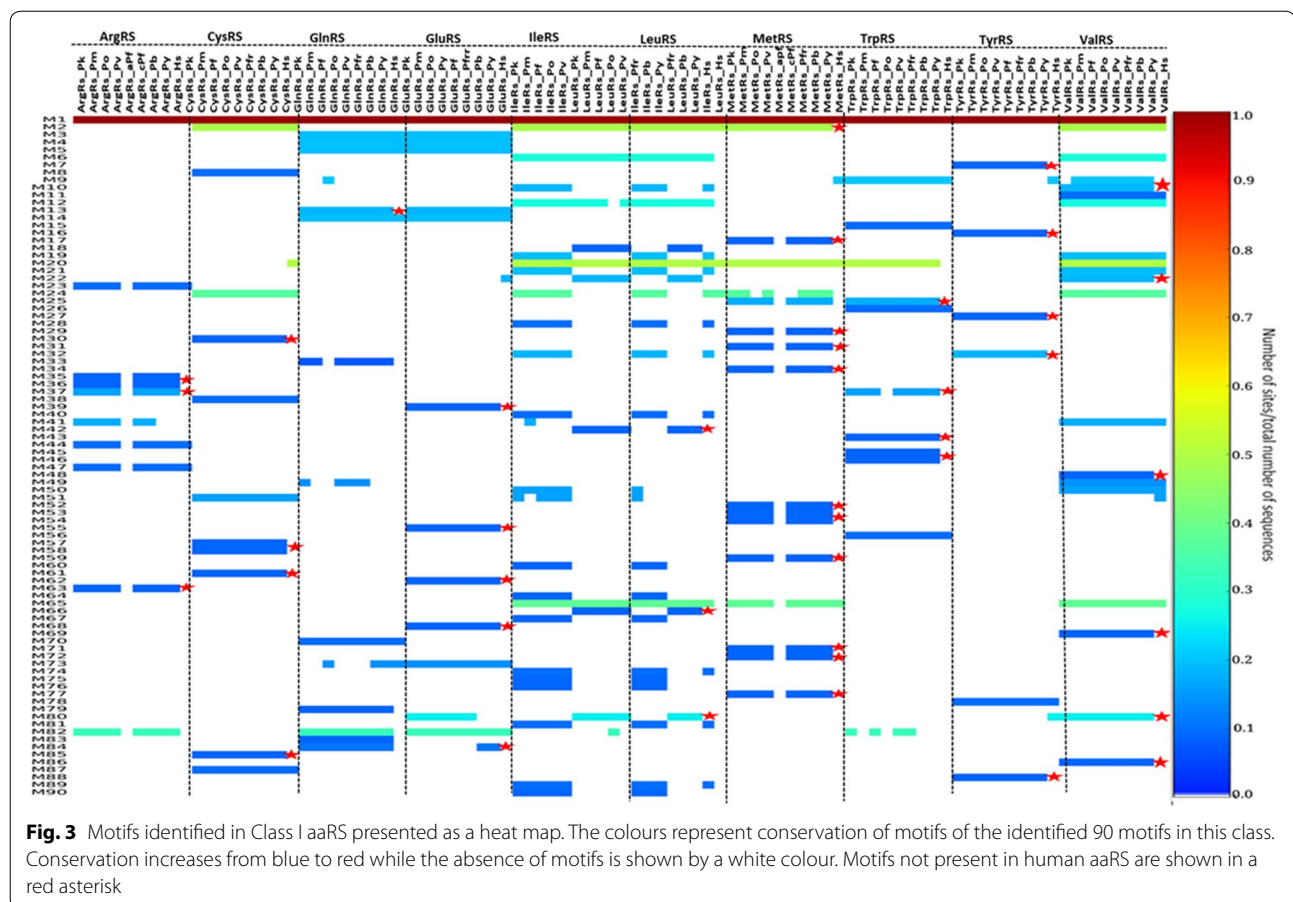
Part 1—sequence-based analyses

Discovery of motifs that are conserved in each AARS class

Motif analysis was done for each aaRS class (Figs. 3 and 4) and for each family (see Additional files 3 and 4). The results were displayed as heatmaps using a Python script and mapped to multiple sequence alignment results and available structures. Motifs discovered for each family varied as shown in Additional files 3 and 4. Motif numbering used in this section is based on the MEME results.

In Class I, 90 motifs were identified as shown in Fig. 3. The start and end positions of highly conserved motifs in this class is shown in Table 1. Motif 1 was conserved in all 92 sequences in this class (Fig. 3). This motif contains conserved residues involved in ATP binding. Motif 2 was present in 45 out of 92 sequences and this motif has also been reported to be important in ATP binding [47]. Class I aaRS enzymes are known to have a Rossmann fold catalytic domain which is characteristic of the highly conserved Motif 1 and 2 [109, 110]. Motif 12, 20 and 65 were also highly conserved among sequences in this class. The other motifs clustered based on the enzyme family but some were conserved across different enzymes within the same class. Motif 3, 4, 5, 13 and 14, for example, was conserved in all GluRS and GlnRS sequences (Fig. 3). These shared motifs show that these two proteins have a high sequence identity and may explain why *Plasmodium* apicoplast GluRS mischarges glutamine specific tRNA with glutamate. In this case, glutamate is then changed to glutamine a reaction catalysed by glutamyl-tRNA amidotransferase enzyme [38, 111].

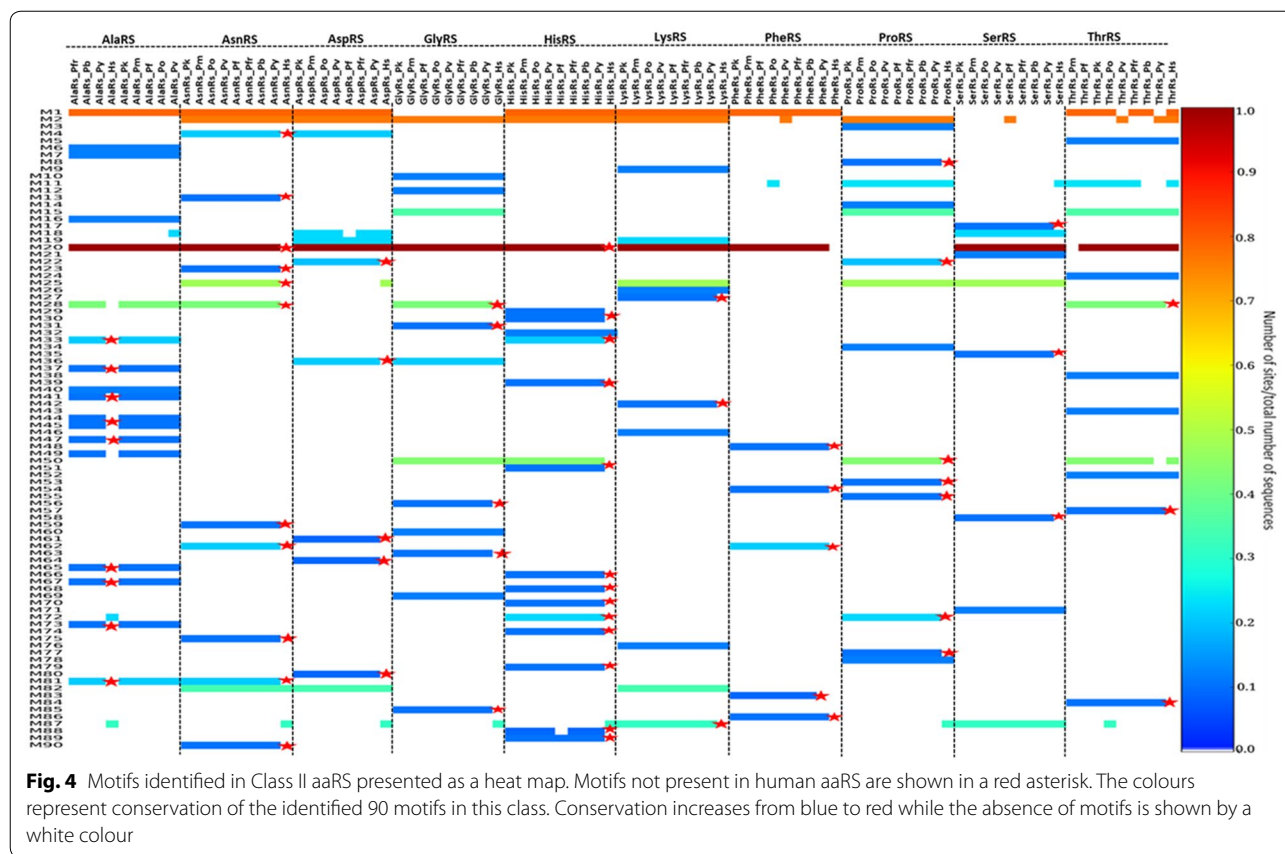
Motif 1 consisting of the HIGH signature which is characteristic of the Rossmann fold was conserved in all Class I aaRS (Fig. 3) [84]. This class also showed high conservation of a Motif 2 containing the KMSKS conserved signature which has also been reported to be part of the RF in this class (Fig. 6 and Additional files 3, 4). The HIGH motif is present in the first half of the RF while the KMSKS motif is present in the second half of the RF domain (Fig. 5 and Additional file 4). Motif conservation of the Rossmann fold reflects the functional importance of this region. This fold is involved in ATP binding and has been reported to be highly conserved in class I proteins [43]. Class I catalytic domain is characteristic of a five strand parallel sheets flanked by α -helices with amino acid and ATP binding sites on opposite sides of a pseudo-2-fold symmetry. The Rossmann fold, in all Class



I proteins has a connective polypeptide I (CPI) insert which is characterized by alpha and beta folds [47]. The conserved Motifs 1 and 2 across the class are present in the catalytic domains [47]. Detailed analysis of each protein family showed conserved motifs specific to each family (Additional file 3). Further, some conserved motifs unique only in the *Plasmodium* proteins were observed (Fig. 4 and Additional files 3, 4).

On mapping the motifs to the multiple sequence alignments, differences at the residue level were observed despite the high level of motif conservation thus these residues can be the basis of drug discovery. Eukaryote specific motifs in ArgRS, MetRS, GluRS WHEP domain and AspRS are important for the association of proteins to a multi- tRNA synthetase complex in eukaryotes [112–114]. In human, nine aaRS form a complex together with non-synthetase p18, p38 and p43 accessory proteins [114–116]. Leucyl, isoleucyl, glutaminyl, lysyl, methionyl, aspartyl, prolyl and glutamyl-tRNA synthetases form the multi-synthetase complex together with the auxiliary proteins in human aaRS but this complex is not present in *Plasmodium* aaRS [41].

These unique motifs may also play important roles other than the canonical catalytic roles [117]. Human LeuRS and GluRS, for example, have been reported to trigger leucine dependent cellular proliferation and glutamine dependent apoptosis by functioning as amino acid binding sensors [118, 119]. Highly conserved motifs specific to each aaRS group are as a result of idiosyncratic insertions at the C-terminal or within or after the Rossmann fold of each protein family in this class [24, 47, 114] (Additional files 3, 4). Methionine, valine, isoleucine and leucine aaRSs are all known to be specific to substrates that have aliphatic side chains and Motifs 20, 24 and 65 that are highly conserved in these four proteins may have a role in this specificity [47]. LeuRS, IleRS, MetRS, ArgRS, ValRS and CysRS have a structurally conserved anticodon binding domain characterized by α -helices and this may explain the conservation of Motifs 2, 20, 44 and 65 among these proteins (Fig. 3) [47]. *Plasmodium* TrpRS has an N-terminal extension which is 227 amino acid residues long that constitute a AlaX-like domain and a linker region that function in binding of tRNA and in aminoacylation activity [120]. This extension is not present



in the human TrpRS and thus explains the unique motifs at the N-terminal of the *Plasmodium* proteins. *Plasmodium* sequences also have a lysine-enriched insertion at the C-terminal end of the KMSKS motif which is 15 residues long in PfTrpRS which is absent in the human sequence [120]. The domain for binding anticodons in Class I is located at the carboxyl terminal except for LeuRS. The structures of this region are highly divergent even within the sub-classes and is known to play an important role in tRNA discrimination [47].

In Class II, there were three highly conserved motifs across the class (Fig. 4, Table 2). In the reporting of motif results of this class, motif names are based on MEME results and not on previous literature. Motif 1 was present in 60 sequences, Motif 2 was present in 58 sequences while motif 20 was present in 76 sequences out of 89 sequences (Fig. 4). Motif 1, Motif 2 and Motif 19 discovered in Class II identified in this study contain the conserved signatures of Class II proteins (motif III, motif II and motif I respectively) reported by Chaliotis et al. [43]. In Class II, motifs also clustered based on the protein family. Motif conservation among proteins may mean that these regions play a specific function in the proteins. Motif discovery was then done for each protein

family to determine conserved motifs within homologous sequences of each protein and the results presented as heat maps (Additional file 3).

Class II aaRS have a highly conserved catalytic domain that occurs as β -sheet strands with α -helices on either side. This domain binds ATP, amino acid and the tRNA during aminoacylation. Motif 1 has been reported previously (as motif III) to be part of the active site forming α -helices and β -strands [43, 121]. Motif 2, (Fig. 4) also found at the catalytic site of proteins in this group forms β strands in pairs joined by a loop [42]. Motif I plays a role in binding of ATP while Motif 2 couples ATP, tRNA and amino acid binding [42, 122]. Another weakly conserved motif in the active site of these proteins forms an α -helix that is linked to a β -strand with a proline residue at the end (Motif 19, Fig. 4). This motif is known to be crucial in formation of dimers in most proteins of this class [123].

