Journal of Molecular Graphics and Modelling 28 (2010) 401–412

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Template-based structure prediction and molecular dynamics simulation study of two mammalian Aspartyl-tRNA synthetases

Zaheer Ul-Haq [a,*], Waqasuddin Khan [a], Shamshad Zarina [b], Rabia Sattar [a], Syed Tarique Moin [a]

[a] Dr. Panjwani Center for Molecular Medicine and Drug Research, International Center for Chemical and Biological Sciences, University of Karachi, Karachi-75270, Pakistan
[b] Department of Biochemistry, University of Karachi, Karachi-75270, Pakistan

ARTICLE INFO

Article history:
Received 23 May 2009
Received in revised form 14 September 2009
Accepted 23 September 2009
Available online 1 October 2009

Keywords:
Aminoacyl-tRNA synthetases (aaRSs)
Aspartyl-tRNA synthetase (AspRS)
Molecular dynamics (MD)
Root mean square deviation (RMSD)
Root mean square fluctuation (RMSF)

ABSTRACT

Aminoacyl-tRNA synthetases (aaRSs) catalyze the esterification of a specific amino acid. There are two classes of aminoacyl-tRNA synthetases. Class I usually exists as a monomeric or dimeric form and has two highly conserved sequence motifs. Functionally, it aminoacylates at the 2′-OH of an adenosine nucleotide. While, class II normally exists as a dimeric or tetrameric form and consists of three highly conserved sequence motifs. It aminoacylates at the 3′-OH of the same adenosine. Aspartyl-tRNA synthetase (AspRS) belongs to class II aaRSs, is not only important to sustain the mechanism of protein fidelity by specifically recognizing its cognate amino acid; but also equally significant in the aminoacylation of $tRNA^{Asp}$. Several crystal structures of AspRS have been reported yet but no structural information is available for mammalian AspRS. In this study, we have applied template-based modeling/structure prediction to elucidate structural details of two mammalian AspRS from *Homo sapiens* and *Mus musculus*. The resultant models showed excellent stereochemistry similar to the crystal structure of yeast. A 5 ns molecular dynamics (MD) simulation was also performed to study the conformational changes occur in the flipping loop region (279–285). The root mean square fluctuation (RMSF) graph shows movements mostly in the catalytic site and in the flipping loop region while the main secondary structure maintained fairly stable conformations.

1. Introduction

Different catalytic strategies adapted by enzymes make proteins diverse in carrying out enormous biological function. One such enzyme is "Aminoacyl-tRNA synthetases (aaRSs)", whose commitment to genetic translation has become a topic of extensive research. aaRSs constitute a family of cytosolic enzymes of class ligases that play a vital role in protein biosynthesis. Generally, there are 20 aaRSs in each living cell. Each enzyme is specific for one amino acid and one or more corresponding tRNAs. In a small number of organisms, sometimes two aaRS's co-exist for the same amino acid specificity {e.g., LysRS in *Escherichia coli* [1] or AspRS in *Thermus thermophillus* [2]}, or sometimes when some aaRS are absent [e.g., GlnRS in many bacteria [3,4].

aaRSs diverge greatly in their subunit sizes and oligomeric structures, and show very limited sequence similarity. On the basis of considerable differences in primary and tertiary structures and in reaction mechanisms, aminoacyl-tRNA synthetases have been divided into two classes [5,6]. Class I includes three enzymes: MetRS, TyrRS, GlnRS with known 3D structures, and seven others; GluRS, ArgRS, ValRS, IleRS, LeuRS, TrpRS, and CysRS which certainly share one recognizable 3D structure domain with the unique sequence signatures [HIGH] [7] and [KMSKS] [8] associated with Rossmann fold. This domain contains the active site built around a right-handed beta/alpha/beta motif, a phosphate binding consensus sequence, GXGXXG, six small hydrophobic residues, a basic residue (Arg or Lys) at the beginning of the first beta strand, and an acidic residue (Glu or Asp) at the end of the second beta strand. The ten other enzymes compose the class II synthetases: AlaRS, AsnRS, AspRS, GlyRS, HisRS, LysRS, PheRS, ProRS, SerRS, and ThrRS. Among class II synthetases, SerRS and AspRS are more diverse and both share three structural motifs representing a common sequence signature. In addition, the active site is also present around an antiparallel β-sheet.

The accurate translation of the genetic code depends on the precise ligand recognition by the synthetases [9]. During the first stage of protein synthesis, aaRSs are responsible for the accurate aminoacylation (or esterification) of their cognate tRNAs in the ribosome-dependent protein synthesis [10]. The catalytic action depends on the recognition of three substrates. The amino acid residue (first substrate) is attached to its respective aaRS in the presence of ATP (the second substrate). ATP complexed with

* Corresponding author. Tel.: +92 21 111222292x309 fax: +92 21 4819018 19.
E-mail address: zaheer.qasmi@iccs.edu (Z. Ul-Haq).

1093-3263/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2009.09.006

magnesium forms an intermediate product, aminoacyl-adenylate (aminoacyl-AMP). Then, aminoacyl-adenylate is transferred to the ribose of terminal adenine nucleotide at the 3′-end of tRNA.

$$\text{Aminoacid} + \text{ATP} \rightarrow \text{Aminoacyl-AMP} + \text{PPi}$$

$$\text{Aminoacyl-AMP} + \text{tRNA} \rightarrow \text{Aminoacyl-tRNA} + \text{AMP}$$

Since bacterial and yeast synthetases occur as free soluble enzymes in cytoplasm under normal conditions, the organization of aaRS's in mammalian cells as a supramolecular assemblage may reveal the evolutionary pressure on the organization of protein biosynthetic machinery [11]. Assessment of the complex and free forms of synthetases indicated that the synthetases complex may provide stabilization against thermal or chemical activation [12] and probably against intracellular proteolysis [13] and to control the synthetase activity. This ubiquitous assembly consists of 11 polypeptide subunits of M_r ranging from 18 to 160 kDa which can be purified to homogeneity [14]. This complex comprises 9 aminoacyl-tRNA synthetases particular for their corresponding amino acid of class I and class II tRNA synthetases [15,16], that are monomers (IleRS, LeuRS, MetRS, GlnRS, ArgRS), and dimers (LysRS, AspRS), and a bifunctional polypeptide (GluProRS), which recently was found out to exist as fusion protein [17] and three auxiliary proteins of 18, 38 and 43 kDa [18,19].

The AspRS offers a comprehensive explanation of the diverse states that exists for an aminoacylation system of class II synthetases. The yeast Aspartyl-tRNA synthetase is a homodimeric enzyme, each chain consists of 557 amino acids, having a molecular weight of 125 kDa [20]. This dimer has a diamond shape with diagonal dimensions of 110 × 90 Å, and to a certain extent flat; it is only 40 Å thick at the local two-fold axis, but is bent along the main diagonal which results in a total thickness of about 60 Å. This dimer is the association of two extended subunits. Each subunit exhibits two distinct domains linked by a small hinge module [21,22]. The total buried surface or contact area of one tRNA molecule is 2500 Å [22].

The crystal structure of yeast Aspartyl-tRNA synthetase-$tRNA^{Asp}$ was revealed by Ruff et al. [21]. The binding of the small ligands, such as, ATP was elucidated in the presence of $tRNA^{Asp}$ in the same organism [23]. Important study of this enzyme has been embarked on by site-directed mutagenesis shared by crystallographic analysis, or its complex with $tRNA^{Asp}$.

Sufficient sequence similarity among the divergent species can be found throughout, apart from at the *N*-termini between mammalian AspRS and yeast AspRS. Deletion of the *N*-terminal extension up to 70 residues in yeast AspRS has inherently no effect on the synthetase activity, while deletion of 5 residues at the *C*-terminus reduces 80% aminoacylation activity [24]. Further biochemical experiments revealed that the first 70 residues are not required for the catalytic activity and they also hinder the crystal formation of this enzyme. In rat and human AspRSs, the *N*-terminal extension is also not necessary for catalytic activity [25,26].

The *N*-terminal domain (residues 68–204) is involved in the identification of anticodon of the $tRNA^{Asp}$. This domain contains a five-stranded β-barrel and is closed to the anticodon loop of the $tRNA^{Asp}$. These β-strands are numbered in the order of their sequences that are S1, S4, S5, S3 and S2. All these are antiparallel, except S5 and S3. An α-helix is located in between strands S3 and S4. As no noteworthy homology with other aaRSs could be identified, the *N*-terminal domain may be distinctive of the subgroup specificity. The hinge domain (residues 205–240), linking the *N*-terminal and *C*-terminal domains of the protein, is a distinctive globular module consisting four short helices.

The active site shaped by a six-stranded antiparallel β-sheet is the main characteristic of *C*-terminal domain (residues 241–557), partially closed by loops and helices at entrance, completed by a seventh parallel strand form the floor of the active site. This domain encloses three signature motifs of class II aaRSs that make contact with the acceptor stem of the $tRNA^{Asp}$ and carry out catalysis. Motif 1 is concerned to the dimer interface, motifs 1 and 2 contribute to the positioning of the acceptor stem by forming a catalytic cavity while motifs 2 and 3 hold the ATP [23,21]. The general order of organization is as follows.

Motif 1 covers residues 258–275. It has an α-helix which leads to distorted β-strand, S1 parallel over a few residues to the first strand S4 of the antiparallel β-sheet. Motif 2 comprises residues 315–348 and is formed by two first strands (S4 and S5) of the six-stranded antiparallel β-sheet. The variable loop of motif 2 (327–334) interacts with the acceptor stem of $tRNA^{Asp}$. Motif 3 starts from residues 517–549 contains strand S10 and helix H9 which is connected to the carboxyl terminus of the protein.

