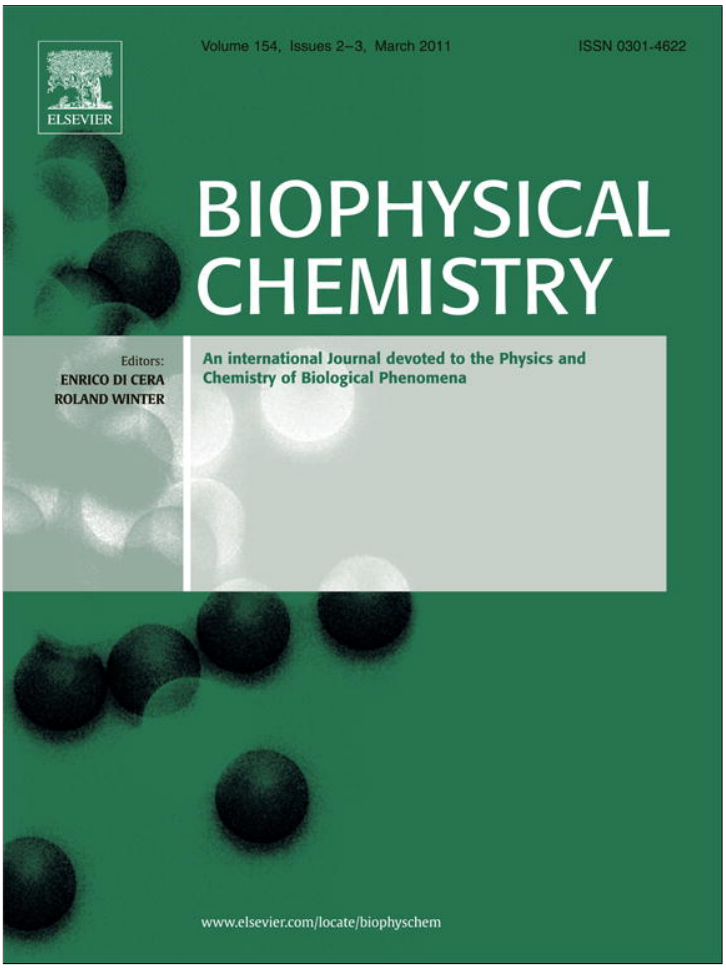

Biophysical Chemistry 154 (2011) 90–98

Contents lists available at ScienceDirect

Biophysical Chemistry

journal homepage: http://www.elsevier.com/locate/biophyschem

Interdomain compactization in human tyrosyl-tRNA synthetase studied by the hierarchical rotations technique

S.O. Yesylevskyy [a,*], O.V. Savytskyi [b], K.A. Odynets [b], A.I. Kornelyuk [b]

[a] Department of Physics of Biological Systems, Institute of Physics, National Academy of Sciences of Ukraine, Prospect Nauki, 46, Kiev-03039, Ukraine
[b] Department of Protein Engineering and Bioinformatics, Institute of Molecular Biology and Genetics, National Academy of Sciences of Ukraine, Akademika Zabolotnogo Str., 150, Kiev-03143, Ukraine

ARTICLE INFO

Article history:
Received 27 October 2010
Received in revised form 21 December 2010
Accepted 17 January 2011
Available online 22 January 2011

Keywords:
Tyrosyl-tRNA synthetase
Protein domain
Dynamic domain
GNM
HCCP
HIEROT
Domain interface
Domain motion

ABSTRACT

Aminoacyl-tRNA synthetases are key enzymes of protein biosynthesis which usually possess multidomain structures. Mammalian tyrosyl-tRNA synthetase is composed of two structural modules: N-terminal catalytic core and an EMAPII-like C-terminal domain separated by long flexible linker. The structure of full-length human cytoplasmic tyrosyl-tRNA synthetase is still unknown. The structures of isolated N-terminal and C-terminal domains of the protein are resolved, but their compact packing in a functional enzyme is a subject of debates. In this work we studied putative compactization of the N- and C-terminal modules of human tyrosyl-tRNA synthetase by the coarse-grained hierarchical rotations technique (HIEROT). The large number of distinct types of binding interfaces between N- and C-terminal modules is revealed in the absence of enzyme substrates. The binding propensities of different residues are computed and several binding "hot spots" are observed on the surfaces of N and C modules. These results could be used to govern atomistic molecular dynamics simulations, which will sample preferable binding interfaces effectively.

1. Introduction

Aminoacyl-tRNA synthetases (aaRSs) are key enzymes of protein biosynthesis, which catalyze the aminoacylation of tRNAs by corresponding amino acids during the realization of the genetic code [1,2]. These enzymes usually possess multidomain structures due to the fusion of their catalytic domains with various additional domains. Eukaryotic aaRSs contain certain specific protein modules, absent in their prokaryotic counterparts, which may be related to additional non-canonical functions of these enzymes acquired during evolution [2–4].

Tyrosyl-tRNA synthetases (TyrRSs) are homodimers and belong to class I aaRSs which comprises catalytic domain with the Rossmann-fold. Earlier it was found that mammalian tyrosyl-tRNA synthetases could be isolated either as full-length enzymes or as truncated N-terminal catalytic modules, which retains full enzymatic activity *in vitro* [5,6]. It was found that non-catalytic C-terminal domain of mammalian TyrRSs reveals unexpected homology with a novel EMAP II cytokine [7,8]. EMAP II (endothelial monocyte-activating polypeptide II) stimulates endothelial-dependent coagulation *in vitro* and apparently plays an important role in inflammation, apoptosis and angiogenesis in tumor tissues [9,10]. The full-length TyrRS does not have any cytokine activity, but isolated N-terminal catalytic module reveals IL8-like activity and non-catalytic C-terminal domain acts as EMAP II-like cytokine after proteolytic cleavage [11–13].

Recently the mutations in human TyrRS (*Hs*TyrRS) were recognized as a source of severe inherited human disorder, which makes this protein important from the biomedical point of view [14,15]. Despite extremely high biological and medical importance, the complete crystal structure of full-length mammalian cytoplasmic TyrRS is not solved at the moment.

N-terminal catalytic module of *Hs*TyrRS (residues 1–342) and non-catalytic EMAP II-like C module (residues 360–528) are separated by long flexible linker (residues 343–359). Two catalytic N modules form a rather inflexible strongly bound dimeric structure, while two C modules possess significant conformational freedom due to unstructured linkers. Currently high-quality crystal structures of individual N and C modules of *Hs*TyrRS are known [16,17], but the crystallization of the full-length protein is very challenging, probably due to its extreme conformational mobility.

Assembling complete structure of *Hs*TyrRS from known crystal structures of N and C modules appears to be quite complicated. Complete crystal structure of eubacterial *Thermus thermophilus* TyrRSs [18] cannot be used as a template for *Hs*TyrRS because of unrelated C modules in these enzymes. There is no direct information about the native interface between N and C modules, thus their spatial

* Corresponding author. Tel.: +380 44 5257991; fax: +380 44 5251589.
E-mail address: yesint3@yahoo.com (S.O. Yesylevskyy).

