Aminoacyl tRNA Synthetases as Malarial Drug Targets: A Comparative Bioinformatics Study

Dorothy Wavinya Nyamai, Özlem Tastan Bishop*

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

*Correspondence to o.tastanbishop@ru.ac.za

Abstract

Treatment of parasitic diseases has been challenging due to the development of drug resistance by parasites, and thus there is need to identify new class of drugs and drug targets. Protein translation is important for survival of plasmodium and the pathway is present in all the life cycle stages of the plasmodium parasite. Aminoacyl tRNA synthetases are primary enzymes in protein translation as they catalyse the first reaction where an amino acid is added to the cognate tRNA. Currently, there is limited research on comparative studies of aminoacyl tRNA synthetases as potential drug targets. The aim of this study is to understand differences between plasmodium and human aminoacyl tRNA synthetases through bioinformatics analysis. *Plasmodium falciparum*, *P. fragile*, *P. vivax*, *P. ovale*, *P. knowlesi*, *P. bergei*, *P. malariae* and human aminoacyl tRNA synthetase sequences were retrieved from UniProt database and grouped into 20 families based on amino acid specificity. Despite functional and structural conservation, multiple sequence analysis, motif discovery, pairwise sequence identity calculations and molecular phylogenetic analysis showed striking differences between parasite and human proteins. Prediction of alternate binding sites revealed potential druggable sites in PfArgRS, PfMetRS and PfProRS at regions that were weakly conserved when compared to the human homologues. These differences provide a basis for further exploration of plasmodium aminoacyl tRNA synthetases as potential drug targets.

Keywords

Aminoacyl tRNA synthetases; motif analysis; phylogenetic tree calculations, homology modelling

Abbreviations

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

Introduction

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 come up with 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 [6,7]. 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 [8–10].

Plasmodium parasite causes malaria disease, which is a major public concern due to its high mortality and morbidity rates [10,11]. There are five plasmodium species that cause malaria in human, and these are *Plasmodium falciparum* (*P. falciparum*), *Plasmodium vivax (P. vivax), Plasmodium knowlesi* (*P. knowlesi), Plasmodium malariae* (*P. malariae)* and *Plasmodium ovale* (*P. ovale*) [12]. Plasmodium has three genomes; cytoplasm, mitochondrial and apicoplast, and each of them needs a functional protein translation mechanism for growth and survival [10,13,14]. 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 [13,15–17].

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 [8,9,18]. 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 [19]. 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 [20–22]. *P. 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 [23,24]. *P. 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 [25,26].

Of the five human malaria parasites, *P. falciparum*, known to be highly pathogenic, causes the most severe forms of malaria, and is responsible for most of the malaria mortality cases reported across the world [27]. *P. 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 aminoacyl synthetase (PheRS) is encoded in the mitochondria [28–30]. *P. falciparum* protein translation in the mitochondria relies on enzymes imported from the cytoplasm including aaRSs [28]. 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 [15,17,30,31]. 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 [29,30]. 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 [29]. *P. falciparum* cytoplasm has genes that encode ProRS, AspRS, IleRS, LysRS, HisRS, PheRS, AsnRS, ArgRS, GlnRS, SerRS, TrpRS, ValRS, MetRS, LeuRS, GluRS and TyrRS [31].

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 [8]. The human mitochondria acquires nuclear-encoded aaRSs with the aid of translation signals within the aaRSs proteins to carry out protein synthesis [32]. 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 [32]. 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 [8,32]. 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 [32,33].

Generally, aaRSs proteins are classified into two distinct classes based on key features of the catalytic site architecture and the manner of charging tRNA [18,20]. Class I aaRSs include IleRS, LeuRS, MetRS, CysRS, GlnRS, GluRS, TrpRS, ValRS, ArgRS and TyrRS. Proteins

in this class have a catalytic domain (Figure 1A) characterized by a Rossmann fold (RF) located near the N-terminal [34]. 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 [35,36] as shown in Figure 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 [8]. 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 [37]. 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 [8]. With the exception of TyrRS, MetRS and TrpRS all Class I enzymes are monomeric [8]. 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 [34,38]. 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 [39]. The editing domain proofreads the aminoacylation process through pre-transfer or post-transfer editing [8]. 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 [8]. The ABD of proteins in Class I occurs at the C-terminal which binds the anticodons in the cognate tRNA [40].

Class I enzymes binds to the tRNA acceptor end through the minor groove and these enzymes aminoacylate the 2'-OH group of adenosine nucleotide [8,40]. Proteins in this class can further be classified into five subclasses based on sequence similarity and

physicochemical properties of their substrates [41,42]. 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 [8,43]. 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 (Figure 1A) [8,40]. 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 [44].

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 [45,46]. Class IIa proteins exist as dimers and includes ProRS, SerRS, GlyRS, ThrRS, HisRS and all have the aminoacylation domain at the N-terminal [8]. Members of this subclass have an ABD at the C-terminal (figure 1B). The anticodon binding domain is absent in SerRS as this protein does not require an anticodon to discriminate its cognate tRNA [47,48]. 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 (Figure 1B) [40]. 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 (Figure 1B). Class IIc includes PheRS, AlaRS and GlyRS and all exist in tetrameric conformation [8,46]. 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 (Figure 1B) [40].

Class II enzymes possess a catalytic site domain characterized by seven β-sheet strands connected by α-helices [49]. This domain, just like the Class I catalytic domain couples amino acid, ATP and tRNA 3'-terminus during catalytic reactions [40,50]. Class II catalytic domain has three weakly conserved motifs (Figure 1B, Figure 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 [51]. 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 [36]. 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 [52] while Class II LysRS occurs in most bacteria and all eukaryotes [8].

Figure 1. Key domains of aminoacyl tRNA synthetases. **A)** Class I aaRS showing the Catalytic Domain (CD) and the anticodon binding domain (ABD). The CD has a CPI insert in all the proteins in this Class. The CPI insert (orange) in IleRS, LeuRS and ValRS is enlarged to form an editing domain while in TyrRS and TrpRS it functions in the formation of dimers. ArgRS has an Add1 domain (cyan) at the N-terminus which is involved in tRNA recognition. **B)** Class II aaRS showing the catalytic domain with the three conserved motifs (I, II and III). In GlyRS, HisRS and ProRS the anticodon binding domain is at the C-terminal. In AspRS, AsnRS and LysRS it is at the N-terminal, while in AspRS, LysRS and AsnRS proteins the ABD occurs at the C-terminal. Dimer interfaces are shown by a magenta color and are characterized by motif I. ProRS has an editing domain that occurs between motif I and II at the catalytic site while in ThrRS the editing domain is located at the N-terminal. AlaRS has a C-Ala domain (gold) at the C-terminal that functions in dimer formation.

Figure 2 A) The apo structure of PfTyrRS (Class I). The catalytic domain (residues 22-260) is shown as cyan (cartoon) while the anticodon binding domain (residues 261-370) is shown in grey. The highly conserved KMSKS motif (red) and the HIGH motif (yellow) are shown in the structure. ATP and tyrosine binding sites are shown as blue dotted ellipses. Asp61, His70, Ala72, Gln73, Gln210, His235, Met237, Leu238, Met248, Lys250 are involved in ATP binding while residues Tyr60, Glu64, Ala96, Phe99, Ile172, Tyr188, Gln192 and Asp195 are involved in tyrosine binding [53]. B) **Apo structure of PfLysRS (Class II)**. The anticodon binding domain (residues 77-226) is shown in grey cartoon while the catalytic domain (residues 227-583) is shown in cyan. Motif I (red), motif II (yellow) and motif III (magenta) are shown by arrows. The ATP binding site and lysine binding site are shown by the blue dotted ellipses. Residues Arg330, His338, Asn339, Phe342, Glu500, Asn503,

Gly556 and Arg559 are implicated in binding of ATP while residues Glu308, Asn330, Glu346, Tyr348, Asn503, Tyr505, and Glu507 are involved in binding of lysine [54].

Protein translation has been explored as a target in the development of antimalarial drugs with most drugs interfering with the ribosome [55]. Recently, there has been increased interest in exploring *P. falciparum* aaRSs as potential drug targets [15,25,31,55–57]. 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, glyburide and TCMDC-124506 [58–61]. 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 [57,62,63]. 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 [58,64,65]. 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 [25,59]. Cladosporin, an adenosine analogue binds at the ATP binding site of PfLysRS [25,59]. 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 [61]. 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 [56]. 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 [60]. 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.

Due to these shortcomings of the current compounds that target aaRSs and the ever-increasing antimalarial drug resistance [6,15,16,19,27,66], 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 [66] and the sexual stages of the parasites thus terminating the infection cycle would help in malaria eradication [67]. 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 [68–70]. High conservation of aaRS across plasmodium and the human host may hinder development of parasite specific inhibitors [10,71,72]. Comparative studies between host and parasite sequences and structures are important in identifying differences that can be exploited for drug development [71,73–75]. 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 [76]. This study included *P. falciparum, P. bergei, P. malariae, P. ovale, P. yoelii, P. vivax, P. knowlesi, P. fragile*, human and other mammalian sequences. Sequences from toxoplasma and cryptosporidium species which belong to the apicomplexa family and other prokaryotes were also included for molecular phylogenetic tree calculations (Additional file 1). 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 [77]; thus, in this

241 study we focus mainly on the cytosolic aaRSs. The sequences were classified into two groups
242 based on differences in structure of their catalytic domain and further into the different aaRS
243 families based on their amino acid substrates [18,20,78]. The study was divided into two
244 parts. First, sequence-based analysis which involved motif search, multiple sequence
245 alignment and phylogenetic tree calculations was carried out. Secondly, structure-based
246 analysis was carried out which involved modeling of 3D structures of proteins, mapping of
247 identified motifs to these structures and identification of probable allosteric drug targeting
248 sites on the 3D models. The results showed striking differences in motifs and at residue level
249 between parasite and human proteins. The results from this study thus form a basis for further
250 research on aaRS as potential antimalarial drug targets and other parasitic diseases.

251

Methods

Sequence retrieval

Plasmodium falciparum aminoacyl tRNA synthetases (PfaaRS) were retrieved from NCBI-Protein database [79]. 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 [80]. 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. bergei*, *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 [81,82]. Crystal structures for human and *P. falciparum* ArgRS, TrpRS, MetRS, TyrRS, LysRS and ProRS proteins were retrieved from Protein Data Bank (PDB) [83].

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 [84]. 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 [85]. 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 *P. falciparum*, *H. sapiens*, *P. bergei*, *P. malariae*, *P. knowlesi*, *P. fragile*, *P. yoelii*, *P. ovale* and *P. vivax* proteins were built by homology modelling using MODELLER v9.15 [86]. Templates were identified using HHpred and PRotein Interactive MOdeling (PRIMO) webservers [87,88] 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 [89]; for MetRS - 4DLP [90]; for TrpRS - 4J75 [91]; for TyrRS - 5USF [92]; for ProRS - 4NCX [93] and for LysRS - 4DPG [94] was used. 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 [95], Verify3D [96] and Qualitative Model Energy Analysis (QMEAN) [97] 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 [98,99]. Visualization and editing of the alignments were done using the Jalview vs. 2.10 software [100]. 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

299 calculations via a Python script. The sequence identity results were translated into heatmaps
300 using a Matlab script.

301 **Molecular phylogenetic analysis**

302 Phylogenetic tree calculations were carried out for each family of aaRSs to study
303 evolutionary relationships within the protein families using Molecular Evolutionary Genetic
304 Analysis (MEGA) vs7.0 tool [101]. For sequence alignment of each family, three gap
305 deletion options - 90%, 95% and 100% - were used to calculate the models, and the best three
306 models for each deletion option were selected based on the lowest Bayesian information
307 criterion (BIC) scores. Maximum Likelihood (ML) statistical method was used to infer
308 evolutionary relationship while calculating trees for the top three models for each gap
309 deletion option for each protein families [102]. Total of 180 (3x3x20) trees were calculated.
310 Nearest-Neighbor-Interchange search was performed for all the constructed trees. BioNJ and
311 Neighbor Join algorithms were used for a matrix of pairwise distances calculated using JTT
312 model to obtain the initial trees for the heuristic search and the topology with the highest log
313 likelihood selected [103]. A strong branch swap filter and 1000 bootstrap replicates were
314 used for each tree calculation. The trees were then compared to the bootstrap consensus trees
315 to ensure that branching patterns were accurate and the best model and gap deletion for each
316 case was, then, chosen.

317 **Prediction of alternate druggable sites**

318 Structure-based drug design and development requires understanding of the structure and
319 function of the binding sites of the target protein. Identification of new drug targeting sites
320 different from the validated active sites is key in development of new classes of drugs. In this
321 study, probable druggable sites of our protein models were determined using FTMap
322 webserver [104] and SiteMap [105,106]. 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 [104]. 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 [107]. Clusters of low energy conformations are calculated and ranking of the probes is done based on the average energy [104]. 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 [105,106]. 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 [105]. The predicted binding sites are then ranked based on a SiteScore calculated using a linear combination of these factors [105].

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 (Figure 3 and 4) and for each family (see Additional file 3 & 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 file 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 Figure 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 (Figure 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 [40]. Class I aaRS enzymes are known to have a Rossmann

366 fold catalytic domain which is characteristic of the highly conserved Motif 1 and 2 [108,109].
367 Motif 12, 20 and 65 were also highly conserved among sequences in this class. The other
368 motifs clustered based on the enzyme family but some were conserved across different
369 enzymes within the same class. Motif 3, 4, 5, 13 and 14, for example, was conserved in all
370 GluRS and GlnRS sequences (Figure 3). These shared motifs show that these two proteins
371 have a high sequence identity and may explain why plasmodium apicoplast GluRS
372 mischarges glutamine specific tRNA with glutamate. In this case, glutamate is then changed
373 to glutamine a reaction catalysed by glutamyl-tRNA amidotransferase enzyme [110,111].

374 Motif 1 consisting of the HIGH signature which is characteristic of the Rossman fold was
375 conserved in all Class I aaRS (Figure 3) [81]. This class also showed high conservation of a
376 Motif 2 containing the KMSKS conserved signature which has also been reported to be part
377 of the RF in this class (Figure 6 & Additional file 3 & 4). The HIGH motif is present in the
378 first half of the RF while the KMSKS motif is present in the second half of the RF domain
379 (Figure 5 & Additional file 4). Motif conservation of the Rossman fold reflects the functional
380 importance of this region. This fold is involved in ATP binding and has been reported to be
381 highly conserved in class I proteins [36]. Class I catalytic domain is characteristic of a five
382 strand parallel sheets flanked by α-helices with amino acid and ATP binding sites on opposite
383 sides of a pseudo-2-fold symmetry. The Rossmann fold, in all Class I proteins has a
384 connective polypeptide I (CPI) insert which is characterized by alpha and beta folds [40]. The
385 conserved Motifs 1 and 2 across the class are present in the catalytic domains [40]. Detailed
386 analysis of each protein family showed conserved motifs specific to each family (Additional
387 file 3). Further, some conserved motifs unique only in the plasmodium proteins were
388 observed (Figure 4 & Additional file 3 & 4).

389 On mapping the motifs to the multiple sequence alignments, differences at the residue level
390 were observed despite the high level of motif conservation thus these residues can be the

391 basis of drug discovery. Eukaryote specific motifs in ArgRS, MetRS, GluRS WHEP domain
392 and AspRS are important for the association of proteins to a multi- tRNA synthetase complex
393 in eukaryotes [112–114]. In human, nine aaRSs form a complex together with non-synthetase
394 p18, p38 and p43 accessory proteins [114–116]. Leucyl, isoleucyl, glutaminyl, lysyl,
395 methionyl, aspartyl, prolyl and glutamyl-tRNA synthetases form the multi-synthetase
396 complex together with the auxiliary proteins in human aaRS but this complex is not present
397 in plasmodium aaRSs [34].