The flipping loop (residues 279–285), adopts an open conformation in the free yeast AspRS which can be visualized in PDB ID code: 1EOV. This region adopts a closed conformation in the presence of aspartic acid or aspartic acid-adenylate, or in the absence of $tRNA^{Asp}$.

Structural details are important for mammalian AspRS in order to elucidate the discriminative features against yeast AspRS. These highlighted features may be used for designing a drug against yeast infections e.g., candidiases, since only one 3D crystal structure is available (1ASZ) for fungi so far which could be used as a target in drug therapy. Fortunately, the sequence similarity between AspRS of *Candida albicans*, the causative agent of candidiases, and of yeast AspRS is 68%. That is why, it is very usual that the conformations of active site residues of 1ASZ has the same as that of *C. albicans* AspRS and 1ASZ can be used to build two homologs to reveal the structural diversity in drug design process against candidiases.

While frequently recognized as homology modeling and more recently, comparative modeling, the in silico approach to predict an unknown structure using a close structural homolog belonging to the same protein family as the target sequence, is better explained as template-based modeling/structure prediction (TBM). This is now the well-established terminology in the protein structure prediction community. In this study, we have predicted the 3D structures of *Homo sapiens* aspartyl-tRNA synthetase (*hs*AspRS) and *Mus musculus* aspartyl-tRNA synthetase (*mm*AspRS). In order to understand the mechanism of ATP-induced conformational changes, molecular dynamics (MD) simulation was carried out. Analysis at different time frames revealed the persistence of motions in flipping loop region (227–233) in both mammalian species during ATP-binding.

2. Methods

2.1. Computational resources

Computational studies were carried on Intel® Xeon Quad core 2.66 GHz equipped with LINUX environment running under openSUSE10.2. Multiple sequence alignment was done by ClustalW [27]. The homology modeling was performed by SYBYL® (version 7.3, TRIPOS, St. Louis, MO). Validation of protein structures was determined by Procheck [28], Verify3D [29] and Errat plots [30]. MD simulation and trajectory analysis were achieved by AMBER 9.0 [31] program package. Visualization was carried out by VMD and CHIMERA software.

2.1.1. Template-based structure prediction of mammalian AspRSs

The amino acid sequences of *hs*AspRS (acc. no.: NP_001340.2) and *mm*AspRS (acc. no.: NP_803228.1) were retrieved from NCBI protein sequence database. The target sequences were 95% identical to each other with only 10 residues differing between

the two. The procedure to predict template, was started with searching an appropriate template for target sequences from a database by using a threading method. The sequences were submitted to NCBI PSI-BLAST [32], a database similarity search tool, which searches Brookhaven PDB database [33] to detect distant evolutionary relationships by identifying the closest homologous structure. A template structure for a given target sequence is recognized by taking into account the significance of the score which indicates fitness of the target to the template, e.g., the E-value, or the sequence identity are the two most common criteria. The first two closest structural homologs for *hs*AspRS and *mm*AspRS were the AspRS from *Saccharomyces cerevisiae* (yeast) with its two different crystal structures. Both structures were 57% identical to the target sequences in which one is complexed with its bound $tRNA^{Asp}$-ATP substrates (1ASZ) while the other is present in its apo from. Other structural homologs are AspRS from *Sulfolobus tokodaii* (36% identical), *Pyrococcus kodakaraensis* (35% identical) and *T. thermophillus* (35% identical). Since one of our tasks is to study the interactions between ATP bound to its AspRS, we considered the crystal structure of yeast AspRS-$tRNA^{Asp}$-ATP complex (1ASZ) as our template which shows the similar occupancy pattern. After obtaining the template structure, the multiple sequence alignment by ClustalW of template and target sequences was also carried out to identify the conserved and variable regions in the target sequences.

In principle, the probability of error in template recognition step is very high. Wrong templates with a different fold but in the correct fold class are often recognized in threading. To ensure the sensitivity and accuracy of our selected template, FUGUE (Find Homologs of Uncharacterized Gene Products Using Environment—specific substitution tables) program [34] was also used to find distant homologs by sequence–structure comparison between a target sequence and its template structure. FUGUE utilizes the curated database of structure-based alignments for homologous protein families, HOMSTRAD (Homologous Structure Alignment Database—release date December 8, 2006) [35]. HOMSTRAD has an all known protein structure which is clustered into homologous families (i.e., common ancestry), and the sequences of representative members of each family are aligned on the basis of their 3D structures using the programs MNYFIT, STAMP and COMPARER. From each FUGUE search, the top HOMSTRAD family with the rank of recommended cut-off score (Z-score > 6.0) was chosen for homology modeling. FUGUE generates different search results in contrast to NCBI-BLAST but the closest homolog remains the same. This method gives the same template (1ASZ) with the Z-score of 35.72 while the second template candidate was 1C0A with the Z-score of 24.29 (Table 1; see also Fig. S1A and B). The sequence–structure alignment that was generated by FUGUE is represented by an annotation program, JOY [36]. FUGUE and HOMSTRAD are supplied by Tripos and the sequence/structure alignment produced by FUGUE may be used in ORCHESTRAR (http:\\www.tripos.com) to generate a comparative model for a target sequence.

Table 1
Data for the closest homolog of AspRS from *Homo sapiens* (NP_001340.2) and *Mus musculus* (NP_803228.1) by NCBI-BLAST search.

PBD ID:	Protein:	% Identity[a]
1ASZ	Asp-tRNA synthetase complexed with $tRNA^{Asp}$-ATP *Saccharomyces cerevisiae*	57%
1EOV	Free Asp-tRNA synthetase *Saccharomyces cerevisiae*	57%
1N9W	Asp-tRNA synthetase from *Sulfolobus tokodaii*	36%
1B8A	Asp-tRNA synthetase *Pyrococcus kodakaraensis*	35%
1WYD	Non-discriminating and archaeal-type Asp-tRNA synthetase *Thermus thermophilus*	35% **34%**[a]

[a]*Note:* All % identity values are same for both species except the one in bold indicative for *Mus musculus*.

The Biopolymer module of SYBYL has ORCHESTRAR suite of applications which is specifically designed for homology modeling. ORCHESTRAR uses a combination of knowledge-based and ab initio approaches. Within ORCHESTRAR, BATON [37] program takes input data from FUGUE to realign the target sequence to the new structure alignment. The BATON structural alignment served as input to the CHORAL [38] program which creates all-atom protein backbone (structurally conserved regions—SCRs) from Cα trace of template using a least squares superposition. Structurally variable regions (loops) were obtained from the database using Biopolymer's loop search capabilities. Three loop search methods were used: loops of length 1–3 was searched by BRIDGE program that eventually close such small gaps between the conserved regions generated by CHORAL, it is also known as homology search method; D (fragment database—FREAD)-method created from HOMSTRA is appropriate for 3–12 residue loops and A (ab initio database—PETRA)-method of 10^8 secondary structure fragments is appropriate for 3–8 residues loops. Loop solutions found by all these three methods were ranked on the basis of their energy (E_k) when they are to be placed in the model. Bridge uses the TUNER [39] algorithm to incorporate loop into the model. The side chains were modeled by borrowing the side chains conformations, or dihedral information from templates, while avoiding steric clashes, but also keeping the existing side chains. The program ANDANTE [40] is involved in adding side chains. Hydrogen atoms were added in the modeled proteins. Both protein models were minimized by the AMBER7 force field99, a distance-dependent dielectric constant of 1.0R and a non-bonded atom cut-off of 8.0 Å by using the Powell method. This method was used to relieve steric clashes and overlaps of side chains. After minimization, the hydrogen atoms were removed. The stereochemical properties of minimized structures were evaluated by Ramacandran plot calculations computed with Procheck program. Verify3D further confirms the environment of side chains. Besides this, Errat plot examines the misfolded regions. RMSD fit was calculated by alignment of structural homology feature in SYBYL.

2.1.2. Molecular dynamics (MD) simulations of mammalian AspRSs

MD simulations were conducted for modeled systems in explicit solvent using the AMBER 9.0 package. To obtain the neutrality of the system, ten and nine Na^+ ions were added in *hs*AspRS and *mm*AspRS, respectively. Now, both models were solvated by 24,840 and 24,561 water molecules, respectively, in a rectangular box around the solute unit. The sizes of the cells were 105.118 Å × 97.951 Å × 96.459 Å consisted of 32,207 atoms, and 105.118 Å × 97.409 Å × 95.884 Å consisted of 32,117 atoms for *hs*AspRS and *mm*AspRS, respectively. Xleap interface of Leap was used to create the rectangular solvate box. The solvated protein systems were subjected to thorough energy minimizations before MD simulation. First, restrain minimization of water molecules was done while holding the solute fixed (500 steps using the steepest descent algorithm followed by 500 steps of conjugate gradient minimizations of the whole system). This was done to remove the steric conflicts between a protein and water molecules and also to relax the system. An unrestrained minimization was then carried out by using the same procedure as for restrained minimization. Bond lengths involving hydrogen were constrained with SHAKE [41] algorithm with harmonic restraints of 25 kcal/mol Å. The simulated system was first subjected to a gradual temperature increase from 0 to 300 K over 60 ps, and then equilibrated for 100 ps at 300 K, followed by production runs of 5 ns. Constant temperature (298 K) and constant pressure (1 atm) were controlled by the Berendsen coupling algorithm with a time constant for heat-bath coupling of 0.2 ps. The dielectric constant was set to 1.0. The cut-off distance was 9.0 Å. Long-range electrostatic calculations were carried out by particle mesh Ewald

method [42]. The resulting trajectories were analyzed by PTRAJ module of the AMBER package.

