

Homology modeling of structure of NH₂-terminal module of mammalian (*Bos taurus*) tyrosyl-tRNA synthetase

K. A. Odynets, O. E. Bazylevskiy, A. I. Kornelyuk

Institute of Molecular Biology and Genetics, National Academy of Science of Ukraine
150 Acad. Zabolotny Street, Kyiv, 03143, Ukraine

E-mail: kornelyuk@imbg.org.ua

Mammalian tyrosyl-tRNA synthetase (TyrRS) is composed of two structural modules: the NH₂-terminal catalytic core and cytokine-like COOH-terminal module. In order to elucidate the structural bases for the N-module functions we have used the computational prediction of its three-dimensional (3D) structure by comparative modeling approach. A model of the bovine TyrRS N-module represents the Rossmann nucleotide-binding fold (RF) which is linked to the α -helical domain (α HD). The RF domain forms a single β -sheet containing 5 parallel and one attached antiparallel β -strands surrounded by α -helices. The connective polypeptide, CP1, inserted between β 3- and β 4-strands of the RF domain is perturbed from the domain core. Comparative analysis of the multiple sequence alignment of known TyrRSs and the obtained model structure reveals the conservative surface elements, which could potentially form the tRNA^{Tyr}-binding surface. This putative surface includes some exposed amino acid residues of CP1, e. g. essential Lys146 and Lys147 residues, which were identified earlier by site-directed mutagenesis.

Introduction. Tyrosyl-tRNA synthetase (TyrRS, tyrosine — tRNA ligase, E. C. 6.1.1.1) catalyzes highly specific attachment of L-tyrosine to cognate tRNA^{Tyr} [1]. In mammalia (particularly in bovine, *Bos taurus*) two functionally active forms of this enzyme have been discovered. During intracellular proteolysis as well as after apoptosis-induced secretion, the C-module may be cleaved from the «minimal» TyrRS (N-module) within protease-sensitive linking segment [4–7]. In bovine TyrRS, the N-module is composed of about 345 amino acid residues (aa) and consists of a class I-defining Rossmann fold (RF) domain (about 235 aa) and a second anticodon-recognition α -helical domain (α HD, about 109 aa) [8].

As an approach to elucidate the structural bases of N-module functions we have used the computational prediction of its three-dimensional (3D) structure by the comparative modeling approach. 3D structures of three eubacterial TyrRSs and one tryptophanyl-tRNA synthetase (TrpRS), determined experimentally [9–12], were used as structural templates for the homology modeling procedure.

Materials and Methods. Sequence of *Bt*TyrRS was deposited to GenBank/GenPept and Swiss-Prot databases (accession numbers AAC82467, Q29465). Search for homologous sequences was performed with iterative PSI-BLAST service (this and other programs and servers are described in our review [13]). TyrRS of 15 eukaryotic organisms (*Bos taurus*, *Homo sapiens*, *Mus musculus*, *Fugu rubripes*, *Drosophila melanogaster*, *Anopheles gambiae*, *Caenorhabditis elegans*, *Pneumocystis carinii*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Encephalitozoon cuniculi*, *Candida albicans*, *Plasmodium falciparum*, *Arabidopsis thaliana* and *Nicotiana tabacum*) and 17 archaeobacteria were analyzed. Multiple sequence alignment was carried out with Clustal W and secondary structure elements were predicted with the PHD and the multiprediction the NPS@ server. The coordinates files were obtained from PDB for crystallographic structures of three eubacterial TyrRSs: *Thermus thermophilus* (PDB ID codes 1H3E, 1H3F); *Bacillus (Geobacillus) stearothermophilus* (2TS1, 3TS1, 4TS1); *Staphylococcus aureus* (1JII, 1JIJ, 1JIK, 1JIL); and tryptophanyl-tRNA synthetase from *B. stearothermophilus* (1D2R, 1I6K, 1I6L, 1I6M). Three

ading servers Bioinbgu and FUGUE [14] were used to search for the similar structures. The complete N-module model was built using Modeller 6.2 program suite [15]. Total 20 initial models from Modeller were built and optimized using «refine-5» protocol with default parameters for optimization procedures. Energy calculation of the models has been performed using Energy option of Modeller.