398 These unique motifs may also play important roles other than the canonical catalytic roles
399 [117]. Human LeuRS and GluRS, for example, have been reported to trigger leucine
400 dependent cellular proliferation and glutamine dependent apoptosis by functioning as amino
401 acid binding sensors [118,119]. Highly conserved motifs specific to each aaRS group are as a
402 result of idiosyncratic insertions at the C-terminal or within or after the Rossmann fold of
403 each protein family in this class [21,40,114] (Additional file 3 & 4). Methionine, valine,
404 isoleucine and leucine aaRSs are all known to be specific to substrates that have aliphatic side
405 chains and Motifs 20, 24 and 65 that are highly conserved in these four proteins may have a
406 role in this specificity [40]. LeuRS, IleRS, MetRS, ArgRS, ValRS and CysRS have a
407 structurally conserved anticodon binding domain characterized by α-helices and this may
408 explain the conservation of Motifs 2, 20, 44 and 65 among these proteins (Figure 3) [40].
409 Plasmodium TrpRS has an N-terminal extension which is 227 amino acid residues long that
410 constitute a AlaX-like domain and a linker region that function in binding of tRNA and in
411 aminoacylation activity [120]. This extension is not present in the human TrpRS and thus
412 explains the unique motifs at the N- terminal of the plasmodium proteins. Plasmodium
413 sequences also have a lysine-enriched insertion at the C-terminal end of the KMSKS motif
414 which is 15 residues long in PfTrpRS which is absent in the human sequence [120]. The
415 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 [40].

Figure 3. Motifs identified in Class I aaRS presented as a heat map. The colours represent conservation of motifs of the identified 90 motifs in this class. Conservation increases from blue to red while the absence of motifs is shown by a white colour. Motifs not present in human aaRS are shown in a red asterisk.

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

	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

In Class II, there were three highly conserved motifs across the class (Figure 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 (Figure 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 (2016) [36]. 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 to determine conserved motifs within homologous sequences of each protein and the results presented as heat maps (Additional file 3).

Figure 4. **Motifs identified in Class II aaRS presented as a heat map.** Motifs not present in human aaRS are shown in a red asterisk. The colours represent conservation of the identified 90 motifs in this class. Conservation increases from blue to red while the absence of motifs is shown by a white colour.

Table 2. **Highly conserved motifs in Class II aaRS**. 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.

	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

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

448 aminoacylation. Motif 1 has been reported previously (as motif III) to be part of the active
449 site forming α-helices and β-strands [36,121]. Motif 2, (Figure 4) also found at the catalytic
450 site of proteins in this group forms β strands in pairs joined by a loop [35]. Motif I plays a
451 role in binding of ATP while Motif 2 couples ATP, tRNA and amino acid binding [35,122].
452 Another weakly conserved motif in the active site of these proteins forms an α-helix that is
453 linked to a β-strand with a proline residue at the end (Motif 19, Figure 4). This motif is
454 known to be crucial in formation of dimers in most proteins of this class [123].

455 Further, subclasses in this class have conserved motifs within each subclass (Figure 4). For
456 example, Ser, Thr, Gly, Pro and His aaRSs all belong to the Class IIa and have anticodon
457 binding domains that are specific to the subclass [48]. These proteins are specific to small and
458 hydrophobic amino acids and have motifs that are conserved among them as shown in the
459 heatmap (Figure 4). The anticodon binding domain comprises of three α-helices five and β-
460 stranded sheets and occurs in the C-terminus of this sub-class [48,124,125]. The anticodon
461 binding domain is absent in SerRS as this protein does not require an anticodon to
462 discriminate its cognate tRNA [47,48]. Subclass IIb which comprises of AsnRS, LysRS and
463 AspRS have a unique anticodon binding domain at the N-terminal and share conserved
464 motifs (Figure 4) [50,126,127]. This subclass of enzymes is specific to large polar and
465 charged amino acid substrates and are similar in structural organization. AspRS is capable of
466 catalysing aminoacylation of aspartate and asparagine and thus it can be classified as
467 discriminating and non-discriminating protein just like GluRS [128,129]. Non-discriminating
468 AspRS is only present in bacteria and archaea but not in eukaryotes [130]. Family specific
469 motifs, can be attributed to the diversity in accessory domains found at the N- and C-
470 terminal or within loops in the core domain [131].

471

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 [100]. 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].

493 **Figure 5**: Motifs discovered in TyrRS family mapped to the multiple sequence alignment
494 results. Motif numbering is based on MEME results. A purple colour shows motifs conserved
495 in all sequences while motifs only present in mammalian sequences are shown in blue. The
496 highly conserved HIGH and KMSKS motifs in Class II aaRSs are shown in a red and yellow
497 box respectively.

498

499 CysRS sequence alignment and motif mapping showed a highly conserved core domain and
500 weakly conserved N- and C-terminal domains. The highly conserved HIGH signature was
501 found in Motif 2 of this family occurring as HLGH in plasmodium and HMGH in the
502 mammalian sequences (Additional file 4B). Motif 8, 10, 12, 13, 18 and 19 were conserved
503 only in mammalian sequences while Motif 11 and 15 were only conserved in plasmodium
504 sequences analysed in this family (Additional file 4B). GlnRS alignment also showed low
505 conservation on both termini with inserts observed for the mammalian sequences at the N-
506 terminal (Additional file 4C). Only two plasmodium specific motifs were found at the core
507 domain, Motif 23 at the N-terminal end of the highly conserved HIGH signature (Motif 2)
508 and Motif 29. Motif 8, 9, 11 and 13 were found only in the mammalian species (Additional
509 file 4C). *P. falciparum* is reported to have Glutathione-S-transferase (GST)-like domains
510 though their function in the malarial parasite has not been reported [34]. These domains are
511 important in formation of multi-synthetase complex through protein-protein interactions in
512 eukaryotes [21,22,117]. GST-like domains have also been reported in MetRS though just like
513 in GlnRS, the function of these domains in plasmodium is not known unlike in eukaryotes
514 where they play a role in protein-protein interactions [19].

515 The GluRS family also showed low conservation at the N-terminal with Motif 16 present in
516 mammalian sequences at this terminal (Additional file 4D). The HIGH signature was found
517 in Motif 3 as HIGH in all sequences analysed except for PfGluRS where it was HVGH
518 (Additional file 4). *P. falciparum* GluRS sequence has a glutamine rich N- terminal from
519 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 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

545 mammalian sequences. Some unique motifs were only present in plasmodium MetRS but
546 were absent in mammalian sequences (Additional file 4G). The highly conserved HIGH
547 signature was observed in Motif 8 which was conserved in all sequences analysed while the
548 KMSKS signature was found in Motif 14 conserved in plasmodium and Motif 6 in
549 mammalian sequences (Additional file 4G). The catalytic domain of MetRS was weakly
550 conserved with only Motif 1, 2, 4, 8, and 15 being conserved in all sequences at this region.
551 The C-terminal showed mammalian and plasmodium specific motifs. Motif 5 and 9 found at
552 the N-terminal were conserved in all analysed sequences in this family (Additional file 4G).

553 TrpRS alignment revealed a plasmodium specific extension at the N-terminal characterised
554 by Motif 8, 9, 10 and 14 (Additional file 4H). This extension plays a role in aminoacylation
555 and tRNA binding as reported in *P. falciparum* [34]. In *P. falciparum*, this extension
556 comprises of a linker region and an AlaX-like domain that plays a role in tRNA binding but
557 does not edit mis-acylations as observed with *Pyrococcus horikoshii* [120]. The core domain
558 and the C-terminal of TrpRS family showed highly conserved motifs in all the sequences
559 with only a short Motif 18 present in mammalian sequences (Additional file 4H). Alignment
560 and mapping motifs discovered in TyrRS sequences showed high conservation of motifs at
561 the core domain (Figure 5). Alignment of sequences in this family showed an extension at the
562 C-terminal of the mammalian TyrRS which was missing in all plasmodium sequences (Figure
563 5). This extension was characterised by Motifs 6, 8, 9, 11 and 20 which were conserved in all
564 the mammalian sequences analysed (Figure 5). This extension in human TyrRS is an
565 endothelial monocyte-activating polypeptide II (EMAPII) domain that has cytokine-like
566 functions like angiogenesis and inflammation [22,140]. Motif discovery showed that the core
567 domain is highly conserved across the mammalian and plasmodium TyrRS sequences (Figure
568 5). The catalytic domain of the human sequence is also different from the malarial parasites
569 in that it has a buried tripeptide cytokine motif (Glu-Leu-Arg) while in plasmodium this motif

570 is on the surface [22,53]. ValRS alignment showed a N-terminal extension for the
571 mammalian sequences that was absent in all plasmodium sequences comprising of Motifs 14,
572 16, 18, 22 and 25 (Additional file 4J). Mapping of motifs showed that the catalytic domain of
573 proteins analysed in this family are highly conserved though a few plasmodium specific
574 motifs were observed. The highly conserved HIGH signature was found in Motif 2 of this
575 family while the KMSKS signature was in Motif 7 (Additional file 4J). The N-terminal
576 domain showed Motif 16, 33, 34 and 35 which were conserved only in plasmodium
577 sequences (Additional file 4J). Motifs 20, 30 and 38 that were specific to mammalian
578 sequences were also observed at the N-terminal (Additional file 4J).

579 Alignment of AlaRS sequences showed a N-terminal extension of varying lengths in the
580 plasmodium species which was absent in mammalian AlaRS (Additional file 4K). The C-
581 terminal of the proteins in this family showed Motifs 20, 21 and 29 that were only conserved
582 in plasmodium sequences and not in human as well as mammalian specific motifs (Motif 8,
583 14, 17 and 18). AsnRS, LysRS, and AspRS alignment and motif discovery showed low
584 conservation at the N-terminal while core domains and the C-terminal showed high
585 conservation. The anticodon binding domain of these proteins is located at the highly variable
586 N-terminal and thus drugs that specifically bind to the parasite tRNA binding site can be
587 designed [14,141]. Motif 11, 12 and 17 were conserved in plasmodium sequences of AsnRS
588 family at the N-terminal while in this region, Motif 5, 6 and 13 conserved in mammalian
589 sequences were observed (Additional file 4L). In AspRS both the catalytic domain and the C-
590 terminal were highly conserved with the presence of two short Motifs (16 and 20) conserved
591 only in mammals (Additional file 4M). GlyRS, HisRS, ProRS, ThrRS families belong to the
592 subclass IIa and have a highly conserved tRNA binding region at the C-terminal as seen in
593 the alignments and motifs in this region (Additional file 4 N, O, R and T). HisRS family
594 showed a N-terminal extension for all plasmodium sequences analysed but absent in the

595 mammalian sequences (Additional file 4O). This extension was characterised by Motifs 11,
596 12, 14, 15, 17, 18, 19 and 23 (Additional file 4O). However, SerRS which also belongs to this
597 subclass does not need an anticodon to discriminate its substrate and thus lacks this domain
598 [48] and the C-terminal of this family showed low conservation (Additional file 4S). ProRS
599 showed Motif 17 and 20 which were conserved only in plasmodium sequences analysed
600 (Additional file 4R). Plasmodium ProRS has a Ybak domain at the N-terminal which edits
601 mischarged Pro-tRNA^{Ala} and Pro-tRNA^{Ser} and this may explain the plasmodium specific
602 motifs at the N-terminal [15,34,57]. The mammalian sequences analysed for this family were
603 of the cytosolic bifunctional Glu/ProRS proteins and this explains the mammalian specific
604 motifs observed at the N-terminal which is believed to be the region responsible for
605 glutamate aminoacylation (Additional file 4R).

606 PheRS motif discovery and alignment showed that the plasmodium sequences are highly
607 variable when compared to mammalian PheRS (Additional file 4Q). Motifs 9, 10, 11 and 13
608 were conserved only in plasmodium while Motifs 5, 6, 8 and 14 were conserved in
609 mammalian sequences in this family (Additional file 4Q). Only Motifs 1, 2, 3 and 4 were
610 conserved across all the sequences in this family (Additional file 4Q). Plasmodium PheRS
611 has a nuclear localization signal and DNA binding domains and thus in addition to
612 aminoacylation, this enzyme mediates cellular processes by binding DNA [142]. Despite high
613 conservation at the aaRS active sites, differences were noted at the residue level after the
614 sequences were aligned. For example, in LysRS family, *P. falciparum* ATP binding pocket at
615 positions Val328 and Ser344 corresponds to Gln321 and Thr338 respectively in the human
616 protein (Figure 6). Residues with a large side chain at this position like observed in human
617 LysRS do not favour binding of cladosporin a known inhibitor for PfLysRS [25]. These two
618 residues are thus believed to be responsible for selective binding of cladosporin to *P.*
619 *falciparum* and not human LysRS [25]. 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 [14]. 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.

Figure 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 results. 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.

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 (Figure 7, Figure 8 and Additional file 6). 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. bergei* 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 we will name them as “conservation boxes”. *P. bergei* and *P. yoelii* are rodent malaria parasites and are used to study human malaria [143,144]. *P. 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 [145,146]. *P. knowlesi* has been reported to have a close phylogenetic relationship to *P. vivax* [147] and the two showed a sequence identity above 95% in TyrRS (Figure 8). *P. fragile*-monkey models can thus be used to study parasite-host-system for the immunological response of the falciparum-like parasite both *in vivo* and *in vitro* [148].

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 & 6.1). 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 & 6.11). 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 & 6.2). Motif 27 was present only in PyCysRS and PbCysRS and these two sequences showed a 90% sequence identity Additional file 3 & 6.2). PfrGlnRS, PkGlnRS and PvGlnRS clustered together and Motif 34, 35 and 37 were only present in these sequences (Additional file 3 & 6.3). In this family, *E. coli*, human and *S. cerevisiae* sequences formed an outgroup showing they are the oldest aaRS (Additional file 6.3). GluRS plasmodium sequences had above 75% sequence identity and shared all identified motifs (Additional file 3 & 6.4). PbIleRS and PyIleRS shared Motif 38 and 39 and showed 95% sequence identity (Additional file 3 & 6.5). 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 6.5 & 6.7). PvLeuRS, PfrLeuRS and PkLeuRS had 80% sequence identity and shared Motif 39 (Additional file 3 & 6.6). In TrpRS, Motif 21 and 23 were only identified in PbTrpRS and

673 PyTrpRS which had 90% sequence identity. In all families in Class I, human sequences
674 showed low sequence identities (below 40%) compared to the plasmodium sequences
675 (Additional file 6).

676
677 **Figure 7. A)** TyrRS family phylogenetic tree. Maximum Likelihood method was used to
678 infer evolutionary history using Le_Gascuel_2008 model at 95% site coverage [149].
679 Phylogenetic tree calculations were done using MEGA7 [101]. The tree that had the highest
680 log likelihood (-2978.09) is shown. Initial tree(s) for the heuristic search were obtained by
681 using BioNJ and Neighbor-Join algorithms to a matrix of pairwise distances calculated using
682 a JTT model, and then selecting the topology with higher log likelihood value. A Gamma
683 distribution was used to calculate evolutionary rate differences among sites (5 categories (+*G*,
684 parameter = 0.4355)). Nine amino acid sequences were used for this analysis. There were 343
685 positions after the calculations. **B)** TyrRS pairwise sequence calculations. The sequence
686 identity values of the sequences in the TyrRS family is shown. The heatmap shows the
687 identity scores as a color-coded matrix for every aaRS sequence versus every aaRS sequence
688 in this family. Conservation increases from blue to red in the heat map.
689

690

691 In Class II aaRSs, ProRS family showed the highest sequence identity with plasmodium
692 sequences having above 80% sequence identity (Additional file 6). The high sequence
693 identity among plasmodium sequences was also reflected in motif identification where all the
694 sequences shared the identified motifs (Additional file 3). In Class IIa GlyRS showed the
695 least conservation with most of the sequences having less than 65% sequence identity
696 (Additional file 6.16). GlyRS family showed low conservation with sequence identity less
697 than 70% for all sequences except for PfrGlyRS, PvGlyRS and PkGlyRS which formed a
698 conservation box with a sequence identity of about 75% (Additional file 6.16). This
699 clustering was also seen in motif identification whereby PfrGlyRS, PvGlyRS and PkGlyRS
700 had Motif 24 which was absent in all other plasmodium sequences in this family. PbGlyRS
701 and PyGlyRS had a sequence identity of 90% and shared Motif 27, 30 and 34 (Additional file
702 3 & 6.16). *P. falciparum* ThrRS had a low sequence identity compared to other plasmodium
703 sequences and it also branched separately in the phylogenetic tree. In SerRS family, human,
704 *T. brucei*, *C. albicans*, *T. gondii* and *C. 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 6.20). 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 & 6.20). 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. bergei* with a sequence identity of 90% (Additional file 6.19). 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 6.14). 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 6.13 & 6.15). 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, *C. ubiquitum* showed a higher sequence identity to *S. typhi* sequence than to *T. 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 & 6.15). 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 (Figure 8). PfrLysRS, PkLysRS and PvLysRS also formed a

conservation box and these three proteins shared Motif 15 which was absent in other plasmodium sequences (Figure 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 & 6). 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 6). Plasmodium sequences in AlaRS family showed a sequence identity above 70% (Additional file 6.12). In this family, *P. vivax*, *P. fragile* and *P. knowlesi* also formed a conservation box while *P. yoelii* and *P. bergei* 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 6).