3. Results

3.1. Multiple sequence alignment

For multiple sequence alignment, only those sequences of AspRSs are selected that have already experimentally determined 3D crystal structures. This strategy helped us to trace the exact template at the sequence level. Archaeal-type *Thermus thermophilus* (1N9W), *S. tokodaii* (1WYD), *S. cerevisiae* (1ASZ), *Thermococcus kodakarensis* (1B8A), *E. coli* (1C0A) and *T. thermophilus* (1EFW) have resolved 3D crystal structures. All these sequence are aligned with the sequences of human and mouse (targets), and rat AspRSs and revealed overall an excellent sequence similarity (Figs. S2 and 1). In both target sequences and their respective modeled structures, amino acid residues are numbered according to their primary sequences.

Both target sequences are 57% identical to the template sequence (Fig. S3). Excellent sequence similarity is found in all three the three motif regions of AspRS. Besides this, sequences are also more similar in the remaining *C*-terminal domain as compared to other regions. Since AspRS belongs to class II synthetases, the three characteristic motif regions are illustrated at this point on the basis of their evolutionary substitution.

3.1.1. Motif 1

Tyr261, Ala263, Thr264, Lys266, Thr268, Val270, His271 and Leu275 in the motif 1 of template are evolutionary changed in both target species. Val270, His271 and Leu275 are replaced by Ile218, Gln219 and Ile223, respectively. Only one residue is semi-conserved, i.e., Thr268 which is replaced by Val216.

3.1.2. Motif 2

The motif 2 has six conserved substitutions, i.e., Arg316, Tyr318, Asn328, Met335, Met343 and Gln348 are replaced by Lys264, Phe266, Asp276, Leu283, Ile291 and Asn296, respectively in both target sequences. One semi-conserved substitution is observed in which Thr339 is replaced by Val287. In yeast, three methionine residues are present at positions 335, 343 and 345 in motif 2 while only one methionine residue is present at position 293 in *hs*AspRS and *mm*AspRS in the same motif, it indicates less chances to form a thioester interaction. Thus, it can be predicted that methionine at positions 335 and 343 could be used as drug target against yeast infections, as these are replaced by longer hydrocarbon chain amino acid residues in mammals.

As it is mentioned earlier that motif 2 has dual role, one is the positioning of acceptor stem of $tRNA^{Asp}$ by forming a catalytic cavity and second is to hold the ATP. By designing a competitive drug that impede with the natural Met335 and ATP interactions by forming a thioester group would be a novel strategy to eliminate the yeast infections. The substitution of Asn328 by Asp276 in both targets could make motif 2 less reactive towards $tRNA^{Asp}$, as the amide group of Asn is replaced by the carboxylic group of aspartic acid. This substitution may decrease the activity of AspRS in both species as the carboxylic group of Asp has lesser chance to form hydrogen bond with the negatively charged phosphate group.

3.1.3. Motif 3

The motif 3 has eight conserved substitutions. Two most striking conserved substitutions are Arg544 and Ala545 which are replaced by Gln488 and Thr489. The inclusion of amide and hydroxyl group side chains may cause some hindrance towards attractive effects to ATP, especially negatively charged hydroxyl group might repel the negatively charged phosphate group of ATP which comes in its vicinity. This may have some decreasing effects on the overall enzyme kinetics of mammalian AspRS. A nucleotide triphosphate binding domain (residues 468–482) is also conserved in both mammalian AspRS.

3.2. 3D Structural details of Aspartyl-tRNA synthetases

The structures of AspRSs produced by SYBYL revealed an excellent agreement between the predicted secondary structure of AspRSs and the experimentally determined 3D structure of AspRS from yeast. The overall topology and secondary structural elements are quite conserved in the class II aspartate sub-family, i.e., the *N*-terminal domain and the *C*-terminal domain. The *N*-terminal domain (residues 22–152) is composed of a five-stranded β barrel sheet while the *C*-terminal (residues 189–501) comprises six-stranded antiparallel β-sheets. These domains are stabilized by the combined effects of non-covalent interactions (mainly, hydrogen bonds and salt bridges). This domain comes close to the anticodon loop of the $tRNA^{Asp}$ on the side of the major groove. The hinge region (residues 153–188) is composed of a turn and a α-helix (Fig. 2A–C). Cα-based superimposition RMSD values between the template, and *hs*AspRS and *mm*AspRS are 0.62 and 0.80 Å, respectively (Fig. 3).

Motif 1 spans from residues Phe206-Ile223. Motif 2 comprises of residues Glu263-Asn296 and formed by the two first strands (S4 and S5) of the six-stranded antiparallel β-sheet. The variable loop of motif 2 (Glu275-His282) comes close to the $tRNA^{Asp}$. Motif 3 starts from residue Phe461 and ends at Phe492 containing strand

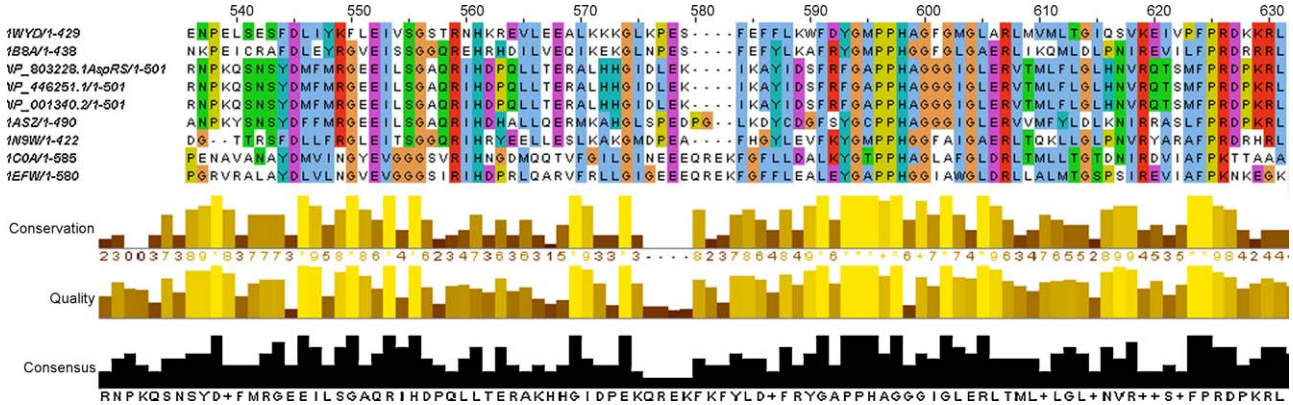

Fig. 1. Multiple sequence alignment of AspRSs generated by the ClustalW algorithm; two target sequences, *Homo sapiens* (NP_001340.2) and *Mus musculus* (NP_803228.1) AspRSs, *Rattus norvegicus*' sequence of AspRS (NP_446251.1) and six AspRSs sequences that have resolved 3D structures. The multiple sequence alignment image was generated by using JalView showing the part of *C*-terminal domain of AspRSs containing catalytic amino acid residues (see also Figs. S2 and S3).