0301-4622/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2011.01.005

arrangement is unknown. It is tempting to assume that any realistic reconstructed full-length structure of *Hs*TyrRS should be compact with contacting N and C modules. It was proposed earlier that the formation of compact form of *Hs*TyrRS and direct contacts between N and C modules are responsible for the absence of cytokine activity of the full-length *Hs*TyrRS [16,17]. Furthermore, indirect data of fluorescence spectroscopy of the closely related full-length bovine TyrRS suggest that the C module is close to the active site, which is located in the N module [19].

Another indirect evidence of domain compactization in TyrRS comes from the crystal structures of full-length TyrRS of other organisms, which also possess multidomain structure. For instance, bacterial tyrosyl-tRNA synthetases possess the C-terminal domain of ~80 residues, which is connected to N-terminal domain by long flexible linker. However, bacterial C domain is quite different from mammalian EMAPII-like C domain. The structure of *T. thermophilus* TyrRS revealed that two C-terminal domains of the dimer are crystallized in quite different orientations in the absence of bound tRNA [18]. In the complex with cognate tRNA the C domain of *T. thermophilus* TyrRS becomes stabilized in a fixed orientation and the linker peptide (residues 345–351) becomes ordered [18]. Despite these conformational changes and the heterogeneity of positions of the C modules in the absence of tRNA both crystal structures of bacterial enzyme are compact.

There are two possible approaches to assembling the complete compact structure of TyrRS. The first one is the direct docking of C module on the surface of the dimer of N modules followed by the filtering of the best structures based on the possible conformations of the linker. However, such modeling ignores real dynamics of the C modules and could easily provide the structures, which are not accessible kinetically. The second approach is a construction of the non-compact structure and simulating the dynamics of C modules in the course of enzyme compactization. This approach will not enumerate all possible compact structures exhaustively but produces the structures, which are most accessible kinetically. In this work we used the second approach in order to explore putative compactization of human tyrosyl-tRNA synthetase.

The ultimate goal of the modeling studies of TyrRS is revealing interactions between the protein modules, which are significant from the biological point of view. Such interactions should be resolved on the atomistic level of details, preferably by means of molecular dynamics (MD) simulations. It is clear, however, that the configurational space available for nearly free moving C modules is too large for being sampled in atomistic simulations. The hierarchical modeling strategy could be used to resolve this problem. The regions of preferable binding between C and N modules could be found in the course of low-cost coarse-grained simulations, which sample significant part of available configurational space. Subsequent MD simulations could elucidate the atomic details of these interfaces and provide more important biological information. In this work we focus on the coarse-grained stage of this procedure.

The compactization of large highly mobile domains of TyrRS is a specific case of the large-scale slow conformational dynamics of proteins, which is one of the most challenging problems in modern computational biophysics. In our case, it is sufficient to determine the general character of slow motions of the protein. This allows using coarse-grained models of the protein, which speeds up the simulations at the cost of sacrificing atomic details. The extreme case of the coarse graining is the subdivision of the protein into small number of rigid blocks (such as N and C modules of TyrRS), which move on mechanical hinges. The most challenging problem with this approach is an adequate and physically consistent identification of the moving blocks. This problem is solved in this work by using the concept of hierarchically organized dynamic domains.

The dynamic domain could be defined as a part of protein, which shares certain pattern of correlated motions and moves relatively independent from the other parts of the protein globule [20,21]. The dynamic domains are hierarchical in nature — each domain could be subdivided into smaller subdomains, which exhibit motions with larger degree of internal correlations. As a result the whole protein could be viewed as system of small quite rigid structural blocks, which are combined into larger but more flexible units, which are, in turn, combined into larger aggregates and so one until the level of the whole globule is reached [21,22].

In this work we used the simulation technique, which accounts for hierarchical rotations of dynamic domains. This method (called HIEROT from HIErarchical ROTations) provides highly simplified coarse-grained model of the protein, which allows sampling large part of its conformational space quickly and efficiently [20,21]. Several non-compact initial structures with different placement of the C modules and different conformations of the interdomain linker were studied. Several compact structures of the *Hs*TyrRS were obtained by the HIEROT technique and compared systematically. The interfaces between N and C modules in the compact structures observed in all performed simulations are classified into several distinct structural classes. Obtained compact structures could not be considered the high quality final models of the compact state of *Hs*TyrRS, but they could serve as reasonable starting points for detailed atomistic simulations.

2. Methods

2.1. Initial structures

The structure of the whole *Hs*TyrRS monomer was constructed from the crystal structures of N- and C-terminal modules of the protein (PDB codes 1N3L:A [16], residues A3–P342 and 1NTG:A [17] residues P360–S528, respectively) using Modeller 9.7 software [23,24]. Missed N-terminal residues M1–D3, the residues of the catalytic loop K222–E228 and the linker residues D343–E359 were added using loops reconstruction option in Modeller 9.7. Five ensembles of 100 structures each were generated in Modeller with the parameters derived from the predefined protocol for homology modeling with multiple templates. Best structures were selected using the Modeller Objective Function (molpdf), the Discrete Optimized Protein Energy (DOPE) score and the normalized DOPE score [25]. Selected structures were verified using the MolProbity web-server [26] to ensure the absence of sterical clashes, unusual rotamers and bonds, etc.

The whole monomers generated by Modeller were assembled into the dimers using the crystallized dimer of N-terminal modules (PDB code 1N3L) as a template for structural alignment. Finally, ten best structures were used for HIEROT simulations.

2.2. Motions of the dynamic domains

The simulation protocol used to study the motions of dynamic domains is the following:

1) The Gaussian Network Model (GNM) is applied to the initial structure of the studied protein. The GNM normal modes are computed and the matrix of the residue–residue correlations is constructed.
2) The Hierarchical Clustering of the Correlation Patterns (HCCP) technique is used to identify the hierarchy of dynamic domains using this matrix.
3) The hierarchical rotations (HIEROT) technique is used to simulate the motion of dynamic domains governed by the residue-level empirical potential.

2.3. The Gaussian Network Model

The Gaussian Network Model (GNM) [27–31] is a popular method of determining the pattern of large-scale motions in globular proteins.

GNM describes the protein as a network of identical harmonic springs, which connect the C_α atoms of the residues located within the cut-off distance r_c (typically 7–10 Å) regardless of their positions in the sequence. Normal modes of such network could be computed easily. It was shown, that GNM describes harmonic motions of many proteins surprisingly well and produces the results, which are often indistinguishable from those of the all-atom normal mode calculations [28,32]. The matrix of the pair-wise correlation of motions of the residues c_{ij} can be calculated easily using computed normal modes [27,28,33].