Figure 8. A) LysRS family phylogenetic tree. Maximum Likelihood method was used to infer evolutionary history using Le_Gascuel_2008 model at 90% site coverage [149]. Phylogenetic tree calculations were done using MEGA7 [101]. The tree that had the highest log likelihood (-6116.25) is shown. Initial tree(s) for the heuristic search were obtained by using BioNJ and Neighbor-Join algorithms to a matrix of pairwise distances calculated using a JTT model, and then selecting the topology with higher log likelihood value. A Gamma distribution was used to calculate evolutionary rate differences among sites (5 categories (+*G*, parameter = 0.6075)). Eleven amino acid sequences were used for this analysis. There were 503 positions after the calculations. **B)** LysRS pairwise sequence calculations. The sequence identity values of the sequences in the LysRS family is shown. The heatmap shows the identity scores as a color-coded matrix for every aaRS sequence versus every aaRS sequence in this family. Conservation increases from blue to red in the heat map.

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 [150]. 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 [151]. 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 [151–153]. 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 [151,154]. Models with lowest z-DOPE were selected and model quality assessment was done using Verify 3D [96], ProSA [95] and QMEAN [97] webservers. 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 [96]. 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 [95].

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 [97]. Descriptors of solvent accessibility and the agreement between calculated and predicted structures are also used in calculating the score [97]. 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 [89] 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 4DLP [90]. Plasmodium TrpRS models were built with 4J75 [155] 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 [92] 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 [60] 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 (Figure 9, Figure 10 and Additional file 5). 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 in HsArgRS (Additional file 5A, Figure 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 5A). 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 (Figure 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 5C). 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 5C). 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 (Figure 9 & Additional file 5D).

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 5E). Mapped motif in ProRS were conserved in all analysed secondary structures (Figure 10 & Additional file 5F).

Figure 9: Class I motifs mapped to the homology models of *P. falciparum* and the human homologue. Motifs are numbered according to the MEME results.

Figure 10: Class II motifs mapped to the homology models of *P. falciparum* and the human homologues. Motifs are numbered according to the MEME results.

840 **Table 3**: Starting and ending positions of motifs identified in PfArgRS, PfMetRS, PfTrpRS, PfTyrRS, PfLysRS and PfProRS as well as the
841 human homologues. Dashes show where the motif was not present.

	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	----	439-459	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

842 **New potential druggable sites in *P. falciparum* aaRSs are identified**

843 FTMap provides information on binding hot spots and the druggability of these sites using
844 probes from fragment libraries [107]. These fragment hits can be used in identification of hits
845 from larger ligands. On the other hand, SiteMap predicts possible binding sites using an
846 algorithm that assigns site points using geometric and energetic properties [105,106]. The site
847 points are then grouped to give sites which are ranked based on a SiteScore computed based
848 on size, hydrophobicity, exposure to the solvent and the ease of donating or accepting
849 hydrogens. Both FTMap and SiteMap showed consistency in prediction of probable binding
850 sites. In all the six modelled proteins, FTMap and SiteMap were able to predict the known
851 active sites which consists of the ATP and amino acid binding sites as the highest ranked site
852 (Figure 11 & 2). Alternative sites were also predicted in PfArgRS, PfMetRS, PfProRS, and
853 HsProRS that can be targeted for design of new drug classes using both FTMap and SiteMap
854 (Figure 11, 12 & 13). Since two tools show consistency in prediction of possible binding
855 sites, we only discuss the results from FTMap in this study.

856 The identified potential druggable site in PfArgRS is in a region located at the anticodon
857 binding domain characterized by Motif 4 and 6 but the site is not present in HsArgRS (Figure
858 9 & 11, Table 3). Probes at this site interact with residues in the ABD – His515, Lys518,
859 Ile522, Lys534, Glu537, Asp541 and Tyr34 located in the N-terminal domain. Motif results
860 showed low conservation of these residues with His515 corresponding to Cys577, Lys518 to
861 Arg580, Ile522 to Ile584, Lys534 to Thr592, Glu537 to Asp595, Asp541 to Glu599 and
862 Tyr34 to Ser105 in the human homologue (Figure 11D). These residues are, however, highly
863 conserved in the other plasmodium sequences studied (Additional file 4A). This region can
864 thus be potentially targeted for inhibitor design with high selectivity to the plasmodium
865 protein as indicated by the low conservation in the human homologue.

Figure 11: Homology models of PfArgRS and HsArgRS and prediction of potential ligand binding sites. The catalytic domain of the models is shown in cyan, the anticodon binding domain (ABD) in grey and the N-terminal domains of PfArgRS and HsArgRS are shown in a light orange colour. The HIGH and KMSKS motifs highly conserved in Class I are shown in red and yellow respectively. Known druggable sites are shown by the red dotted ellipses while the predicted site in PfArgRS by FTMap is shown in purple dotted ellipses. **A)** PfArgRS homology model. **B)** Insert – zoomed view of the predicted druggable site in PfArgRS with the residues interacting with probes represented as magenta sticks. **C)** HsArgRS homology model. No probable druggable sites were predicted in human homologue. **D)** Motif 4 and 6 logos showing conservation of residues in this family and PfArgRS residues interacting with probes at the predicted site.

The predicted hotspot in PfMetRS is in a pocket formed by Motifs 5, 9, 14, 20 and the loop region of Motif 4 (Figure 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 (Figure 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 of that bind selectively to PfMetRS. A study by Hussain *et al* [19] 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 [19]. 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].

Figure 12: Homology models of PfMetRS and HsMetRS and prediction of potential ligand binding sites. The catalytic domain of the models is shown in cyan and the anticodon binding domain (ABD) in grey. The HIGH and KMSKS motifs are shown in red and yellow respectively. **A)** PfMetRS homology model. Known druggable sites are shown by the red dotted ellipses while the predicted site in PfMetRS by FTMap is shown in purple dotted ellipses. **B)** Insert – zoomed view of the predicted druggable pocket in PfMetRS showing stick representation (magenta) of residues interacting with probes at this site. The predicted site is located at the catalytic domain **C)** HsMetRS homology model. HsMetRS had no probable druggable sites predicted by FTMap. **D)** Motif 5, 9 and 20 logos showing conservation of residues in this family and PfMetRS residues interacting with probes predicted site.

The identified potentially druggable site in PfProRS occurs at a region characterised by Motif 1, 5 and 11 which are also present in HsProRS (Figure 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 (Figure 13D & Additional file 4R). A previous study by Hewitt *et al* [60] reported selective binding of glyburide and TCMDC-124506 at the PfProRS predicted site. This pocket is located at a region formed by α5 (residues 513-524), α9 (residues 261-272) and β-hairpin 1 and 2 (residues 276-287). We observed interaction of Phe262 to the FTMap probes and Tyr746 (Figure 13) which is also reported to interact with glyburide and TCMDC-124506 [60]. Inhibition of PfProRS by the two compounds is known to be through distortion of the ATP binding site [60]. 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 [60]. The unique predicted sites in PfArgRS, PfMetRS and PfProRS can thus be targeted through high throughput screening to identify new inhibitors.

Figure 13: Build homology models of ProRS and prediction of potential ligand binding sites. The catalytic domain is shown in cyan, anticodon binding domain in grey and the C-

terminal zinc-binding like domain is shown in light pink colour. Motif 2 located at the catalytic domain is shown in yellow. Known druggable sites are shown by the red dotted ellipses while other predicted sites by FTMap are shown in purple dotted ellipses. A) The homology model of PfProRS. B) Insert – zoomed view of the predicted site in PfProRS showing residues interacting with probes at this site as magenta sticks. These probes were interacting with residues – Tyr746, Thr397, Phe262, Arg401 and Lys394. C) The homology model of HsLysRS showing a probable druggable site with residues Thr1164, Phe1161, Thr1277, Leu1162, Thr1276 and Arg1278 interacting with probes. **D)** Motif 1 and 11 logos showing conservation of residues in these motifs and PfProRS residues interacting with probes at the predicted site.

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. bergei* were seen to cluster in trees in most of the aaRS families showing that these proteins

954 are closely related, and this was also depicted by the high sequence identity and shared motifs
955 among them. *P. fragile*, *P. knowlesi* and *P. vivax* aaRSs were also seen to share evolutionary
956 history and had high sequence identity. Prediction of additional druggable sites identified hot
957 spots in PfArgRS, PfMetRS and PfProRS. The identified sites showed low conservation and
958 variation of identified motifs between *P. falciparum* proteins and the human homologues.
959 The identified sites can thus be targeted to develop drugs that only selectively bind to
960 plasmodium proteins. As per the results in this study, it is evident that despite structural
961 conservation, plasmodium aaRS have key features that differentiate them from human
962 proteins. These differences can be targeted to develop antimalarial drugs with less toxicity to
963 the host.

964

965 **Author contributions**

966 Ö.T.B designed the study. D.W.N acquired the data, performed data analysis and wrote the
967 initial draft. All authors contributed in interpretation and discussion of results and writing of
968 the manuscript.

969 **Acknowledgements**

970 This work is supported by the National Research Foundation (NRF) South Africa (Grant
971 Number 105267). The content of this publication is solely the responsibility of the authors
972 and does not necessarily represent the official views of the funders. D.W.N thanks Rhodes-
973 Henderson Bursary and National Research Foundation South Africa for financial support.
974 D.W.N thanks Vuyani Moses for his guidance in performing the calculations and data
975 analysis.

976

977

978

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

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.

Additional file 6

Phylogenetic trees and pairwise sequence calculations for aaRS families: Molecular Phylogenetic calculations were performed using MEGA7. Sequence identity calculations

1004 were done using an in-house python script and results displayed as heatmaps. Conservation

1005 increases from blue to red.

1006

References

1. 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.

2. Fevre EM, Wissmann B V, Welburn SC, Lutumba P. The burden of human African Trypanosomiasis. PLoS Negl Trop Dis [Internet]. 2008;2. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed8&AN=2009223996

3. Gething PW, Patil AP, Smith DL, Guerra C a., Elyazar IRF, Johnston GL, et al. A new world malaria map: Plasmodium falciparum endemicity in 2010. Malar J [Internet]. 2011;10:378. Available from: http://eprints.soton.ac.uk/344413/

4. 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.

5. Prichard RK, Basanez MG, Boatin BA, McCarthy JS, Garcia HH, Yang GJ, et al. PLOS NEGLECTED TROPICAL DISEASES. A Res. Agenda Helminth Dis. Humans Interv. Control Elimin. 2012. p. 14.

6. Baird JK, J.K. B. Drug therapy: effectiveness of antimalarial drugs. N Engl J Med [Internet]. 2005;352:1565. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed7&NEWS=N&AN=2005172567%5Cnhttp://content.nejm.org/

7. Nayyar GML, Breman JG, Newton PN, Herrington J. Poor-quality antimalarial drugs in southeast Asia and sub-Saharan Africa. Lancet Infect. Dis. 2012. p. 488–96.

8. Rajendran V, Kalita P, Shukla H, Kumar A, Tripathi T. Aminoacyl-tRNA synthetases: Structure, function, and drug discovery. Int J Biol Macromol [Internet]. Elsevier B.V.; 2018;111:400–14. Available from: https://doi.org/10.1016/j.ijbiomac.2017.12.157

9. Fang P, Guo M. Evolutionary Limitation and Opportunities for Developing tRNA Synthetase Inhibitors with 5-Binding-Mode Classification. Life [Internet]. 2015;5:1703–25. Available from: http://www.mdpi.com/2075-1729/5/4/1703

10. 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 [Internet]. Elsevier Current Trends; 2018 [cited 2018 Jun 12]; Available from: https://www.sciencedirect.com/science/article/pii/S135964461730538X

11. World Health Organization W, WHO, World Health Organization W. World Malaria Report 2015. World Health. 2015;1–280.

12. Antinori S, Galimberti L, Milazzo L, Corbellino M. Biology of human malaria plasmodia including Plasmodium knowlesi. Mediterr J Hematol Infect Dis. 2012;4.

13. Jackson KE, Habib S, Frugier M, Hoen R, Khan S, Pham JS, et al. Protein translation in Plasmodium parasites. Trends Parasitol. 2011. p. 467–76.

14. 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. p. 1–13.

15. 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 [Internet]. 2011;1:188. Available from: http://www.nature.com/articles/srep00188

16. Pino P, Aeby E, Foth BJ, Sheiner L, Soldati T, Schneider A, et al. Mitochondrial translation in absence of local tRNA aminoacylation and methionyl tRNAMet formylation in Apicomplexa. Mol Microbiol. 2010;76:706–18.

17. 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 [Internet]. Australian Society for Parasitology Inc.; 2012;42:177–86. Available from: http://dx.doi.org/10.1016/j.ijpara.2011.11.008

18. Ibba M, Söll D. Aminoacyl-tRNA Synthesis. Annu Rev Biochem [Internet]. 2000;69:617–50. Available from: http://www.annualreviews.org/doi/10.1146/annurev.biochem.69.1.617

19. 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.

20. Park SG, Schimmel P, Kim S. Aminoacyl tRNA synthetases and their connections to disease. Proc Natl Acad Sci U S A [Internet]. 2008;105:11043–9. Available from: http://europepmc.org/articles/PMC2516211/?report=abstract

21. Guo M, Schimmel P, Yang XL. Functional expansion of human tRNA synthetases achieved by structural inventions. FEBS Lett. 2010. p. 434–42.

22. Paul M, Schimmel P. Essential nontranslational functions of tRNA synthetases. Nat Chem Biol [Internet]. 2013;9:145–53. Available from: http://www.nature.com/doifinder/10.1038/nchembio.1158

23. Bhatt TK, Khan S, Dwivedi VP, Banday MM, Sharma A, Chandele A, et al. Malaria parasite tyrosyl-tRNA synthetase secretion triggers pro-inflammatory responses. Nat

1076 Commun [Internet]. 2011;2:530. Available from:
1077 http://www.nature.com/doifinder/10.1038/ncomms1522

1078 24. Wolf YI, Aravind L, Grishin N V., Koonin E V. Evolution of Aminoacyl-tRNA
1079 synthetases-analysis of unique domain architectures and phylogenetic trees reveals a complex
1080 history of horizontal gene transfer events. Genome Res. 1999;9:689–710.

1081 25. Sharma A, Khan S, Sharma A, Belrhali H, Yogavel M. Structural basis of malaria
1082 parasite lysyl-tRNA synthetase inhibition by cladosporin. J Struct Funct Genomics.
1083 2014;15:63–71.

1084 26. Sherma A, Yogavel M, Sharma A. Structural and functional attributes of malaria parasite
1085 diadenosine tetraphosphate hydrolase. Sci Rep [Internet]. Nature Publishing Group;
1086 2016;6:1–12. Available from: http://dx.doi.org/10.1038/srep19981

1087 27. Miller LH, Baruch DI, Marsh K, Doumbo OK. The pathogenic basis of malaria. Nature.
1088 2002;415:673–9.

1089 28. Sharma A, Sharma A. *Plasmodium falciparum* mitochondria import tRNAs along with an
1090 active phenylalanyl-tRNA synthetase. Biochem J [Internet]. 2015;465:459–69. Available
1091 from: http://www.biochemj.org/bj/465/bj4650459.htm

1092 29. Jackson KE, Pham JS, Kwek M, De Silva NS, Allen SM, Goodman CD, et al. Dual
1093 targeting of aminoacyl-tRNA synthetases to the apicoplast and cytosol in Plasmodium
1094 falciparum. Int J Parasitol [Internet]. Pergamon; 2012 [cited 2018 Jun 12];42:177–86.
1095 Available from:
1096 https://www.sciencedirect.com/science/article/pii/S0020751911002992?via%3Dihub

1097 30. Pham JS, Sakaguchi R, Yeoh LM, De Silva NS, McFadden GI, Hou Y-M, et al. A dual-
1098 targeted aminoacyl-tRNA synthetase in Plasmodium falciparum charges cytosolic and

1099 apicoplast tRNACys. Biochem J [Internet]. 2014;458:513–23. Available from:
1100 http://www.ncbi.nlm.nih.gov/pubmed/24428730

1101 31. Manickam Y, Chaturvedi R, Babbar P, Malhotra N, Jain V, Sharma A. Drug targeting of
1102 one or more aminoacyl-tRNA synthetase in the malaria parasite Plasmodium falciparum.
1103 Drug Discov Today [Internet]. Elsevier Ltd; 2018;0:26–33. Available from:
1104 http://dx.doi.org/10.1016/j.drudis.2018.01.050

1105 32. Bonnefond L, Fender A, Rudinger-Thirion J, Giegé R, Florentz C, Sissler M. Toward the
1106 full set of human mitochondrial aminoacyl-tRNA synthetases: Characterization of AspRS and
1107 TyrRS. Biochemistry. 2005;44:4805–16.