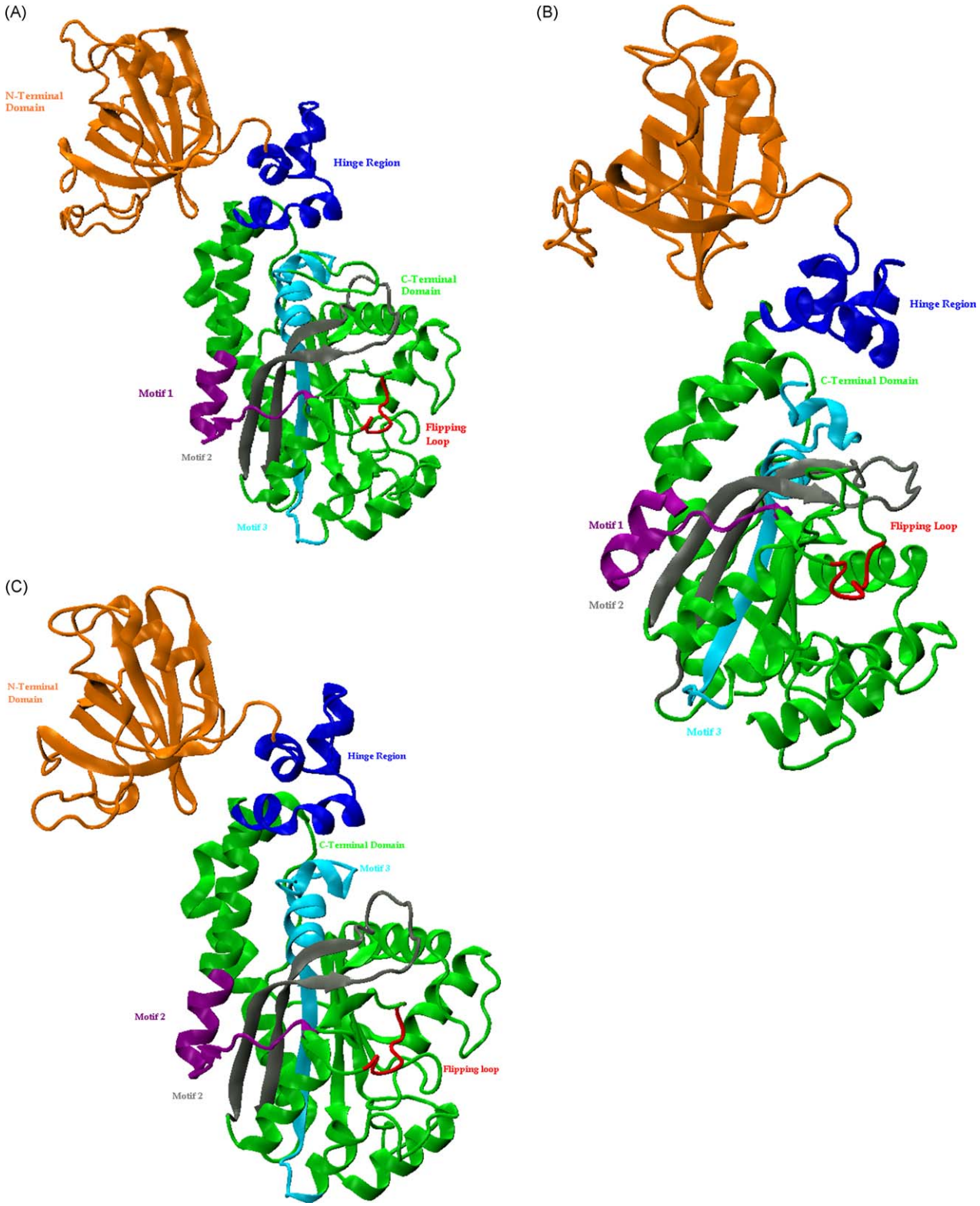

Fig. 2. Cartoon representation of yeast AspRS with removal of bound substrates; showing the domain architecture. The *N*-terminal domain (orange, residues 68–204), hinge region (blue, 205–240) and the *C*-terminal domain (green, 241–557). Within the *C*-terminal domain, motif 1 (purple, 258–275), motif 2 (grey, 315–348) and motif 3 (cyan, 517–548). Flipping loop (red, 279–285). The β-strands are represented as arrows and the α helices as rods. Same coloring scheme is used for (B) *Homo sapiens* AspRS and (C) *Mus musculus* AspRS. The span of all three domains is as follows: *N*-terminal domain (22–152), hinge region (153–188), and the *C*-terminal domain (189–501). Within the *C*-terminal domain, motif 1 (206–223), motif 2 (263–296) and motif 3 (461–492) are present. Flipping loop (227–233) is also colored in red.

S10 and helix H9, connected to the carboxyl terminus of the protein. However, the RMSD values between motifs 1, 2 and 3 homolog models to its crystal structure are 0.25, 0.24 and 0.20 Å, respectively (Fig. 4A–C).

The class II specific active site residues of AspRS which includes Phe286, Arg475, Leu283, Arg273, Ile425 and Gly472 have different topological orientations. Phe286, Ile425 and Gly472 have extended β conformations. This conformation helps these conserved amino

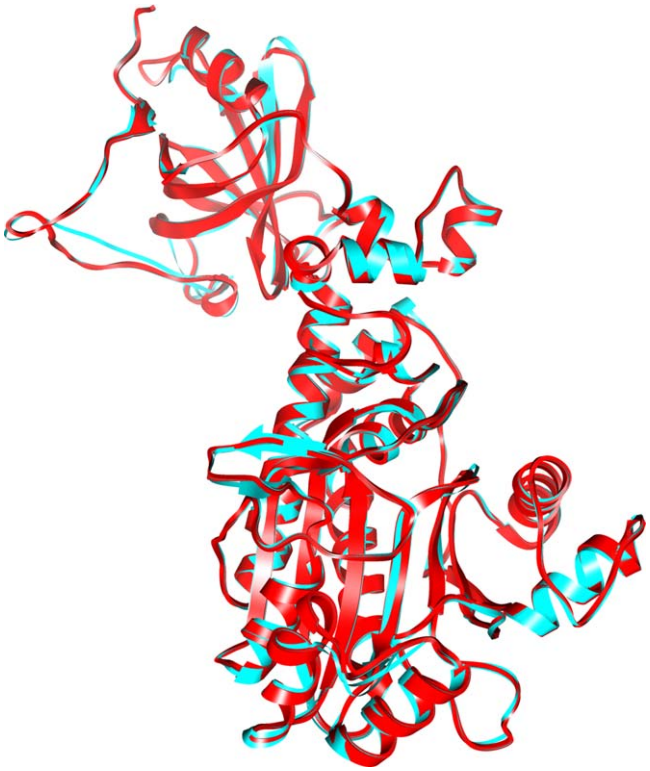

Fig. 3. Superimposition of Cα atoms of yeast (red), human and mouse (cyan) AspRSs. The homolog structures aligned with the crystal structure of yeast AspRS showed an average RMSD value of 0.71 Å.

acid residues to form interplay of hydrogen bondings and provide appropriate orientation that invites ATP to come and bind. Arg475 has a α conformation due to which it may slightly disrupt the overall α-helix conformation. The adenine base of ATP has same electrostatic interactions with Phe286 on one plane and Arg475 on the other side. With the help of two hydrogen bonds that is the main chain carbonyl and nitrogen atoms of Leu283, the adenine base is linked to the enzyme. The γ-phosphate of ATP also makes contact with the Arg475. The second main affix position of ATP is the α-phosphate that stabilizes this complex by forming a salt bridge with Arg273, a conserved residue of class II synthetases residue. The hydroxyl groups of the ribose forms hydrogen bonds with the main chain atoms of the Ile425 and Gly472 (Fig. 5).

The exterior of the antiparallel β-sheet of modeled proteins also contains the same amino acid residues as template which makes contact with ATP in the template, involving large numbers of Gly residues at 288, 428, 468, 470 and 472 positions. Gly472, close to the ribose of ATP, is almost conserved in all sequences of class II synthetases. Phe286 is very crucial to its position because if any other side chain is present instead of Phe conjugated system, the stacking of the adenine moiety of ATP to that side chain is absolutely unfavorable. The entrance of the active site is enclosed by essentially conserved Gly446, 230 and 281 residues.

Some other important amino acid residues of both targets AspRSs that play an important role by interacting with ATP are Gln251, Arg431, Glu292, Lys254, Gln255 and Asp290. These are all stringently conserved in both targets and the mutations of any one of them may leave the enzyme inactive.

3.2.1. Flipping loop

Loops invariably lie on the surface of proteins and thus often participate in interaction between protein and other molecules. The presence of proline, a completely nonpolar amino acid residue, in our template structure causes a constraint in the flipping loop. It is unusual because its side chain curls back to the main chain and

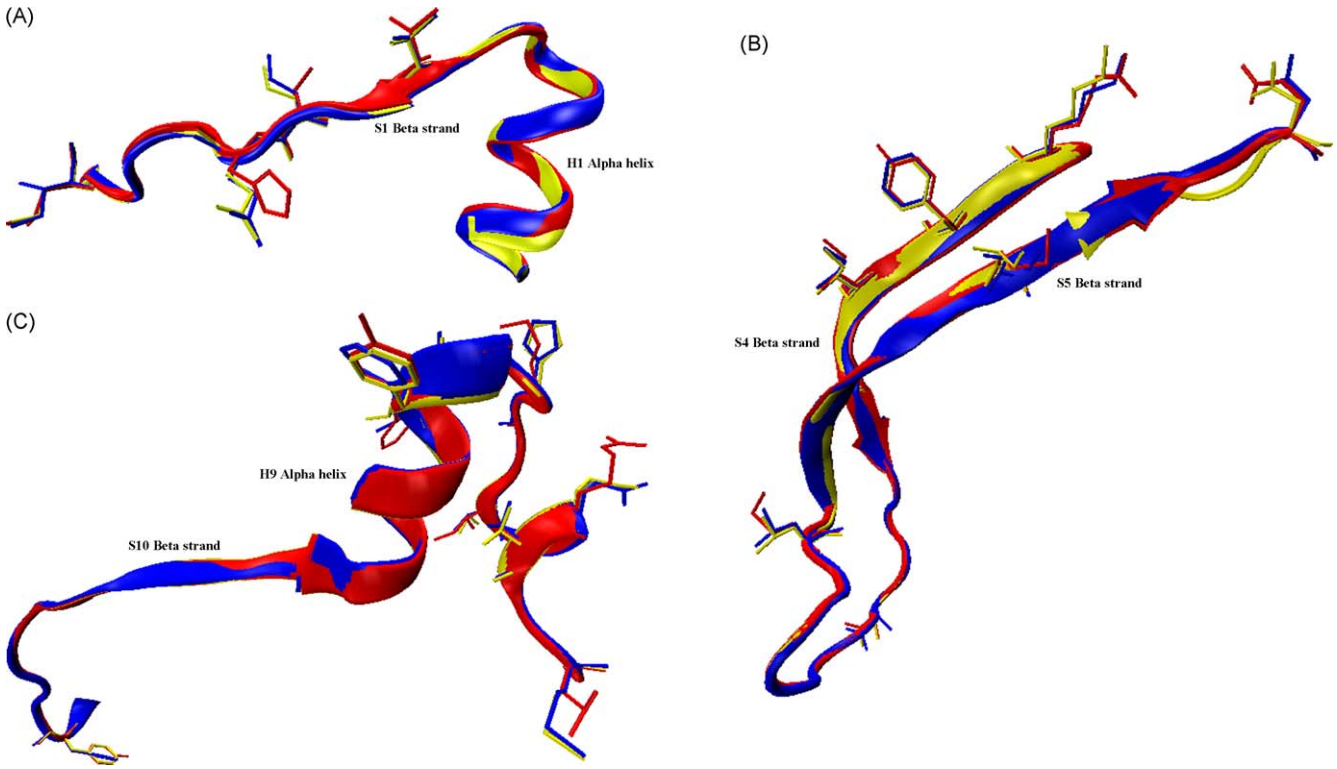

Fig. 4. Cα-based superimposed template and target models. Red, yellow and blue ribbon represented yeast, *Homo sapiens* and *Mus musculus* AspRSs, respectively. Conserved substitution for (A) motif 1, (B) motif 2, and (C) motif 3 are represented by CPK model.