2.4. The HCCP technique

The Hierarchical Clustering of the Correlation Patterns (HCCP) method was designed as a technique, which identifies the dynamic domains reliably regardless of their spatial position and orientation in the complex proteins [20,21]. HCCP utilizes the matrix of residue–residue correlations of motion c_{ij} obtained from the GNM calculations. The c_{ij} contains the correlations of *infinitesimal harmonic* displacements of C_α atoms from their equilibrium positions. Such displacements are not directly related to the large-amplitude motions of domains, which we study in the present work [34]. They are used for domain identification only, but not for the prediction of direction and amplitude of domain motions.

In order to eliminate the sensitivity to small variations of the input data, the *correlations of the correlation patterns* p_{ij} are used instead of the pair-wise correlations c_{ij}. The p_{ij} matrix is of dimension $N \times N$ and its elements show to what extent the columns i and j of c_{ij} are similar in terms of linear correlation. The matrix p_{ij} provides much more robust way of comparing the motions of residues than does the conventional correlation matrix c_{ij} [20].

The residues with similar correlation patterns can be combined to clusters, which share the same character of motion. Several such clusters can be further combined as having weaker motion similarities and so on. The clustering algorithm was revised significantly in the current work in comparison with the previous implementations [21,35]. The scheme of the algorithm is the following:

1. Each residue of the protein is assigned to be the simplest cluster of size 1.
2. The pair of residues i,j with the largest correlation p_{ij} is found.
3. Residues i and j are merged into a single cluster. The matrix p_{ij} is recalculated by the following rule:

$$p_{ij} = \frac{1}{m_i m_j} \sum_{k \in \{M_i\}} \sum_{l \in \{M_j\}} p_{kl} \quad (1)$$

   where m_i and m_j are the numbers of elements in clusters i and j; M_i and M_j are the vectors of sizes m_i and m_j, respectively, which contain the indexes of the residues in these clusters. In other words the average correlation of all inter-cluster pairs is calculated.
4. The procedure of the Intercalating Segments Elimination (ISE) is applied. ISE is described in details in [21].
5. Steps 2 and 3 are continued until all the residues are merged and the whole protein becomes a single cluster.

2.5. The HIEROT technique

The HCCP technique subdivides the protein into the hierarchy of clusters, defined on different levels of correlations of infinitesimal harmonic displacements around local energy minimum. The clusters of particular hierarchical level are assumed to move as a whole relatively independently from one another. Thus, internal degrees of freedom of the clusters can be considered frozen at the given level of hierarchy and the clusters are approximated as rigid bodies. The only remaining motion is relative displacement of the clusters. The technique of Hierarchical Rotations (HIEROT), which simulates the motions of the rigid hierarchically organized clusters in a physically consistent way, was developed in our previous work [34].

The clusters, which are physically linked by the polypeptide chain, can move relative to one another as the rigid bodies if the number of such linkers is one or two. If there is only one linker it serves as a pivot for free rotation, characterized by two coordinates. If there are two linkers they form a hinge axis characterized by a single angle of rotation.

The following procedure was used to determine the number of the clusters, which are allowed to move. We start from the first level of hierarchy in HCCP, which corresponds to the largest clusters (dynamic domains), and determine the linkers and the type of motions. Then, we consider the next level of hierarchy (subdomains of dynamic domains). The subdomain is not allowed to move if it shares all linkers with any domain of the higher level. If the subdomain is smaller than M residues, its motion is ignored. This procedure is continued for the third level of hierarchy (the subdomains of subdomains) and so on for the desired number of hierarchical levels N_L. The values $M = 30$, $N_L = 50$ were used.

Our procedure transforms the protein into the mechanical system of small number of rigid bodies, connected by pivots and axes. We model the motion of this system by means of standard Metropolis Monte-Carlo simulations. The reader is referred to our previous work [34] for the details of the HIEROT technique and the simulation algorithm.

2.6. Residue-level potentials

Residue-level knowledge-based DFIRE-SCM potentials are used in the Monte-Carlo simulations. These potentials are proved to be quite accurate in determining the native states of various proteins among decoys [36,37]. Their accuracy is comparable to those of empirical all-atom potentials used in molecular dynamics simulations [36,37]. The centers of geometry of the side chains are used as the force centroids, which means that only one interaction center per residue exists. It is necessary to note, that DFIRE potentials are based on the structures of real proteins in the crystal environment. This means that the solvation effects are ignored and that the most compact conformation is likely to have the lowest energy. This allows an efficient search for the compact states of the proteins starting from the less compact ones, but not *vice-versa*.

2.7. Analysis of domain interfaces

In order to compare the resulting compact structures obtained in several HIEROT simulations the following procedure was used:

1. The interfaces of both C-terminal modules of *Hs*TyrRS with the dimer of N-terminal modules were determined at the residue level. The atoms of C and N modules were considered to be in contact if the distance between their centers was smaller than the cut-off distance. The residues were considered to be in contact if at least one pair of atoms from them was in contact in any trajectory frame. The lists of contacting residues were constructed for all studied trajectories.
2. Each studied trajectory is considered to be a cluster of size 1.
3. The similarity matrix is computed as $s_{ij} = a_{ij}/b_{ij}$, where a_{ij} is the number of contacts shared by all $n_i + n_j$ trajectories in clusters i and j (n_i is the number of trajectories in the cluster i); b_{ij} is the number of contacts, shared by any two or more trajectories in clusters i and j.
4. The largest matrix element s_{ij} is found and clusters i and j are merged.

5. Steps 3 and 4 are repeated while $s_{ij}<c$, where c is the critical ratio of the shared contacts ($c=20\%$ in this study).

As a result, all trajectories are subdivided into clusters, which share more than 20% of the residue–residue contacts between one of the C modules and the dimer of N modules. The trajectories inside the same cluster could be classified as having the domain interface of the same type, while the domain interfaces of the trajectories from different clusters do not show any significant similarity.

Two different values of the cut-off are used. The cut-off of 2.5 Å corresponds to the closest and strongest non-bound contacts between the domains, while the cut-off of 3.5 Å covers all contacts, including weak interactions.

2.8. Conservation of residues

The conservation of the residues was obtained using the ConSurf server [38]. Separate PDB structures of N and C modules (1N3L:A and 1NTG:A respectively) were used. The conservation information for these structures was extracted from the ConSurf-DB database [39], converted to the machine-readable form and mapped into the structure of the whole *Hs*TyrRS monomer. Conservation of the linker residues was not considered. The discrete conservation score adopted in ConSurf where used (referred as *S* below). In this scheme the conservation ranges from 1 (highly variable residue) to 9 (conserved residue).

2.9. Simulation details

All custom simulation software for this work is written in C++ using the Pteros molecular modeling library (http://www.sourceforge.net/projects/pteros/). The GNM calculations are among the core functionality of this library. The modified HCCP algorithm is re-implemented and tested against the previous Fortran-95 version. The HIEROT algorithm is implemented from scratch.