1108 33. Antonellis A, Green ED. The Role of Aminoacyl-tRNA Synthetases in Genetic Diseases.
1109 Annu Rev Genomics Hum Genet [Internet]. 2008;9:87–107. Available from:
1110 http://www.annualreviews.org/doi/10.1146/annurev.genom.9.081307.164204

1111 34. Bhatt TK, Kapil C, Khan S, Jairajpuri MA, Sharma V, Santoni D, et al. A genomic
1112 glimpse of aminoacyl-tRNA synthetases in malaria parasite Plasmodium falciparum. BMC
1113 Genomics. 2009;10:644.

1114 35. Eriani G, Cavarelli J, Martin F, Ador L, Rees B, Thierry JC, et al. The class II aminoacyl-
1115 tRNA synthetases and their active site: Evolutionary conservation of an ATP binding site. J
1116 Mol Evol. 1995;40:499–508.

1117 36. Chaliotis A, Vlastaridis P, Mossialos D, Ibba M, Becker HD, Stathopoulos C, et al. The
1118 complex evolutionary history of aminoacyl-tRNA synthetases. Nucleic Acids Res [Internet].
1119 2016;45:gkw1182. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27903912

1120 37. Chen JF, Guo NN, Li T, Wang ED, Wang YL. CP1 domain in Escherichia coli leucyl-
1121 tRNA synthetase is crucial for its editing function. Biochemistry. 2000;39:6726–31.

1122 38. Doublié S, Bricogne G, Gilmore C, Carter, CW. Tryptophanyl-tRNA synthetase crystal
1123 structure reveals an unexpected homology to tyrosyl-tRNA synthetase. Structure. 1995;3:17–
1124 31.

1125 39. Yadavalli SS, Ibba M. Quality control in aminoacyl-tRNA synthesis: Its role in
1126 translational fidelity. Adv. Protein Chem. Struct. Biol. 2012. p. 1–43.

1127 40. Perona JJ, Hadd A. Structural diversity and protein engineering of the aminoacyl-tRNA
1128 Synthetases. Biochemistry. 2012;51:8705–29.

1129 41. Cusack S. Aminoacyl-tRNA synthetases Stephen Cusack. Curr Opin Struct Biol
1130 [Internet]. 1997;7:881–9. Available from: http://biomednet.com/elecref/O959440X00700881

1131 42. Perona JJ, Gruic-Sovulj I. Synthetic and editing mechanisms of aminoacyl-tRNA
1132 synthetases. Top. Curr. Chem. 2014. p. 1–41.

1133 43. Ibba M, Losey HC, Kawarabayasi Y, Kikuchi H, Bunjun S, Söll D. Substrate recognition
1134 by class I lysyl-tRNA synthetases: a molecular basis for gene displacement. Proc Natl Acad
1135 Sci U S A. 1999;96:418–23.

1136 44. Bennett EJ, Shaler T a, Woodman B, Ryu K-Y, Zaitseva TS, Becker CH, et al. Anticodon
1137 recognition and discrimination by the alpha-helix cage domain of class I lysyl-tRNA
1138 synthetase. J Biol Chem [Internet]. 2007;282:11033–8. Available from:
1139 http://www.ncbi.nlm.nih.gov/pubmed/17540772%5Cnhttp://www.pubmedcentral.nih.gov/arti
1140 clerender.fcgi?artid=2816034&tool=pmcentrez&rendertype=abstract%5Cnhttp://www.ncbi.n
1141 lm.nih.gov/pubmed/17567576%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/17581591%5Cnhtt
1142 p://www

1143 45. Pouplana LR, Buechter DD, Davis MW, Schimmel P. Idiographic representation of
1144 conserved domain of a class II tRNA synthetase of unknown structure. Protein Sci.

1145 1993;2:2259–62.

1146 46. Smith TF, Hartman H. The evolution of Class II Aminoacyl-tRNA synthetases and the
1147 first code. FEBS Lett [Internet]. Federation of European Biochemical Societies;
1148 2015;589:3499–507. Available from: http://dx.doi.org/10.1016/j.febslet.2015.10.006

1149 47. Normanly J, Ollick T, Abelson J. Eight base changes are sufficient to convert a leucine-
1150 inserting tRNA into a serine-inserting tRNA. Proc Natl Acad Sci U S A [Internet].
1151 1992;89:5680–4. Available from:
1152 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=49356&tool=pmcentrez&rendert
1153 ype=abstract

1154 48. Yao P, Fox PL. Aminoacyl-tRNA synthetases in medicine and disease. EMBO Mol. Med.
1155 2013. p. 332–43.

1156 49. Martinez-Rodriguez L, Erdogan O, Jimenez-Rodriguez M, Gonzalez-Rivera K, Williams
1157 T, Li L, et al. Functional class I and II amino acid-activating enzymes can be coded by
1158 opposite strands of the same gene. J Biol Chem. 2015;290:19710–25.

1159 50. Ruff M, Krishnaswamy S, Boeglin M, Poterszman a, Mitschler a, Podjarny a, et al.
1160 Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA
1161 synthetase complexed with tRNA(Asp). Science. 1991;252:1682–9.

1162 51. Bhatt T, Kapil C, Khan S, Jairajpuri M, Sharma V, Santoni D, et al. A genomic glimpse
1163 of aminoacyl-tRNA synthetases in malaria parasite Plasmodium falciparum. BMC Genomics
1164 [Internet]. 2009;10:644. Available from:
1165 http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-10-644

1166 52. Ibba M, Bono JL, Rosa P a, Söll D. Archaeal-type lysyl-tRNA synthetase in the Lyme
1167 disease spirochete Borrelia burgdorferi. Proc Natl Acad Sci U S A. 1997;94:14383–8.

1168 53. Bhatt TK, Khan S, Dwivedi VP, Banday MM, Sharma A, Chandele A, et al. Malaria
1169 parasite tyrosyl-tRNA synthetase secretion triggers pro-inflammatory responses. Nat
1170 Commun [Internet]. 2011;2:530. Available from:
1171 http://www.ncbi.nlm.nih.gov/pubmed/22068597

1172 54. Khan S, Garg A, Camacho N, Van Rooyen J, Kumar Pole A, Belrhali H, et al. Structural
1173 analysis of malaria-parasite lysyl-tRNA synthetase provides a platform for drug development.
1174 Acta Crystallogr Sect D Biol Crystallogr [Internet]. 2013 [cited 2017 May 22];69:785–95.
1175 Available from: http://scripts.iucr.org/cgi-bin/paper?S0907444913001923

1176 55. Saint-Léger A, Sinadinos C, Ribas de Pouplana L. The growing pipeline of natural
1177 aminoacyl-tRNA synthetase inhibitors for malaria treatment. Bioengineered. 2016;7:60–4.

1178 56. Ebere, Palencia A, Guo D, Ahyong V, Dong C, Li X, et al. Antimalarial Benzoxaboroles
1179 Target. 2016;60:4886–95.

1180 57. Jain V, Kikuchi H, Oshima Y, Sharma A, Yogavel M. Structural and functional analysis
1181 of the anti-malarial drug target prolyl-tRNA synthetase. J Struct Funct Genomics.
1182 2014;15:181–90.

1183 58. Keller TL, Zocco D, Sundrud MS, Hendrick M, Edenius M, Yum J, et al. Halofuginone
1184 and other febrifugine derivatives inhibit prolyl- tRNA synthetase. 2012;8:311–7.

1185 59. Hoepfner D, McNamara CW, Lim CS, Studer C, Riedl R, Aust T, et al. Selective and
1186 specific inhibition of the plasmodium falciparum lysyl-tRNA synthetase by the fungal
1187 secondary metabolite cladosporin. Cell Host Microbe [Internet]. Elsevier Inc.; 2012;11:654–
1188 63. Available from: http://dx.doi.org/10.1016/j.chom.2012.04.015

1189 60. Hewitt SN, Dranow DM, Horst BG, Abendroth JA, Forte B, Hallyburton I, et al.
1190 Biochemical and Structural Characterization of Selective Allosteric Inhibitors of the

Plasmodium falciparum Drug Target, Prolyl-tRNA-synthetase. ACS Infect Dis [Internet]. 2017 [cited 2017 May 22];3:34–44. Available from: http://pubs.acs.org/doi/abs/10.1021/acsinfecdis.6b00078

61. 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 [Internet]. American Society for Microbiology; 2016 [cited 2018 Jun 12];60:4886–95. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27270277

62. 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 [Internet]. Elsevier Ltd; 2015;23:819–29. Available from: http://dx.doi.org/10.1016/j.str.2015.02.011

63. 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 [Internet]. Elsevier Ltd.; 2017;25:1495–1505.e6. Available from: http://dx.doi.org/10.1016/j.str.2017.07.015

64. 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.

65. Zhou H, Sun, Litao Yang X-L, And Schimmel P. ATP-Directed Capture of Bioactive Herbal-Based Medicine on Human tRNA Synthetase. 2015;2:147–85.

66. 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.

67. Baker DA. Malaria gametocytogenesis. Mol Biochem Parasitol [Internet]. Elsevier B.V.;

2010;172:57–65. Available from: http://dx.doi.org/10.1016/j.molbiopara.2010.03.019

68. 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. p. 1–13.

69. 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 U S A. 2000;97:12153–7.

70. Pouplana LR De, Schimmel P. Visions & Reflections A view into the origin of life : aminoacyl-tRNA synthetases. Nature. 2000;57:865–70.

71. Faya N, Penkler DL, Tastan Bishop Ö. Human, vector and parasite Hsp90 proteins: A comparative bioinformatics analysis. FEBS Open Bio. 2015;5:916–27.

72. Pink R, Hudson A, Mouriès MA, Bendig M. Opportunities and challenges in antiparasitic drug discovery. Nat Rev Drug Discov. 2005;4:727–40.

73. Bunjun S, Stathopoulos C, Graham D, Min B, Kitabatake M, Wang a L, et al. A dual-specificity aminoacyl-tRNA synthetase in the deep-rooted eukaryote Giardia lamblia. Proc Natl Acad Sci U S A. 2000;97:12997–3002.

74. Hatherley R, Clitheroe CL, Faya N, Tastan Bishop Ö. Plasmodium falciparum Hop: Detailed analysis on complex formation with Hsp70 and Hsp90. Biochem Biophys Res Commun. 2015;456:440–5.

75. 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.

76. Shibata S, Gillespie JR, Kelley AM, Napuli AJ, Zhang Z, Kovzun K V., et al. Selective inhibitors of methionyl-tRNA synthetase have potent activity against Trypanosoma brucei

1237 infection in mice. Antimicrob Agents Chemother. 2011;55:1982–9.

1238 77. Dahl EL, Rosenthal PJ. Multiple antibiotics exert delayed effects against the Plasmodium
1239 falciparum apicoplast. Antimicrob Agents Chemother. 2007;51:3485–90.

1240 78. Schimmel P. Development of tRNA synthetases and connection to genetic code and
1241 disease. Protein Sci. 2008;17:1643–52.

1242 79. Richa Agarwala, Tanya Barrett, Jeff Beck, Dennis A. Benson, Colleen Bollin, Evan
1243 Bolton, Devon Bourexis, J. Rodney Brister, Stephen H. Bryant, Kathi Canese, Karen Clark,
1244 Michael DiCuc- cio, Ilya Dondoshansky, Scott Federhen, Michael Feolo, Kathryn Funk, L
1245 KZ. Database resources of the National Center for Biotechnology Information. Nucleic Acids
1246 Res [Internet]. 2014;41:1–12. Available from:
1247 http://www.ncbi.nlm.nih.gov/pubmed/25398906

1248 80. The UniProt Consortium. UniProt: a hub for protein information. Nucleic Acids Res
1249 [Internet]. 2015;43:D204-12. Available from:
1250 http://nar.oxfordjournals.org/cgi/content/long/43/D1/D204

1251 81. Eriani G, Delarue M, Poch O, Gangloff J, Moras D. Partition of tRNA synthetases into
1252 two classes based on mutually exclusive sets of sequence motifs. Lett to Nat. 1990;347:203–
1253 6.

1254 82. Cusack S. Aminoacyl-tRNA synthetases. Curr Opin Struct Biol. 1997;7:881–9.

1255 83. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein
1256 Databank. Nucleic Acids Res. 2000;28:235–42.

1257 84. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME Suite. Nucleic Acids Res.
1258 2015;43:W39–49.

1259 85. Bailey TL, Gribskov M. Combining evidence using p-values: application to sequence

1260 homology searches. Bioinformatics [Internet]. 1998;14:48–54. Available from:
1261 https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/14.1.48

1262 86. Kelm S, Shi J, Deane CM. MEDELLER: Homology-based coordinate generation for
1263 membrane proteins. Bioinformatics. 2010;26:2833–40.

1264 87. Söding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology
1265 detection and structure prediction. Nucleic Acids Res. 2005;33.

1266 88. Hatherley R, Brown DK, Glenister M, Bishop ÖT. PRIMO: An interactive homology
1267 modeling pipeline. PLoS One. 2016;11:1–20.

1268 89. Jain V, Yogavel M, Sharma A. Dimerization of Arginyl-tRNA Synthetase by Free Heme
1269 Drives Its Inactivation in Plasmodium falciparum. Structure. 2016;24:1476–87.

1270 90. Ojo KK, Ranade RM, Zhang Z, Dranow DM, Myers JB, Choi R, et al. Brucella melitensis
1271 Methionyl-tRNA-Synthetase (MetRS), a potential drug target for brucellosis. PLoS One.
1272 2016;11.

1273 91. Koh CY, Kim JE, Napoli AJ, Verlinde CLMJ, Fan E, Buckner FS, et al. Crystal structures
1274 of Plasmodium falciparum cytosolic tryptophanyl-tRNA synthetase and its potential as a
1275 target for structure-guided drug design. Mol Biochem Parasitol. 2013;189:26–32.

1276 92. Barros-Álvarez X, Kerchner KM, Koh CY, Turley S, Pardon E, Steyaert J, et al.
1277 Leishmania donovani tyrosyl-tRNA synthetase structure in complex with a tyrosyl adenylate
1278 analog and comparisons with human and protozoan counterparts. Biochimie. 2017;138:124–
1279 36.

1280 93. Hewitt SN, Dranow DM, Horst BG, Abendroth JA, Forte B, Hallyburton I, et al.
1281 Biochemical and structural characterization of selective allosteric inhibitors of the
1282 Plasmodium falciparum drug target, prolyl-tRNA-synthetase. ACS Infect Dis. 2017;3:34–44.

1283 94. Ofir-Birin Y, Fang P, Bennett SP, Zhang H-M, Wang J, Rachmin I, et al. Structural
1284 Switch of Lysyl-tRNA Synthetase between Translation and Transcription. Mol Cell
1285 [Internet]. 2013 [cited 2017 May 23];49:30–42. Available from:
1286 http://linkinghub.elsevier.com/retrieve/pii/S1097276512008635

1287 95. Wiederstein M, Sippl MJ. ProSA-web: Interactive web service for the recognition of
1288 errors in three-dimensional structures of proteins. Nucleic Acids Res. 2007;35.

1289 96. Eisenberg D, Lüthy R, Bowie JU. VERIFY3D: Assessment of protein models with three-
1290 dimensional profiles. Methods Enzymol. 1997;277:396–406.