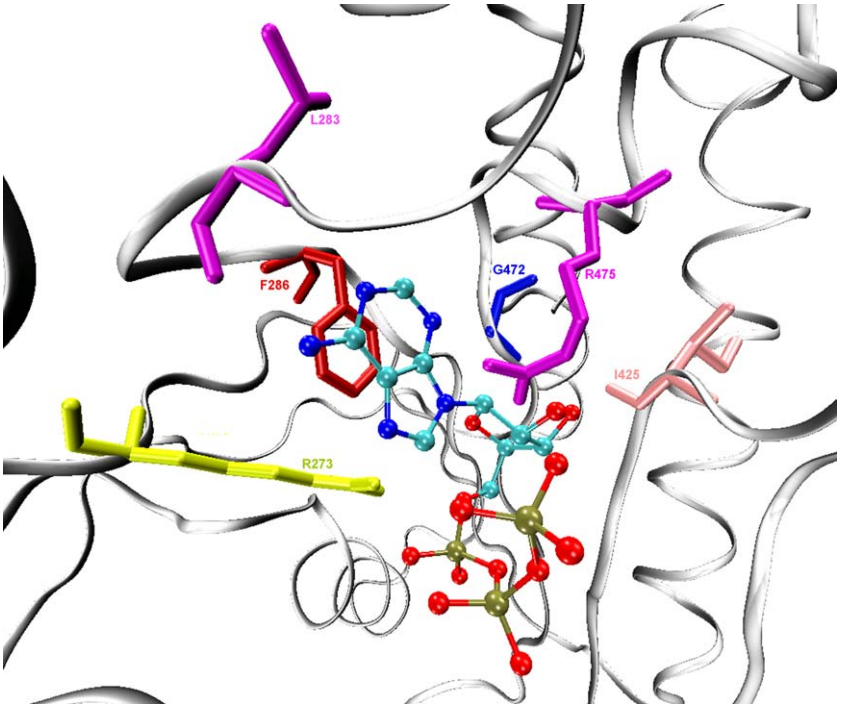

Fig. 5. Active site of the mammalian AspRSs showing the class II specific key amino acid residues with its bound substrate ATP. The ATP is shown in ball and stick model. The amino acid residues (represented as their color IDs) are involved in the substrate positioning (see text of hydrogen bonding pattern).

seizes it. This produces an almost rigid side chain, where only different ring puckers are allowed, that is, C2′ endo conformation of ribose moiety in case of ATP. Furthermore, it fixes the dihedral angle between Cα and the peptide nitrogen to a small range of about ±20°. Thus, a proline residue introduces a hindrance in the flipping loop movement. In addition, the nitrogen atom has no substituent hydrogen to participate in hydrogen bonding with other residues. The proline residue at 279 is replaced by alanine 227 in both mammalian targets. This semiconserve substitution leads to more free movement to adopt open and close conformations with regard to the substrates. Superimposition of Cα and main chain atom of these flipping loops show a deviation of 0.27 Å. On comparing, this loop structure to the free AspRS of yeast PDB ID: IEOV, the deviation reaches to 3.84 Å for *hs*AspRS and 4.02 Å for *mm*AspRS (Fig. 6).

3.3. Protein models validity

The geometrical and structural consistencies of both the modeled proteins were evaluated by different approaches. The structural validation of homologs was carried out by the Procheck program. The ϕ and φ distributions of Ramachandran plots are summarized in Table 2. This analysis showed that both the modeled structures had only slight violations of stereochemistry for 1.4% of the residues, which are only 6 in number, present in the disallowed region of the Ramachandran plot. Gln155 and Glu229 are also lying in the disallowed region in target species as Gln133 and Leu207 in the template structure, respectively. Reported Ramachandran plot analysis of template with $tRNA^{Asp}$ and ATP, with no bound substrates (1EOV) along with the two homologs in this study shows vast differences indicating that the protein molecule undergoes a large conformational change between two different states, that is, the bound and the unbound form. Overall, the Ramachandran plot of both homologs also has the same distribution of sterically allowed main chain atoms as the template had.

In addition, two more protein evaluation programs were utilized to check the stereochemistry of our 3D models. Verify3D scores the compatibility between amino acid sequence and the environment of the amino acid side chains in the model. It assesses the environment of side chain based on the solvent accessibility and fraction of the side chain covered by polar atoms. Errat assesses the arrangement of different types of atoms with respect to one another in the protein models. Errat is a sensitive technique, which is good for identifying incorrectly folded regions in preliminary protein models. However, it is limited as it does not have any parameterization regarding surface polarity—shown to be an important factor in determining the validity of protein structure. Hence, any misfolded proteins displaying surface areas and volumes comparable to their native forms, will not be identified by Errat as misfolded. In verify 3D plot, the amino acid residues which have scores lower than zero are not well enough suited according to their environment, especially a cluster of eighteen amino acid residues (39–56) located at the *N*-terminal domain in *hs*AspRS and *mm*AspRS, respectively. Although, these amino acid residues are not involved in any type of interactions with the substrate. Errat plot gives overall quality factor of 93.89% and 89.31% for *hs*AspRS and *mm*AspRS, respectively (Fig. 7A and B). However, a few amino acid residues of *C*-terminal domain of both modeled proteins are not folded correctly. However, error rate remains negligible for those regions that account for the binding and catalytic activity.

The above-mentioned protocol to evaluate the homolog structures is the trivial one. To further validate the accuracy of the homolog structures and the method used to generate the 3D models of target sequences, a novel approach is used. In this strategy, the template sequence is considered as a target while the homolog structures are supposed as template structures. The same outline is used to elucidate the re-coordinates of template. The generated template by the ORCHESTRAR interface of SYBYL has almost the same Procheck value as the real PDB coordinates of 1ASZ are (Table 3). The RMSD values between the PDB coordinates of 1ASZ and the one generated by ORCHESTRAR are 1.73 Å (template–*hs*AspRS) and 1.71 Å (template–*mm*AspRS) that not only shows the reliability of our target homolog but also authenticates the procedure used to build the homologs.

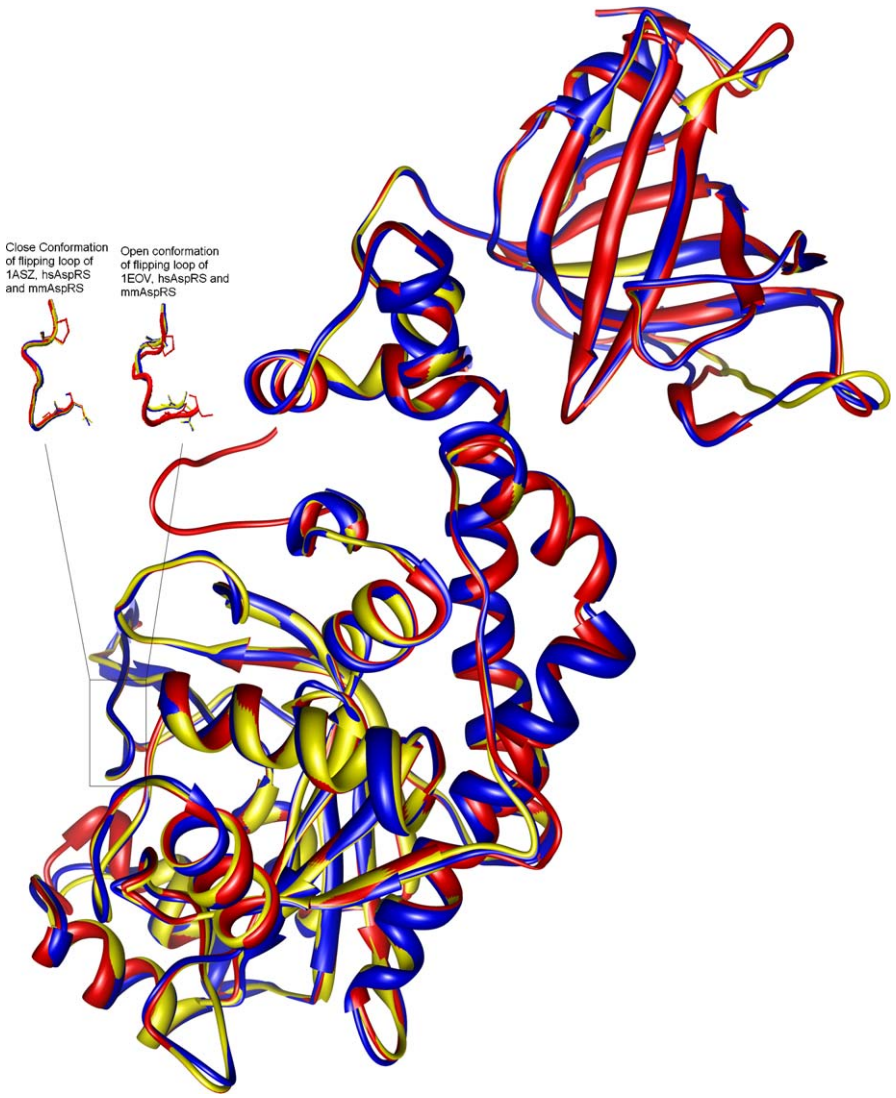

Fig. 6. Flipping loop movement deviation between the 1ASZ (closed state—red ribbon) and 1EOV (open state—red ribbon: see enlarged view) to the predicted 3D models of AspRSs (closed state—yellow; *Homo sapiens*, blue; *Mus musculus*). The amino acid residues represented by CPK model show conserved substitution. 1EOV has been omitted for the sake of clarity from superimposed image.