Further, subclasses in this class have conserved motifs within each subclass (Fig. 4). For example, Ser, Thr, Gly, Pro and His aaRSs all belong to the Class IIa and have anticodon binding domains that are specific to the subclass [55]. These proteins are specific to small and hydrophobic amino acids and have motifs that are conserved

Table 1 Highly conserved motifs in Class I aaRS

	Motif 1	Motif 2	Motif 6	Motif 12	Motif 20	Motif 65
PfArgRS	134–152	–	–	–	–	–
HsArgRS	197–215	–	–	–	–	–
PfCysRS	131–149	382–419	–	–	–	–
HsCysRS	53–71	405–442	–	–	700–714	–
PfGlnRS	284–302	–	–	–	–	–
HsGlnRS	266–284	–	–	–	–	–
PfGluRS	313–331	–	–	–	–	–
HsGluRS	200–218	–	–	–	–	–
PfIleRS	127–145	782–819	623–651	222–271	167–181	91–111
HsIleRS	44–62	599–636	444–472	139–188	94–108	63–83
PfLeuRS	141–159	1119–1156	688–716	219–268	181–195	160–180
HsLeuRS	49–67	715–752	–	–	89–103	–
PfMetRS	228–246	514–551	–	–	268–282	247–267
HsMetRS	269–287	–	–	–	310–324	289–309
PfTrpRS	306–324	–	–	–	534–548	–
HsTrpRS	159–177	–	–	–	–	–
PfTyrRS	61–79	–	–	–	–	–
HsTyrRS	53–71	–	–	–	–	–
PfValRS	86–104	628–665	454–482	181–230	126–140	105–125
HsValRS	340–358	861–898	708–736	435–484	380–394	359–379

Motif 1, 2, 6, 12, 20 and 65 in Class I *Plasmodium falciparum* aaRS and the human homologues. Motif positions in the sequences are indicated and dashes are used where the motif is not present

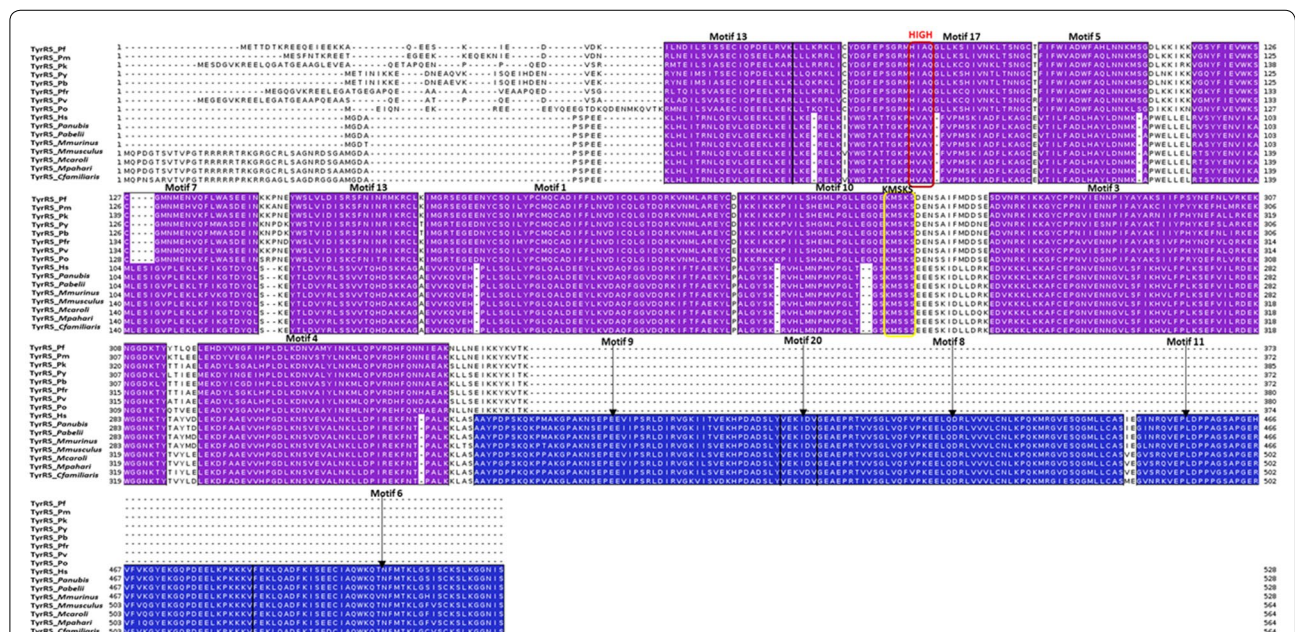


Fig. 5 Motifs discovered in TyRS family mapped to the multiple sequence alignment results. Motif numbering is based on MEME results. A purple colour shows motifs conserved in all sequences while motifs only present in mammalian sequences are shown in blue. The highly conserved HIGH and KMSKS motifs in Class II aaRSs are shown in a red and yellow box, respectively

Table 2 Highly conserved motifs in Class II aaRS

	Motif 1	Motif 2	Motif 19	Motif 20
PfAlaRS	620–640	–	–	942–962
HsAlaRS	236–256	–	–	23–43
PfAsnRS	574–594	369–389	–	343–363
HsAsnRS	512–532	313–333	249–289	–
PfAspRS	590–610	382–402	317–357	360–380
HsAspRS	465–485	264–284	199–239	242–262
PfGlyRS	–	469–489	–	173–193
HsGlyRS	–	322–342	–	130–150
PfHisRS	918–938	688–708	–	658–678
HsHisRS	378–398	148–168	–	118–138
PfLysRS	549–569	321–341	254–294	362–382
HsLysRS	543–563	314–334	247–287	–
PfPheRS	313–333	–	–	110–130
HsPheRS	308–328	–	–	–
PfProRS	–	463–483	–	–
HsProRS	–	1225–1245	–	–
PfSerRS	–	–	–	202–222
HsSerRS	–	–	–	192–212
PfThrRS	653–673	–	–	358–378
HsThrRS	444–464	–	–	151–171

Motif 1, 2, 19 and 20 in Class II *P. falciparum* aaRS and the human homologues. Motif positions in the sequences are indicated and dashes are used where the motif is not present

among them as shown in the heatmap (Fig. 4). The anticodon binding domain comprises of three α -helices five and β -stranded sheets and occurs in the C-terminus of this sub-class [55, 124, 125]. The anticodon binding domain is absent in SerRS as this protein does not require an anticodon to discriminate its cognate tRNA [54, 55]. Subclass IIb which comprises of AsnRS, LysRS and AspRS have a unique anticodon binding domain at the N-terminal and share conserved motifs (Fig. 4) [57, 126, 127]. This subclass of enzymes is specific to large polar and charged amino acid substrates and are similar in structural organization. AspRS is capable of catalysing aminoacylation of aspartate and asparagine and thus it can be classified as discriminating and non-discriminating protein just like GluRS [128, 129]. Non-discriminating AspRS is only present in bacteria and archaea but not in eukaryotes [130]. Family specific motifs, can be attributed to the diversity in accessory domains found at the N- and C-terminal or within loops in the core domain [131].

Multiple sequence alignment and motif mapping

Plasmodium and mammalian sequences for every aaRS family were aligned using TCOFFEE as indicated in the methodology. The alignment results were visualized using Jalview software and motifs discovered for each

family mapped to these alignments [101]. A purple colour was used for the motifs that were conserved in all *Plasmodium* and mammalian sequences, blue colour for only motifs conserved in mammalian species and green colour for motifs conserved only in *Plasmodium* sequences (Additional file 4). On carrying out motif analysis and sequence alignment of Class I aaRSs, it was observed that not all families had the KMSKS signature though all proteins had the HIGH signature (Additional file 4). Alignment of ArgRS showed inserts in mammalian ArgRS at both the C- and N-termini that are not present in *Plasmodium* sequences (Additional file 4A). The highly conserved HIGH signature in Class I aaRSs catalytic domain was observed in Motif 1 of this family (HVGH) (Additional file 4A). Motifs 10 and 12 which were conserved only in mammalian sequences were observed in the N-terminal. Human ArgRS has a basic 72 residue extension at the N-terminal which is characteristic of mammalian ArgRS and plays a role in interaction with accessory proteins like p43 to form the multi-synthetase complex [116, 132]. Mammals also have an ArgRS isoform that lacks this extension and is believed to be important in ubiquitin dependent protein degradation where it forms Arg-tRNA^{Arg} which is transferred to ArgRS which then adds the arginine to all acidic N-terminal amino acids [133, 134].

CysRS sequence alignment and motif mapping showed a highly conserved core domain and weakly conserved N- and C-terminal domains. The highly conserved HIGH signature was found in Motif 2 of this family occurring as HLGH in *Plasmodium* and HMGH in the mammalian sequences (Additional file 4B). Motif 8, 10, 12, 13, 18 and 19 were conserved only in mammalian sequences while Motif 11 and 15 were only conserved in plasmodium sequences analysed in this family (Additional file 4B). GlnRS alignment also showed low conservation on both termini with inserts observed for the mammalian sequences at the N-terminal (Additional file 4C). Only two *Plasmodium* specific motifs were found at the core domain, Motif 23 at the N-terminal end of the highly conserved HIGH signature (Motif 2) and Motif 29. Motif 8, 9, 11 and 13 were found only in the mammalian species (Additional file 4C). *P. falciparum* is reported to have Glutathione-S-transferase (GST)-like domains though their function in the malarial parasite has not been reported [41]. These domains are important in formation of multi-synthetase complex through protein–protein interactions in eukaryotes [24, 25, 117]. GST-like domains have also been reported in MetRS though just like in GlnRS, the function of these domains in *Plasmodium* is not known unlike in eukaryotes where they play a role in protein–protein interactions [22].

The GluRS family also showed low conservation at the N-terminal with Motif 16 present in mammalian sequences at this terminal (Additional file 4D). The HIGH signature was found in Motif 3 as HIGH in all sequences analysed except for PfGluRS where it occurs as HVGH (Additional file 4). *P. falciparum* GluRS sequence has a glutamine rich N-terminal from residue 68 as opposed to other *Plasmodium* species. In mammals, including human, this enzyme is a bifunctional protein acting both as GluRS and ProRS. Thus it catalyses aminoacylation of both proline and glutamate [135]. On alignment with *Plasmodium* GluRS, the mammalian sequences showed a C-terminal extension indicating that it is the N-terminal end that catalyses glutamate aminoacylation. The human enzyme contains three motifs that link the two catalytic domains that function in formation of the multicomplex synthetase and play a role in protein-nucleic acid interactions [135, 136]. Similar motifs have been reported in other aaRS like GlyRS, HisRS and TrpRS though they occur at the N- or C-termini of the core domains as a single copy as opposed to the Glu/ProRS where they occur as tandem repeats linking the two catalytic domains [135–137]. Human IleRS has an extension at the C-terminal which was absent in *Plasmodium* sequences, but the core domain of this family was highly conserved (Additional file 4E). Motif 19, 20 and 26 were conserved in the C-terminal of mammalian IleRS sequences but absent in *Plasmodium* sequences. The three tandem motifs in the human bifunctional Glu/ProRS have been shown to interact with two repeated motifs in IleRS at the C-terminal extension [138]. In IleRS, the HIGH signature was found in Motif 1 while the KMSKS signature was in Motif 3 occurring as HYGH and KMSKR, respectively (Additional file 4E). Alignment and motif discovery of LeuRS family showed that this family of protein has low conservation even at the core domain (Additional file 4F). Motif 21, 25 and 27 were conserved in *Plasmodium* sequences. Only Motifs 3, 5, 6, 26 and 36 were conserved through all mammalian and *Plasmodium* sequences (Additional file 4F).

The other motifs were conserved only in mammalian sequences. The highly conserved Motif 6 had the HIGH signature occurring as HVGH for PfLeuRS, PmLeuRS, PoLeuRS, PyLeuRS, HMGH for PfrLeuRS, PvLeuRS and PkLeuRS and HLGH in the analysed mammalian sequences (Additional file 4F). Anticodon binding domain in LeuRS is located at the C-terminal which had a low conservation as seen in Additional file 4F and this may provide specific targets for drug discovery [139]. Motif discovery and alignment of MetRS showed high conservation of mammalian sequences. Some unique motifs were only present in *Plasmodium* MetRS but were absent in mammalian sequences (Additional file 4G). The

highly conserved HIGH signature was observed in Motif 8 which was conserved in all sequences analysed while the KMSKS signature was found in Motif 14 conserved in *Plasmodium* and Motif 6 in mammalian sequences (Additional file 4G). The catalytic domain of MetRS was weakly conserved with only Motif 1, 2, 4, 8 and 15 being conserved in all sequences at this region. The C-terminal showed mammalian and *Plasmodium* specific motifs. Motif 5 and 9 found at the N-terminal were conserved in all analysed sequences in this family (Additional file 4G).

TrpRS alignment revealed a *Plasmodium* specific extension at the N-terminal characterised by Motif 8, 9, 10 and 14 (Additional file 4H). This extension plays a role in aminoacylation and tRNA binding as reported in *P. falciparum* [41]. In *P. falciparum*, this extension comprises of a linker region and an AlaX-like domain that plays a role in tRNA binding but does not edit mis-acylations as observed with *Pyrococcus horikoshii* [120]. The core domain and the C-terminal of TrpRS family showed highly conserved motifs in all the sequences with only a short Motif 18 present in mammalian sequences (Additional file 4H). Alignment and mapping motifs discovered in TyrRS sequences showed high conservation of motifs at the core domain (Fig. 5). Alignment of sequences in this family showed an extension at the C-terminal of the mammalian TyrRS which was missing in all *Plasmodium* sequences (Fig. 5). This extension was characterised by Motifs 6, 8, 9, 11 and 20 which were conserved in all the mammalian sequences analysed (Fig. 5). This extension in human TyrRS is an endothelial monocyte-activating polypeptide II (EMAPII) domain that has cytokine-like functions like angiogenesis and inflammation [25, 140]. Motif discovery showed that the core domain is highly conserved across the mammalian and *Plasmodium* TyrRS sequences (Fig. 5). The catalytic domain of the human sequence is also different from the malarial parasites in that it has a buried tripeptide cytokine motif (Glu-Leu-Arg) while in *Plasmodium* this motif is on the surface [25, 26]. ValRS alignment showed a N-terminal extension for the mammalian sequences that was absent in all *Plasmodium* sequences comprising of Motifs 14, 16, 18, 22 and 25 (Additional file 4J). Mapping of motifs showed that the catalytic domain of proteins analysed in this family are highly conserved though a few *Plasmodium* specific motifs were observed. The highly conserved HIGH signature was found in Motif 2 of this family while the KMSKS signature was in Motif 7 (Additional file 4J). The N-terminal domain showed Motif 16, 33, 34 and 35, which were conserved only in *Plasmodium* sequences (Additional file 4J). Motifs 20, 30 and 38 that were specific to mammalian sequences were also observed at the N-terminal (Additional file 4J).

