

Molecular modelling of FtsZ proteins based on their homology in *Escherichia coli* and *Mycobacterium tuberculosis* as the key stage of rational design of new antituberculous compounds

Molekularno modeliranje proteinov FtsZ na osnovi njihove homologije v *Escherichia coli* in *Mycobacterium tuberculosis* kot ključna stopnja racionalnega oblikovanja novih protituberkuloznih komponent

Oleh Demchuk ^a, Pavel Karpov ^a, Peter Raspor ^b, Yaroslav Blume ^{a*}

^aDepartment of Genomics and Molecular Biotechnology, Institute of Food Biotechnology and Genomics, Natl. Academy of Sciences of Ukraine, Osipovskogo str, 2a, Kyiv, 04123, Ukraine;

^bBiotechnical Faculty, University of Ljubljana, Jamnikarjeva, 101, 1000 Ljubljana, Slovenia

*correspondence: yablume@univ.kiev.ua

Abstract: The analysis of the quality of X-ray structures from *Mycobacterium tuberculosis* FtsZ proteins, which are deposited in the ProteinDataBank, gave a possibility to select a 2Q1Y (Chain A) structure as a template for future *in silico* research. Also several spatial models of FtsZ protein from *Escherichia coli* were reconstructed with on-line servers »SWISS-MODEL Workspace« and I-TASSER, than the most appropriate structure was selected. Basing on complex bioinformatic study, the model, which was rebuilt by SwissModel server from 2Q1Y (chain A) template, was supposed as the most significant.

Keywords: FtsZ, *Escherichia coli*, *Mycobacterium tuberculosis*, 3D-structure modelling and verification, *in silico*

Izvleček: Analiza struktur proteinov FtsZ z X žarki iz *Mycobacterium tuberculosis* deponiranih v »ProteinDataBank« je dala možnost izbora strukture 2Q1Y (veriga A) kot matrice za nadaljno *in silico* raziskavo. Nekaj prostorskih modelov proteina FtsZ iz bakterije *Escherichia coli* je bilo rekonstruiranih na on-line serverju SwissModel in I-Tasser, kateremu je sledil izbor najprimernejše strukture. Na osnovi celovite bioinformacijske preverbe kaže, da je model narejen na platformi serverja SwissModel iz matrice 2Q1Y (veriga A) najbolj sprejemljiv za uporabo.

Ključne besede: FtsZ, *Escherichia coli*, *Mycobacterium tuberculosis*, 3D-strukturno modeliranje in preverjanje, *in silico*

Introduction

Tuberculosis is the leading cause of death in the world from a single infectious disease, claiming over three million lives each year (Huang et al. 2006). Furthermore, poor patient compliance

and inadequate control programs have lead to the emergence of multidrug-resistant strains of *M. tuberculosis* (Raviglione 2000). Bacterial resistance to three or more 'second-line' antibiotics is classified as extremely drug-resistant tuberculosis. Therefore, there is an urgent need

for the development of new anti-tuberculosis drugs with novel mechanism of action(s), which are active against drug-resistant as well as drug-sensitive *M. tuberculosis* strains. FtsZ (Filamentous temperature-sensitive protein Z) is an essential cell division protein has been shown to be a bacterial homolog of the mammalian cytoskeleton protein tubulin (Kumar et al. 2010). Accordingly, FtsZ protein is a very promising target for new antimicrobial drug development, and especially compounds effective against drug-resistant *M. tuberculosis* strains (Kumar et al. 2011). It is the perspective target for such numerous and diverse groups of low molecular weight compounds as benzimidazoles (Ohashi et al. 1999, Kumar et al. 2011), naphthalenesulfonates (Yu and Margolin 1998), azithromycins (Margalit et al. 2004), ethyl carbamates (White et al. 2000, 2002), diterpenoid phenols (Jaiswal et al. 2007) etc. At the same time, modern rational design of new compounds with antibacterial activity is impossible without stage of virtual screening, with application of accurate three-dimensional models of target FtsZ-proteins. The last is very important due to FtsZ – tubulin structure and function similarity, places particularly high demands on quality of 3D-models used for *in silico* molecular docking and virtual screening.

