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.