3.4. Hydrogen bond pattern

Hydrogen bond is a key force to sustain the binding between the ligand and a receptor molecule. Most of these hydrogen bonds are very strong, some even with an occurrence of more than 90% (as in our case). These bonds make the complex steadier than the ligand-free conformation. Table 4 shows the H-bonds between the ATP and the *hs*AspRSs. The Cartesian coordinates of ATP are extracted from the template.

Table 2
Ramachandran plot calculations on 3D predicted models of *Homo sapiens* AspRS and *Mus musculus* AspRS.

Homo sapiens AspRS	
% of residues in generously allowed regions	1.2
% of residue in disallowed regions	1.4
Number of non-glycine and non-proline residues	424
Mus musculus AspRS	
% of residues in most favored regions	85.2
% of residues in most additional allowed regions	12.9
% of residues in generously allowed regions	0.5
% of residues in disallowed regions	1.4
Number of non-glycine and non-proline residues	425

3.5. Molecular dynamics simulation study

To gain insight into the stability and dynamic properties of the homolog structures, explicit solvent MD simulation is performed. The total energy of whole system and root mean square deviation (RMSD) from the starting structure are essential to determine the sustainability and convergence of MD simulation. Fig. 8 shows RMSD of Cα backbone with respect to the initial structure. The graph clearly shows that the RMSD reached below approximately 3 Å in case of *hs*AspRS and about 3.5 Å in case of *mm*AspRS. These results suggested that a relatively stable conformation of both the proteins is achieved through the MD simulation.

3.5.1. Root mean square fluctuation analysis

In order to locate the flexible regions, the root mean square fluctuation (RMSF) was examined for the Cα atoms of each of the residue representing the average displacement of these atoms. As shown in Fig. 9, it is cleared that the $RMSF_{avg}$ of *hs*AspRS and *mm*AspRS are 3.2 and 3.5 Å, respectively. These values are somewhat larger than those obtained from the X-ray crystal structure refinement of its template when $tRNA^{Asp}$ is bound to it. The reported $RMSF_{avg}$ value of yeast AspRS in this state is 1.2 Å in

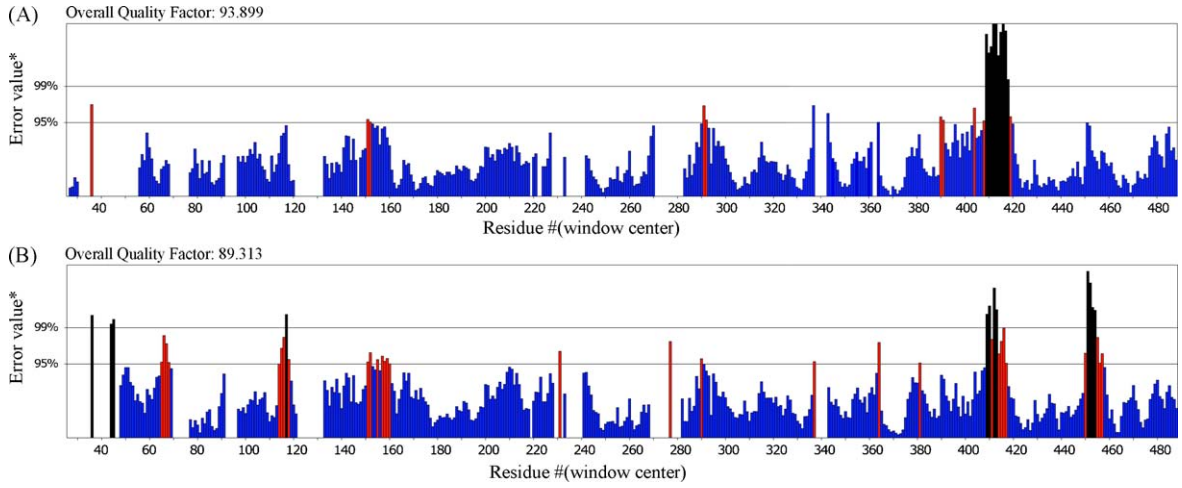

Fig. 7. Errat plot for (A) *Homo sapiens* AspRS and (B) *Mus musculus* AspRS. Black bars show the misfolded region located distantly from the active site, Red bars demonstrate the error region between 95 and 99%, blue bars indicate the region having less error rate for protein folding.

Table 3
Cross-validation of protocol used by ORCHESTRAR; Ramachandran plot.

Lot statistics	1ASZ (chains A and B)		1ASZ (chain A)		ORCHESTRAR build structures of 1ASZ			
	No. of residues	Percentage	No. of residues	Percentage	Template–homolog structure of *hs*AspRS		Template–homolog structure of *mm*AspRS	
					No. of residues	Percentage	No. of residues	Percentage
Residues in most favored regions	666	76.6	326	74.9	343	79.0	344	79.1
Residues in most allowed regions	176	20.2	92	21.1	79	18.2	77	17.7
Resides in most generously allowed regions	13	1.5	11	2.5	2	0.5	4	0.9
Residues in disallowed regions	15	1.7	6	1.4	10	2.3	10	2.3
Total number of non-glycine and non-proline residues	870	–	435	–	434	–	435	–
Total no. of residues	980	–	490	–	489	–	490	–

Analysis of template–1ASZ as re-coordinate by ORCHESTRAR.

the catalytic domain, 2.6 Å in the hinge region, and 1.7 Å (monomer A) and 2.1 Å (monomer B) in the anticodon binding domain [43]. Our $RMSF_{avg}$ values for all these regions are 4.0, 2.5 and 2.0 Å for *hs*AspRS and 4.3, 2.4 and 3.0 Å for *mm*AspRS, respectively. The amino acid residues of *N*-terminal domain of *mm*AspRS have more fluctuation as compared to the *hm*AspRS, since the *N*-terminal domain has to rotate in order to bind the anticodon bonding domain of $tRNA^{Asp}$ [43]. Hinge region of both systems remain well stabilized as compared to both domains.

The fluctuation in the C-terminal domain to accommodate substrates is the most interesting aspect of AspRS. The flipping loop (227–233), which entraps aspartyl-adenylate complex in the active site of AspRS, flips back in its open position upon binding of the cognate $tRNA^{Asp}$ to allow the nucleophilic attack of the activated α-COOH by $tRNA^{Asp}$. In both systems, the flipping loop adopts a closed conformation as in the binary AspRS-aspartyl-adenylate complex, because the resultant homologs structures were modeled from the complex structures; however, in apo AspRS, the loop adopts an open conformation allowing the aspartic acid to reach its binding site. As a function of time, the conformation of flipping loop does not change so much as expected. Differences in the flipping loop residues occur but they are negligible (Fig. 10). The opening of the active site through a lift of the flipping loop takes place when $tRNA^{Asp}$ is bound to the enzyme. The loop is then responsible for the recognition and correct positioning of the terminal adenine. But in our case, the flipping loop goes more towards the closed form as both simulated systems demonstrate the closing of flipping in the absence of $tRNA^{Asp}$, further confirms the fact that flipping loop acts as a door that only opens when the right tRNA is bound to the synthetase. The closed conformation of flipping loop remained stable over the entire simulation which seems to be more energetically preferred.

Table 4
ATP–*Homo sapiens* AspRS interatomic distances (hydrogen bonding (Å)).

ATP contacts	Protein contacts	Donor–acceptor distances (predicted this work)		Predicted (Cavarelli and co-workers [23])	Accuracy (%)	
		Homo sapiens AspRS	*Mus musculus* AspRS		*Homo sapiens* AspRS	*Mus musculus* AspRS
ATP OP2α	Arg273 Nη1	2.64	2.21	2.64	100	84
ATP N1	Leu283 NH	3.39	3.38	3.34	98.5	100
ATP N6	Leu283 CO	2.72	2.52	2.83	96	89
ATP O3′	Ile425 CO	2.74	1.82	2.73	100	66.6
ATP O3′	Gly472 NH	3.68	3.65	3.80	97	96.0
ATP O2′	Arg475 N	2.55	2.30	2.77	92	83.0

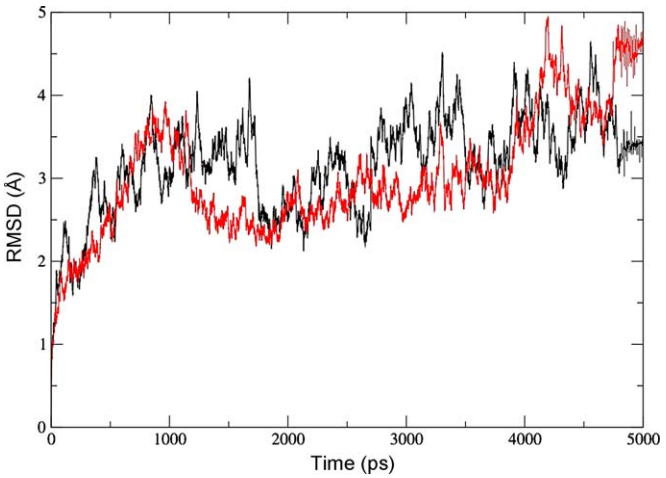

Fig. 8. Root mean square deviation (RMSD) of Cα backbone as a function of time for the two systems (black line; *Homo sapiens* AspRS, and red line; *Mus musculus* AspRS).