Secondary structure assignment was done using DSSP and STRIDE algorithms. Search for structurally related domains and their superimposition (both in pair and multiple structure alignment modes) were carried out using CE and VAST. The SCOP server 1.61 release was used to find other similar structures from related families and superfamilies. Quality of the optimized models was verified using Eval123D, ANOLEA and ERRAT servers and the best model was chosen for the subsequent optimization. The models optimization with a simulated molecular dynamics (MD) approach was carried out with Gromacs 3.1.4 program [16].

The model structures were placed into a periodic rectangular box, filled with SPC water layer. The minimal distances between the protein and boundaries of the box were 0.7 nm. Energy minimization was performed on the proteins using a steepest-descent algorithm with 2 fs integration time step and a tolerance of $0.01 \text{ kJ} \cdot \text{mol}^{-1} \cdot \text{nm}^{-1}$ during 10 ps. Position restrained MD was done to distribute water molecules, and actual MD was simulated during 1 ns. The temperature was controlled by coupling to an external bath of 300 K with coupling time constant of 10 fs. The GROMACS force field was used in this work.

Solvent accessible surface areas of the proteins were calculated using GetArea 1.1 server. Pockets and cavities search and analysis was performed using castP. Distribution of surface conservative residues was done using ConSurf 2.0. Structure visualization and analysis were performed with Deep View (Swiss-PDB Viewer) 3.7b2 and Protein Explorer 1.299.

Results and Discussion. The iterative PSI-BLAST search for the M1-P344 segment of *Bt*TyrRS has found 32 homologous sequences of the N-modules from other eukaryotic and archaeobacterial TyrRSs, and about 30 sequences of TyrRSs and TrpRSs homologous to its α HD domain. The N-module sequences can be divided into five parts according to their distinct substructures: the N-subdomain (residues M1-E33); the first half of Rossmann fold (R34-K119); CPI insertion (G120-L177); the second half of RF with junked «KMSSS» catalytic loop (K178-L235); and α HD domain (D236-P344). The highest local homology between *Bt*TyrRS and *Tt*TyrRS is in

the N-subdomain and within L-tyrosine binding H3 α -helix and «KMSSS» loop.

We have modeled the *Bt*TyrRS N-module, containing 340 aa residues (L5-P344) with molecular mass of 38.5 kDa. The 3D structure of the most similar (69 identical residues for M1-P344 region, 22.5 % identity) *T. thermophilus* TyrRS (PDB code 1H3E), determined with 2.9 Å resolution, was used as the best template structure. The target/template pair sequence alignment has been extracted from multiple sequence alignment and corrected manually to optimize the positions of insertions. The localization of insertions/deletions was also deduced from inspection of the template secondary structure elements. As the most significant errors in models are often due to misalignment of target and template sequences, we performed careful manual editing of the alignment, as well as the iterative realignment and model building. There are eleven indels in the *Bt*TyrRS N-module in comparison with the template, and their exact locations may vary slightly. There are only four short insertions within the RF domain of *Bt*TyrRS (¹⁰⁶ESIG, Y129, ¹⁴⁸AG and E227) and seven short insertions in the template *Tt*TyrRS structure.

Unlike the RF domain, a weak sequence homology between the α HD domains of eu- and prokaryotic TyrRSs reveals their significant divergence after the separation of eukaryotes and archaeobacteria from eubacteria. Different threading methods allowed us to select the *B. stearrowthermophilus* TrpRS structure (PDB code 1D2R, chain A) as an alternative template for the *Bt*TyrRS α HD modeling because of its higher structural similarity. For example, sequence-structure search with Bioinbgu server predicts *Bs*TrpRS as better template than *Bs*TyrRS (Z-scores are 39.0 and 4.3 respectively). That sequence-structure alignment gives only two insertions (²⁵⁸NGVLAFIRHVL and ³⁰²EV) within the bovine sequence compared to the *Bs*TrpRS α HD domain. We have used the *Bs*TrpRS α HD domain (K192-D297 region) as an alternative structure template for the α HD structure modeling. The initial models of the α HD domain were built with the Modeller program and connected with the RF domain into complete two-domain N-module structure by overlapping their common KMSSS-loop segments. The best from 20 initial models were optimized in water environment using restricted molecular dynamics simulation techniques with the Gromacs 3.1.4 program.