Alignment of AlaRS sequences showed a N-terminal extension of varying lengths in the *Plasmodium* species which was absent in mammalian AlaRS (Additional file 4K). The C-terminal of the proteins in this family showed Motifs 20, 21 and 29 that were only conserved in *Plasmodium* sequences and not in human as well as mammalian specific motifs (Motif 8, 14, 17 and 18). AsnRS, LysRS, and AspRS alignment and motif discovery showed low conservation at the N-terminal while core domains and the C-terminal showed high conservation. The anticodon binding domain of these proteins is located at the highly variable N-terminal and thus drugs that specifically bind to the parasite tRNA binding site can be designed [17, 141]. Motif 11, 12 and 17 were conserved in *Plasmodium* sequences of AsnRS family at the N-terminal while in this region, Motif 5, 6 and 13 conserved in mammalian sequences were observed (Additional file 4L). In AspRS, both the catalytic domain and the C-terminal were highly conserved with the presence of two short Motifs (16 and 20) conserved only in mammals (Additional file 4M). GlyRS, HisRS, ProRS, ThrRS families belong to the subclass IIa and have a highly conserved tRNA binding region at the C-terminal as seen in the alignments and motifs in this region (Additional file 4 N, O, R and T). HisRS family showed a N-terminal extension for all *Plasmodium* sequences analysed but absent in the mammalian sequences (Additional file 4O). This extension was characterised by Motifs 11, 12, 14, 15, 17, 18, 19 and 23 (Additional file 4O). However, SerRS which also belongs to this subclass does not need an anticodon to discriminate its substrate and thus lacks this domain [55] and the C-terminal of this family showed low conservation (Additional file 4S). ProRS showed Motif 17 and 20 which were conserved only in *Plasmodium* sequences analysed (Additional file 4R). *Plasmodium* ProRS has a Ybak domain at the N-terminal which edits mischarged Pro-tRNA^{Ala} and Pro-tRNA^{Ser} and this may explain the *Plasmodium* specific motifs at the N-terminal [18, 41, 62]. The mammalian sequences analysed for this family were of the cytosolic bifunctional Glu/ProRS proteins and this explains the mammalian specific motifs observed at the N-terminal which is believed to be the region responsible for glutamate aminoacylation (Additional file 4R).

PheRS motif discovery and alignment showed that the *Plasmodium* sequences are highly variable when compared to mammalian PheRS (Additional file 4Q). Motifs 9, 10, 11 and 13 were conserved only in *Plasmodium* while Motifs 5, 6, 8 and 14 were conserved in mammalian sequences in this family (Additional file 4Q). Only Motifs 1, 2, 3 and 4 were conserved across all the sequences in this family (Additional file 4Q). *Plasmodium* PheRS has

a nuclear localization signal and DNA binding domains and thus in addition to aminoacylation, this enzyme mediates cellular processes by binding DNA [142]. Despite high conservation at the aaRS active sites, differences were noted at the residue level after the sequences were aligned. For example, in LysRS family, *P. falciparum* ATP binding pocket at positions Val328 and Ser344 corresponds to Gln321 and Thr338, respectively in the human protein (Fig. 6). Residues with a large side chain at this position like observed in human LysRS do not favour binding of cladosporin a known inhibitor for PflLysRS [28, 143]. These two residues are thus believed to be responsible for selective binding of cladosporin and its analogues to *P. falciparum* and not human LysRS [28, 143]. Discovery of drugs that have high specificity to parasitic proteins has for a long time been a challenge resulting in drug toxicity in human cells [17]. The alignment results showed striking differences at the sequence level of *Plasmodium* and human aaRSs that can further be explored for the design and development of drugs with few side effects.

Phylogenetic tree calculations and pairwise sequence identity calculations agree in grouping sequences

On conducting phylogenetic tree analysis, all *Plasmodium* species clustered together, and this was also seen on performing all versus all pairwise sequence identity calculations (Figs. 7, 8 and Additional file 5). In this study, numbering of sequences in sequence identity heatmaps was based on the branching of phylogenetic trees. In Class I, *Plasmodium* sequences in TyrRS family showed the highest sequence identity (above 85%) while GlnRS *Plasmodium* sequences showed the lowest sequence (below 75%) identity among *Plasmodium* families. In most of the families, *P. yoelii* and *P. berghei* sequences were clustered together in the trees. *P. vivax*, *P. fragile* and *P. knowlesi* were also clustered together in many families, indicating that they are highly conserved and share evolutionary history. These similarities were also captured in sequence identity calculations, and reflected as imaginary boxes in heat maps. Here they will be named "conservation boxes". *Plasmodium berghei* and *P. yoelii* are rodent malaria parasites and are used to study human malaria [144, 145]. *Plasmodium fragile* infects simians and studies have shown that human red blood cells do not support the growth of this parasite, but it showed a high sequence identity to *P. knowlesi* whose natural vertebrate host is *Macaca fascicularis*, but has been reported to infect human in some parts of Southeast Asia [146, 147]. *Plasmodium knowlesi* has been reported to have a close phylogenetic relationship to *P. vivax* [148] and the two showed a sequence identity above 95% in TyrRS (Fig. 8). *Plasmodium fragile*-monkey models can thus be

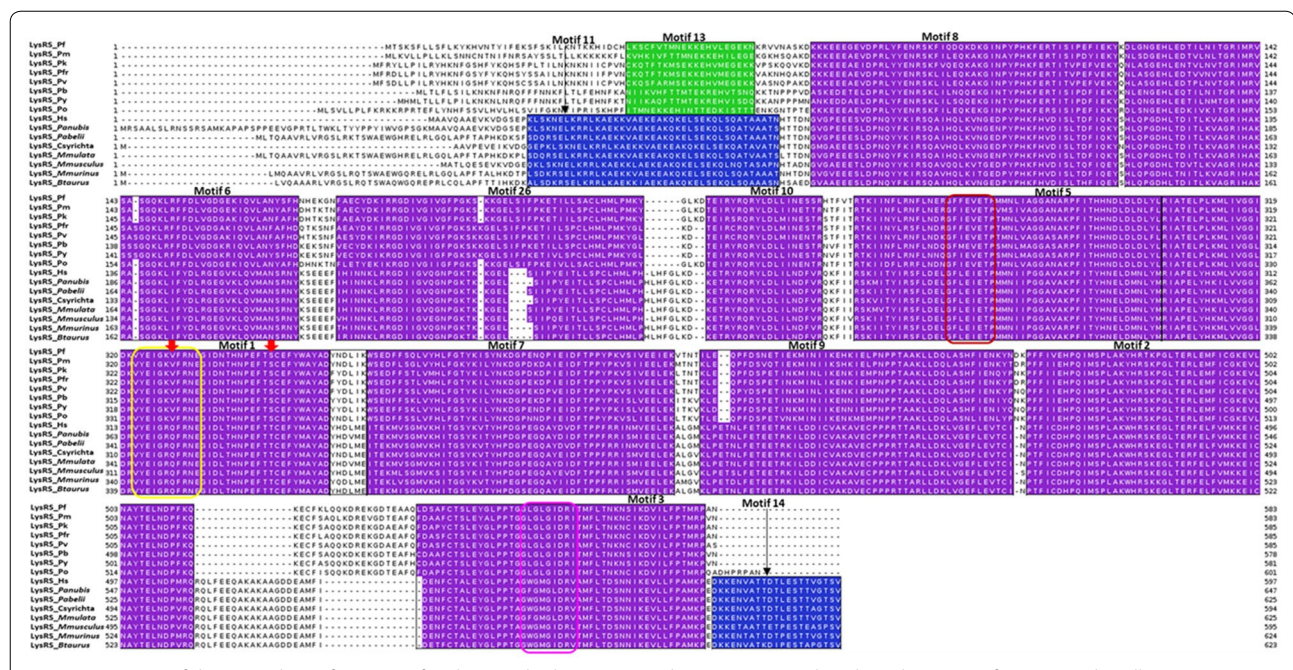


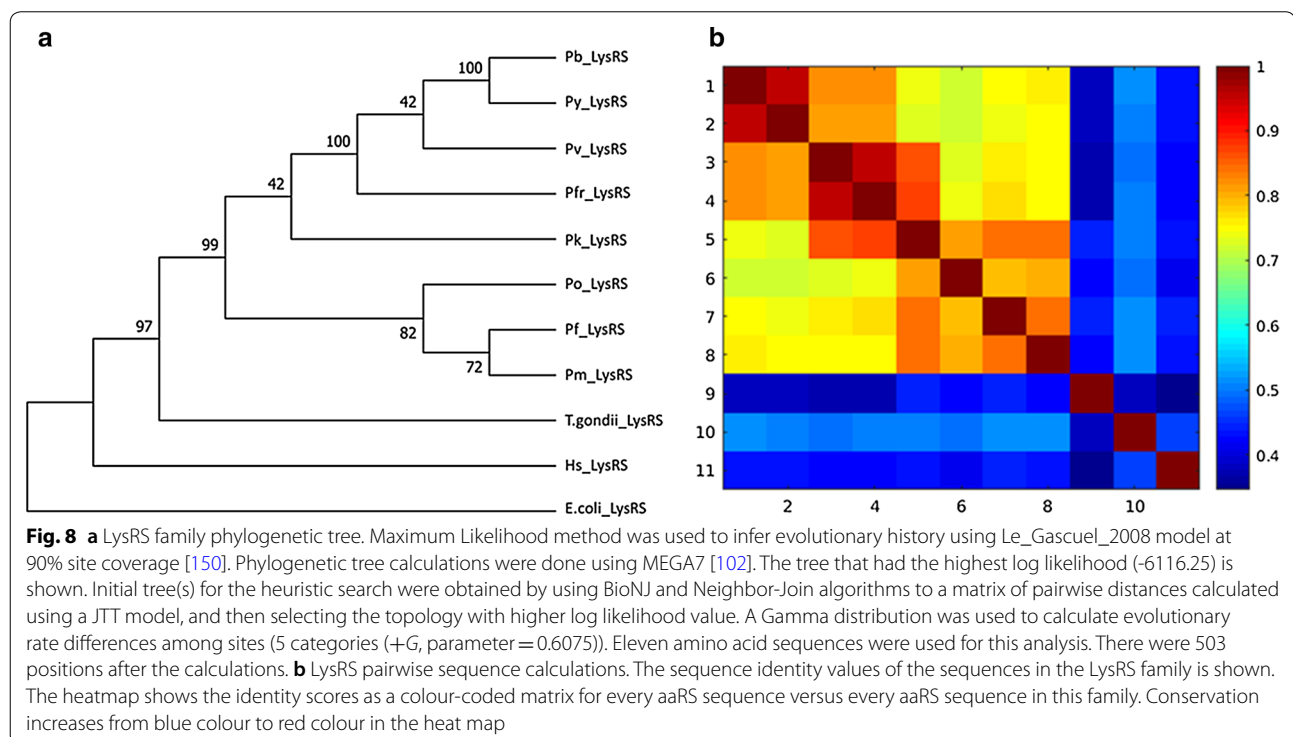
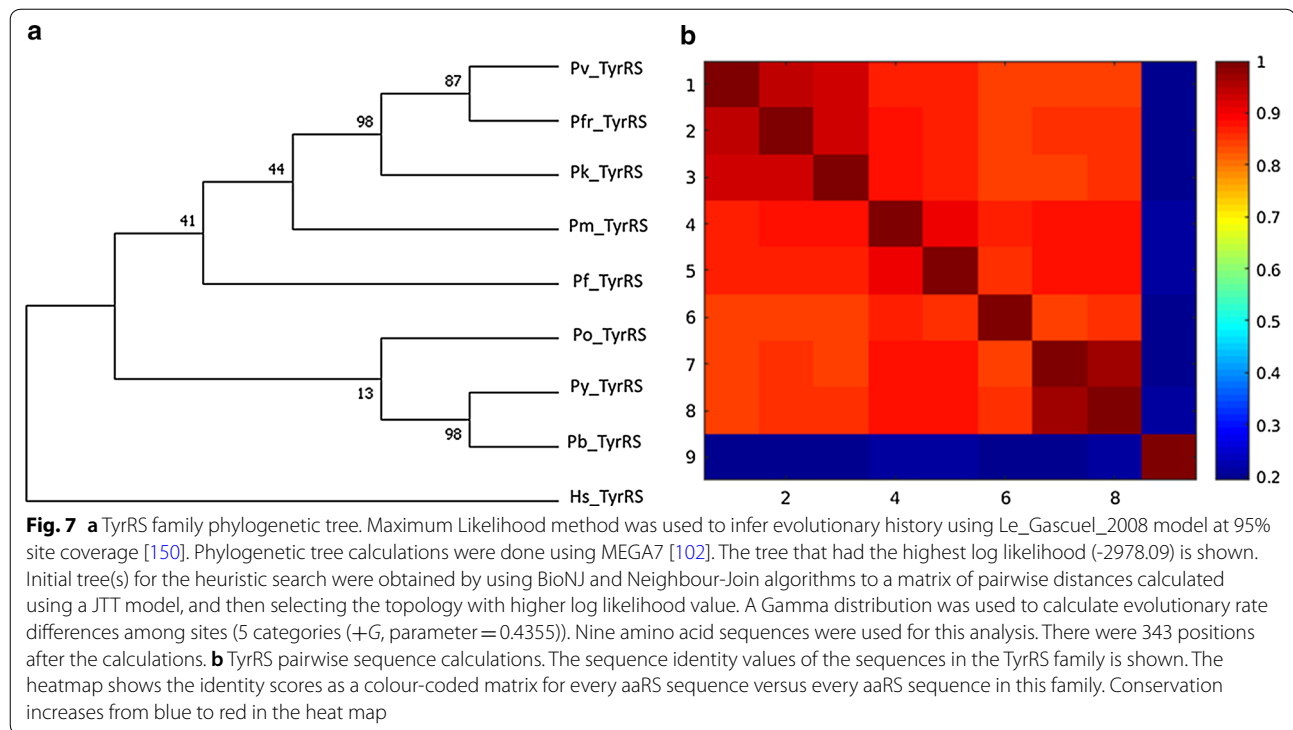
Fig. 6 Mapping of discovered motifs in LysRS family to multiple sequence alignment. A purple colour shows motifs conserved in all sequences while motifs only present in mammalian sequences are shown in blue. One motif conserved only in *Plasmodium* species is shown in green. Motif numbering is based on MEME software. The three conserved signatures in Class II aaRSs are shown in red, yellow and pink boxes. The red arrows show residues Val328 and Ser344 in *P. falciparum* which are key residues in binding of ATP

used to study parasite-host-system for the immunological response of the falciparum-like parasite both in vivo and in vitro [149].