The interfaces of C- and N-terminal modules for each trajectory were determined using custom Tcl scripts for VMD [40] molecular analysis program. The clustering of the interfaces was performed by the stand-alone custom Tcl script. The figures are prepared with VMD [40].

The simulation results are referenced in the text as "*i*–*n*:*c*", where *i* is the code number of initial structure (according to Modeller rank), *n* is the number of trajectory generated from this initial structure, *c* (1 or 2) is the C-terminal module under consideration. Initial structures are referred as *i* using the same nomenclature.

The model temperature 0.75 was used in all HIEROT simulations. Twenty independent simulations of 500,000 Monte-Carlo steps each were performed for each structure. The trajectory frames were written each 1000 steps.

Table 1
Characteristics of initial structures. Modeller scores are shown for one monomer, while the radius of gyration is for assembled dimer.

Initial structure	molpdf	DOPE	Normalized DOPE	Radius of gyration, Å
1	2513.18	−59421.80	−1.04	37.86
2	2419.23	−59379.84	−1.03	41.59
3	2188.63	−59468.36	−1.04	42.48
4	2208.57	−59350.75	−1.03	51.71
5	2699.56	−59232.81	−1.01	42.66
6	2616.82	−59199.10	−1.01	42.24
7	2690.47	−59098.57	−1.00	43.22
8	2455.57	−59026.08	−0.99	38.88
9	2449.47	−58922.91	−0.97	38.51
10	2476.73	−58806.62	−0.96	60.52

The protein modeling was partially performed in Grid environment using the infrastructure of the MolDynGrid virtual laboratory (https://moldyngrid.org/) [41].

2.10. Limitations and alternatives

Our approach possesses several limitations. Particularly, the linker is modeled by the rigid segments connected by few pivot points, while in reality the whole unstructured linker is highly flexible. This restricts the motion of C modules to some extent because of sterical constraints imposed by the rigid segments of the linker. However, the existence of several points of free rotation in the linker provides enough conformational freedom for C module to expose almost any part of its surface to the N modules. Simplified modeling of the linker keeps the number of degrees of freedom small and makes the whole problem tractable.

Another limitation is imposed by the DFIRE potential, which is biased towards the compact structures and often leads to trapping in local energy minima [34]. However, the DFIRE potentials are very fast and quite accurate for the compact structures, which justifies their usage in our coarse-grained conformational sampling. More detailed but also much more complex residue level potentials, such as UNRES [42], could also be used.

As it was already mentioned in the Introduction, there are other ways of finding the compact structure of *Hs*TyrRS. Direct docking of N and C modules followed by the reconstruction of the linker is one of them. Such technique allows exhaustive enumeration of possible binding interfaces between the modules, but ignores their kinetic accessibility. Another way is using molecular dynamics with the coarse-grained force field, such as UNRES [43] or MARTINI [44], in order to simulate the compactization of domains. Coarse-grained force field will reduce the computational burden significantly in comparison to all-atom simulations allowing much better sampling of the conformational space.

3. Results

3.1. General properties of initial structures and interfaces

The characteristics of initial structures used in the simulations are summarized in Table 1. It is clearly seen that Modeller scores of all initial structures are quite similar, which means that all of them are of approximately the same quality. However, the radius of gyration varies significantly between the structures, which reflect their very different compactness. Visual inspection shows that the placement of the C modules vary dramatically in different starting structures as well as the conformation of the linker.

All simulations lead to the stable well-equilibrated compact structures where both C modules form strong contacts with the dimer of the N modules. A pronounced asymmetry between the positions and orientations of the C modules is observed, which is caused by independent motions of the C modules. The contacts between the C modules of different subunits are observed for several starting structures. The significance of such contacts is unclear, thus they are not analyzed in details.

3.2. Binding energies

It is possible to estimate the binding energy of each of C modules in each trajectory. This is a total DFIRE-SCM energy of interaction between given C module and the dimer of the N modules averaged over equilibrated part of HIEROT trajectory.

The binding energies from all 200 simulations were analyzed, which gives 400 energy values (two for each simulation because of two C modules). The spectrum of the binding energies is very broad, which suggests that the C modules show no tendency to form a

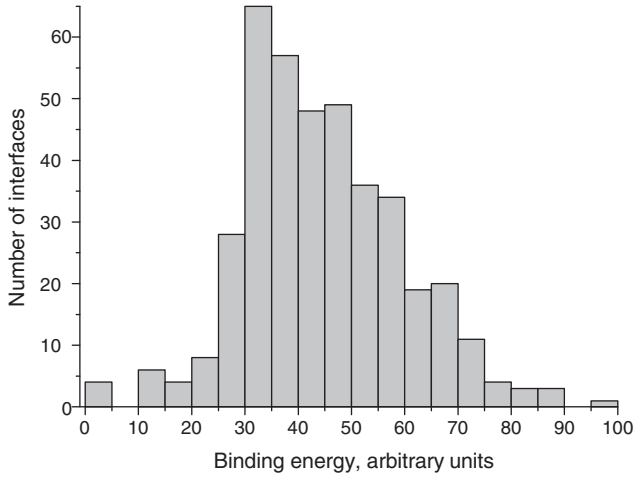

Fig. 1. The distribution of the binding energies of individual C-terminal modules on the dimers of N-terminal modules of *Hs*TyrRS.

unique interface with the dimer of N modules (Fig. 1). The visual inspection shows that the peak of the distribution does not correspond to any well-defined interface but rather marks the mean binding energy of a variety of different interfaces (data not shown).

3.3. Types of interfaces

Inspection of the equilibrated compact structures shows that there is a wide variety of domain interfaces, which are obviously dependent on initial position and orientation of the C-module and the conformation of the linker. The clustering analysis of 400 available interfaces revealed 90 distinct types of interfaces, which share more than 20% of the residue–residue contacts in the case of 2.5 Å cut-off. 172 interfaces share no similarity with any of these types and to each other ("lone" interfaces). Lone interfaces are most likely formed by random unspecific contacts and could be ignored in further analysis. In the case of 3.5 Å cut-off the number of the distinct types is 107 and the number of lone interfaces is 133.

There is only one type of interface, which is formed in simulations with *different* initial structures in the case of 2.5 Å cut-off. In the case of 3.5 Å cut-off the number of such types is two (Table 2). All other types of interfaces are observed exclusively in the simulations with the same initial structures.

It is clearly seen in Table 2 that the clusters of trajectories, which lead to similar interfaces, contain only two trajectories. In all clusters, only one of the C modules forms similar contacts in all simulations. Another C module form very different interfaces (Fig. 2).

Different initial structures lead to different number of the interface types (Table 3). This number depends on the cut-off used to compute the interfaces, but the general trend remains the same for both cut-off values.

There is a pronounced negative correlation between the number of the interface types and the radius of gyration of initial structures. In contrast, the correlation between the number of the lone interfaces and the radius of gyration is strongly positive (Table 3). The possible reason of this correlation is discussed below.