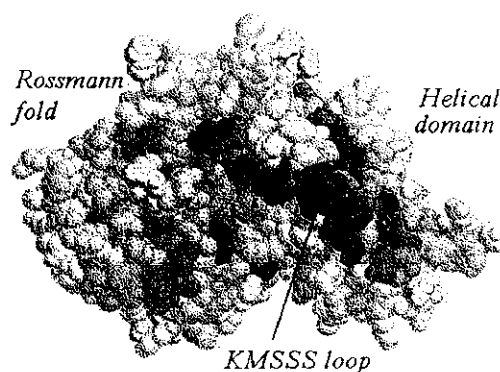
To test the validity of the initial and optimized models, a combination of evaluation criteria was used to discriminate between the correct and incorrect models. For the models verification such criteria were used as interatomic clashes, stereochemical properties

(bond lengths, angles and dihedral angles, peptide bonds planarity, C_{α} tetrahedral distortion etc.), position of residues in the Ramachandran plot, root-mean-square deviation (RMSD) for C_{α} - C_{α} atoms of model/template structure pairs. Using several evaluation servers we analyzed the following structures: 1) the templates (1H3E and 1D2R); 2) initial and refined models of the N-module obtained from Modeller; and 3) the optimized best models obtained by molecular dynamics simulation using the Gromacs program. ANOLEA server performs energy calculations on a protein chain, evaluating the non-local environment of each heavy atom in the molecule. The energy of each pairwise interaction in this non-local environment is taken from a distance-dependent knowledge-based mean force potential. Server ERRAT analyzes the statistics of non-bonded interactions between different atom types, and a single output plot gives the value of the error function vs. position of a 9-residue sliding window. The error values give confidence limits that are extremely useful in making decisions about model reliability. Regions of candidate protein structures that are mistraced or misregistered can then be identified by analysis of the pattern of nonbonded interactions from each window.

The best model obtained after MD simulation represents a two-domain protein with solvent accessible surface area 16.3 \AA^2 and 41.7 \AA^3 volume (Figure). The RMSD between the model and template structures is 1.65 \AA for 756 backbone atoms.

A number of analyses were carried out to predict the elements responsible for tRNA binding ability of N-module. The RF and α HD domains are arranged into the N-module forming common tRNA binding surface, which is located on the same side for all dimeric «minimal» TyrRSs and TrpRSs. The comparative analysis of multiple sequence alignment of known TyrRSs and the obtained model structure reveals the conservative surface elements, which could potentially form the tRNA^{Tyr}-binding surface. This putative tRNA-contacting region includes CPI (residues T121-L177) inserted into RF domain. We have analyzed the solvent exposed amino acid residues. The most exposed residues in the CPI of *Bt*TyrRS N-module are: D122, L125, K127, E128, L131, Y134, R135, S137, S138, T146, Q142, H143, K146, K147, K154, Q155 and V156.

The secondary structure elements were defined with the DSSP and STRIDE programs. The RF domain of the *Bt*TyrRS N-module (residues M1-L235) forms a single β -sheet containing 5 parallel and one attached antiparallel β -strands arrange as (β 0)- β 5- β 4- β 1- β 2- β 3 and surrounded by α -helices. The β -sheet adopts additional structural elements. The RF



Accessible surface view of the model structure for N-module of bovine tyrosyl-tRNA synthetase. The N-module is shown from the side of its active site cavity, where catalytic KMSSS loop corresponds to conservative dark-colored amino acid residues. Two domains of N-module are the Rossmann fold (left) and the α -helical domain (right). Colors of surface residues correspond to their conservativity obtained from the multiple sequence alignment analysis

domain includes the additional N-terminal segment of the first 33 residues containing α - β 0- β element, which is characteristic of all known TyrRSs but is not homologous to TrpRSs and other class I synthetases. The characteristic CPI insertion (residues G120-L177) is located between β 3- and β 4-strands and perturbs from the core of the RF domain.