In ArgRS sequence identity calculations, *Plasmodium* sequences had above 80% sequence identity and motif discovery showed that all motifs identified were conserved in all sequences (Additional file 3 and Additional file 5: 5.1). ValRS *Plasmodium* sequences showed 80% sequence identity with PvValRS, PkValRS and PfrValRS clustering together with a 90% sequence identity. In this family, PyValRS and PbValRS showed above 95% sequence identity, clustered together in the phylogenetic tree and shared Motif 36 which was absent in the other *Plasmodium* sequences (Additional file 3 and Additional file 5: 5.10). PvCysRS, PkCysRS and PfrCysRS clustered together with a 90% sequence identity and shared Motif 22 which was missing in other *Plasmodium* sequences (Additional file 3 and Additional file 5: 5.2). Motif 27 was present only in PyCysRS and PbCysRS and these two sequences showed a 90% sequence identity (Additional file 3 and Additional file 5: 5.2). PfrGlnRS, PkGlnRS and PvGlnRS clustered together and Motif 34, 35 and 37 were only present in these sequences (Additional file 3 and Additional file 5: 5.3). In this family, *Escherichia coli*, human and *Saccharomyces cerevisiae* sequences formed an outgroup showing they are the oldest aaRS

(Additional file 5: 5.3). GluRS *Plasmodium* sequences had above 75% sequence identity and shared all identified motifs (Additional file 3 and Additional file 5: 5.4). PblLeRS and PyLeRS shared Motif 38 and 39 and showed 95% sequence identity (Additional file 3 and Additional file 5: 5.5). *Cryptosporidium* and *Toxoplasma* belong to the Apicomplexan family together with *Plasmodium* and their sequences showed about 50% sequence identity to *Plasmodium* sequences in IleRS and MetRS family (Additional file 5: 5.5 and 5.7). PvLeuRS, PfrLeuRS and PkLeuRS had 80% sequence identity and shared Motif 39 (Additional file 3 and Additional file 5: 5.6). In TrpRS, Motif 21 and 23 were only identified in PbTrpRS and PyTrpRS which had 90% sequence identity. In all families in Class I, human sequences showed low sequence identities (below 40%) compared to the *Plasmodium* sequences (Additional file 5).

In Class II aaRSs, ProRS family showed the highest sequence identity with *Plasmodium* sequences having above 80% sequence identity (Additional file 5). The high sequence identity among *Plasmodium* sequences was also reflected in motif identification where all the sequences shared the identified motifs (Additional file 3). In Class IIa GlyRS showed the least conservation with most of the sequences having less than 65% sequence identity (Additional file 5: 5.16). GlyRS family



showed low conservation with sequence identity less than 70% for all sequences except for PfrGlyRS, PvGlyRS and PkGlyRS which formed a conservation box

with a sequence identity of about 75% (Additional file 5: 5.16). This clustering was also seen in motif identification whereby PfrGlyRS, PvGlyRS and PkGlyRS had Motif 24

which was absent in all other *Plasmodium* sequences in this family. PbGlyRS and PyGlyRS had a sequence identity of 90% and shared Motif 27, 30 and 34 (Additional file 3 and Additional file 5: 5.16). *Plasmodium falciparum* ThrRS had a low sequence identity compared to other *Plasmodium* sequences and it also branched separately in the phylogenetic tree. In SerRS family, human, *Trypanosoma brucei*, *Candida albicans*, *Toxoplasma gondii* and *Cryptosporidium parvum* also formed a conservation box but with a sequence identity of about 65%. *P. vivax*, *P. fragile* and *P. knowlesi* in this family had a high sequence identity forming a conservation box and clustered together in the phylogenetic tree (Additional file 5: 5.20). PfrThrRS, PvThrRS and PkThrRS shared Motif 24, 27, 29 and 33 showing these sequences are closely related as depicted by trees and sequence identity calculations (Additional file 3 and Additional file 5: 5.19). In SerRS family, *Plasmodium* sequences formed a conservation box with about 75% sequence identity with each other except for *P. yoelii* which was more identical to *P. berghei* with a sequence identity of 90% (Additional file 5: 5.19). In motif identification, *P. yoelii* shared Motif 20 and 22 which were all absent in all other *Plasmodium* sequences explaining the high sequence conservation (Additional file 3). PkSerRS and PvSerRS branched together and the two shared Motif 19 showing that the sequences are closely related (Additional file 3). In HisRS family, all *Plasmodium* sequences formed a conservation box showing more than 70% sequence identity to each other except for PfHisRS (Additional file 5: 5.14). This difference was also seen in motifs identified in this family where Motif 21 was present in all *Plasmodium* sequences but absent in PfHisRS (Additional file 3).

In Class IIb, AsnRS sequences were highly conserved with above 80% sequence identity while AspRS was the least conserved with about 65% sequence identity (Additional file 5: 5.13 and 5.15). The high sequence conservation in AsnRS was also seen in motif discovery where all *Plasmodium* sequences shared identified motifs (Additional file 3). In AsnRS family, *Candida ubiquitum* showed a higher sequence identity to *Salmonella typhi* sequence than to *Toxoplasma gondii* which belongs to the same phylum. PvAspRS and PkAspRS branched together in tree calculation and these two proteins shared Motif 22, 26 and 28 showing they are closely related (Additional file 3 and Additional file 5: 5.15). In LysRS family, *Plasmodium* sequences showed a sequence identity of above 75% with PbLysRS and PyLysRS forming a conservation box with about 95% sequence identity (Fig. 8). PfrLysRS, PkLysRS and PvLysRS also formed a conservation box and these three proteins shared Motif 15, which was absent in other *Plasmodium* sequences (Fig. 8, Additional file 3).

Overall, PheRS was the least conserved family in Class II with *Plasmodium* sequences with only about 50% sequence identity and this was seen during motif discovery where only a few motifs were conserved across species (Additional file 3 & 5). In the AlaRS family, *P. falciparum* (sequence 5 in the heatmap) was less conserved compared to other *Plasmodium* sequences as seen in the (Additional file 5). *Plasmodium* sequences in AlaRS family showed a sequence identity above 70% (Additional file 5: 5.12). In this family, *P. vivax*, *P. fragile* and *P. knowlesi* also formed a conservation box while *P. yoelii* and *P. berghei* also formed a conservation box indicating that these sequences are highly conserved compared to other *Plasmodium* sequences. PfrAlaRS, PvAlaRS and PkAlaRS shared Motif 31 which was absent in all other *Plasmodium* sequences but present in mammalian sequences (Additional file 3). In all the families in Class II, human sequences in this class branched out as an out group and this is supported by the low sequence identity (below 40%) shown in the conservation heatmaps (Additional file 5).

Part 2—structural analyses

Accurate 3D protein models are calculated for Class I and Class II aaRSs

In the PDB, there are only four Class I (ArgRS, MetRS, TrpRS, TyrRS) and two Class II (LysRS and ProRS) structures that were available with reasonable quality. As a first step, each of these crystal structures was remodelled to eliminate the missing residues, except PfTyrRS, as this structure does not have missing residues. It was previously shown that homology modelling with a very high sequence template identity (or remodelling itself) does not introduce modelling errors [151]. As a next step, these models were used to model the 3D structures of the homologues (see Additional file 2 for further information).

For each protein, 100 homology models were calculated, and the three best models selected based on z-DOPE scores. DOPE score is an atomic statistical potential which depends on a native protein structure [152]. It is highly accurate in assessment of the quality of protein models as it accounts for the spherical and finite shape of the protein native structure [152–154]. It depends on the number of atom pairs considered and thus the number of all possible pairs of heavy atoms in the protein are normalized to get the z-DOPE score [152, 155]. Models with lowest z-DOPE were selected and model quality assessment was done using Verify 3D [97], ProSA [96] and QMEAN [98] web servers. Verify 3D assesses the compatibility of the 3D structure with the amino acid sequence (1D) and assigns a class to the structure based on the local environment, location and

secondary structure and compares this to known native structures [97]. At least 80% of the amino acid residues should have a score greater than or equal to 0.2 in the 3D/1D profile for the structure to be considered of good quality. ProSA-web is a tool for checking errors in a 3D model and displays the quality score as graphical presentation. Areas of the model that are not accurate are identified by a plot of local quality scores which are then mapped on the 3D structure using colour codes [96].

QMEAN score describes the major geometrical aspects of protein models using five structural descriptors. The overall status of residues is described by a solvation potential, long-range interactions are assessed by secondary structure-specific pairwise residue-level potential that is dependent on distance and a torsion angle potential is used to determine the local geometry which is calculated over three consecutive residues [98]. Descriptors of solvent accessibility and the agreement between calculated and predicted structures are also used in calculating the score [98]. All the calculated models passed the quality evaluation tests from these three tools (Additional file 2).

The models for the *Plasmodium* ArgRS were built using 5JLD [91] as a template while 4ZAJ was used for the human homologue. The ArgRS models consist of the N-terminal, catalytic domain and the anticodon binding domain. All the models for MetRS, which included the catalytic domain and the anticodon binding domain, were calculated using *Brucella melitensis* MetRS crystal structure (4DLP) [95]. *Plasmodium* TrpRS models were built with 4J75 [92] while 1R6T [156] was used for HsTrpRS. It was possible to model the N-terminal, catalytic and anticodon binding domains for this family. The crystal structure 5USF [93] was used for the calculation of *Plasmodium* TyrRS while 1Q11 [156] consisting only the catalytic and anticodon binding domain was used to model the HsTyrRS. The catalytic and anticodon binding domains of LysRS were built using 4DPG [94] as the starting structure while 4NCX [65] was used for building ProRS models which included a zinc-binding like domain at the C-terminal.

The 3D models were, then, used for mapping identified motifs to structures as well as for the search of alternate druggable sites in *P. falciparum* homologues.

Motif mapping to homology models

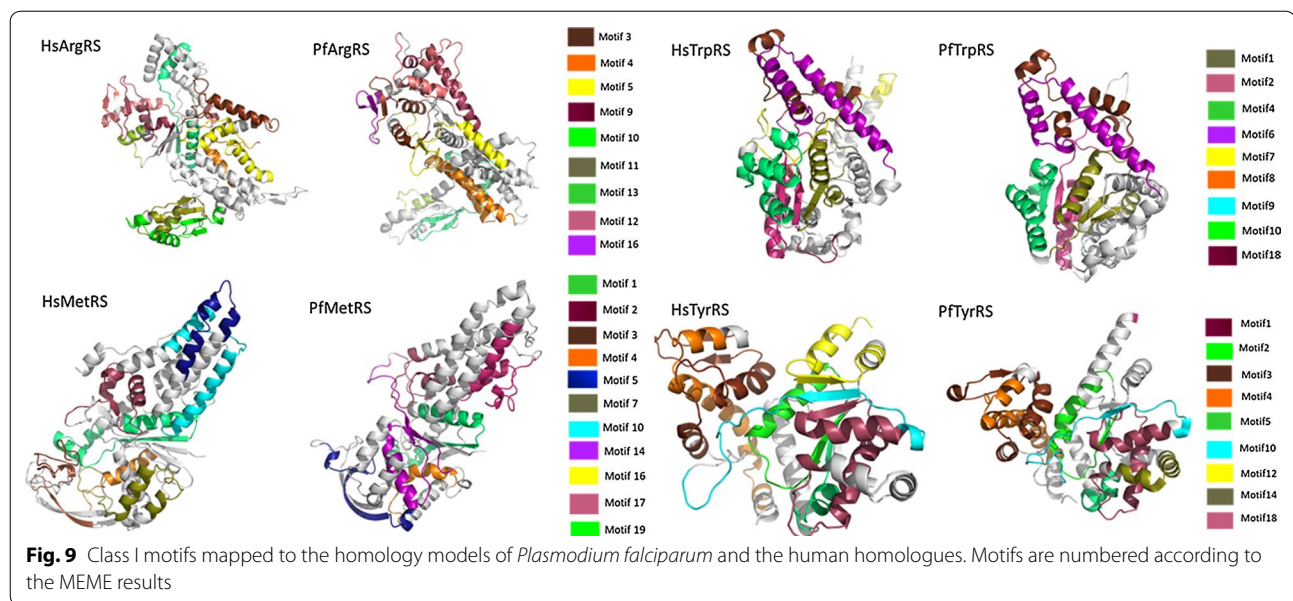
Out of all identified motifs (Additional file 3), the motifs of the six families with structures were mapped into the 3D structures (Fig. 9, Fig. 10 and Additional file 6). The start and end residues for motifs identified in the six families are shown for *P. falciparum* and the human homologues (Table 3). In ArgRS family, motifs were conserved in all analysed structures except Motif 16 which was

present only in the *Plasmodium* sequences but absent in HsArgRS (Additional file 6A, Fig. 9). HsArgRS N-terminal had Motifs 10 and 11 which were absent in *Plasmodium* structures. Motif 13 was not positionally conserved in the analysed structures. In *Plasmodium* it occurs in the anticodon binding domain and the N-terminal while in HsArgRS it occurs in catalytic and the anticodon binding domains (Additional file 6A). In HsMetRS, Motif 5 was in the anticodon binding domain while in *Plasmodium* structures this motif was mapped to the catalytic site. The motif occurs in an alpha helix region in HsMetRS while in PfMetRS the site consists of beta sheets. Motif 14 occurring in the catalytic site and a loop region in PfMetRS was missing in HsMetRS structure (Fig. 9). Motif 10 was present in HsMetRS anticodon binding domain but absent in *Plasmodium*. Other motifs in this family were conserved across all analysed structures. In TrpRS, Motif 7 was only present in HsTrpRS but absent in all *Plasmodium* structures. Motif 8, 9 and 10 were present only in PyTrpRS (Additional file 6C). Motif 1 and Motif 4 were mapped at the catalytic domain in all structures except in PyTrpRS where they are in the anticodon binding domain (Additional file 6C). Motif 2 was present at the catalytic domain of all the TrpRS homology model structures but absent in PyTrpRS. In TyrRS family, Motif 14 was conserved in PfTyrRS, PkTyrRS, PmTyrRS, PvTyrRS and PyTyrRS while Motif 12 was only present in human (Fig. 9 and Additional file 6D).

In Class II, in LysRS, Motif 9 occurs at the catalytic domain in a region consisting of alpha helices and loops in all structures except PfrLysRS and PmLysRS where it mapped in a region consisting of both beta strands and alpha helices. PfrLysRS and PmLysRS did not have Motif 4 present in the anticodon binding domain of all other structures. Motif 8 mapped in a region consisting of alpha helices in all structures except in PfrLysRS and PmLysRS where the region consisted of beta sheets and alpha helices (Additional file 6E). Mapped motifs in ProRS were conserved in all analysed secondary structures (Fig. 10 and Additional file 6F).