3.4. Regions of preferred binding

Although there is no unique binding interface between C and N modules in our simulations, some residues possess higher propensity of forming contacts than the other. We computed the probability of each residue from both C and N modules to be in contact with any residue from another module as $P_c = N_c/N_{sim}$, where N_c is the number of simulations, where given residue forms at least one contact, N_{sim} is the total number of simulations. P_c was computed for each monomer separately and then averaged for both of them in order to obtain symmetric picture for identical monomers.

The results are shown in Fig. 3, where the residues are colored according to their P_c value. It is clearly seen that majority of the residues of C module form contacts, but one side of the C module is much more likely to be in contact with the dimer of N modules then the other. There is rather extended "hot spot" on the surface of the C module, which contains the residues with $P_c > 0.45$ (ILE 445–ILE 448, GLU 489–LYS 496, GLN 504–GLN 507, LYS 523, and ASP 526). The approximate center of the hot spot is ASP 494 with $P_c = 0.57$.

The residues of the N module are much more heterogeneous in their probability of forming contacts. There is a saddle-like contacting region in the center of the dimer of N modules with three well-defined hot spots, which contain the residues with $P_c > 0.4$. The main hot spot on the N terminal module contains the residues from TYR 79 to LEU 89. The residue TRP 87, which is the center of this hot spot, possesses the largest P_c value (0.55) in the whole N-module. The second hot spot contains the residues PRO 200–TYR 204 and has P_c values around 0.4. This region of the second monomer is very close in space to the first hot spot from the first monomer (Fig. 3B). The last hot spot contains the residues LYS 335, SER 338 and ALA 339 with P_c values around 0.41, which are located near the beginning of the flexible interdomain linker.

3.5. Conservation of residues in the binding hot spots

Although the binding hot spots do not include known functionally important residues of *Hs*TyrRS it is interesting to analyze the evolutionary conservation of residues in these regions. Mapping the conservation data from ConSurf-DB to our structure allowed comparing the P_c values and conservation scores S directly. There is no correlation between these parameters at the level of whole protein (correlation coefficient <0.1), however the comparison for individual residues revealed interesting features.

In general the hot spots are rather heterogeneous in terms of S. All values from 1 to 9 are observed without any systematic correlation with P_c values. The single hot spot of C module and the main hot spot of N module possess highly conservative residues ($S = 9$) PHE 495 and ILE 497 in the C module and ASN 82 in the N module. There are also slightly less conservative residues ($S = 8$) ILE 445, GLU 489, LYS 490, ALA 493, and LYS 496 in the C module and ALA 85 in the N module. Two remaining hot spots consist of neutral ($S = 4$–6) or variable ($S = 1$–3) residues.

4. Discussion

The problem of modeling full-length flexible structure of human cytoplasmic tyrosyl-tRNA synthetase is very challenging. It is not known whether the compactization of C and N modules of this enzyme occurs in its free state or at the presence of tRNA and other ligands. There are no direct experimental findings, which could clarify this question. In our work we modeled the compactization of domains in tyrosyl-tRNA synthetase in the ligand-free state, however similar approach could be used in the future to study the complexes with the substrates, especially with tRNA.

Table 2
Types of interfaces formed in simulations with different initial structures. The interfaces, which are present for both values of the cut-off are shown in bold.

Interface type	Interfaces	Shared contacts
Cut-off 2.5 Å		
1	**6–14:2, 5–5:2**	42.0%
Cut-off 3.5 Å		
1	4–13:1, 8–7:1	23.7%
2	**6–14:2, 5–5:2**	48.4%

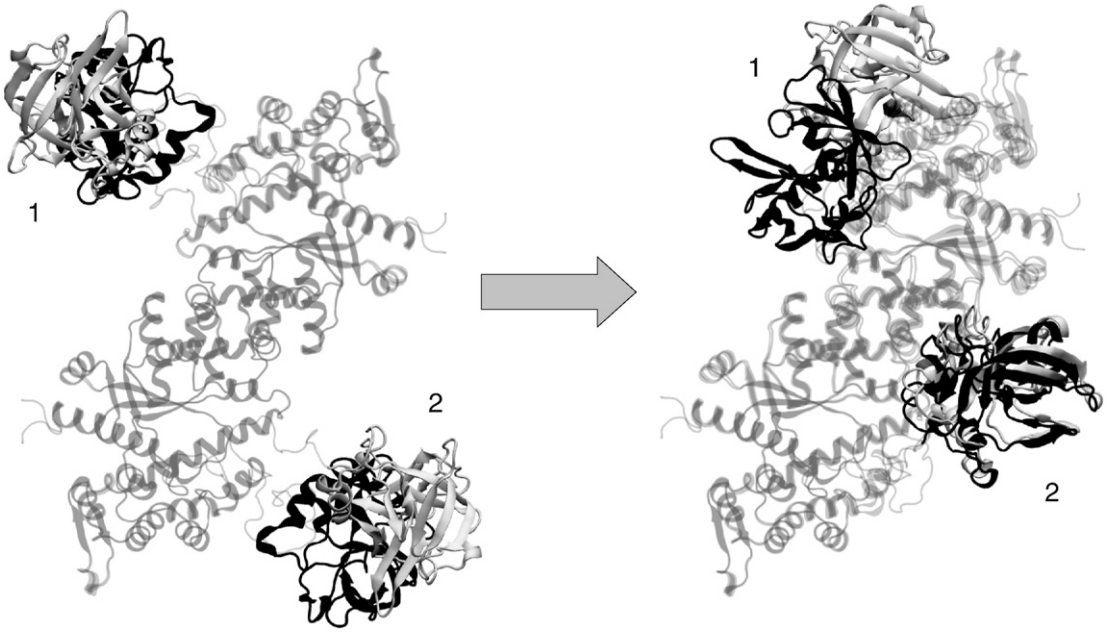

Fig. 2. Initial (left) and final (right) structures of the simulations 6–14 and 5–5 aligned by their N-terminal modules (transparent). C-terminal modules are black for 6–14 and gray for 5–5. Numbers show the first and the second C-module respectively.

Separate crystal structures of the N and C modules of human tyrosyl-tRNA synthetase are available [42], however long and flexible connecting linker allows almost infinite number of their relative positions and orientations. The problem of finding the most probable compact structure is similar to conventional docking problem, however the presence of the linker makes it nontrivial. Mechanical restraint imposed by the linker narrows the conformational space available for the C-terminal module dramatically but prevents us from using conventional docking algorithms.

In this work, we used the method of hierarchical rotations to study interdomain compactization in *Hs*TyrRS. The simplified nature of the force field prevents us from obtaining precise atomistic details of domain interactions but allows effective scan of possible compact structures, which are kinetically accessible from given starting configuration of the protein. Preferable compact structures found by the HIEROT technique could be further simulated using atomistic molecular dynamics technique, which is able to reveal significant biological details.