1291 97. Benkert P, Künzli M, Schwede T. QMEAN server for protein model quality estimation.
1292 Nucleic Acids Res [Internet]. 2009;37:W510–4. Available from:
1293 http://www.ncbi.nlm.nih.gov/pubmed/19429685%5Cnhttp://www.pubmedcentral.nih.gov/arti
1294 clerender.fcgi?artid=PMC2703985%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcg
1295 i?artid=2703985&tool=pmcentrez&rendertype=abstract

1296 98. Di Tommaso P, Moretti S, Xenarios I, Orobitg M, Montanyola A, Chang JM, et al. T-
1297 Coffee: A web server for the multiple sequence alignment of protein and RNA sequences
1298 using structural information and homology extension. Nucleic Acids Res. 2011;39.

1299 99. Pei J, Kim BH, Grishin N V. PROMALS3D: A tool for multiple protein sequence and
1300 structure alignments. Nucleic Acids Res. 2008;36:2295–300.

1301 100. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2-A
1302 multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25:1189–
1303 91.

1304 101. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis
1305 Version 7.0 for Bigger Datasets. Mol Biol Evol. 2016;33:1870–4.

1306 102. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular
1307 evolutionary genetics analysis using maximum likelihood, evolutionary distance, and
1308 maximum parsimony methods. Mol Biol Evol. 2011;28:2731–9.

1309 103. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices
1310 from protein sequences. Bioinformatics. 1992;8:275–82.

1311 104. Ngan CH, Bohnuud T, Mottarella SE, Beglov D, Villar EA, Hall DR, et al. FTMAP:
1312 Extended protein mapping with user-selected probe molecules. Nucleic Acids Res. 2012;40.

1313 105. Halgren T. New method for fast and accurate binding-site identification and analysis.
1314 Chem Biol Drug Des. 2007;69:146–8.

1315 106. Halgren TA. Identifying and characterizing binding sites and assessing druggability. J
1316 Chem Inf Model. 2009;49:377–89.

1317 107. Kozakov D, Grove LE, Hall DR, Bohnuud T, Mottarella SE, Luo L, et al. The FTMap
1318 family of web servers for determining and characterizing ligand-binding hot spots of proteins.
1319 Nat Protoc. 2015;10:733–55.

1320 108. Moras D. Structural and functional relationships between aminoacyl-tRNA synthetases.
1321 Trends Biochem Sci [Internet]. 1992;17:159–64. Available from:
1322 papers3://publication/uuid/7E1AF307-2980-4379-933C-27969EA880B7

1323 109. Fourmy D, Mechulam Y, Blanquet S. Crucial role of an idiosyncratic insertion in the
1324 Rossman fold of class 1 aminoacyl-tRNA synthetases: the case of methionyl-tRNA
1325 synthetase. Biochemistry [Internet]. 1995;34:15681–8. Available from:
1326 http://www.ncbi.nlm.nih.gov/pubmed/7495798

1327 110. Mailu BM, Ramasamay G, Mudeppa DG, Li L, Lindner SE, Peterson MJ, et al. A
1328 nondiscriminating glutamyl-tRNA synthetase in the plasmodium apicoplast: The first enzyme

in an indirect aminoacylation pathway. J Biol Chem. 2013;288:32539–52.

111. Mailu BM, Li L, Arthur J, Nelson TM, Ramasamy G, Fritz-Wolf K, et al. Plasmodium apicoplast Gln-tRNAGln biosynthesis utilizes a unique GatAB amidotransferase essential for erythrocytic stage parasites. J Biol Chem. 2015;290:29629–41.

112. Kyriacou S V., 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 [Internet]. 2016;15:203. Available from: http://malariajournal.biomedcentral.com/articles/10.1186/s12936-016-1247-0

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. p. 985–93.

116. Lee SW. Aminoacyl-tRNA synthetase complexes: beyond translation. J Cell Sci [Internet]. 2004;117:3725–34. Available from: http://jcs.biologists.org/cgi/doi/10.1242/jcs.01342

117. Guo M, Yang X-L, Schimmel P. New functions of aminoacyl-tRNA synthetases beyond translation. Nat Rev Mol Cell Biol [Internet]. 2010;11:668–74. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3042954&tool=pmcentrez&rende

rtype=abstract

118. Ko YG, Kim EK, Kim T, Park H, Park HS, Choi EJ, et al. Glutamine-dependent Antiapoptotic Interaction of Human Glutaminyl-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.

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 A. 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 [Internet]. 1994;13:327–37. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=394812&tool=pmcentrez&rendertype=abstract

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 [Internet]. 1995;14:4156–67. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=394498&tool=pmcentrez&rendertype=abstract

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 U S A [Internet]. 2002;99:2678–83. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=122407&tool=pmcentrez&rendertype=abstract

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 [Internet]. 1995;5:48–

55. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7773747

132. Guigou L, Shalak 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 [Internet]. 1987;326:808–11. Available from: http://www.ncbi.nlm.nih.gov/pubmed/3033511

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, Kyung Hee Kim, Min Jung Kim, 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 U S A. 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 (80-) [Internet]. 1999;284:147–51. Available from: http://www.sciencemag.org/cgi/doi/10.1126/science.284.5411.147

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 [Internet]. 2000;19:2371–80. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=384352&tool=pmcentrez&rendertype=abstract

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 [Internet]. 2001;305:451–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11152603

143. de Oca MM, Engwerda C, Haque A. Plasmodium berghei ANKA (PbA) infection of C57BL/6J mice: a model of severe malaria. [Internet]. Methods Mol. Biol. 2013. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23824903

144. Otto TD, Böhme U, Jackson AP, Hunt M, Franke-Fayard B, Hoeijmakers WAM, et al. A comprehensive evaluation of rodent malaria parasite genomes and gene expression. BMC Biol [Internet]. 2014;12:86. Available from: http://bmcbiol.biomedcentral.com/articles/10.1186/s12915-014-0086-0

145. Mitsui H, Arisue N, Sakihama N, Inagaki Y, Horii T, Hasegawa M, et al. Phylogeny of Asian primate malaria parasites inferred from apicoplast genome-encoded genes with special emphasis on the positions of Plasmodium vivax and P. fragile. Gene. 2010;450:32–8.

146. Pain A, Böhme U, Berry AE, Mungall K, Finn RD, Jackson AP, et al. The genome of

1444 the simian and human malaria parasite Plasmodium knowlesi. Nature. 2008;455:799–803.

1445 147. Carlton JM, Adams JH, Silva JC, Bidwell SL, Lorenzi H, Caler E, et al. Comparative
1446 genomics of the neglected human malaria parasite Plasmodium vivax. Nature. 2008;455:757–
1447 63.

1448 148. Höglind A, Areström I, Ehrnfelt C, Masjedi K, Zuber B, Giavedoni L, et al. Systematic
1449 evaluation of monoclonal antibodies and immunoassays for the detection of Interferon-γ and
1450 Interleukin-2 in old and new world non-human primates. J Immunol Methods. 2017;441:39–
1451 48.

1452 149. Le SQ, Gascuel O. An improved general amino acid replacement matrix. Mol Biol Evol.
1453 2008;25:1307–20.

1454 150. Bishop ÖT, Kroon M. Study of protein complexes via homology modeling, applied to
1455 cysteine proteases and their protein inhibitors. J Mol Model. 2011;17:3163–72.

1456 151. Shen MY, Sali A. Statistical potential for assessment and prediction of protein structures
1457 [Internet]. Protein Sci. 2006. p. 2507–24. Available from:
1458 http://www.ncbi.nlm.nih.gov/pubmed/17075131

1459 152. Eramian D, Shen M, Devos D, Melo F, Sali A, Marti-Renom MA. A composite score
1460 for predicting errors in protein structure models. Protein Sci [Internet]. 2006;15:1653–66.
1461 Available from: http://doi.wiley.com/10.1110/ps.062095806

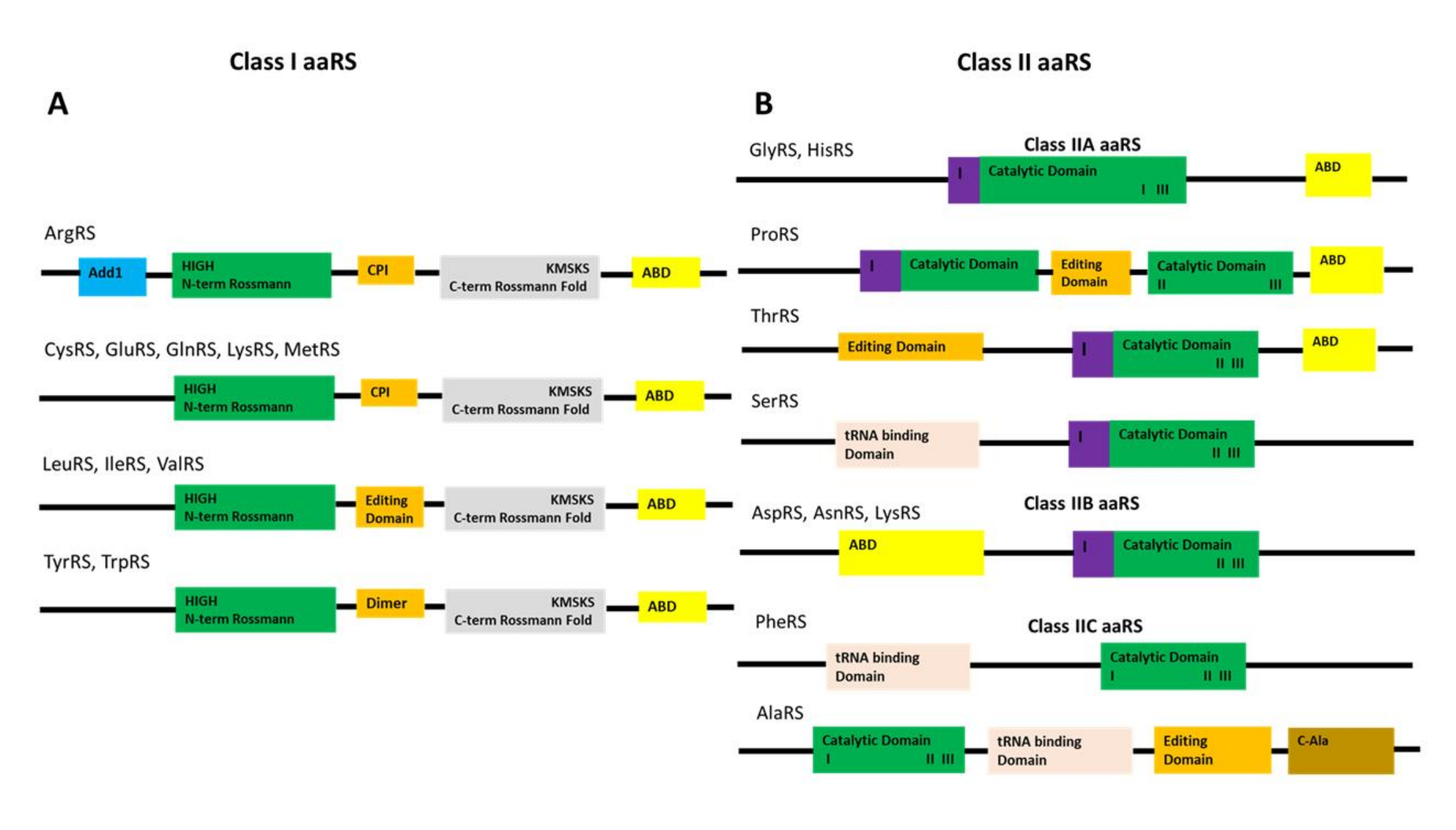
1462 153. Marko AC, Stafford K, Wymore T. Stochastic pairwise alignments and scoring methods
1463 for comparative protein structure modeling. J Chem Inf Model [Internet]. 2007;47:1263–70.
1464 Available from: http://www.ncbi.nlm.nih.gov/pubmed/17391002

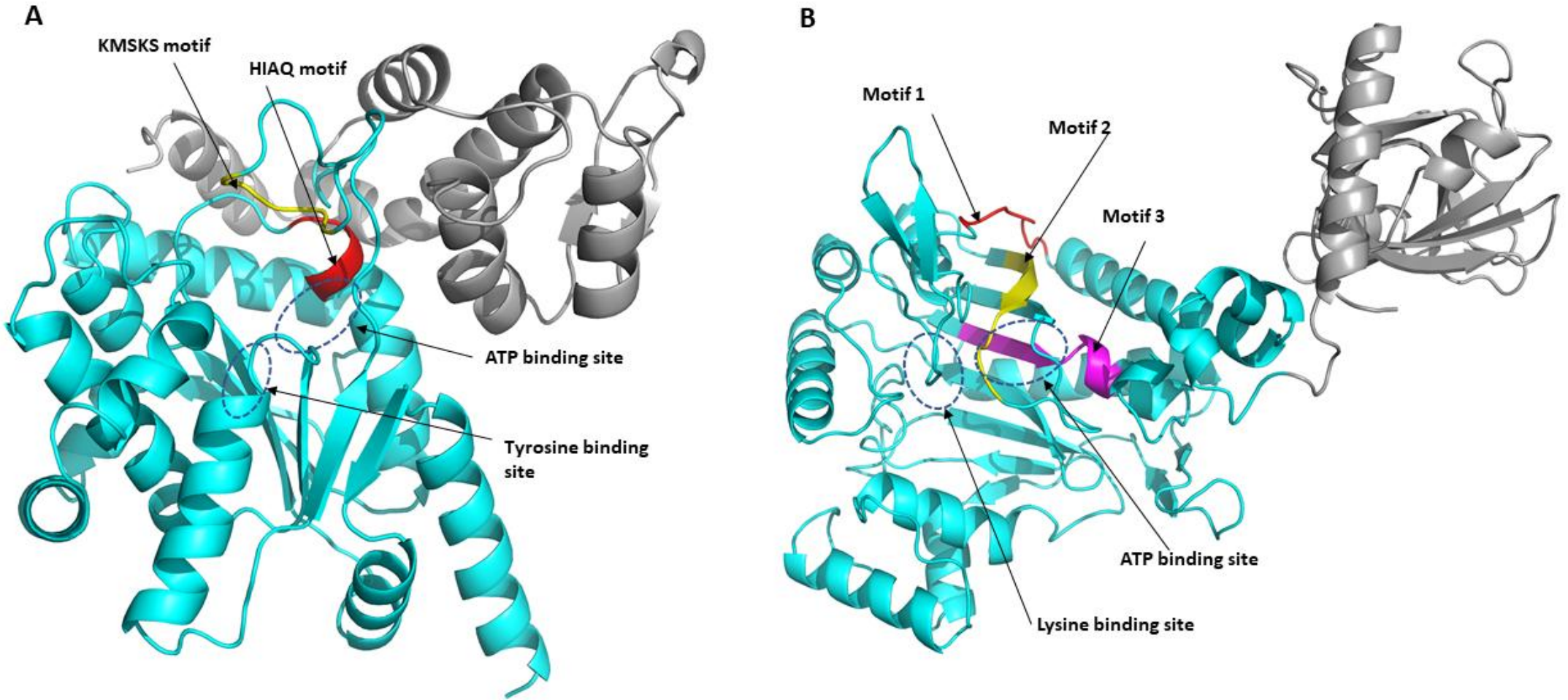
1465 154. Chen H, Kihara D. Estimating quality of template-based protein models by alignment
1466 stability. Proteins Struct Funct Genet. 2008;71:1255–74.