New potential druggable sites in *Plasmodium falciparum* aaRSs are identified

FTMap provides information on binding hot spots and the druggability of these sites using probes from fragment libraries [108]. These fragment hits can be used in identification of hits from larger ligands. On the other hand, SiteMap predicts possible binding sites using an algorithm that assigns site points using geometric and energetic properties [106, 107]. The site points are then grouped to give sites which are ranked based on a SiteScore computed based on size, hydrophobicity, exposure to the solvent and the ease of donating or accepting



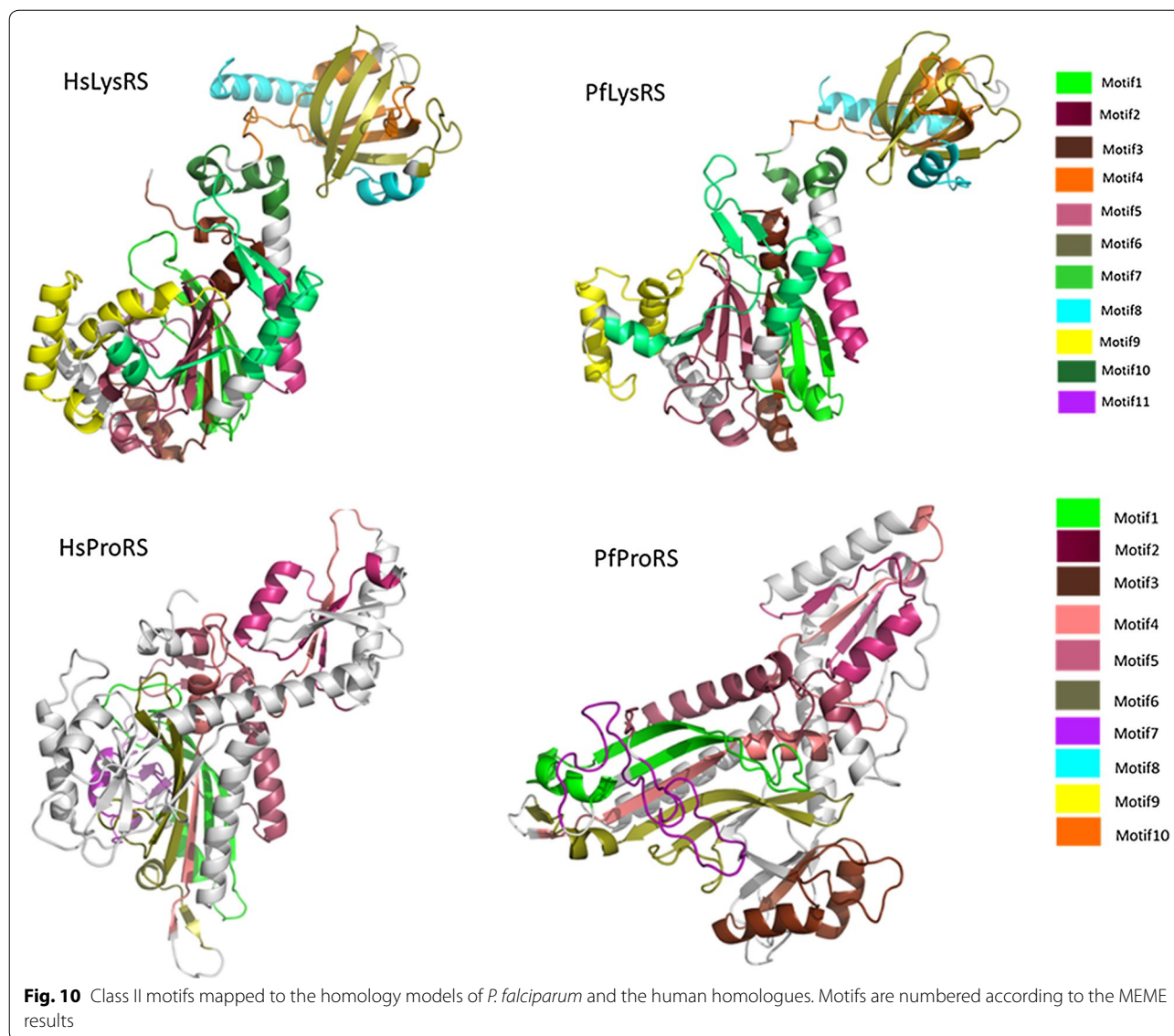
hydrogens. Both FTMap and SiteMap showed consistency in prediction of probable binding sites. In all the six modelled proteins, FTMap and SiteMap were able to predict the known active sites which consists of the ATP and amino acid binding sites as the highest ranked site (Figs. 2, 11). Alternative sites were also predicted in PfArgRS, PfMetRS, PfProRS, and HsProRS that can be targeted for design of new drug classes using both FTMap and SiteMap (Figs. 11, 12 and 13). Since the two tools show consistency in prediction of possible binding sites, we only discuss the results from FTMap in this study.

The identified potential druggable site in PfArgRS is in a region located at the anticodon binding domain characterized by Motif 4 and 6 but the site is not present in HsArgRS (Fig. 9, 11, Table 3). Probes at this site interact with residues in the ABD—His515, Lys518, Ile522, Lys534, Glu537, Asp541 and Tyr34 located in the N-terminal domain. Motif results showed low conservation of these residues with His515 corresponding to Cys577, Lys518 to Arg580, Ile522 to Ile584, Lys534 to Thr592, Glu537 to Asp595, Asp541 to Glu599 and Tyr34 to Ser105 in the human homologue (Fig. 11d). These residues are, however, highly conserved in the other *Plasmodium* sequences studied (Additional file 4A). This region can thus be potentially targeted for inhibitor design with high selectivity to the *Plasmodium* protein as indicated by the low conservation in the human homologue.

The predicted hotspot in PfMetRS is in a pocket formed by Motifs 5, 9, 14, 20 and the loop region of Motif 4 (Figs. 10, 12). Motif 5 is present in HsMetRS,

but this motif occurs in the anticodon binding domain while Motif 14 is not present in HsMetRS (Figs. 10, 12). HsMetRS, however has a Motif 7 present in this site which is absent in the PfMetRS. Probes at the PfMetRS predicted site were interacting with residues Trp481, Ala421, Asp422, Arg415, Pro419, Met385, Leu420, Leu423 and Tyr353. Tyr353, Leu420, Asp422 and Ala421 located in Motif 4 and 9 corresponds to Ala733, Val50, Gln52 and Leu51, respectively in the human homologue. The low conservation of residues in these two motifs may explain why the probes only docked to PfMetRS and not HsMetRS. This difference in conservation at residue level in the predicted site can thus be targeted for the potential development of drugs that bind selectively to PfMetRS. A study by Hussain et al. [22] reported an auxiliary binding site different from ATP and methionine binding sites in PfMetRS. Inhibitors at this site interacted with residues Phe482, Ile231, His483, Tyr454, Trp447, Ile479 and Leu451 [22]. These residues map to Motif 4 and 14 located at the predicted site by FTMap in PfMetRS homology model (Table 3). An auxiliary binding pocket has also been reported in *Trypanosoma brucei* MetRS [157].

The identified potentially druggable site in PfProRS occurs at a region characterised by Motif 1, 5 and 11 which are also present in HsProRS (Figs. 9, 13). In PfProRS, residues Tyr746, Thr397, Phe262, Arg401 and Lys394 were interacting with probes docked at this site while in human, Thr1164, Phe1167, Thr1277, Leu1162, Arg1278 and Thr1276 were interacting with the probes.



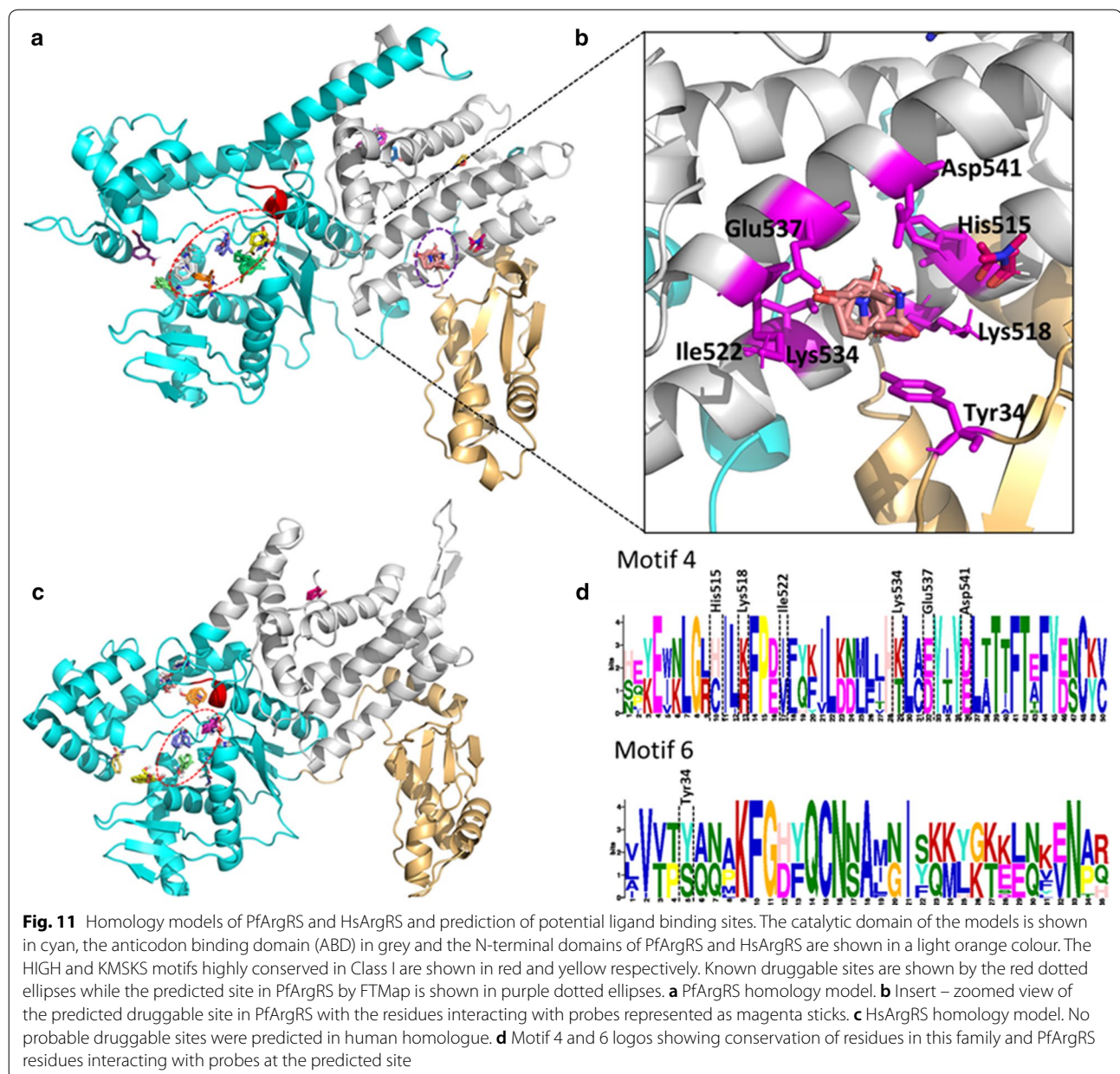
All residues implicated in the interaction of probes in PfProRS were conserved in all the studied sequences in this family except Thr397 which corresponds to Gln1159 in the human homologue (Fig. 13D and Additional file 4R). A previous study by Hewitt et al. [65] reported selective binding of glyburide and TCMDC-124506 at the PfProRS predicted site. This pocket is located at a region formed by $\alpha 5$ (residues 513–524), $\alpha 9$ (residues 261–272) and β -hairpin 1 and 2 (residues 276–287). FTMap showed interactions between Phe262 and Tyr746 and

the probes (Fig. 13) which were also reported to interact with glyburide and TCMDC-124506 [65]. Inhibition of PfProRS by the two compounds is known to be through distortion of the ATP binding site [65]. Binding of glyburide and TCMDC-124506 causes movement of a loop between Val389 and Glu404 displacing Phe405, Arg401 and Arg390 which are key residues in ATP binding [65]. The unique predicted sites in PfArgRS, PfMetRS and PfProRS can thus be targeted through high throughput screening to identify new inhibitors.

Table 3 Starting and ending positions of motifs identified in PfArgRS, PfMetRS, PfTrpRS, PfTyrRS, PfLysRS and PfProRS as well as the human homologues

	Motif 1	Motif 2	Motif 3	Motif 4	Motif 5	Motif 6	Motif 7	Motif 8	Motif 9	Motif 10	Motif 11	Motif 12	Motif 13	Motif 14	Motif 15	Motif 16	Motif 17
PfArgRS	135-184	441-490	308-357	506-555	364-413	30-64	230-279	70-119	186-226	-	-	-	15-29	558-578	420-434	286-306	-
HsArgRS	198-247	499-548	371-420	568-617	428-477	101-135	293-342	142-191	249-289	50-99	620-660	8-48	346-360	478-498	549-563	-	363-370
PfMetRS	243-292	519-559	-	439v459	331-360	-	-	193-242	405-433	-	-	562-611	-	468-517	301-329	-	657-706
HsMetRS	285-334	598-638	-	510-530	711-740	-	460-509	234-283	35-63	-	-	104-153	159-208	-	343-371	839-888	-
PfTrpRS	301-350	447-496	512-561	249-298	392-441	562-611	-	74-123	-	124-173	1-41	203-243	-	43-71	354-368	612-626	174-199
HsTrpRS	154-203	304-353	354-403	102-151	247-296	404-453	48-97	-	-	-	222-242	-	6-46	-	206-220	454-468	-
PfTyrRS	172-221	60-88	265-314	320-360	91-108	-	116-144	-	-	223-251	-	-	30-50	149-169	254-264	362-469	51-58
HsTyrRS	150-199	39-67	240-289	295-335	69-86	488-528	97-125	395-444	339-388	200-228	447-487	1-29	129-149	-	229-239	89-96	30-37
PfLysRS	304-353	465-514	532-581	177-226	254-303	121-170	360-409	70-119	414-463	228-248	-	515-529	40-60	-	-	354-359	-
HsLysRS	297-346	459-508	526-575	169-218	247-296	114-163	353-402	63-112	409-458	221-241	-	509-523	577-597	-	-	347-352	1-15
PfProRS	362-411	268-317	662-711	502-551	570-606	447-496	319-359	-	-	-	718-746	-	-	418-446	615-648	-	109-149
HsProRS	1124-1173	1030-1079	332-381	1266-1315	1338-1374	1209-1258	1081-1121	447-496	666-715	392-441	1484-1512	191-240	257-306	1180-1208	1383-1416	1443-1483	497-537