A detailed understanding of the protein function requires the identification of some conserved amino acid residues at the protein surface, which may be responsible for the protein function. The ConSurf server was used to identify such important regions, based on the phylogenetic relations between homologous proteins from their multiple sequence alignment. Conservativity of the exposed amino acid residues of the eukaryotic-type N-module is represented on Figure as color-coded surface of the *Bt*TyrRS N-module. The majority of exposed (more than 50 %) and strongly conservative residues are localized in and around the described surface region.

The functional role of lysine residues K146 and K147 located in CPI of *Bt*TyrRS has been previously studied by site-directed mutagenesis [17]. The replacement of both residues with Asn and Tyr, respectively, as well as the substitution of K147 alone, caused the inactivation of mutant TyrRS in the tRNA^{Tyr} aminoacylation reaction. In our model both Lys residues are exposed and may form contacts with tRNA^{Tyr}.

Acknowledgements. This work was supported by NATO Linkage grant HTECG.LG 974684 and NATO Computer networking supplement No. 976022 to NATO linkage grant.

Note added in proof. When this article was

submitted, the article of Yang et al., (2002) [18] was published where crystallographic structure of a human «minimal» TyrRS was described with resolution 1.18 Å.

К. О. Одиноць, О. Є. Базилевський, О. І. Корнелюк

Моделювання за гомологією структури NH₂-кінцевого модуля тирозил-тРНК синтетази ссавців (*Bos taurus*)

Резюме

Тирозил-тРНК синтетаза ссавців (*TyrRS*) складається з двох структурних модулів: каталітичного NH₂-кінцевого кора та цитокіноподібного COOH-кінцевого модуля. Для вивчення структурних основ функціонування каталітичного N-модуля здійснено комп'ютерне моделювання його просторової (3D) структури. Модель структури N-модуля *TyrRS* являє собою нуклеотидзв'язуючий фолд — згортку Россмана (RF), до якого приєднаний α-спіральный домен (αHD). RF-домен формує β-згортку, яка містить п'ять паралельних та один антипаралельний β-стренди, оточені α-спіралями. З'єднувальний поліпептид, CP1, розміщений між β3- та β4-стрендами RF-домену. Аналіз моделі виявив консервативні елементи, які можуть формувати область зв'язування гомологічної tRNA^{Tyr}. Ця область містить деякі амінокислотні залишки CP1, наприклад, функціонально важливі Lys146 та Lys147, які були ідентифіковані раніше методом сайт-спрямованого мутагенезу.

К. А. Одынец, А. Е. Базилевский, А. И. Корнелюк

Моделирование по гомологии структуры NH₂-концевого модуля тирозил-тРНК синтетазы млекопитающих (*Bos taurus*)

Резюме

Тирозил-тРНК синтетаза млекопитающих (*TyrRS*) состоит из двух структурных модулей: каталитического NH₂-концевого кора и цитокинноподобного COOH-концевого модуля. Для изучения структурных основ функционирования каталитического N-модуля проведено компьютерное моделирование его пространственной (3D) структуры. Модель структуры N-модуля *TyrRS* представляет собой нуклеотидсвязывающий фолд — свертку Россмана (RF), к которому присоединен α-спиральный домен (αHD). RF-домен формирует β-лист, содержащий пять параллельных и один антипараллельный β-стрэнды, окруженные α-спиралями. Соединительный полипептид, CP1, размещен между β3- и β4-стрэндами RF-домена. Анализ модели выявил консервативные элементы, способные формировать область связывания гомологичной tRNA^{Tyr}. Эта область включает некоторые аминокислотные остатки CP1, например, функционально важные Lys146 и Lys147, которые были идентифицированы ранее методом сайт-направленного мутагенеза.