In all HIEROT simulations C module always binds to the dimer of N modules and unbinding events were never observed. This is expectable because of the nature of DFIRE residue-level force field. This force field ensures that the compact structures have much lower energy then any non-compact conformations, which speeds up compactization but prevents us from observing realistic binding-unbinding dynamics.

In general 90 unrelated types of interfaces were found for small 2.5 Å cut-off and 107 for larger 3.5 Å cut-off. The difference between two cut-off values is insignificant taking into account the fact, that much larger number of weak interactions is included in the case of 3.5 Å cut-off.

Some simulations starting from each initial structure converge to a limited number of distinct interface types, which are unique for particular initial structure. Remaining simulations lead to lone interfaces. The number of the interface types and the number of lone interfaces depend strongly on the radius of gyration of the corresponding initial structure (Table 3). This dependence could be explained easily. The starting structures, with large radii of gyration possess extended linkers and allow great conformational freedom for their C modules. Obviously, this enormous conformational space could not be sampled effectively even in the coarse-grained simulations. As a result, the probability of forming similar interfaces in different runs is low, the majority of the runs lead to accidental lone interfaces and the number of distinct interface types is small. In contrast, the conformational space of the C modules is restricted in the structures with small radii of gyration. This leads to much better sampling and higher probability of convergence between different runs. As a result, the number of the interface types is larger and the number of the lone interfaces is small. This strong trend (the correlation between the radius of gyration and the number of the lone interfaces reaches 0.87) is observed for both cut-off values.

The number of the interface types for particular staring structure increases with the increase of cut-off, but this increase is remarkably small for the majority of initial structures (Table 3). This convinces us that the cut-off used to define the contacts has only marginal influence on the results.

In the case of small 2.5 Å cut-off, only one interface type includes the trajectories from different starting structures. In the case of larger 3.5 Å cut-off, this number increases to two, which is still very small in

Table 3
Statistics of the interface types for each initial structure. N_{types} is the number of the interface types; N_{lone} is the number of lone interfaces; r_{gyr} is the radius of gyration from Table 1.

Initial structure	2.5 Å cut-off		3.5 Å cut-off	
	N_{types}	N_{lone}	N_{types}	N_{lone}
1	9	12	11	8
2	9	22	10	19
3	13	12	13	11
4	4	30	8	24
5	14	8	14	7
6	9	12	10	9
7	11	15	14	8
8	10	10	13	5
9	9	17	11	12
10	3	34	5	30
Correlation with r_{gyr}	−0.72	0.84	−0.77	0.87

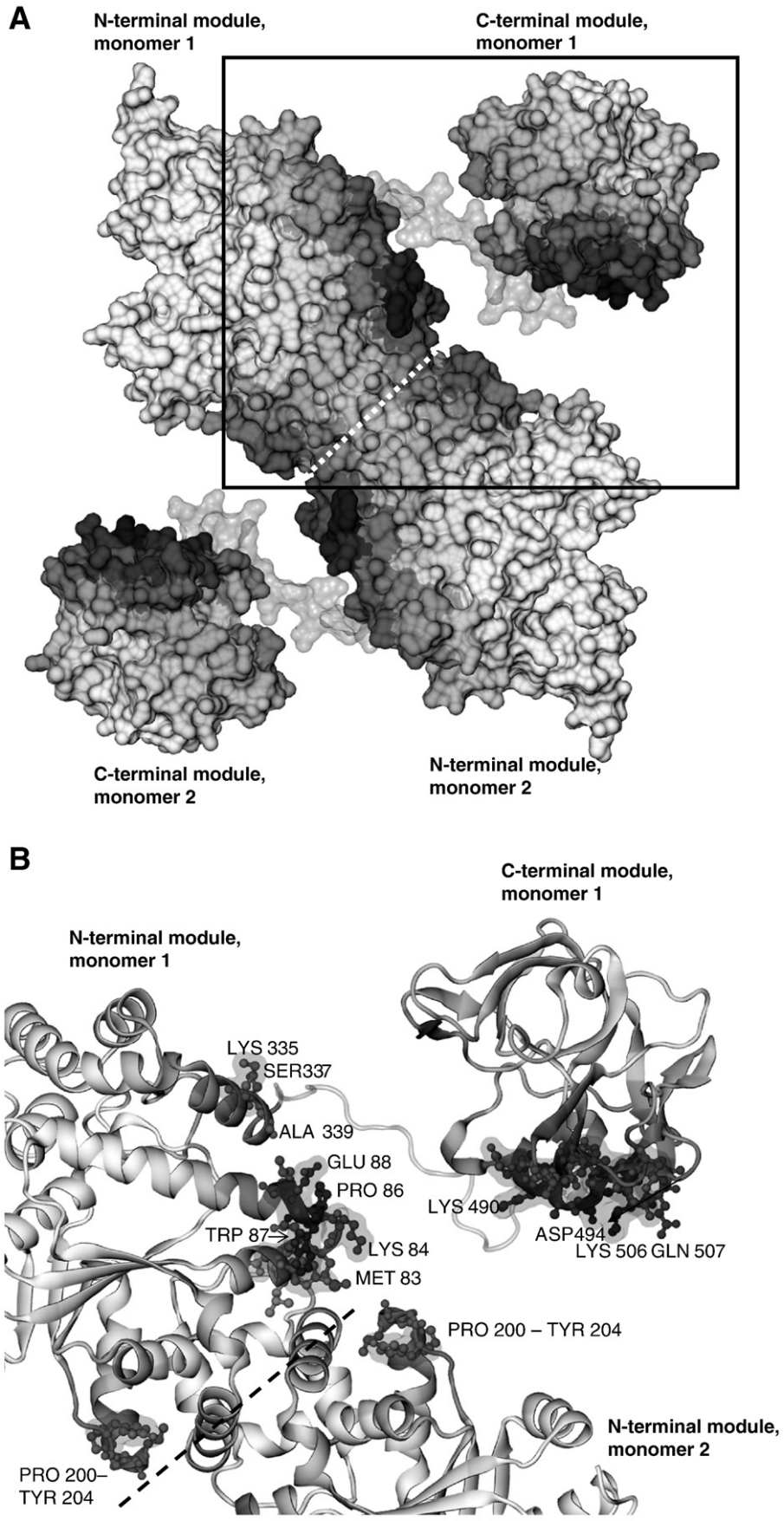

Fig. 3. The structure of the HsTyRS colored according to P_c. Black corresponds to maximal P_c, white to zero. (A) The whole dimer shown in the surface representation. Linkers are shown transparent. White dashed line shows approximate boundary of two monomers. The square encloses the area shown in (B). (B) Part of the dimer with binding hot spots. The protein is shown in cartoon representation, while the hot spots are shown as balls and sticks. Residue labels are shown. Only few of the hot spot residues are labeled in the C-terminal module for clarity.

comparison to the total number of the interface types (around 2%). These exceptional interface types demonstrate that some rare universal binding interfaces, which are accessible from different initial structures, exist in the system. However, the number of such interfaces is not known. Extrapolating our results it is possible to conclude that thousands of independent HIEROT simulations are needed to reproduce any particular type of binding interface reliably and much larger number is required to obtain the probabilities of their occurrence.