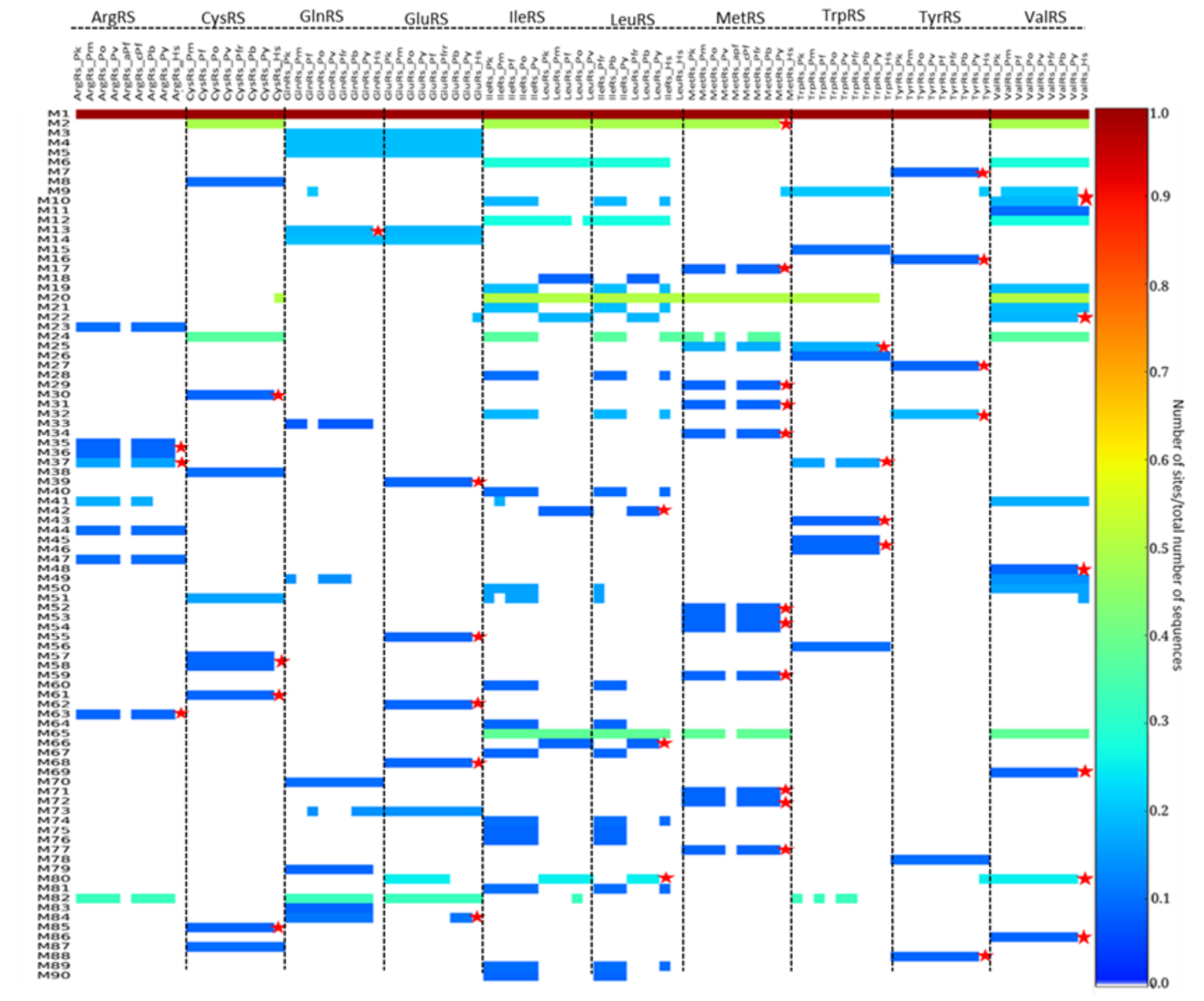
155. Koh CY, Kim JE, Napoli AJ, Verlinde CLMJ, Fan E, Buckner FS, et al. Crystal structures of Plasmodium falciparum cytosolic tryptophanyl-tRNA synthetase and its potential as a target for structure-guided drug design. Mol Biochem Parasitol [Internet]. Elsevier B.V.; 2013;189:26–32. Available from: http://dx.doi.org/10.1016/j.molbiopara.2013.04.007

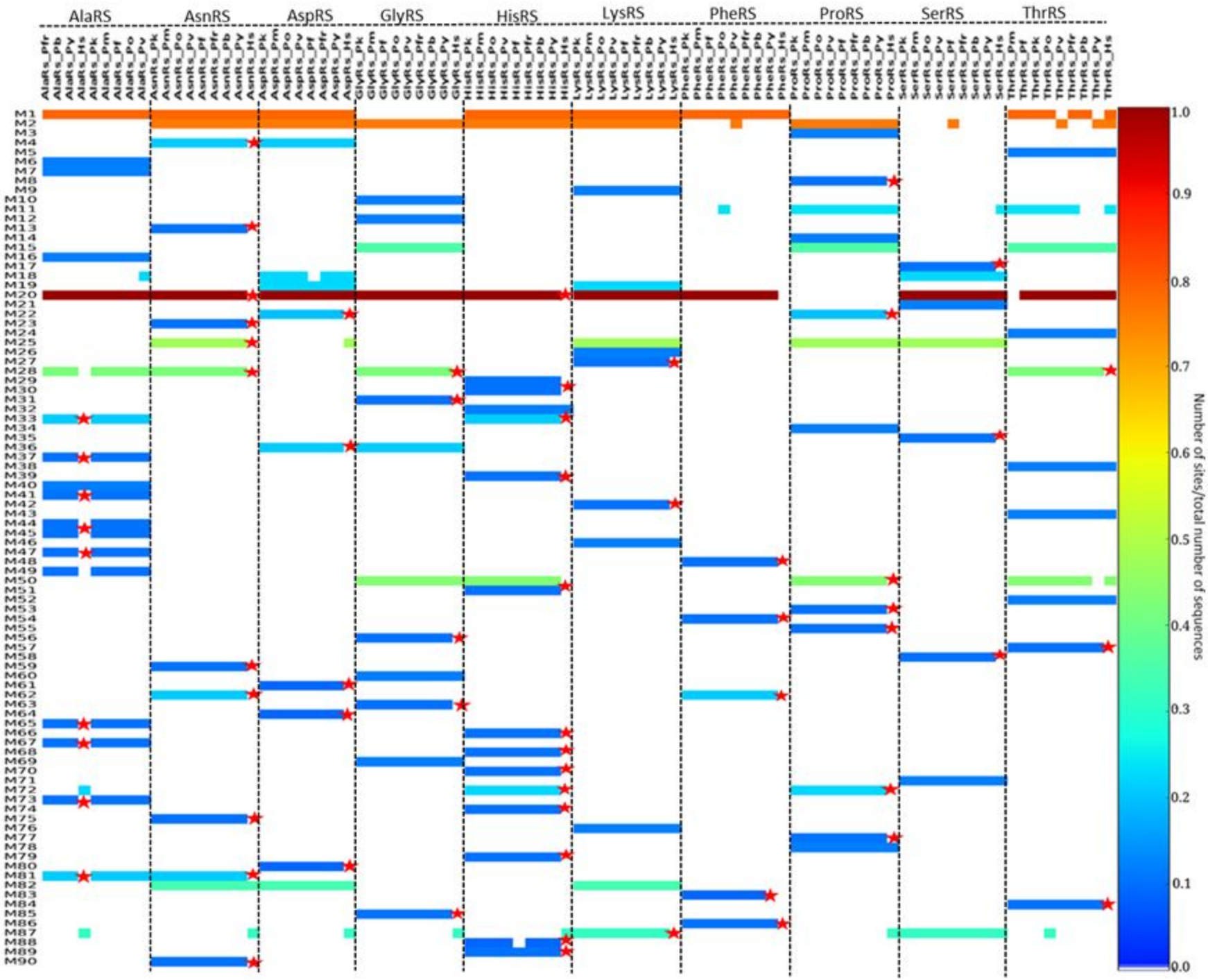
156. Yang X-L, Otero FJ, Skene RJ, McRee DE, Schimmel P, Ribas de Pouplana L. Crystal structures that suggest late development of genetic code components for differentiating aromatic side chains. Proc Natl Acad Sci [Internet]. 2003;100:15376–80. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.2136794100

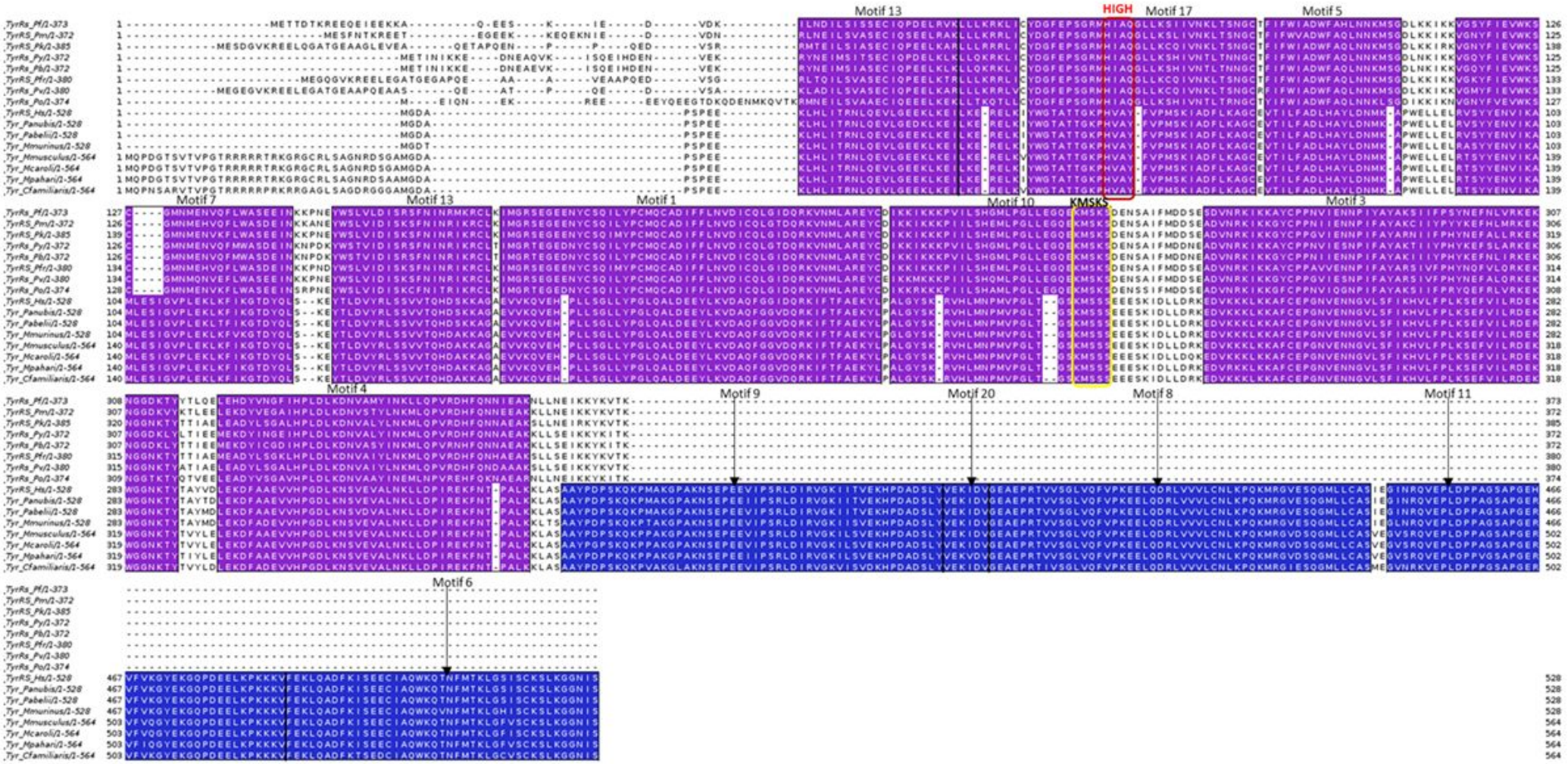
157. Koh CY, Kim JE, Shibata S, Ranade RM, Yu M, Liu J, et al. Distinct states of methionyl-tRNA synthetase indicate inhibitor binding by conformational selection. Structure. 2012;20:1681–91.

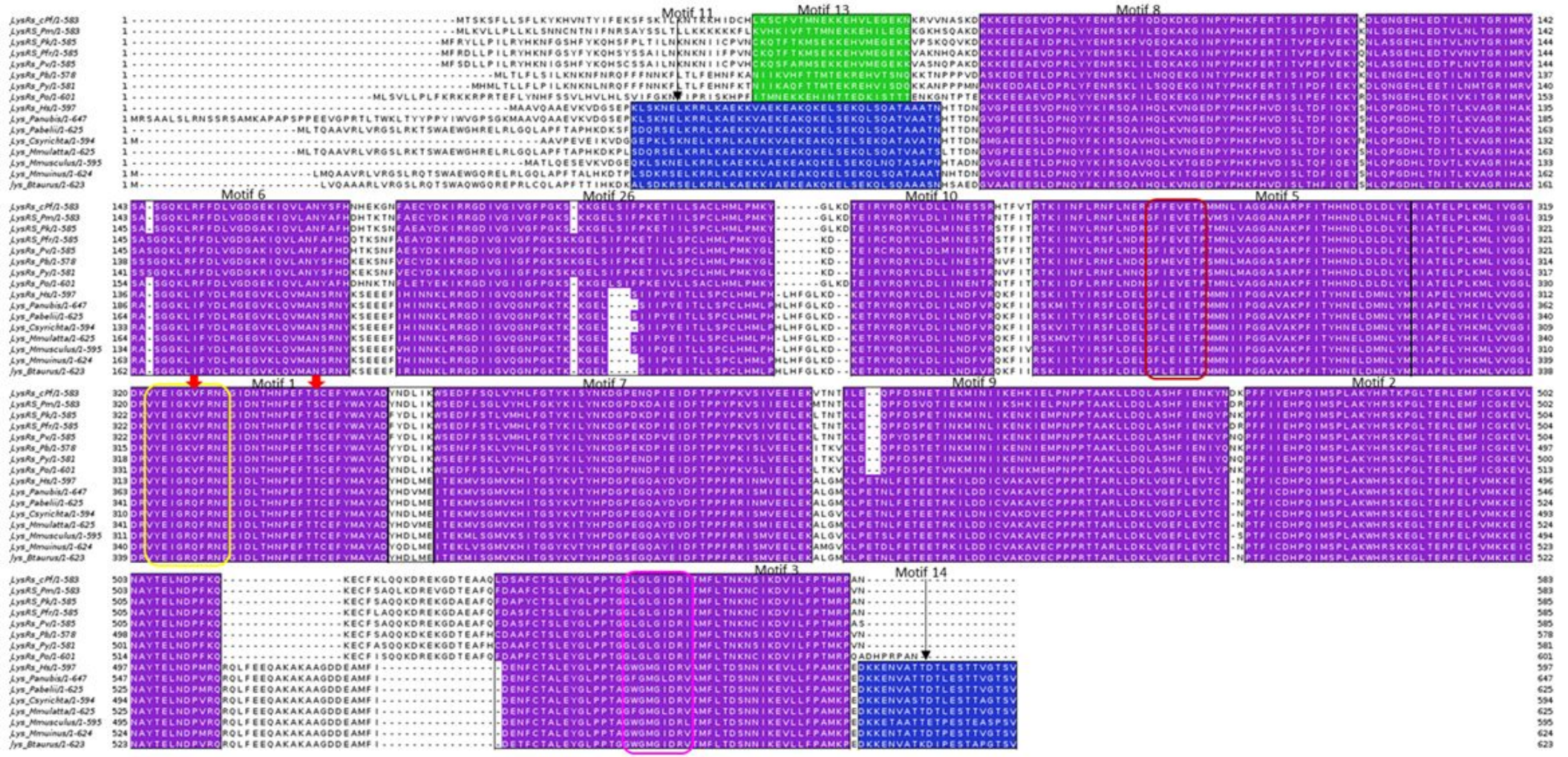

Class I aaRS
A
ArgRS
Add1
HIGH
N-term Rossmann
CPI
KMSKS
C-term Rossmann Fold
ABD
CysRS, GluRS, GlnRS, LysRS, MetRS
HIGH
N-term Rossmann
CPI
KMSKS
C-term Rossmann Fold
ABD
LeuRS, IleRS, ValRS
HIGH
N-term Rossmann
Editing
Domain
KMSKS
C-term Rossmann Fold
ABD
TyrRS, TrpRS
HIGH
N-term Rossmann
Dimer
KMSKS
C-term Rossmann Fold
ABD
Class II aaRS
B
GlyRS, HisRS
Class IIA aaRS
I
Catalytic Domain
I III
ABD
ProRS
I
Catalytic Domain
Editing
Domain
Catalytic Domain
II III
ABD
ThrRS
Editing Domain
I
Catalytic Domain
II III
ABD
SerRS
tRNA binding
Domain
I
Catalytic Domain
II III
AspRS, AsnRS, LysRS
Class IIB aaRS
ABD
I
Catalytic Domain
II III
PheRS
Class IIC aaRS
tRNA binding
Domain
Catalytic Domain
I II III
AlaRS
Catalytic Domain
I II III
tRNA binding
Domain
Editing
Domain
C-Ala