Dashes show where the motif was not present

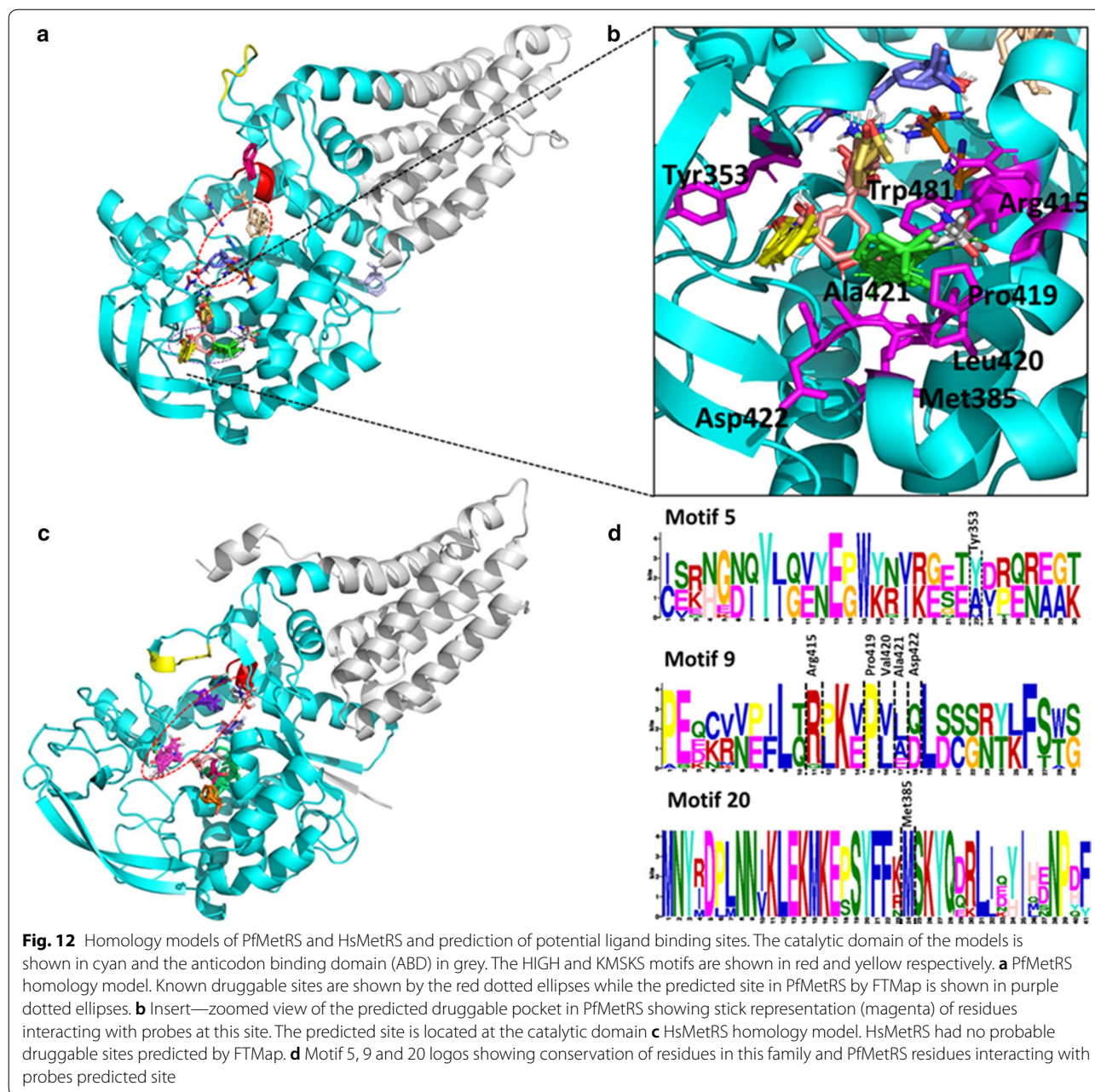


Conclusion

Resistance and selectivity remain a challenge when designing anti-parasitic drugs. This study aimed at getting insights on the differences at sequence and structure level between *Plasmodium* and human aaRS. Motif analysis of the two aaRSs classes showed family specific motifs. Further, analysis of motifs for each family showed *Plasmodium* specific and also mammalian specific motifs. Multiple sequence alignments and motif analysis of aaRS families showed high conservation of the core domains while N- and C- termini of most families showed low conservation. Interestingly, the core domain of LeuRS

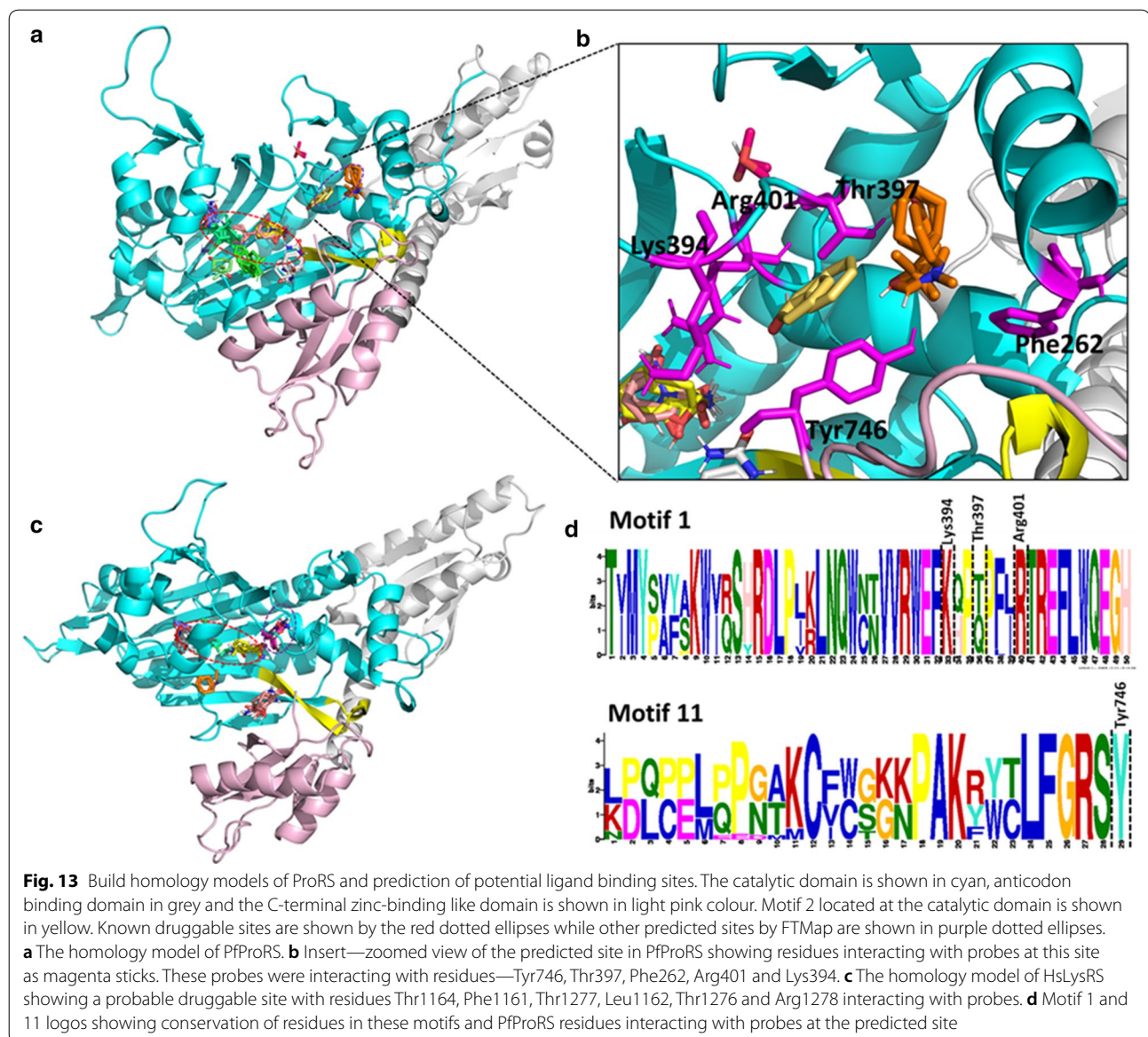
sequences showed low conservation despite functional conservation. ArgRS sequence alignment showed mammalian specific inserts at the N- and C-termini while mammalian TyrRS and ValRS had N-terminal extension not present in *Plasmodium* sequences. Inhibitors can be designed to target the highly variable ABD located either at the N-terminal or the C-terminal.

On doing pairwise sequence identity calculations, ProRS was the most conserved aaRS family while GlyRS was the least conserved. Phylogenetic studies showed that human proteins had different evolutionary history to *Plasmodium* proteins with *Plasmodium* sequences



clustering together. *Plasmodium* sequences also showed high sequence identity compared to the human homologues, which had below 40% sequence identity. *P. yoelii* and *P. berghei* were seen to cluster in trees in most of the aaRS families showing that these proteins are closely related, and this was also depicted by the high sequence identity and shared motifs among them. *Plasmodium fragile*, *P. knowlesi* and *P. vivax* aaRSs were also seen to share evolutionary history and had high

sequence identity. Prediction of additional druggable sites identified hot spots in PfArgRS, PfMetRS and PfProRS. The identified sites showed low conservation and variation of identified motifs between *P. falciparum* proteins and the human homologues. The identified sites can thus be targeted to develop drugs that only selectively bind to *Plasmodium* proteins. From the results of this study, it is evident that despite structural conservation, *Plasmodium* aaRS have key features that



differentiate them from human proteins. These differences can be targeted to develop anti-malarial drugs with less toxicity to the host.

Additional files

Additional file 1. A table showing the data set used in the study with Blast details and crystal structures retrieved from the Protein Data Bank. The species, E-value, identity, accession number, PDB ID and sequence lengths are given.

Additional file 2. Homology model validation results obtained for Verify 3D, QMEAN and ProSA webservers. The z-DOPE scores for each model and the templates used for modelling are also shown.

Additional file 3. Motifs discovered for the 20 aminoacyl tRNA synthetase families using MEME software. The default motif width of 6–50 residues was used. The Mast tool was used to identify overlapping motifs. The number of motifs run for each family varied and motif conservation was presented as number of sites divided by total number of class sequences and results displayed as heatmaps. Motif conservation increases from blue to red.

Additional file 4. Results on mapping of discovered motifs on multiple sequence alignments for the 20 aaRS families. Multiple sequence alignment was performed using TCOFFEE software with default parameters.

Additional file 5. Phylogenetic trees and pairwise sequence calculations for aaRS families: Molecular Phylogenetic calculations were performed using MEGA7. Sequence identity calculations were done using an in-house python script and results displayed as heatmaps. Conservation increases from blue to red.

Additional file 6. Mapping of unique motifs to homology models in *Plasmodium* ArgRS, MetRS, TrpRS, TyrRS, LysRS and ProRS families and the respective human homologues. Motif numbering for each protein is based on the MEME results.

Abbreviations

aaRS: aminoacyl tRNA synthetases; RF: Rossmann fold; CPI: connective peptide I; MSA: multiple sequence alignment; CD: catalytic domain; ABD: anticodon binding domain.

Authors' contributions

ÖTB conceived the project. DWN acquired the data, performed data analysis and wrote the initial draft. All authors contributed in interpretation and discussion of results and writing of the manuscript. Both authors read and approved the final manuscript.

Acknowledgements

D.W.N thanks Rhodes University and National Research Foundation South Africa for financial support. Authors thank Dr. Vuyani Moses for fruitful discussions.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analysed during this study are included in this published article. Protein models are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Not applicable.

Funding

This work is supported by the National Research Foundation (NRF) South Africa (Grant Number 105267) and Rhodes University Henderson Bursary. The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of the funders.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 18 October 2018 Accepted: 27 January 2019