Although the variety of possible interface types is not sampled exhaustively in this work, the general picture is clearly visible. The energy landscape of binding is extremely rugged with the large number of local energy minima and, probably, kinetic traps. The larger the conformational space available for the C-terminal module the larger the number of the variants of binding observed and the smaller the probability of convergence to a particular type of the interface.

The large number of simulations allowed us to compute the probability of contact formation for different residues. It appears that the C module possesses an extended binding hot spot, which cover one of its sides. Certain residues in the hot spot, such as ASP 494, possess binding probabilities larger then 0.5, which means that they participate in interface formation in every second simulation. The surface of the dimer of N-modules possesses three distinct regions of preferred binding, which are located at the sides, which faces the C-modules (Fig. 3). The binding probabilities in these hot spots are larger than 0.4 and reach 0.55 for the residue TRP 87.

None of the hot spot residues corresponds to known active sites of the *Hs*TyrRS, which means that interdomain compactization does not involve any residues directly involved in protein functioning in our simulations. However, the hot spot of the C module and the main hot spot of the N module contain several highly conserved residues. It is possible to speculate that evolutionary conservation of these residues is caused by their binding properties, but other reasons are possible as well.

The existence of distinct hot spots does not contradict the fact that no unique binding interface was found in our simulations. The residues of particular hot spot form contacts with a broad range of the residues from another module rather randomly. Exact pairs of contacting residues are very diverse, however the residues from the hot spots are found in many of such pairs.

Although our results are inevitably biased due to a limited number of starting structures it is obvious that the binding propensities of different residues are very different. The identification of distinct binding hot spots may help in performing targeted molecular dynamics simulations, which would sample only those regions of the conformational space, which correspond to strong preferred binding. Such hierarchical modeling strategy could save a lot of simulation time because the regions of preferable binding are determined in advance by the coarse-grained HIEROT technique.

5. Conclusion

Our HIEROT simulations revealed a number of possible binding interfaces between C- and N-terminal modules of human tyrosyl-tRNA synthetase in the absence of enzyme substrates. It was shown that there are several regions of preferable binding (hot spots) on the surface of both modules, which contain the residues with high binding propensities. These data could be used for targeting atomistic molecular dynamics simulations into the most interesting regions of configurational space, especially for the complexes of tyrosyl-tRNA synthetase with tyrosine, ATP and cognate tRNA.

References

[1] M. Ibba, D. Soll, Aminoacyl-tRNA synthesis, Annu. Rev. Biochem. 69 (2000) 617–650.

[2] M. Mirande, Aminoacyl-tRNA synthetase family from prokaryotes and eukaryotes: structural domains and their implications, Prog. Nucleic Acid Res. Mol. Biol. 40 (1991) 95–142.

[3] B. Cahuzac, E. Berthonneau, N. Birlirakis, E. Guittet, M. Mirande, A recurrent RNA-binding domain is appended to eukaryotic aminoacyl-tRNA synthetases, EMBO J. 19 (2000) 445–452.

[4] M. Guo, X.L. Yang, P. Schimmel, New functions of aminoacyl-tRNA synthetases beyond translation, Nat. Rev. Mol. Cell Biol. 11 (2010) 668–674.

[5] A.I. Kornelyuk, I.V. Kurochkin, G.K. Matsuka, Tyrosyl-tRNA synthetase from the bovine liver. Isolation and physico-chemical properties. Molekulyarnaya Biologiya (Moscow) 22 (1988) 176–186.

[6] D.V. Gnatenko, I.V. Kurochkin, T.A. Ribkinska, A.I. Kornelyuk, G.K. Matsuka, Purification and characterization of functionally active proteolytically modified form of tyrosyl-tRNA synthetase from bovine liver, Ukr. Biokhim. Zhurn. (Kiev) 63 (1991) 61–67.

[7] T.A. Kleeman, D. Wei, K.L. Simpson, E.A. First, Human tyrosyl-tRNA synthetase shares amino acid sequence homology with a putative cytokine, J. Biol. Chem. 272 (1997) 14420–14425.

[8] O.V. Levanets, V.G. Naidenov, K.A. Odynets, M.I. Woodmaska, G.K. Matsuka, A.I. Kornelyuk, Homology of C-terminal non-catalytic domain of mammalian tyrosyl-tRNA synthetase with cytokine EMAP II and non-catalytic domains of methionyl- and phenylalanyl-tRNA synthetases, Biopolym. Cell (Kiev) 13 (1997) 474–478.

[9] A.C. Berger, G. Tang, H.R. Alexander, S.K. Libutti, Endothelial monocyte-activating polypeptide II, a tumour-derived cytokine that plays an important role in inflammation, apoptosis, and angiogenesis, J. Immunother. 23 (2000) 519–527.

[10] A.G. Reznikov, L.V. Chaykovskaya, L.I. Polyakova, A.I. Kornelyuk, Antitumor effect of endothelial monocyte-activating polypeptide-II on human prostate adenocarcinoma in mouse xenograft model, Exper. Oncol. 29 (2007) 267–271.

[11] K. Wakasugi, P. Schimmel, Two distinct cytokines released from a human aminoacyl-tRNA synthetase, Science 284 (1999) 147–151.

[12] K. Wakasugi, P. Schimmel, Highly differentiated motifs responsible for two cytokine activities of a split human tRNA synthetase, J. Biol. Chem. 274 (1999) 23155–23159.

[13] A.I. Kornelyuk, M.P.R. Tas, A. Dybrovsky, C. Murray, Cytokine activity of the non-catalytic EMAP-2-like domain of mammalian tyrosyl-tRNA synthetase, Biopolym. Cell (Kiev) 15 (1999) 168–172.

[14] A. Jordanova, J. Irobi, F.P. Thomas, P. Van Dijck, K. Meerschaert, M. Dewil, I. Dierick, A. Jacobs, E. De Vriendt, V. Guergueltcheva, C.V. Rao, I. Tournev, F.A. Gondim, M. D'Hooghe, V. Van Gerwen, P. Callaerts, L. Van Den Bosch, J.P. Timmermans, W. Robberecht, J. Gettemans, J.M. Thevelein, P. De Jonghe, I. Kremensky, V. Timmerman, Disrupted function and axonal distribution of mutant tyrosyl-tRNA synthetase in dominant intermediate Charcot–Marie–Tooth neuropathy, Nat. Genet. 38 (2006) 197–202.