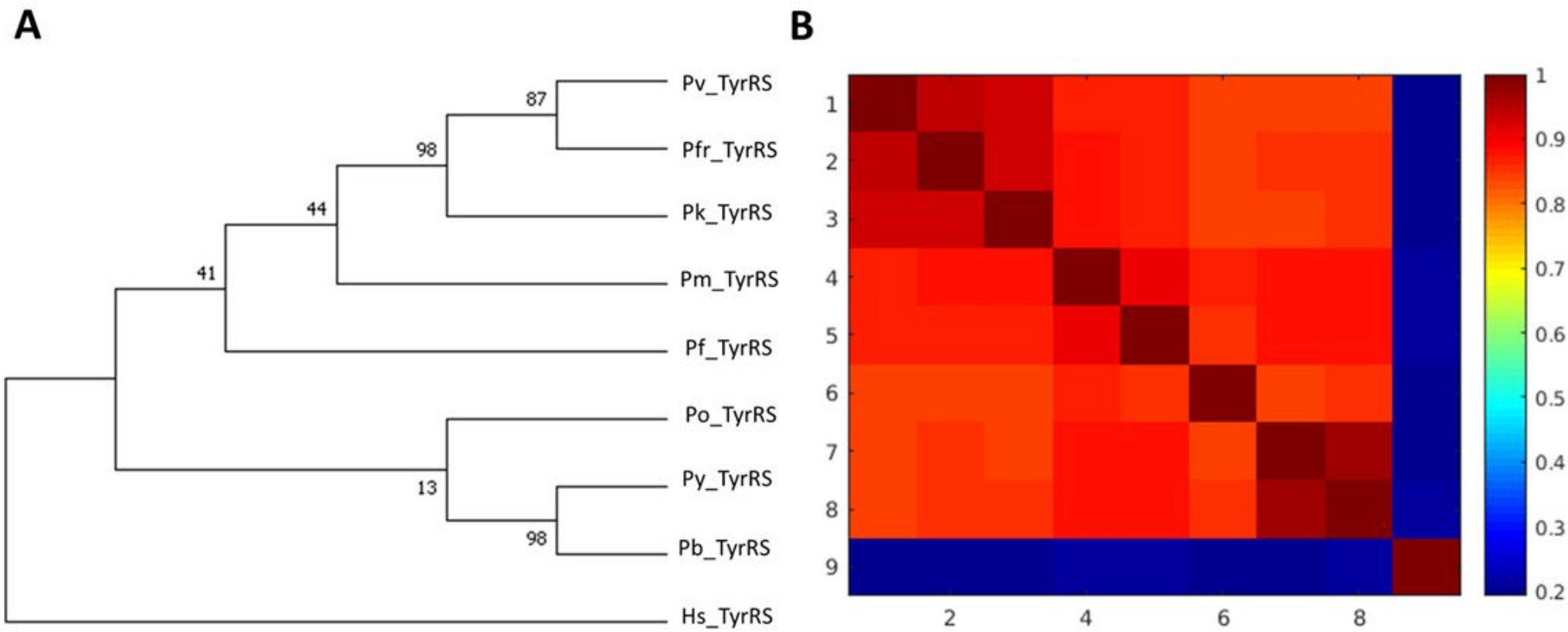

A
KMSKS motif
HIAQ motif
ATP binding site
Tyrosine binding site
B
Motif 1
Motif 2
Motif 3
ATP binding site
Lysine binding site

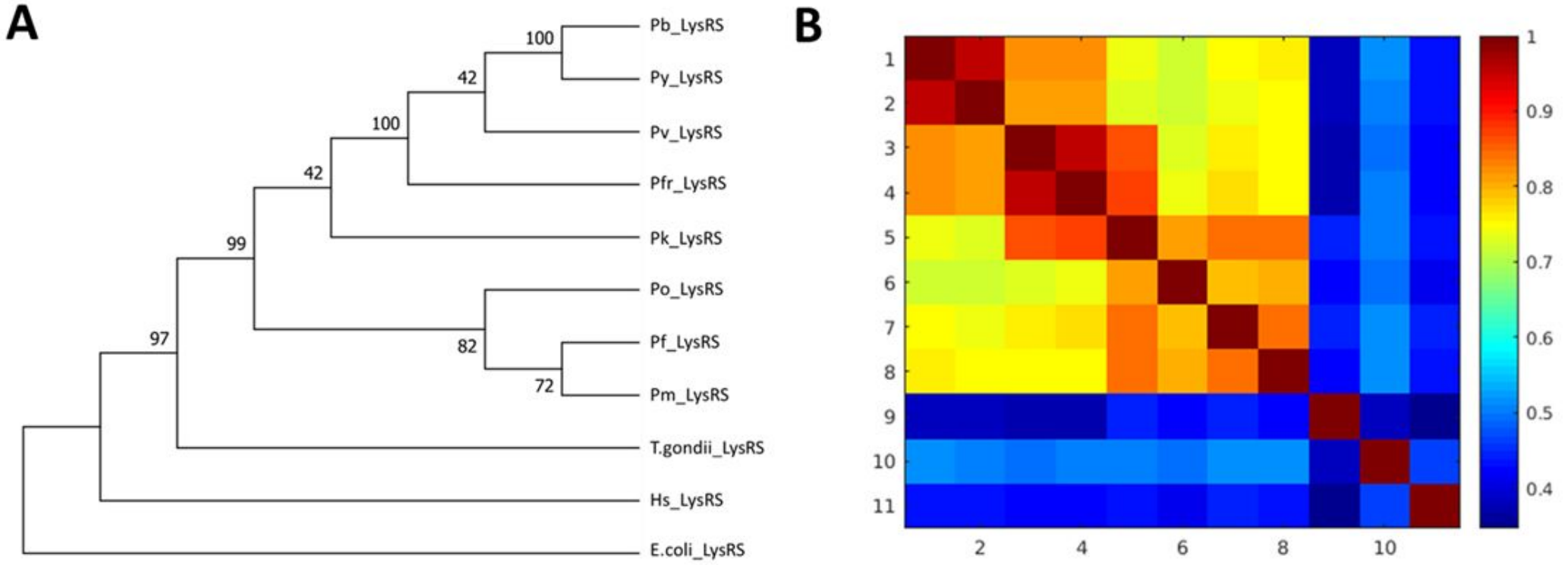

ArgRS
CysRS
GlnRS
GluRS
IleRS
LeuRS
MetRS
TrpRS
TyrRS
ValRS
ArgRS_Pk
ArgRS_Pm
ArgRS_Po
ArgRS_Pv
ArgRS_aPf
ArgRS_cPf
ArgRS_Pb
ArgRS_Py
ArgRS_Hs
CysRS_Pm
CysRS_Pf
CysRS_Po
CysRS_Pv
CysRS_Pfr
CysRS_Pb
CysRS_Py
CysRS_Hs
GlnRS_Pk
GlnRS_Pm
GlnRS_Pf
GlnRS_Po
GlnRS_Pv
GlnRS_Pfr
GlnRS_Pb
GlnRS_Py
GlnRS_Hs
GluRS_Pk
GluRS_Pm
GluRS_Po
GluRS_Py
GluRS_Pf
GluRS_Pfrr
GluRS_Pb
GluRS_Py
GluRS_Hs
IleRS_Pk
IleRS_Pm
IleRS_Pf
IleRS_Po
IleRS_Pv
LeuRS_Pk
LeuRS_Pm
LeuRS_Pf
LeuRS_Po
LeuRS_Pv
IleRS_Pfr
IleRS_Pb
LeuRS_Pfr
LeuRS_Pb
IleRS_Py
LeuRS_Py
LeuRS_Hs
MetRS_Pk
MetRS_Pm
MetRS_Po
MetRS_Pv
MetRS_aPf
MetRS_cPf
MetRS_Pfr
MetRS_Pb
MetRS_Py
MetRS_Hs
TrpRS_Pk
TrpRS_Pm
TrpRS_Pf
TrpRS_Po
TrpRS_Pv
TrpRS_Pfr
TrpRS_Pb
TrpRS_Py
TrpRS_Hs
TyrRS_Pk
TyrRS_Pm
TyrRS_Po
TyrRS_Pv
TyrRS_Pf
TyrRS_Pfr
TyrRS_Pb
TyrRS_Py
TyrRS_Hs
ValRS_Pk
ValRS_Pm
ValRS_Pf
ValRS_Po
ValRS_Pv
ValRS_Pfr
ValRS_Pb
ValRS_Py
ValRS_Hs
M1
M2
M3
M4
M5
M6
M7
M8
M9
M10
M11
M12
M13
M14
M15
M16
M17
M18
M19
M20
M21
M22
M23
M24
M25
M26
M27
M28
M29
M30
M31
M32
M33
M34
M35
M36
M37
M38
M39
M40
M41
M42
M43
M44
M45
M46
M47
M48
M49
M50
M51
M52
M53
M54
M55
M56
M57
M58
M59
M60
M61
M62
M63
M64
M65
M66
M67
M68
M69
M70
M71
M72
M73
M74
M75
M76
M77
M78
M79
M80
M81
M82
M83
M84
M85
M86
M87
M88
M89
M90
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Number of sites/total number of sequences

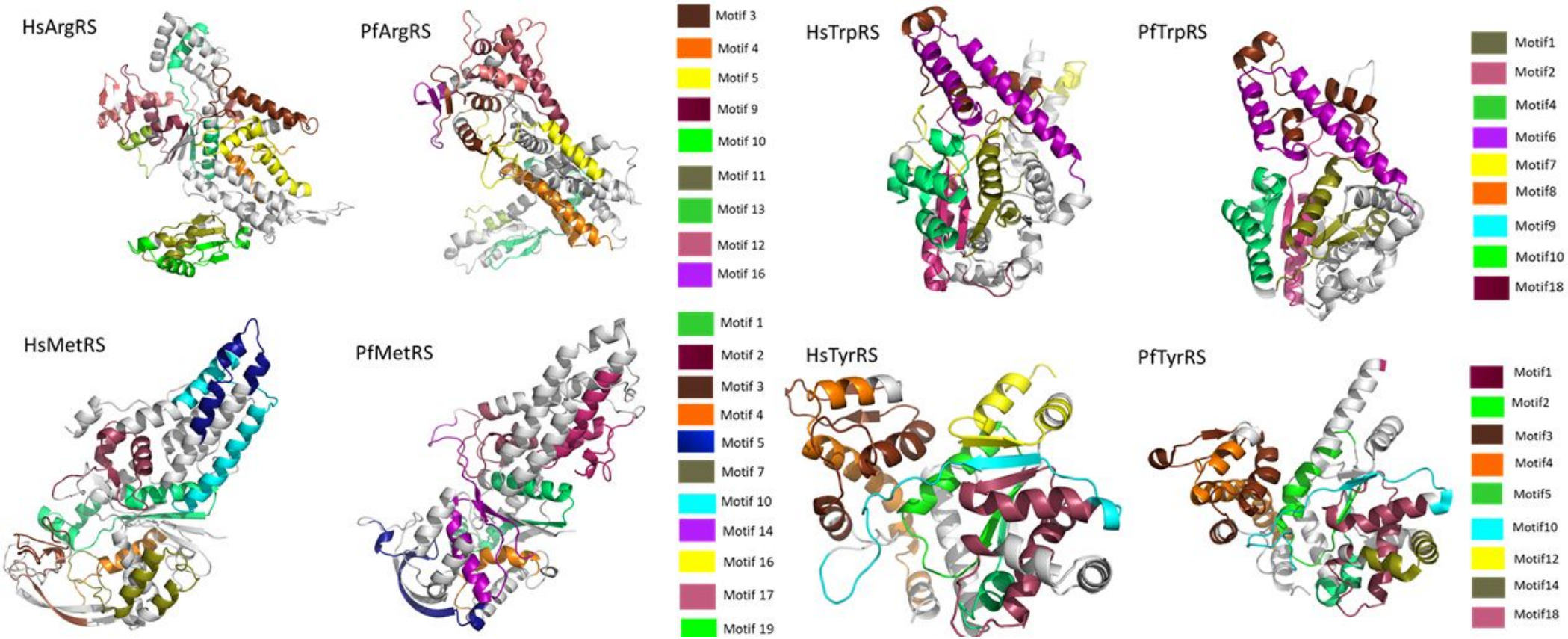

AlaRS
AsnRS
AspRS
GlyRS
HisRS
LysRS
PheRS
ProRS
SerRS
ThrRS
AlaRs_Pfr
AlaRs_Pb
AlaRs_Py
AlaRs_Hs
AlaRs_Pk
AlaRs_Pm
AlaRs_Pf
AlaRs_Po
AlaRs_Pv
AsnRs_Pk
AsnRs_Pm
AsnRs_Po
AsnRs_Pv
AsnRs_Pf
AsnRs_Pfr
AsnRs_Pb
AsnRs_Py
AsnRs_Hs
AspRs_Pk
AspRs_Pm
AspRs_Po
AspRs_Pv
AspRs_Pf
AspRs_Pfr
AspRs_Py
AspRs_Hs
GlyRs_Pk
GlyRs_Pm
GlyRs_Pf
GlyRs_Po
GlyRs_Pv
GlyRs_Pfr
GlyRs_Pb
GlyRs_Py
GlyRs_Hs
HisRs_Pk
HisRs_Pm
HisRs_Po
HisRs_Pv
HisRs_Pf
HisRs_Pfr
HisRs_Pb
HisRs_Py
HisRs_Hs
LysRs_Pk
LysRs_Pm
LysRs_Po
LysRs_Pv
LysRs_Pf
LysRs_Pfr
LysRs_Pb
LysRs_Py
LysRs_Hs
PheRs_Pk
PheRs_Pm
PheRs_Pf
PheRs_Po
PheRs_Pv
PheRs_Pfr
PheRs_Pb
PheRs_Py
PheRs_Hs
ProRs_Pk
ProRs_Pm
ProRs_Po
ProRs_Pv
ProRs_Pf
ProRs_Pfr
ProRs_Pb
ProRs_Py
ProRs_Hs
SerRs_Pk
SerRs_Pm
SerRs_Po
SerRs_Pv
SerRs_Pf
SerRs_Pfr
SerRs_Pb
SerRs_Py
SerRs_Hs
ThrRs_Pm
ThrRs_Pf
ThrRs_Pk
ThrRs_Po
ThrRs_Pv
ThrRs_Pfr
ThrRs_Pb
ThrRs_Py
ThrRs_Hs
M1
M2
M3
M4
M5
M6
M7
M8
M9
M10
M11
M12
M13
M14
M15
M16
M17
M18
M19
M20
M21
M22
M23
M24
M25
M26
M27
M28
M29
M30
M31
M32
M33
M34
M35
M36
M37
M38
M39
M40
M41
M42
M43
M44
M45
M46
M47
M48
M49
M50
M51
M52
M53
M54
M55
M56
M57
M58
M59
M60
M61
M62
M63
M64
M65
M66
M67
M68
M69
M70
M71
M72
M73
M74
M75
M76
M77
M78
M79
M80
M81
M82
M83
M84
M85
M86
M87
M88
M89
M90
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Number of sites/total number of sequences