Published online: 06 February 2019

References

- Brooker S, Akhwale W, Pullan R, Estambale B, Clarke SE, Snow RW, et al. Epidemiology of *Plasmodium*-helminth co-infection in Africa: populations at risk, potential impact on anemia, and prospects for combining control. *Am J Trop Med Hyg*. 2007;77:88–98.
- Fèvre EM, Wissmann BV, Welburn SC, Lutumba P. The burden of human African trypanosomiasis. *PLoS Negl Trop Dis*. 2008;2:e333.
- Gething PW, Patil AP, Smith DL, Guerra CA, Elyazar IRF, Johnston GL, et al. A new world malaria map: *Plasmodium falciparum* endemicity in 2010. *Malar J*. 2011;10:378.
- Boatin BA, Basáñez MG, Prichard RK, Awadzi K, Barakat RM, García HH, et al. A research agenda for helminth diseases of humans: towards control and elimination. *PLoS Negl Trop Dis*. 2012;6:e1547.
- Prichard RK, Basanez MG, Boatin BA, McCarthy JS, Garcia HH, Yang GJ, et al. A research agenda for helminth diseases of humans: interventions for control and elimination. *PLoS Negl Trop Dis*. 2012;6:e1549.
- Baird JK. Drug therapy: effectiveness of antimalarial drugs. *N Engl J Med*. 2005;352:1565–77.
- Nayyar GML, Breman JG, Newton PN, Herrington J. Poor-quality antimalarial drugs in southeast Asia and sub-Saharan Africa. *Lancet Infect Dis*. 2012;12:488–96.
- Yeung S, Pongtavornpinyo W, Hastings IM, Mills AJ, White NJ. Antimalarial drug resistance, artemisinin-based combination therapy, and the contribution of modeling to elucidating policy choices. *Am J Trop Med Hyg*. 2004;71:179–86.
- Fairhurst RM, Nayyar GML, Breman JG, Hallett R, Vennerstrom JL, Duong S, et al. Artemisinin-resistant malaria: research challenges, opportunities, and public health implications. *Am J Trop Med Hyg*. 2012;87:231–41.
- Dondorp AM, Nosten F, Yi P, Das D, Phyo AP, Tarning J, et al. Artemisinin resistance in *Plasmodium falciparum* malaria. *N Engl J Med*. 2009;361:455–67.
- Rajendran V, Kalita P, Shukla H, Kumar A, Tripathi T. Aminoacyl-tRNA synthetases: structure, function, and drug discovery. *Int J Biol Macromol*. 2018;111:400–14.
- Fang P, Guo M. Evolutionary limitation and opportunities for developing tRNA synthetase inhibitors with 5-binding-mode classification. *Life*. 2015;5:1703–25.
- Manickam Y, Chaturvedi R, Babbar P, Malhotra N, Jain V, Sharma A. Drug targeting of one or more aminoacyl-tRNA synthetase in the malaria parasite *Plasmodium falciparum*. *Drug Discov Today*. 2018;23:6.
- WHO. World malaria report 2018. Geneva, World Health Organization, 2018.
- Antinori S, Galimberti L, Milazzo L, Corbellino M. Biology of human malaria plasmodia including *Plasmodium knowlesi*. *Mediterr J Hematol Infect Dis*. 2012;4:e2012013.
- Jackson KE, Habib S, Frugier M, Hoen R, Khan S, Pham JS, et al. Protein translation in *Plasmodium* parasites. *Trends Parasitol*. 2011;27:467–76.
- Pham JS, Dawson KL, Jackson KE, Lim EE, Pasaje CFA, Turner KEC, et al. Aminoacyl-tRNA synthetases as drug targets in eukaryotic parasites. *Int J Parasitol Drugs Drug Resist*. 2014;4:1–13.
- Khan S, Sharma A, Jamwal A, Sharma V, Pole AK, Thakur KK, et al. Uneven spread of cis- and trans-editing aminoacyl-tRNA synthetase domains within translational compartments of *P. falciparum*. *Sci Rep*. 2011;1:188.
- Pino P, Aeby E, Foth BJ, Sheiner L, Soldati T, Schneider A, et al. Mitochondrial translation in absence of local tRNA aminoacylation and methionyl-tRNA^{Met} formylation in apicomplexa. *Mol Microbiol*. 2010;76:706–18.
- Jackson KE, Pham JS, Kwek M, De Silva NS, Allen SM, Goodman CD, et al. Dual targeting of aminoacyl-tRNA synthetases to the apicoplast and cytosol in *Plasmodium falciparum*. *Int J Parasitol*. 2012;42:177–86.
- Ikba M, Söll D. Aminoacyl-tRNA Synthesis. *Annu Rev Biochem*. 2000;69:617–50.
- Hussain T, Yogavel M, Sharma A. Inhibition of protein synthesis and malaria parasite development by drug targeting of methionyl-tRNA synthetases. *Antimicrob Agents Chemother*. 2015;59:1856–67.
- Park SG, Schimmel P, Kim S. Aminoacyl tRNA synthetases and their connections to disease. *Proc Natl Acad Sci USA*. 2008;105:11043–9.
- Guo M, Schimmel P, Yang XL. Functional expansion of human tRNA synthetases achieved by structural inventions. *FEBS Lett*. 2010;584:434–42.
- Paul M, Schimmel P. Essential nontranslational functions of tRNA synthetases. *Nat Chem Biol*. 2013;9:145–53.
- Bhatt TK, Khan S, Dwivedi VP, Bandy MM, Sharma A, Chandele A, et al. Malaria parasite tyrosyl-tRNA synthetase secretion triggers pro-inflammatory responses. *Nat Commun*. 2011;2:530.
- 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.
- Sharma A, Khan S, Sharma A, Belrhali H, Yogavel M. Structural basis of malaria parasite lysyl-tRNA synthetase inhibition by cladosporin. *J Struct Funct Genomics*. 2014;15:63–71.
- Sherma A, Yogavel M, Sharma A. Structural and functional attributes of malaria parasite diadenosine tetraphosphate hydrolase. *Sci Rep*. 2016;6:19981.
- Miller LH, Baruch DI, Marsh K, Doumbo OK. The pathogenic basis of malaria. *Nature*. 2002;415:673–9.
- Sharma A, Sharma A. *Plasmodium falciparum* mitochondria import tRNAs along with an active phenylalanyl-tRNA synthetase. *Biochem J*. 2015;465:459–69.
- Pham JS, Sakaguchi R, Yeoh LM, De Silva NS, McFadden GI, Hou Y-M, et al. A dual-targeted aminoacyl-tRNA synthetase in *Plasmodium falciparum* charges cytosolic and apicoplast tRNACys. *Biochem J*. 2014;458:513–23.

33. Manickam Y, Chaturvedi R, Babbar P, Malhotra N, Jain V, Sharma A. Drug targeting of one or more aminoacyl-tRNA synthetase in the malaria parasite *Plasmodium falciparum*. *Drug Discov Today*. 2018;0:26–33.
34. Herman JD, Pepper LR, Cortese JF, Galinsky K, Zuzarte-luis V, Derbyshire ER, et al. The cytoplasmic prolyl-tRNA synthetase of the malaria parasite is a dual-stage target for drug development. *Sci Transl Med*. 2016;7:288ra77.
35. Hoepfner D, McNamara CW, Lim CS, Studer C, Riedl R, Aust T, et al. Selective and specific inhibition of the *Plasmodium falciparum* lysyl-tRNA synthetase by the fungal secondary metabolite cladosporin. *Cell Host Microbe*. 2012;11:654–63.
36. Sonoiki E, Palencia A, Guo D, Ahyong V, Dong C, Li X, et al. Antimalarial benzoxaboroles target *Plasmodium falciparum* leucyl-tRNA synthetase. *Antimicrob Agents Chemother*. 2016;60:4886–95.
37. Filisetti D, Théobald-Dietrich A, Mahmoudi N, Rudinger-Thirion J, Candolfi E, Frugier M. Aminoacylation of *Plasmodium falciparum* tRNA(Asn) and insights in the synthesis of asparagine repeats. *J Biol Chem*. 2013;288:36361–71.
38. Mailu BM, Ramasamay G, Mudeppa DG, Li L, Lindner SE, Peterson MJ, et al. A nondiscriminating glutamyl-tRNA synthetase in the *Plasmodium* apicoplast: the first enzyme in an indirect aminoacylation pathway. *J Biol Chem*. 2013;288:32539–52.
39. Bonnefond L, Fender A, Rudinger-Thirion J, Giegé R, Florentz C, Sissler M. Toward the full set of human mitochondrial aminoacyl-tRNA synthetases: characterization of AspRS and TyrRS. *Biochemistry*. 2005;44:4805–16.
40. Antonellis A, Green ED. The Role of Aminoacyl-tRNA synthetases in genetic diseases. *Annu Rev Genomics Hum Genet*. 2008;9:87–107.
41. Bhatt T, Kapil C, Khan S, Jairajpuri M, Sharma V, Santoni D, et al. A genomic glimpse of aminoacyl-tRNA synthetases in malaria parasite *Plasmodium falciparum*. *BMC Genomics*. 2009;10:644.
42. Eriani G, Cavarelli J, Martin F, Ador L, Rees B, Thierry JC, et al. The class II aminoacyl-tRNA synthetases and their active site: evolutionary conservation of an ATP binding site. *J Mol Evol*. 1995;40:499–508.
43. Chaliotis A, Vlastaridis P, Mossialos D, Ibba M, Becker HD, Stathopoulos C, et al. The complex evolutionary history of aminoacyl-tRNA synthetases. *Nucleic Acids Res*. 2016;45:1059–68.
44. Chen JF, Guo NN, Li T, Wang ED, Wang YL. CP1 domain in *Escherichia coli* leucyl-tRNA synthetase is crucial for its editing function. *Biochemistry*. 2000;39:6726–31.
45. Doublé S, Bricogne G, Gilmore C, Carter CW. Tryptophanyl-tRNA synthetase crystal structure reveals an unexpected homology to tyrosyl-tRNA synthetase. *Structure*. 1995;3:17–31.
46. Yadavalli SS, Ibba M. Quality control in aminoacyl-tRNA synthesis: its role in translational fidelity. *Adv Protein Chem Struct Biol*. 2012;86:1–43.
47. Perona JJ, Hadd A. Structural diversity and protein engineering of the aminoacyl-tRNA synthetases. *Biochemistry*. 2012;51:8705–29.
48. Cusack S. Aminoacyl-tRNA synthetases Stephen. *Curr Opin Struct Biol*. 1997;7:881–9.
49. Perona JJ, Gruic-Sovulj I. Synthetic and editing mechanisms of aminoacyl-tRNA synthetases. *Top Curr Chem*. 2014;344:1–41.
50. Ibba M, Losey HC, Kawarabayasi Y, Kikuchi H, Bunjun S, Söll D. Substrate recognition by class I lysyl-tRNA synthetases: a molecular basis for gene displacement. *Proc Natl Acad Sci USA*. 1999;96:418–23.
51. Bennett EJ, Shaler TA, Woodman B, Ryu K-Y, Zaitseva TS, Becker CH, et al. Anticodon recognition and discrimination by the alpha-helix cage domain of class I lysyl-tRNA synthetase. *J Biol Chem*. 2007;282:11033–8.
52. Pouplana LR, Buechter DD, Davis MW, Schimmel P. Idiographic representation of conserved domain of a class II tRNA synthetase of unknown structure. *Protein Sci*. 1993;2:2259–62.
53. Smith TF, Hartman H. The evolution of Class II aminoacyl-tRNA synthetases and the first code. *FEBS Lett*. 2015;589:3499–507.
54. Normanly J, Ollick T, Abelson J. Eight base changes are sufficient to convert a leucine-inserting tRNA into a serine-inserting tRNA. *Proc Natl Acad Sci USA*. 1992;89:5680–4.
55. Yao P, Fox PL. Aminoacyl-tRNA synthetases in medicine and disease. *EMBO Mol Med*. 2013;5:332–43.
56. Martinez-Rodriguez L, Erdogan O, Jimenez-Rodriguez M, Gonzalez-Rivera K, Williams T, Li L, et al. Functional class I and II amino acid-activating enzymes can be coded by opposite strands of the same gene. *J Biol Chem*. 2015;290:19710–25.
57. Ruff M, Krishnaswamy S, Boeglin M, Poterszman A, Mitschler A, Podjarny A, et al. Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA(Asp). *Science*. 1991;252:1682–9.
58. Ibba M, Bono JL, Rosa PA, Söll D. Archaeal-type lysyl-tRNA synthetase in the Lyme disease spirochete *Borrelia burgdorferi*. *Proc Natl Acad Sci USA*. 1997;94:14383–8.
59. Khan S, Garg A, Camacho N, Van Rooyen J, Kumar Pole A, Belrhali H, et al. Structural analysis of malaria-parasite lysyl-tRNA synthetase provides a platform for drug development. *Acta Crystallogr Sect D: Biol Crystallogr*. 2013;69:785–95.
60. Saint-Léger A, de Ribas Pouplana L. The growing pipeline of natural aminoacyl-tRNA synthetase inhibitors for malaria treatment. *Bioengineered*. 2016;7:60–4.
61. Baragaña B, Hallyburton I, Lee MCS, Norcross NR, Grimaldi R, Otto TD, et al. A novel multiple-stage antimalarial agent that inhibits protein synthesis. *Nature*. 2015;522:315–20.
62. Jain V, Kikuchi H, Oshima Y, Sharma A, Yogavel M. Structural and functional analysis of the anti-malarial drug target prolyl-tRNA synthetase. *J Struct Funct Genomics*. 2014;15:181–90.
63. Kato N, Comer E, Sakata-Kato T, Sharma A, Sharma M, Maetani M, et al. Diversity-oriented synthesis yields novel multistage antimalarial inhibitors. *Nature*. 2016;538:344–9.
64. Keller TL, Zocco D, Sundrud MS, Hendrick M, Edenius M, Yum J, et al. Halofuginone and other febrifugine derivatives inhibit prolyl-tRNA synthetase. *Nat Chem Biol*. 2012;8:311–7.
65. Hewitt SN, Dranow DM, Horst BG, Abendroth JA, Forte B, Hallyburton I, et al. Biochemical and structural characterization of selective allosteric inhibitors of the *Plasmodium falciparum* drug target, prolyl-tRNA synthetase. *ACS Infect Dis*. 2017;3:34–44.
66. Jain V, Yogavel M, Oshima Y, Kikuchi H, Touquet B, Hakimi MA, et al. Structure of prolyl-tRNA synthetase-halofuginone complex provides basis for development of drugs against malaria and toxoplasmosis. *Structure*. 2015;23:819–29.
67. Jain V, Yogavel M, Kikuchi H, Oshima Y, Hariguchi N, Matsumoto M, et al. Targeting prolyl-tRNA synthetase to accelerate drug discovery against malaria, leishmaniasis, toxoplasmosis, cryptosporidiosis, and coccidiosis. *Structure*. 2017;25(1495–1505):e6.
68. Son J, Lee EH, Park M, Kim JH, Kim J, Kim S, et al. Conformational changes in human prolyl-tRNA synthetase upon binding of the substrates proline and ATP and the inhibitor halofuginone. *Acta Crystallogr Sect D: Biol Crystallogr*. 2013;69:2136–45.
69. Zhou H, Sun L, Yang X-L, Schimmel P. ATP-directed capture of bioactive herbal-based medicine on human tRNA synthetase. *Nature*. 2013;494:121–4.
70. Burrows JN, Chibale K, Wells TNC. The state of the art in anti-malarial drug discovery and development. *Curr Top Med Chem*. 2011;11:1226–54.
71. Baker DA. Malaria gametocytogenesis. *Mol Biochem Parasitol*. 2010;172:57–65.
72. Chihade JW, Brown JR, Schimmel PR, de Pouplana LR. Origin of mitochondria in relation to evolutionary history of eukaryotic alanyl-tRNA synthetase. *Proc Natl Acad Sci USA*. 2000;97:12153–7.
73. De Pouplana LR, Schimmel P. Visions & reflections a view into the origin of life: aminoacyl-tRNA synthetases. *Nature*. 2000;57:865–70.
74. Faya N, Penkler DL, Tasthan Bishop Ö. Human, vector and parasite Hsp90 proteins: a comparative bioinformatics analysis. *FEBS Open Bio*. 2015;5:916–27.
75. Pink R, Hudson A, Mouriès MA, Bendig M. Opportunities and challenges in antiparasitic drug discovery. *Nat Rev Drug Discov*. 2005;4:727–40.
76. Bunjun S, Stathopoulos C, Graham D, Min B, Kitabatake M, Wang AL, et al. A dual-specificity aminoacyl-tRNA synthetase in the deep-rooted eukaryote *Giardia lamblia*. *Proc Natl Acad Sci USA*. 2000;97:12997–3002.
77. Hatherley R, Clitheroe CL, Faya N, Tasthan Bishop Ö. *Plasmodium falciparum* Hop: detailed analysis on complex formation with Hsp70 and Hsp90. *Biochem Biophys Res Commun*. 2015;456:440–5.
78. Hatherley R, Blatch GL, Bishop ÖT. *Plasmodium falciparum* Hsp70-x: a heat shock protein at the host-parasite interface. *J Biomol Struct Dyn*. 2014;32:1766–79.
79. Shibata S, Gillespie JR, Kelley AM, Napuli AJ, Zhang Z, Kovzun KV, et al. Selective inhibitors of methionyl-tRNA synthetase have potent activity