The region ahead of motif 2 also exhibited high fluctuation in the simulation. This loop may be involved in forming a tunnel entrance for the substrates in front of the active site.

3.5.2. Inter-domain motions in hsAspRS mmAspRS

Inter-domain motions in any protein can easily be determined by measuring its dihedral angles. Since upon the binding of $tRNA^{Asp}$, the *N*-terminal domain of yeast AspRS, rotates [43], the MD analysis reveals this motion in an anti-clockwise direction. As shown in Fig. 11, the black line demonstrates the motion of *N*-terminal domain of *hs*AspRS with respect to the hinge domain. All the dihedral angles of this domain are in the negative range, that is, 0 to −180°. The red line corresponds to the *C*-terminal movement of the same protein, indicating the clockwise direction of this portion. Mostly all of its dihedral angles are in the positive range. The same pattern was observed for *mm*AspRS.

The hinge domain that lies between these terminal domains plays a pivotal role in the overall movement of these proteins. Furthermore, future research is needed to discover those amino acid residues that restrict the movement of these domains which in turn accounts for the loss of activity of these enzymes.

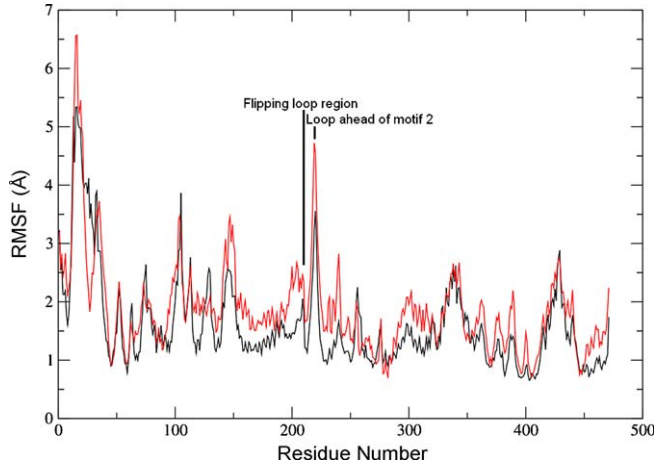

Fig. 9. The calculated root mean square fluctuations of Cα backbone atom (RMSF) of both the simulated systems (black line; *Homo sapiens* AspRS, and red line; *Mus musculus* AspRS). Numbering of amino acid residues starts from 1 (AMBER numbering) rather than from 22 for the first amino acid (Glu).

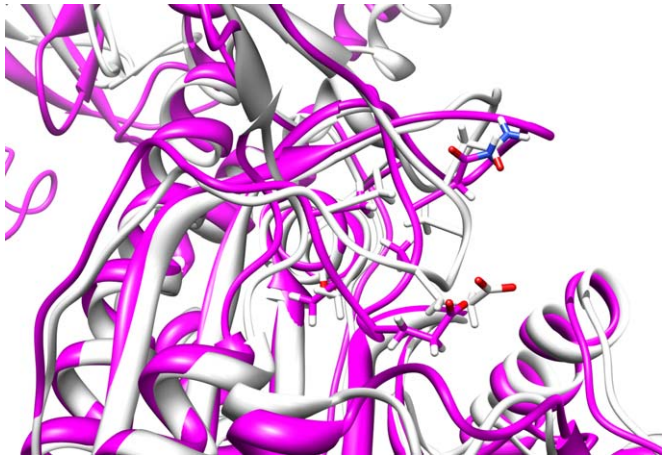

Fig. 10. Superimposition of starting structure (white ribbon) and the final structure (purple ribbon); the flipping loop is shown as stick. The last trajectory shows that the loop is more closed than the starting trajectory indicating the completion of first aminoacylation reaction in the absence of $tRNA^{Asp}$. After the formation of aspartyl-adenylate complex, loop flips back to accommodate $tRNA^{Asp}$ (not shown in this simulation).

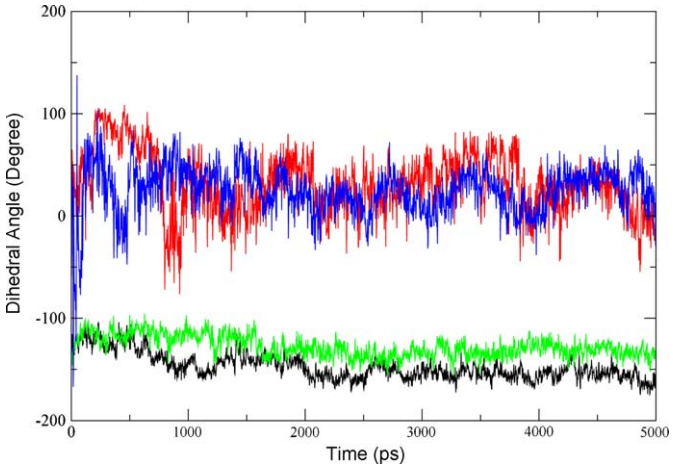

Fig. 11. Hinge twisting motion analysis: dihedral angle for *Homo sapiens* AspRS and *Mus musculus* AspRS with respect to the first snapshot as a function of time for (black) *N*-terminal domain of *Homo sapiens* AspRS; (red) *C*-terminal domain of *Homo sapiens* AspRS; (green) *N*-terminal domain of *Mus musculus* AspRS; (blue) *C*-terminal domain of *Mus musculus* AspRS.

4. Conclusion

This study has presented two homology models of AspRSs of mammalian origin, *H. sapiens* AspRS and *M. musculus* AspRS, in order to provide reliable models with which to design new inhibitors and to investigate the role of some conserved residues between the yeast and human proteins. Due to considerably high sequence identity between the mammalian proteins and the yeast protein, it is necessary to know the distinguishing features of both proteins related to two different species so that one could utilize these features to design a drug against different yeast diseases (Candidiases). This study has much importance because only one crystal structure is available for fungi kingdom (yeast AspRS: 1ASZ) which was used indirectly in this study to elucidate the structural discrepancies between the two species. The model does have acceptable structural profiles for the three programs of structural analysis applied. A nontrivial method that determines the accuracy of the software used and the resultant models from them was also successful. In addition, the molecular dynamics simulation further

improved the general structure of models and stable RMS deviations throughout the simulation were also obtained.

Multiple sequence alignment analysis clearly demonstrates that the sequence identity plays a major role in selecting an appropriate template which in turn effects overall structure prediction. A single template model does provide Cα traceability well enough if we had the sequence identity much above 50%. In our case, the sequence identity is 57%, so the predicted models also have good stereochemical properties. Multiple sequence alignment also allows us to correctly position significant amino acid residues, particularly in the substrate recognition sites.

Besides many comparative model servers and modeling packages available, ORCHESTRAR is chosen for template-based structure prediction because it is able to construct models where other softwares fail; it generates more complete models; it is generally observed that SCR and all-atom models have equivalent or lower RMSD with respect to solved structures; due to loop search algorithms, ORCHESTRAR more precisely predicts the orientation of side chains; produces better stereochemical properties with no D-amino acids; generates more accurate models in the cases where all homologs are <30% identical to the target sequence, predicting backbone conformations better, especially around ligand binding sites and protein–protein interfaces. On this basis, the ORCHESTRAR suite of applications is chosen; however, SWISS-MODEL and ModBase server were also used to build the target sequences (data not shown) but the overall all-atom models from ORCHESTRAR behave well. ORCHESTRAR build the SCR regions completely in the first run of restrained sampling by fitting the backbone on the topology of the template. The side chains conformations that burrowed from the template by ANDANTE program also orient well. In addition, if one not fits well, database sampling procedure was also used that selects conformations (only chi values) from a library of loop fragments that matches the geometrical constraints. Only top scoring rotamers are automatically selected for incorporation in the backbone chain.

In particular, a conserved ATP-binding domain is present in both target sequences. This domain consists of a central sheet, surrounded on both sides by helices. A unique characteristic feature of this domain is a loop between the first strand and the first helix. This loop has an amino acid sequence of Gly X-X-X-Gly-Lys. In mammalian AspRS, this analogy stands for Gly468-Gly-Gly-Ile-Gly-Leu473.

The purpose of this study is to develop a drug discovery strategy against yeast AspRS to cease the uncontrolled growth of yeast causing infections. Candida, a genus of yeast, of which *C. albicans* is the most commonly occurring species, is responsible for Candidiasis. Candidiasis includes a wide range of yeast infections, including mycosis. Recently, the probable sequence of *C. albicans* AspRS is revealed (accession no. XP_713240.1). This sequence has 69% similarity to its yeast homolog, *S. cerevisiae* AspRS, the only protein molecule with its X-ray 3D crystal structure. In this study, the same 3D AspRS was used as a template to guide the exploration of the homolog models. In order to develop a drug targeting yeast AspRS against yeast infections, likeness between eukaryotic AspRSs creates a problem to drug discovery. A drug should be specifically enough intelligent so that it would be able to identify its true target rather than inhibiting the host system. Therefore, it is very crucial to understand the characteristics structural features and distinguishing behavior of key amino acid residues of mammalian AspRS, so that the proposed drug should be capable to discriminate between these two closely related protein molecules. Although, so far, no experimental data exists for aminoacylaion of mammalian AspRSs under physiological conditions. Further studies are needed in this regard for designing selective inhibitors of yeast AspRS; well distinguished from mammalian AspRS, for therapeutic applications by arresting yeast growth.