REFERENCES

- Schimmel P. Aminoacyl tRNA synthetases: general scheme of structure-function relationships in the polypeptides and recognition of transfer RNAs // *Ann. Rev. Biochem.*—1987.—56.—P. 125—158.
- Kornelyuk A. I., Kurochkin I. V., Matsuka G. H. Tyrosyl-tRNA synthetase from beef liver. Purification and physico-chemical properties // *Mol. Biol. (Kiev)*.—1988.—22, N 1.—P. 176—186.
- Kleeman T. A., Wei D., Simpson K. L., First E. A. Human tyrosyl-tRNA synthetase shares amino acid sequence homology with a putative cytokine // *J. Biol. Chem.*—1997.—272, N 22.—P. 14420—14425.
- Kornelyuk A. I., Tas M. P. R., Dybrovsky A., Murray C. Cytokine activity of the non-catalytic EMAP-2-like domain of mammalian tyrosyl-tRNA synthetase // *Biopolimery i kletka (Kiev)*.—1999.—15, N 2.—P. 168—172.
- Levanets O. V., Naidenov V. G., Odynets K. A., Woodmaska M. I., Matsuka G. Kh., Kornelyuk A. I. 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 // *Biopolimery i kletka (Kiev)*.—1997.—13, N 6.—P. 474—478.
- Wakasugi K., Schimmel P. Two distinct cytokines released from a human aminoacyl-tRNA synthetase // *Science*.—1999.—284, N 5411.—P. 147—151.
- Wakasugi K., Schimmel P. Highly differentiated motifs responsible for two cytokine activities of a split human tRNA synthetase // *J. Biol. Chem.*—1999.—274, N 33.—P. 23155—23159.
- Kornelyuk A. I. Structural and functional investigation of mammalian tyrosyl-tRNA synthetase // *Biopolimery i kletka (Kiev)*.—1998.—14, N 4.—P. 349—359.
- Yaremchuk A., Kriklyviy I., Tukalo M., Cusack S. Class I tyrosyl-tRNA synthetase has a class II mode of tRNA recognition // *EMBO J.*—2002.—21, N 14.—P. 3829—3840.
- Brick P., Bhat T. N., Blow D. M. Structure of tyrosyl-tRNA synthetase refined at 2.3 Å resolution. Interaction of the enzyme with the tyrosyl adenylate intermediate // *J. Mol. Biol.*—1989.—208, N 1.—P. 83—98.
- Qiu X., Janson C. A., Smith W. W., Green S. M., McDevitt P., Johanson K., Carter P., Hibbs M., Lewis C., Chalker A., Fosberry A., Lalonde J., Berge J., Brown P., Houge-Frydrych C. S., Jarvest R. L. Crystal structure of *Staphylococcus aureus* tyrosyl-tRNA synthetase in complex with a class of potent and specific inhibitors // *Protein Sci.*—2001.—10, N 10.—P. 2008—2016.
- Ilyin V. A., Temple B., Li G.-P., Vachette Y. P., Carter C. W. JR. 2.9 Å Crystal structure of ligand-free tryptophanyl-tRNA synthetase: domain movements fragment the adenine nucleotide binding site // *Protein Sci.*—2000.—9, N 2.—P. 218—231.
- Одиноць К. О., Корнелюк О. І. Методи аналізу і моделювання просторової структури білків // *Наук. записки НАУКМА. Природн. науки.*—2001.—19.—С. 7—17.
- Shi J., Blundell T. L., Mizuguchi K. FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties // *J. Mol. Biol.*—2001.—310, N 1.—P. 243—257.
- Sanchez R., Sali A. Comparative protein structure modeling. Introduction and practical examples with modeller // *Methods Mol. Biol.*—2000.—143.—P. 97—129.
- Lindahl E., Hess B., van der Spoel D. GROMACS 3.0: a package for molecular simulation and trajectory analysis // *J. Mol. Model.*—2001.—7.—P. 306—317.
- Найденов В. Г., Вудмаска М. И., Корнелюк А. И. Мацука Г. Х. Сайт-направленный мутагенез остатков лизина, локализованных в соединительном пептиде нуклеотидсвязывающего домена (свертки Россмана) тирозил-тРНК синтетазы из печени быка // *Биополимеры и клетка.*—2000.—16, № 4.—С. 275—280.
- Yang X. L., Skene R. J., McRee D. E., Schimmel P. Crystal structure of a human aminoacyl-tRNA synthetase cytokine // *Proc. Nat. Acad. Sci. USA.*—2002.—99, N 243.—P. 15369—15374.

УДК 577.152.6:577.332
Надійшла до редакції 17.09.02