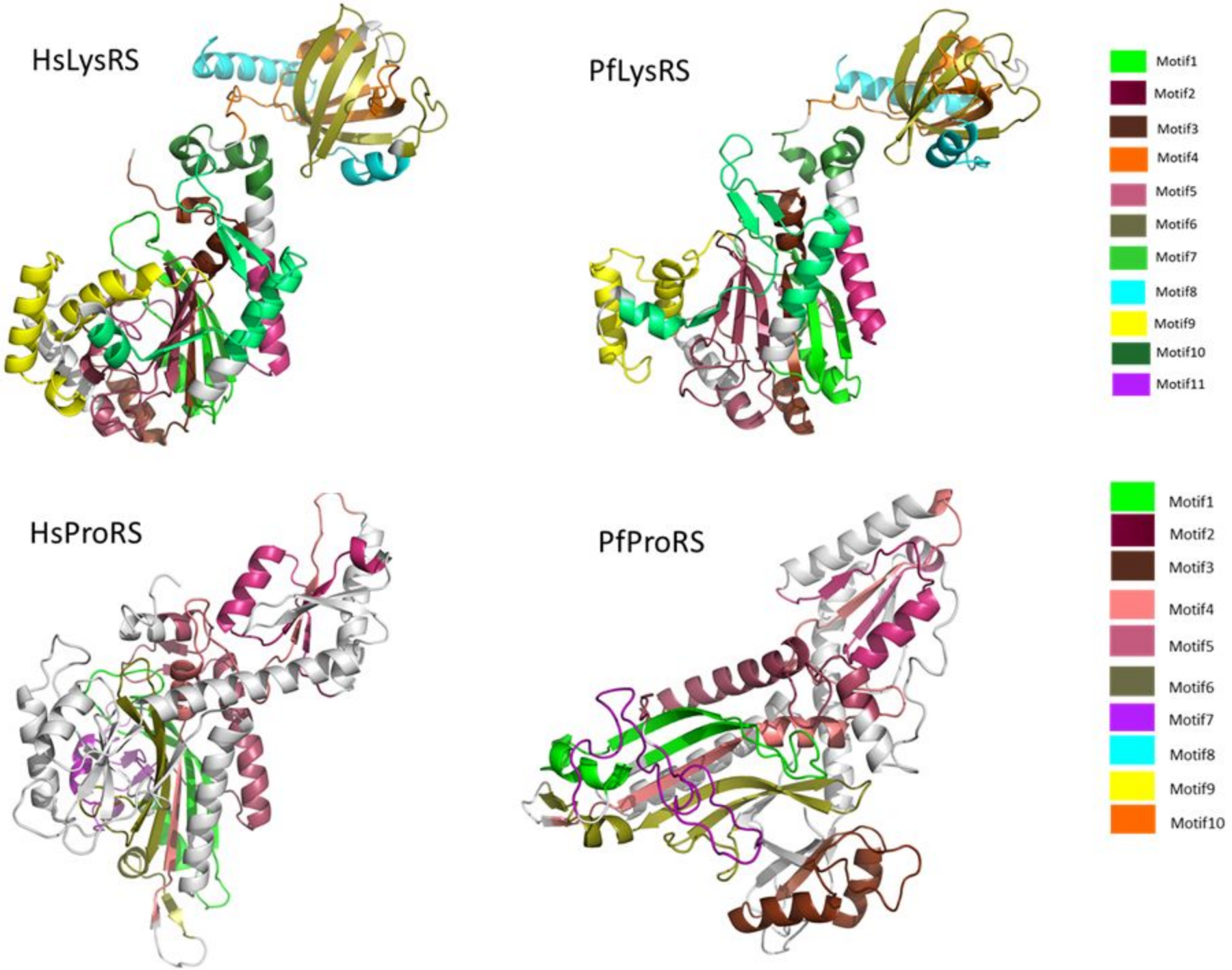

Motif 13
HIGH
Motif 17
Motif 5
Motif 7
Motif 13
Motif 1
Motif 10
KMSKS
Motif 3
Motif 4
Motif 9
Motif 20
Motif 8
Motif 11
Motif 6

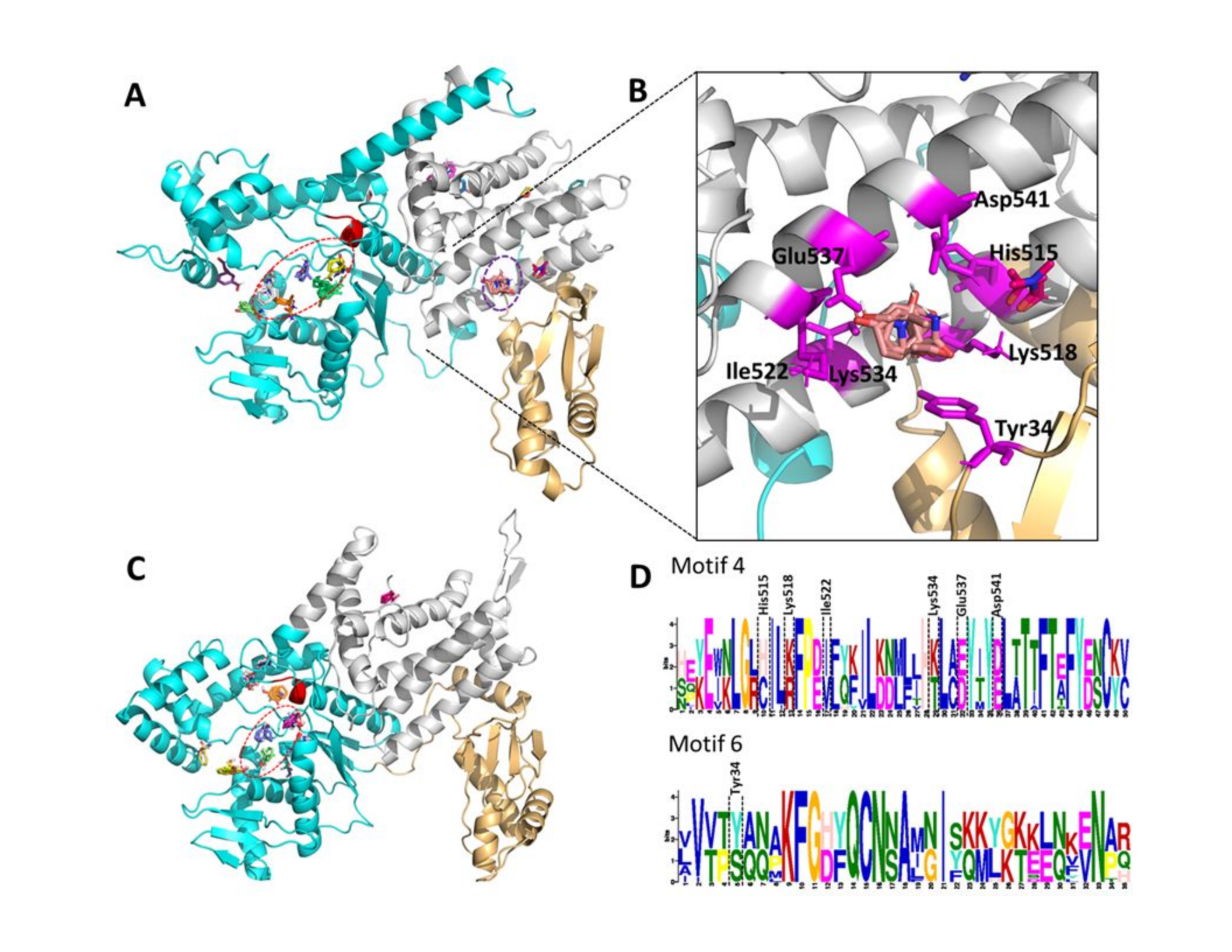

Motif 11
Motif 13
Motif 8
Motif 6
Motif 26
Motif 10
Motif 5
Motif 1
Motif 7
Motif 9
Motif 2
Motif 3
Motif 14

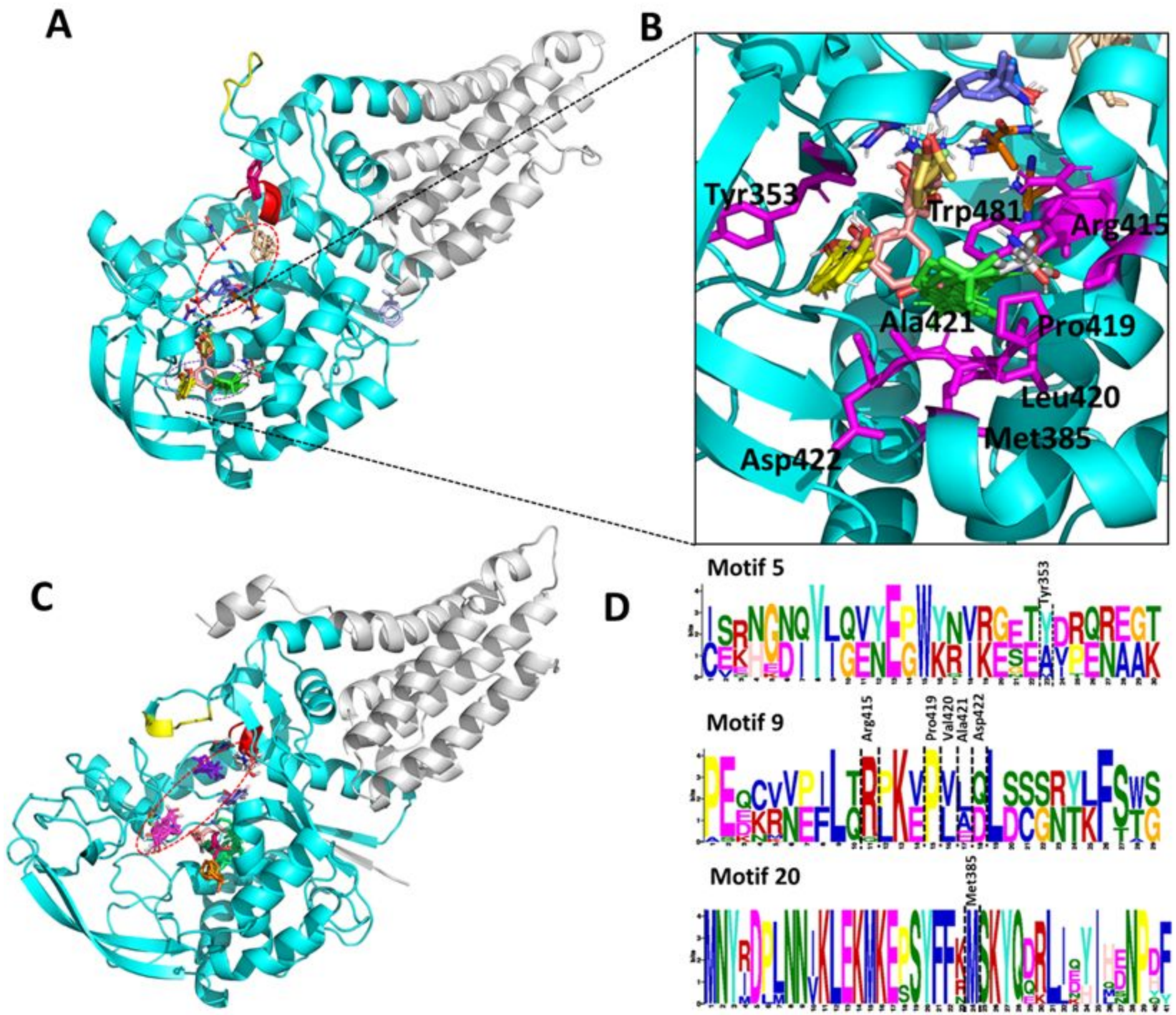

A
B
Pv_TyrRS
Pfr_TyrRS
Pk_TyrRS
Pm_TyrRS
Pf_TyrRS
Po_TyrRS
Py_TyrRS
Pb_TyrRS
Hs_TyrRS
87
98
44
41
13
98
1
2
3
4
5
6
7
8
9
2
4
6
8
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2

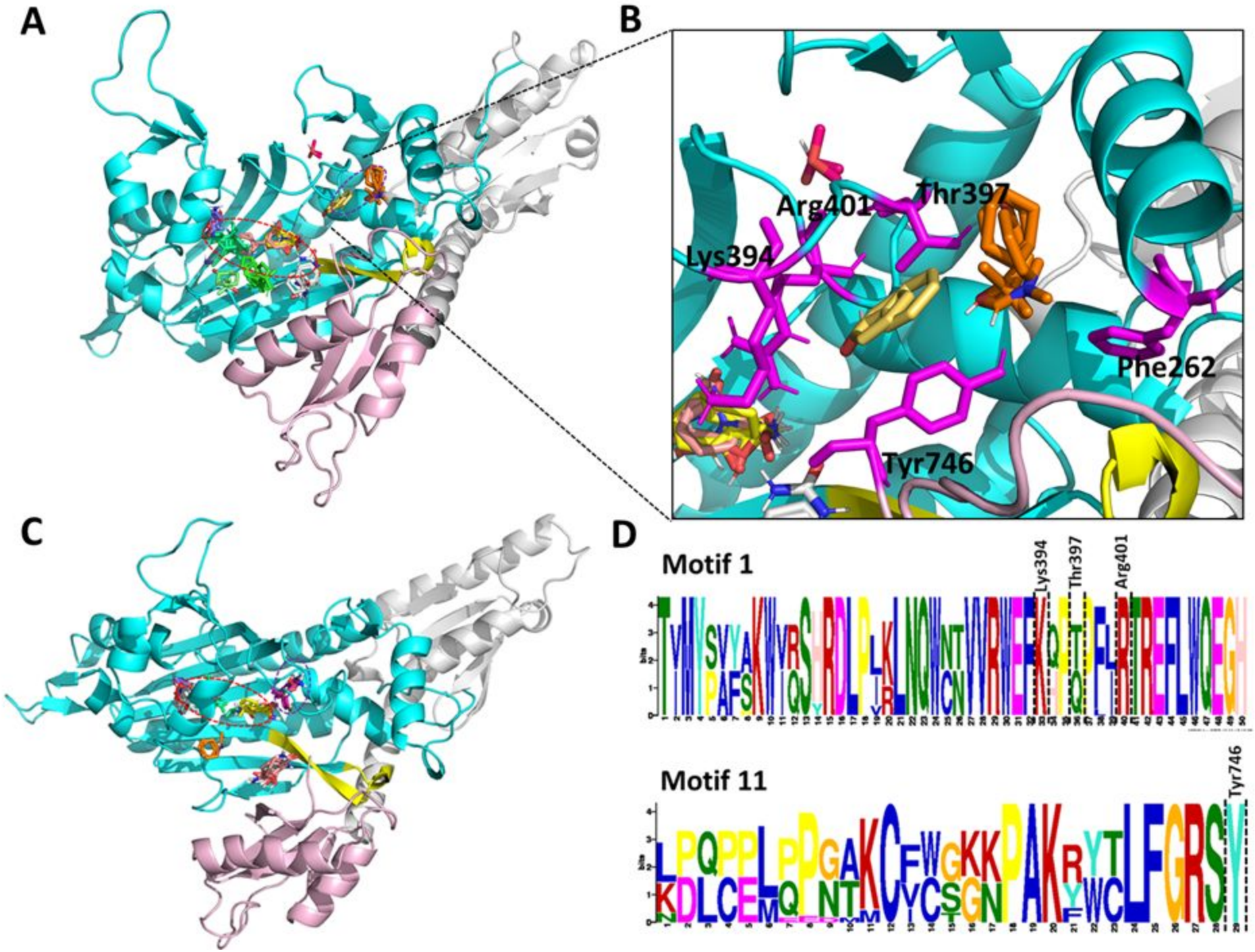

A
Pb_LysRS
Py_LysRS
Pv_LysRS
Pfr_LysRS
Pk_LysRS
Po_LysRS
Pf_LysRS
Pm_LysRS
T.gondii_LysRS
Hs_LysRS
E.coli_LysRS
100
42
100
42
99
97
82
72
B
1
2
3
4
5
6
7
8
9
10
11
2
4
6
8
10
1
0.9
0.8
0.7
0.6
0.5
0.4


HsArgRS
PfArgRS
Motif 3
Motif 4
Motif 5
Motif 9
Motif 10
Motif 11
Motif 13
Motif 12
Motif 16
HsTrpRS
PfTrpRS
Motif1
Motif2
Motif4
Motif6
Motif7
Motif8
Motif9
Motif10
Motif18
HsMetRS
PfMetRS
Motif 1
Motif 2
Motif 3
Motif 4
Motif 5
Motif 7
Motif 10
Motif 14
Motif 16
Motif 17
Motif 19
HsTyrRS
PfTyrRS
Motif1
Motif2
Motif3
Motif4
Motif5
Motif10
Motif12
Motif14
Motif18


HsLysRS
PfLysRS
Motif1
Motif2
Motif3
Motif4
Motif5
Motif6
Motif7
Motif8
Motif9
Motif10
Motif11
HsProRS
PfProRS
Motif1
Motif2
Motif3
Motif4
Motif5
Motif6
Motif7
Motif8
Motif9
Motif10


A
B
Asp541
Glu537
His515
Lys518
Ile522
Lys534
Tyr34
C
D
Motif 4
His515
Lys518
Ile522
Lys534
Glu537
Asp541
Motif 6
Tyr34


A
B
Tyr353
Trp481
Arg415
Ala421
Pro419
Leu420
Met385
Asp422
C
D
Motif 5
Tyr353
Motif 9
Arg415
Pro419
Val420
Ala421
Asp422
Motif 20
Met385


A
B
Arg401
Thr397
Lys394
Phe262
Tyr746
C
D
Motif 1
Lys394
Thr397
Arg401
Motif 11
Tyr746