Note: The coordinate files of both homologs are submitted to the publicly accessible Protein Model Database (PMDB [44]; www.caspur.it/PMDB). The PMDB ID of *hs*AspRS and *mm*AspRS are PM0075852 and PM0075853, respectively

Acknowledgements

We are greatly acknowledged for the technical support provided by Prof. Bernd M. Rode (University of Innsbruck) during this research. The authors are also grateful to the AMBER supporting team for providing us AMBER software and technical support.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2009.09.006.

References

[1] F. Lévéque, P. Plateau, P. Dessen, S. Blanquet, RNA recognition and evolution of determinants in seryl-tRNA synthesis, Nucleic Acids Res. 18 (1990) 305–312.
[2] H.D. Becker, J. Reinbolt, R. Kreutzer, R. Giége, D. Kern, Existence of two distinct aspartyl-tRNA synthetases in *Thermus thermophilus*: structural and biochemical properties of the two enzymes, Biochemistry 36 (1997) 8785–8797.
[3] Y. Gangon, L. Lacoste, N. Champange, J. Lapointe, Widespread use of the glu-trna^{gln} transamidation pathway among bacteria, J. Biol. Chem. 271 (1996) 14856–14863.
[4] V. Lamour, S. Quevillon, S. Diriong, V.C. Nguyen, M. Lipinski, M. Mirande, Evolution of the Glx-tRNA synthetase family: the glutaminyl enzyme as a case of horizontal gene transfer, Proc. Natl. Acad. Sci. 91 (1994) 8670–8674.
[5] S. Cusack, M. Härtlein, R. Leberman, Sequence, structural and evolutionary relationships between class 2 aminoacyl-tRNA synthetases, Nucleic Acids Res. 19 (1991) 3489–3498.
[6] G. Eriani, M. Delarue, O. Poch, J. Gangloff, D. Moras, Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs, J. Biol. Chem. 68 (1982) 6014–6023.
[7] T. Webster, H. Tsai, M. Kula, G.A. Mackie, P. Schimmel, Specific sequence homology and three-dimensional structure of an aminoacyl transfer RNA synthetase, Science 226 (1984) 1315–1317.
[8] C. Hountondji, P. Dessen, P. Blanquet, Sequence similarities among the family of aminoacyl-tRNA synthetases, Biochemistry 68 (1986) 1071–1078.
[9] D. Soll, U.L. RajBhandary, tRNA: Structure, Biosynthesis and Function, Am. Soc. Microbial Press, 1995.
[10] P. Schimmel, Aminoacyl tRNA synthetases: general scheme of structure-function relationships in the polypeptides and recognition of transfer RNAs, Annu. Rev. Biochem. 56 (1987) 125–158.
[11] V.S. Reed, M.E. Wastney, D.C.H. Yang, Mechanisms of the transfer of aminoacyl-tRNA from aminoacyl-tRNA synthetase to the elongation factor 1 alpha, J. Biol. Chem. 52 (1994) 32932–32936.
[12] C.V. Dang, D.C.H. Yang, Disassembly and gross structure of particulate aminoacyl-tRNA synthetases from rat liver. Isolation and the structural relationship of synthetase complexes, J. Biol. Chem. 254 (1979) 5350–5356.
[13] A. Jacobo-Molina, M. Vilia-Gracia, H.C. Chen, D.C.H. Yang, Proteolytic signal sequences (PEST) in the mammalian aminoacyl-tRNA synthetase complex, FEBS Lett. 232 (1988) 65–68.
[14] S.Z. Wahab, D.C.H. Yang, Synthesis of diadenosine 5′, 5‴-P1,P4-tetraphosphate by lysyl-tRNA synthetase and a multienzyme complex of aminoacyl-tRNA synthetases from rat liver, J. Biol. Chem. 260 (1985) 5286–5289.
[15] M. Mirande, Aminoacyl-tRNA synthetase family from prokaryotes and eukaryotes: structural domains and their implications, Prog. Nucl. Acid Res. Mol. Biol. 40 (1991) 95–142.
[16] D.C.H. Yang, Mammalian aminoacyl-tRNA synthetases, Curr. Top. Cell. Reg. 34 (1996) 101–136.
[17] C. Cerini, P. Kerjan, M. Astier, D. Gratecos, M. Mirande, M. Semeriva, A component of the multisynthetase complex is a multifunctional aminoacyl-tRNA synthetase, EMBO J. 10 (1991) 4267–4277.
[18] S. Quevillon, M. Mirande, The p18 component of the multisynthetase complex shares a protein motif with the β and γ subunits of eukaryotic elongation factor 1, FEBS Lett. 395 (1996) 63–67.
[19] S. Quevillon, F. Agou, J.C. Robinson, M. Mirande, The p43 component of the mammalian multi-synthetase complex is likely to be the precursor of the endothelial monocyte-activating polypeptide Ii cytokine, J. Biol. Chem. 272 (1997) 32573–32579.
[20] B. Lorber, D. Kern, H. Mejdoub, Y. Boulanger, J. Reinbolt, R. Giegé, The micro-heterogeneity of the crystallizable yeast cytoplasmic aspartyl-tRNA synthetase, Eur. J. Biochem. 165 (1987) 409–417.

[21] M. Ruff, S. Krishnaswamy, M. Boeglin, A. Poterszman, A. Mitschler, A. Podjarny, B. Rees, J.C. Thierry, D. Moras, Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA (Asp), Science 252 (1991) 1682–1689.
[22] J. Cavarelli, B. Rees, M. Ruff, J.C. Thierry, D. Moras, Yeast $tRNA^{Asp}$ recognition by its cognate class II aminoacyl-tRNA synthetase, Nature 362 (1993) 181–184.
[23] J. Cavarelli, B. Rees, G. Eriani, M. Ruff, M. Boeglin, J. Gangloff, J.C. Thierry, D. Moras, The active site of yeast aspartyl-tRNA synthetase: structural and functional aspects of the aminoacylation reaction, EMBO J. 13 (1994) 327–337.
[24] G. Eriani, G. Prevost, D. Kern, P. Vincendon, G. Dirheimer, J. Gangloff, Cytoplasmic aspartyl-tRNA synthetase from *Saccharomyces cerevisiae*: study of its functional organisation by deletion analysis, Eur. J. Biochem. 200 (1991) 337–343.
[25] F. Agou, J.P. Waller, M. Mirande, Expression of rat aspartyl-tRNA synthetase in *Saccharomyces cerevisiae*, J. Biol. Chem. 271 (1996) 29295–29303.
[26] C. Escalante, D.C. Yang, Expression of human aspartyl-tRNA synthetase in *Escherichia coli*. Functional analysis of the N-terminal putative amphiphilic helix, J. Biol. Chem. 268 (1993) 6014–6023.
[27] F. Jeanmougin, J.D. Thompson, M. Gouy, D.G. Higgins, T.J. Gibson, Multiple sequence alignment with ClustalX, Trends Biochem. Sci. 23 (1998) 403–405.
[28] R.A. Laskowski, et al., PROCHECK: a program to check the stereochemical quality of protein structures, J. Appl. Cryst. 26 (1993) 283–291.
[29] J.U. Bowie, et al., A method to identify protein sequences that fold into a known three-dimensional structure, Science (1991) 253164–253170.
[30] C. Colovos, et al., Verification of protein structures: patterns of nonbonded atomic interactions, Protein Sci. 2 (1993) 1511–1519.
[31] Case, D. et al., AMBER 9 Molecular Dynamics Package, 2006.
[32] S.F. Altschul, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database algorithms, Nucleic Acids Res. 25 (1997) 3389–3402.
[33] F.C. Bernstein, T.F. Koetzle, G.J.B. Williams, E.F. Meyer, M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi, The Protein Data Bank: a computer-based archival file for macromolecular structures, J. Mol. Biol. 112 (1977) 535–542.
[34] J. Shi, T.L. Blundell, K. Mizuguchi, FUGUE: sequence structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties, J. Mol. Biol. 310 (2001) 243–257.
[35] K. Mizuguchi, C.M. Deane, T.l. Blundell, J.P. Overington, Homstrad: a database of protein structure alignments for homologous families, Protein Sci. 7 (1998) 2469–2471.
[36] K. Mizuguchi, C.M. Deane, T.l. Blundell, J.P. Overington, JOY: protein sequence–structure representation and analysis, Bioinformatics 14 (1998) 617–623.
[37] A. Sali, T.L. Blundell, Definition of general topological equivalence in protein structures. A procedure involving comparison of properties and relationships through simulated annealing and dynamic programming, J. Mol. Biol. 212 (1990) 403–428.
[38] R.W. Montalvao, R.E. Smith, S.C. Lovell, T.L. Blundell, L. Chora, A differential geometry approach to the prediction of the cores of protein structures, Bioinformatics 21 (2005) 3719–3725.
[39] TUNER, Tripos International (http://www.tripos.com).
[40] R.E. Smith, S.C. Lovell, D.F. Burke, R.W. Montalvao, T.L. Blundell, Andante: reducing side-chain rotamer search space during comparative modeling using environment-specific substitution probabilities, Bioinformatics 23 (2007) 1099–1105.
[41] J.P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes, J. Comput . Phys. 23 (1977) 327–341.
[42] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle mesh Ewald method, J. Chem. Phys. 103 (1995) 8577–8593.
[43] C. Sauter, B. Lorber, J. Cavarelli, D. Moras, R. Giegé, The free yeast aspartyl-tRNA synthetase differs from the tRNA (Asp)-complexed enzyme by structural changes in the catalytic site, hinge region, and anticodon-binding domain, J. Mol. Biol. 23 (2000) 1313–1324.
[44] T. Castrignano, P.D. De Meo, D. Cozzetto, I.G. Talamo, A. Tramontano, The PMDB Protein Model Database, Nucleic Acids Res. 34 (2006) 306–309.