[15] E. Storkebaum, R. Leitão-Gonçalves, T. Godenschwege, L. Nangle, M. Mejia, I. Bosmans, T. Ooms, A. Jacobs, P. Van Dijck, X.L. Yang, P. Schimmel, K. Norga, V. Timmerman, P. Callaerts, A. Jordanova, Dominant mutations in the tyrosyl-tRNA synthetase gene recapitulate in *Drosophila* features of human Charcot–Marie–Tooth neuropathy, Proc. Natl Acad. Sci. USA 106 (2009) 11782–11787.

[16] X.-L. Yang, R.J. Skene, D.E. McRee, P. Schimmel, Crystal structure of a human aminoacyl-tRNA synthetase cytokine, Proc. Natl Acad. Sci. USA 99 (2002) 15369–15374.

[17] X.-L. Yang, J. Liu, R.J. Skene, D.E. McRee, P. Schimmel, Crystal structure of an EMAP-II-like cytokine released from a human tRNA synthetase, Helv. Chim. Acta 86 (2003) 1246–1257.

[18] A. Yaremchuk, I. Kriklivyi, M. Tukalo, S. Cusack, Class I tyrosyl-tRNA synthetase has a class II mode of cognate tRNA recognition, EMBO J. 21 (2002) 3829–3840.

[19] A.I. Kornelyuk, I.V. Klimenko, K.A. Odynets, Conformational change of mammalian tyrosyl-tRNA synthetase induced by tyrosyl adenylate formation, Biochem. Mol. Biol. Int. 35 (1995) 317–322.

[20] S.O. Yesylevskyy, V.N. Kharkyanen, A.P. Demchenko, Hierarchical clustering of the correlation patterns: new method of domain identification in proteins, Biophys. Chem. 119 (2006) 84–93.

[21] S.O. Yesylevskyy, V.N. Kharkyanen, A.P. Demchenko, Dynamic protein domains: identification, interdependence and stability, Biophys. J. 91 (2006) 670–685.

[22] A.P. Demchenko, S.O. Yesylevskyy, Nanoscopic description of biomembrane electrostatics: results of molecular dynamics simulations and fluorescence probing, Chem. Phys. Lipids 160 (2009) 63–84.

[23] N. Eswar, M.A. Marti-Renom, B. Webb, M.S. Madhusudhan, D. Eramian, M. Shen, U. Pieper, A. Sali, Comparative protein structure modeling with MODELLER, John Wiley & Sons, Inc, 2006.

[24] M.A. Marti-Renom, A. Stuart, A. Fiser, R. Sánchez, F. Melo, A. Sali, Comparative protein structure modeling of genes and genomes, Annu. Rev. Biophys. Biomol. Struct. 29 (2000) 291–325.

[25] M.Y. Shen, A. Sali, Statistical potential for assessment and prediction of protein structures, Protein Sci. 15 (2006) 2507–2524.

[26] V.B. Chen, W.B. Arendall III, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L. W. Murray, J.S. Richardson, D.C. Richardson, MolProbity: all-atom structure validation for macromolecular crystallography, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 12–21.

[27] A.R. Atilgan, S.R. Durell, R.L. Jernigan, M.C. Demirel, O. Keskin, I. Bahar, Anisotropy of fluctuation dynamics of proteins with an elastic network model, Biophys. J. 80 (2001) 505–515.

[28] I. Bahar, A.R. Atilgan, B. Erman, Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential, Fold. Des. 2 (1997) 173–181.

[29] Y. Yildirim, P. Doruker, Collective motions of RNA polymerases. Analysis of core enzyme, elongation complex and holoenzyme, J. Biomol. Struct. Dyn. 22 (2004) 267–280.
[30] O. Keskin, Comparison of full-atomic and coarse-grained models to examine the molecular fluctuations of c-AMP dependent protein kinase, J. Biomol. Struct. Dyn. 20 (2002) 333–345.
[31] P. Doruker, A.R. Atilgan, I. Bahar, Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: application to a-amylase inhibitor, Proteins 40 (2000) 512–524.
[32] I. Bahar, A. Rader, Coarse-grained normal mode analysis in structural biology, Curr. Opin. Struct. Biol. 15 (2005) 586–592.
[33] S.O. Yesylevskyy, A.S. Klymchenko, A.P. Demchenko, Semi-empirical study of two-color fluorescent dyes based on 3-hydroxychromone, J. Mol. Struct. 755 (2005) 229–239.
[34] S.O. Yesylevskyy, V.N. Kharkyanen, A.P. Demchenko, The blind search for the closed states of hinge-bending proteins, Proteins: Struct., Funct., Bioinf. 71 (2007) 831–843.
[35] S.O. Yesylevskyy, V.N. Kharkyanen, A.P. Demchenko, The change of protein intradomain mobility on ligand binding, is it a commonly observed phenomenon? Biophys. J. 91 (2006) 3002–3013.
[36] C. Zhang, S. Liu, H. Zhou, Y. Zhou, An accurate, residue-level, pair potential of mean force for folding and binding based on the distance-scaled, ideal-gas reference state, Protein Sci. 13 (2004) 400–411.
[37] H. Zhou, Y. Zhou, Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction, Protein Sci. 11 (2002) 2714–2726.
[38] H. Ashkenazy, E. Erez, E. Martz, T. Pupko, N. Ben-Tal, ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids, Nucl. Acids Res. 38 (2010) W529–W533.
[39] O. Goldenberg, E. Erez, G. Nimrod, N. Ben-Tal, The ConSurf-DB: pre-calculated evolutionary conservation profiles of protein structures, Nucleic Acids Res. 37 (2009) D323–D327.
[40] W. Humphrey, A. Dalke, K. Schulten, VMD — visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–38.
[41] A.O. Salnikov, I.A. Sliusar, O.O. Sudakov, O.V. Savytskyi, A.I. Kornelyuk, MolDyn-Grid virtual laboratory as a part of Ukrainian academic grid infrastructure, IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, Rende (Cosenza), Italy, 2009, pp. 237–240.
[42] S. Ołdziej, C. Czaplewski, A. Liwo, M. Chinchio, M. Nanias, J.A. Vila, M. Khalili, Y.A. Arnautova, A. Jagielska, M. Makowski, H.D. Schafroth, R. Kaźmierkiewicz, D.R. Ripoll, J. Pillardy, J.A. Saunders, Y.K. Kang, K.D. Gibson, H.A. Scheraga, Physics-based protein-structure prediction using a hierarchical protocol based on the UNRES force field: assessment in two blind tests, Proc. Natl Acad. Sci. USA 102 (2005) 7547–7552.
[43] A.V. Rojas, A. Liwo, H.A. Scheraga, Molecular dynamics with the united-residue (UNRES) force field. Ab initio folding simulations of multi-chain proteins, J. Phys. Chem. B 111 (2008) 293–309.
[44] L. Monticelli, S.K. Kandasamy, X. Periole, R.G. Larson, D.P. Tieleman, S.-J. Marrink, The MARTINI coarse-grained force field: extension to proteins, J. Chem. Theory Comput. 4 (2008) 819–834.