At present, Worldwide ProteinDataBank (wwPDB – www.pdb.org: RCSB PDB (USA)/PDBe (Europe)/PDBj (Japan)) (Berman et al. 2003) contain the row of bacterial FtsZ X-ray structures of different resolution: 2R6R (1.70 Å) (Oliva et al. 2007) and 2R75 (1.40 Å) (Läppchen et al. 2008) from *Aquifex aeolicus*; 2VAM (2.50 Å) (Oliva et al. 2007), 2VXY (1.70 Å) (Haydon et al. 2008), 2RHH (2.00 Å), 2RHJ (1.76 Å), 2RHL (2.45 Å) and 2RHO (2.45 Å) (Raymond et al. 2009) from *Bacillus subtilis*; 1FSZ (2.80 Å) (Löwe and Amos 1998), 1W5B (2.20 Å), 1W5A (2.40 Å), 1W58 (2.50 Å) and 1W59 (2.70 Å) (Oliva et al. 2004) from *Methanococcus jannaschii*; 1RLU (2.08 Å), 1RQ2 (1.86 Å), 1RQ7 (2.60 Å) (Leung et al. 2004), 2Q1X (2.35 Å) and 2Q1Y (2.30 Å) (Respcio et al. database record) from *M. tuberculosis*; 1OFU (2.10 Å) (Cordell et al., 2003) and 2VAW (2.90 Å) (Oliva et al. 2007) from *Pseudomonas aeruginosa*; 1W5F (2.00 Å) (Oliva et al. 2004)

from *Thermotoga maritima*. Most of these species belong to different phyla and subkingdoms of the *Bacteria* kingdom, and one, *M. jannaschii*, to the phylum *Euryarchaeota* (subkingdom *Archaea*). At the same time, there are a number of defects in all deposited in the Protein Data Bank structures. (Höltje et al. 2008) Most of deposited in PDB X-ray structures of bacterial FtsZs characterized by the loss of N- and C-terminal fragments (typically a few tens of residues), presence of gaps in protein globule, as well as absence of certain heavy atoms of side chains of amino acid residues.

Unfortunately, until now, there are no more or less complete X-ray structures of *E. coli* FtsZ protein, model organism also plays an important role in modern biological engineering and industrial microbiology. Currently only the 1F47 (PDB) structure have the last 17 amino acid residues (Lys367-Asp383), forming a short unstructured region, ends with a two-helix turn at the C-terminal end (Mosyak et al. 2000).

However, we are also interested in complete structure of this protein, due to the fact that commercial analytical kits for *in vitro* binding experiments are more available for *E. coli* FtsZ protein analysis than analytical kits for *M. tuberculosis*. *In vitro* and *in silico* modelling of interaction with low-molecular compounds of both, *E. coli* and *M. tuberculosis* FtsZ proteins, such as benzimidazole derivatives, allow us much accurate binding-site identification and analysis. Based on FtsZs structural homology, these experimentally confirmed binding site (or sites), can be extrapolated from *E. coli* protein to the structure of mycobacterial homolog. This allow us more accurate prediction of binding sites of such new and promising anti-TB compounds as benzimidazoles.

Thus, the purpose of the research was *in silico* modelling of three-dimensional structure of *E. coli* FtsZ protein, and qualitative reconstruction of *M. tuberculosis* FtsZ protein model based on comprehensive analysis of X-ray structures deposited in the Protein Data Bank.

Methods

Analysis of Protein Data Bank structures of FtsZ M. tuberculosis

Complete amino acid sequence of *M. tuberculosis* FtsZ (P64170) (Cole et al. 1998, Fleischmann et al. 2002) was downloaded from UniProt (<http://www.uniprot.org/>) database (The UniProt Consortium 2008). Multiple alignments of amino acid sequences of *M. tuberculosis* FtsZ PDB-structures and P64170 were realized in ClustalX 2.0.5 with a set of BLOSSUM matrices (<http://www.clustal.org>, Larkin et al. 2007). The PDB-structures of FtsZ *M. tuberculosis* protein were analyzed using »DeepView – Swiss-PdbViewer 4.0.3« (Guex and Peitsch 1997; <http://www.expasy.org/spdbv/>). In the absence of heavy atoms in the side-chains the program generated a warning notice about the type and location of structural defects. Lack of amino acid residues was detected by using Accelrys Discovery Studio Visualizer 3.0 (Accelrys Software Inc. – <http://accelrys.com/>).