- against *Trypanosoma brucei* infection in mice. *Antimicrob Agents Chemother*. 2011;55:1982–9.
80. Dahl EL, Rosenthal PJ. Multiple antibiotics exert delayed effects against the *Plasmodium falciparum* apicoplast. *Antimicrob Agents Chemother*. 2007;51:3485–90.
 81. Schimmel P. Development of tRNA synthetases and connection to genetic code and disease. *Protein Sci*. 2008;17:1643–52.
 82. Richa A, Tanya B, Jeff B, Dennis AB, Colleen B, Evan B, Devon B, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 2014;41:1–12.
 83. The UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res*. 2015;43:D204–12.
 84. 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.
 85. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Databank. *Nucleic Acids Res*. 2000;28:235–42.
 86. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME suite. *Nucleic Acids Res*. 2015;43:W39–49.
 87. Bailey TL, Gribskov M. Combining evidence using p-values: application to sequence homology searches. *Bioinformatics*. 1998;14:48–54.
 88. Kelm S, Shi J, Deane CM. MEDELLER: homology-based coordinate generation for membrane proteins. *Bioinformatics*. 2010;26:2833–40.
 89. Söding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res*. 2005;33:W244–8.
 90. Hatherley R, Brown DK, Glenister M, Bishop ÖT. PRIMO: an interactive homology modeling pipeline. *PLoS ONE*. 2016;11:e0166698.
 91. Jain V, Yogavel M, Sharma A. Dimerization of Arginyl-tRNA synthetase by free heme drives inactivation in *Plasmodium falciparum*. *Structure*. 2016;24:1476–87.
 92. Koh CY, Kim JE, Napoli AJ, Verlinde CLMJ, Fan E, Buckner FS, et al. Crystal structures of *Plasmodium falciparum* cytosolic tryptophanyl-tRNA synthetase and its potential as a target for structure-guided drug design. *Mol Biochem Parasitol*. 2013;189:26–32.
 93. Barros-Álvarez X, Kerchner KM, Koh CY, Turley S, Pardon E, Steyaert J, et al. Leishmania donovani tyrosyl-tRNA synthetase structure in complex with a tyrosyl adenylate analog and comparisons with human and protozoan counterparts. *Biochimie*. 2017;138:124–36.
 94. Ofir-Birin Y, Fang P, Bennett SP, Zhang H-M, Wang J, Rachmin I, et al. Structural switch of lysyl-tRNA synthetase between translation and transcription. *Mol Cell*. 2013;49:30–42.
 95. Ojo KK, Ranade RM, Zhang Z, Dranow DM, Myers JB, Choi R, et al. *Brucella melitensis* methionyl-tRNA-synthetase (MetRS), a potential drug target for brucellosis. *PLoS ONE*. 2016;11:e0160350.
 96. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res*. 2007;35:W407–10.
 97. Eisenberg D, Lüthy R, Bowie JU. VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol*. 1997;277:396–406.
 98. Benkert P, Künzli M, Schwede T. QMEAN server for protein model quality estimation. *Nucleic Acids Res*. 2009;37:W510–4.
 99. Di Tommaso P, Moretti S, Xenarios I, Orobitg M, Montanyola A, Chang JM, et al. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res*. 2011;39:W13–7.
 100. Pei J, Kim BH, Grishin NV. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res*. 2008;36:2295–300.
 101. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 2009;25:1189–91.
 102. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 2016;33:1870–4.
 103. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011;28:2731–9.
 104. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. *Bioinformatics*. 1992;8:275–82.
 105. Ngan CH, Bohnuud T, Mottarella SE, Beglov D, Villar EA, Hall DR, et al. FTMAP: extended protein mapping with user-selected probe molecules. *Nucleic Acids Res*. 2012;40:W271–5.
 106. Halgren T. New method for fast and accurate binding-site identification and analysis. *Chem Biol Drug Des*. 2007;69:146–8.
 107. Halgren TA. Identifying and characterizing binding sites and assessing druggability. *J Chem Inf Model*. 2009;49:377–89.
 108. Kozakov D, Grove LE, Hall DR, Bohnuud T, Mottarella SE, Luo L, et al. The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat Protoc*. 2015;10:733–55.
 109. Moras D. Structural and functional relationships between aminoacyl-tRNA synthetases. *Trends Biochem Sci*. 1992;17:159–64.
 110. Fourmy D, Mechulam Y, Blanquet S. Crucial role of an idiosyncratic insertion in the Rossman fold of class 1 aminoacyl-tRNA synthetases: the case of methionyl-tRNA synthetase. *Biochemistry*. 1995;34:15681–8.
 111. Mailu BM, Li L, Arthur J, Nelson TM, Ramasamy G, Fritz-Wolf K, et al. *Plasmodium* apicoplast Gln-tRNA^{Gln} biosynthesis utilizes a unique GatAB amidotransferase essential for erythrocytic stage parasites. *J Biol Chem*. 2015;290:29629–41.
 112. Kyriacou SV, Deutscher MP. An important role for the multienzyme aminoacyl-tRNA synthetase complex in mammalian translation and cell growth. *Mol Cell*. 2008;29:419–27.
 113. Khan S, Doerig C, Baker D, Billker O, Blackman M, Chitnis C, et al. Recent advances in the biology and drug targeting of malaria parasite aminoacyl-tRNA synthetases. *Malar J*. 2016;15:203.
 114. Robinson JC, Kerjan P, Mirande M. Macromolecular assemblage of aminoacyl-tRNA synthetases: quantitative analysis of protein-protein interactions and mechanism of complex assembly. *J Mol Biol*. 2000;304:983–94.
 115. Han JM, Kim JY, Kim S. Molecular network and functional implications of macromolecular tRNA synthetase complex. *Biochem Biophys Res Commun*. 2003;303:985–93.
 116. Lee SW. Aminoacyl-tRNA synthetase complexes: beyond translation. *J Cell Sci*. 2004;117:3725–34.
 117. Guo M, Yang X-L, Schimmel P. New functions of aminoacyl-tRNA synthetases beyond translation. *Nat Rev Mol Cell Biol*. 2010;11:668–74.
 118. Ko YG, Kim EK, Kim T, Park H, Park HS, Choi EJ, et al. Glutamine-dependent antiapoptotic interaction of human glutamyl-tRNA synthetase with apoptosis signal-regulating kinase 1. *J Biol Chem*. 2001;276:6030–6.
 119. Han JM, Jeong SJ, Park MC, Kim G, Kwon NH, Kim HK, et al. Leucyl-tRNA synthetase is an intracellular leucine sensor for the mTORC1-signaling pathway. *Cell*. 2012;149:410–24.
 120. Khan S, Garg A, Sharma A, Camacho N, Picchioni D, Saint-Léger A, et al. An appended domain results in an unusual architecture for malaria parasite tryptophanyl-tRNA synthetase. *PLoS ONE*. 2013;8:e66224.
 121. Cusack S, Berthet-Colominas C, Härtlein 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;347:249–55.
 122. Cavarelli J, Eriani G, Rees B, Ruff M, Boeglin M, Mitschler A, et al. The active site of yeast aspartyl-tRNA synthetase: structural and functional aspects of the aminoacylation reaction. *EMBO J*. 1994;13:327–37.
 123. Dignam JD, Guo J, Griffith WP, Garbett NC, Holloway A, Mueser T. Allosteric interaction of nucleotides and tRNA ala with *E. coli* alanyl-tRNA synthetase. *Biochemistry*. 2011;50:9886–900.
 124. Arnez JG, Harris DC, Mitschler A, Rees B, Francklyn CS, Moras D. Crystal structure of histidyl-tRNA synthetase from *Escherichia coli* complexed with histidyl-adenylate. *EMBO J*. 1995;14:4143–55.
 125. Logan DT, Mazauric MH, Kern D, Moras D. Crystal structure of glycyl-tRNA synthetase from *Thermus thermophilus*. *EMBO J*. 1995;14:4156–67.
 126. Onesti S, Miller AD, Brick P. The crystal structure of the lysyl-tRNA synthetase (LysU) from *Escherichia coli*. *Structure*. 1995;3:163–76.
 127. Berthet-Colominas C, Seignovert L, Härtlein M, Grotli M, Cusack S, Leberman R. The crystal structure of asparaginyl-tRNA synthetase from *Thermus thermophilus* and its complexes with ATP and asparaginyl-adenylate: the mechanism of discrimination between asparagine and aspartic acid. *EMBO J*. 1998;17:2947–60.

128. Becker HD, Reinbolt J, Kreutzer R, Giegé R, Kern D. Existence of two distinct aspartyl-tRNA synthetases in *Thermus thermophilus*. Structural and biochemical properties of the two enzymes. *Biochemistry*. 1997;36:8785–97.
129. Min B, Pelaschier JT, Graham DE, Tumbula-Hansen D, Söll D. Transfer RNA-dependent amino acid biosynthesis: an essential route to asparagine formation. *Proc Natl Acad Sci USA*. 2002;99:2678–83.
130. Sheppard K, Yuan J, Hohn MJ, Jester B, Devine KM, Söll D. From one amino acid to another: tRNA-dependent amino acid biosynthesis. *Nucleic Acids Res*. 2008;36:1813–25.
131. Delarue M. Aminoacyl-tRNA synthetases. *Curr Opin Struct Biol*. 1995;5:48–55.
132. Guigou L, Shalav V, Mirande M. The tRNA-interacting factor p43 associates with mammalian arginyl-tRNA synthetase but does not modify its tRNA aminoacylation properties. *Biochemistry*. 2004;43:4592–600.
133. Ferber S, Ciechanover A. Role of arginine-tRNA in protein degradation by the ubiquitin pathway. *Nature*. 1987;326:808–11.
134. Zheng YG, Wei H, Ling C, Xu MG, Wang ED. Two forms of human cytoplasmic arginyl-tRNA synthetase produced from two translation initiations by a single mRNA. *Biochemistry*. 2006;45:1338–44.
135. Jeong EJ, Hwang GS, Kim KH, Kim MJ, Kim S, Kim KS. Structural analysis of multifunctional peptide motifs in human bifunctional tRNA synthetase: identification of RNA-binding residues and functional implications for tandem repeats. *Biochemistry*. 2000;39:15775–82.
136. Rho SB, Lee JS, Jeong EJ, Kim KS, Kim YG, Kim S. A multifunctional repeated motif is present in human bifunctional tRNA synthetase. *J Biol Chem*. 1998;273:11267–73.
137. Fett R, Knippers R. The primary structure of human glutaminyl-tRNA synthetase: a highly conserved core, amino acid repeat regions, and homologies with translation elongation factors. *J Biol Chem*. 1991;266:1448–55.
138. Rho SB, Kim MJ, Lee JS, Seol W, Motegi H, Kim S, et al. Genetic dissection of protein-protein interactions in multi-tRNA synthetase complex. *Proc Natl Acad Sci USA*. 1999;96:4488–93.
139. Hsu JL, Martinis SA. A flexible peptide tether controls accessibility of a unique C-terminal RNA-binding domain in leucyl-tRNA synthetases. *J Mol Biol*. 2008;376:482–91.
140. Wakasugi K. Two distinct cytokines released from a human aminoacyl-tRNA synthetase. *Science*. 1999;284:147–51.
141. Frugier M, Moulinier L, Giegé R. A domain in the N-terminal extension of class IIb eukaryotic aminoacyl-tRNA synthetases is important for tRNA binding. *EMBO J*. 2000;19:2371–80.
142. Dou X, Limmer S, Kreutzer R. DNA-binding of phenylalanyl-tRNA synthetase is accompanied by loop formation of the double-stranded DNA. *J Mol Biol*. 2001;305:451–8.
143. Das P, Babbar P, Malhotra N, Sharma M, Jachak GR, Gonnade RG, et al. Specific stereoisomeric conformations determine the drug potency of cladosporin scaffold against malarial parasite. *J Med Chem*. 2018;61:5664–78.
144. de Oca MM, Engwerda C, Haque A. *Plasmodium berghei* ANKA (PbA) infection of C57BL/6J mice: a model of severe malaria. *Methods Mol Biol*. 2013;1031:203–13.
145. Otto TD, Böhme U, Jackson AP, Hunt M, Franke-Fayard B, Hoesjmakers WAM, et al. A comprehensive evaluation of rodent malaria parasite genomes and gene expression. *BMC Biol*. 2014;12:86.
146. Mitsui H, Arisue N, Sakihama N, Inagaki Y, Horii T, Hasegawa M, et al. Phylogeny of Asian primate malaria parasites inferred from apicomplast genome-encoded genes with special emphasis on the positions of *Plasmodium vivax* and *P. fragile*. *Gene*. 2010;450:32–8.
147. Pain A, Böhme U, Berry AE, Mungall K, Finn RD, Jackson AP, et al. The genome of the simian and human malaria parasite *Plasmodium knowlesi*. *Nature*. 2008;455:799–803.
148. Carlton JM, Adams JH, Silva JC, Bidwell SL, Lorenzi H, Caler E, et al. Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature*. 2008;455:757–63.
149. Höglind A, Arestrom I, Ehrnfelt C, Masjedi K, Zuber B, Giavedoni L, et al. Systematic evaluation of monoclonal antibodies and immunoassays for the detection of interferon- γ and interleukin-2 in old and new world non-human primates. *J Immunol Methods*. 2017;441:39–48.
150. Le SQ, Gascuel O. An improved general amino acid replacement matrix. *Mol Biol Evol*. 2008;25:1307–20.
151. Bishop ÖT, Kroon M. Study of protein complexes via homology modeling, applied to cysteine proteases and their protein inhibitors. *J Mol Model*. 2011;17:3163–72.
152. Shen MY, Sali A. Statistical potential for assessment and prediction of protein structures. *Protein Sci*. 2006;15:2507–24.
153. Eramian D, Shen M, Devos D, Melo F, Sali A, Marti-Renom MA. A composite score for predicting errors in protein structure models. *Protein Sci*. 2006;15:1653–66.
154. Marko AC, Stafford K, Wymore T. Stochastic pairwise alignments and scoring methods for comparative protein structure modeling. *J Chem Inf Model*. 2007;47:1263–70.
155. Chen H, Kihara D. Estimating quality of template-based protein models by alignment stability. *Proteins Struct Funct Genet*. 2008;71:1255–74.
156. Yang X-L, Otero FJ, Skene RJ, DE McRee, Schimmel P, de Ribas Pouplana L. Crystal structures that suggest late development of genetic code components for differentiating aromatic side chains. *Proc Natl Acad Sci USA*. 2003;100:15376–80.
157. Koh CY, Kim JE, Shibata S, Ranade RM, Yu M, Liu J, et al. Distinct states of methionyl-tRNA synthetase indicate inhibitor binding by conformational selection. *Structure*. 2012;20:1681–91.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