Reconstruction of 3D model of E. coli FtsZ protein

Complete amino acid sequence of *E. coli* O157:H7 FtsZ (P0A9A8) (Perna et al. 2001) was downloaded from UniProt (<http://www.uniprot.org/>) database (The UniProt Consortium 2008). A three-dimensional structural modelling was carried out on the I-TASSER server (Roy et al. 2010; <http://zhanglab.cmb.med.umich.edu/I-TASSER>) and with »SwissModel Automatic Modelling Mode« of »SWISS-MODEL Workspace« server (<http://swissmodel.expasy.org/>) (Arnold et al. 2006). Both servers running in automatic mode of PDB structure (template) selection. As a result we generated 5 models of three-dimensional structures with I-TASSER, and and one more model with »SWISS-MODEL Workspace«. Additionally, with »SWISS-MODEL Workspace« server we constructed another model based on 2Q1Y (Chain A) template PDB structure which was specified in the manual mode.

Root mean square deviations (RMSD) of the fitted 3-D structures were calculated using »molecule align« tool of PyMol 1.4 package (www.pymol.org).

Estimation of protein model quality

The 3-D structures quality was assessed by processing the models on the MolProbity server. (Chen et al. 2010, <http://molprobity.biochem.duke.edu/>) This study was performed to estimate the statistics of all-atom contacts (i.e. »all atoms Clashscore«) and protein geometry: defined percentages of poor rotamers, Ramachandran outliers, Ramachandran favoured, residues with bad bonds, residues with bad angles and defined C β deviations >0.25Å and MolProbity score.

We used the Protein Structure and Model Assessment Tools available at »SWISS-MODEL Workspace« server (<http://swissmodel.expasy.org>) to assess the quality of the 3-D models based on Raw-score and Z-score of QMEAN6 (Composite scoring function for model quality estimation) (Benkert et al. 2009) and global model quality estimation based on »DFire energy« (all-atom distance-dependent statistical potential) (Zhou and Zhou 2002).

Results

Selection and quality checking of X-ray Protein Data Bank structures of M. tuberculosis FtsZ

Scanning the Worldwide Protein Data Bank (wwPDB) we revealed several crystal structures of the *M. tuberculosis* cell division protein FtsZ, determined at 1.86 to 2,60 Å by X-ray method. The following PDB structures have been studied: 1RLU (Chains: A, B) and 2Q1Y (Chains: A, B) X-rays of FtsZ-GTPgammaS (5'-guanosine diphosphate monothiophosphate) complexes (Chains: A, B), 1RQ2 (Chains: A, B) and 2Q1X (Chains: A, B) X-rays of FtsZ-citrate complexes (Chains: A, B) and 1RQ7 X-ray of FtsZ-GDP complex (Chains: A, B). Using pairwise sequence alignment of polypeptide chains A and B from X-ray PDB structures with complete sequences from UniProt we tested them on presence of gaps (meaning occurrence of defects in structures). PDB-structures were verified on the presence of such artifacts as deficiency of heavy atoms (carbon, oxygen and nitrogen) in side chains of individual amino acid residues (using DeepView-Swiss-PdbViewer 4.0.3 software package). All gaps in polypeptide

chains and residues with defective side chains were checked and represented in the Figure 1 and Table 1.

As a result, of the ten available X-ray structures (considering chains A and B) the only one chain of *M. tuberculosis* FtsZ protein was selected as a most complete experimentally proved structure and the base of subsequent work on its detailed 3-D reconstruction and *in silico* analysis. Such structure was a chain A of X-ray FtsZ-GTP-gamma-S complex from 2Q1Y (2.30 Å, R-value=0.174,

R-free=0.210) (DOI:10.2210/pdb2q1y/pdb). It has not just first 7 (N-end) and the last 66 (C-end) amino acid residues, but, as opposed to the same A-chain of 1RLU (2.08 Å, R-value=0.182, R-free=0.224) (Leung et al., 2004), has a complete atomic composition of all available amino acids. The 2Q1Y chain A were analyzed with MolProbity server, and the values of evaluation functions (see Table 2), demonstrate its high quality for further *in silico* experiments and modelling its interactions with low molecular weight compounds.

Table 1: Features and identified defects in Protein Data Bank X-ray structures of *Mycobacterium tuberculosis* FtsZ.

Tabela 1: Lastnosti in okvare struktur PDB v proteinu FtsZ iz bakterije *Mycobacterium tuberculosis*.

PDB Structure	Method	Resolution, Å	Chain	Defective regions of molecules		
				N-terminal tile	Tubulin/FtsZ family, GTPase domain	C-terminal tail
1RLU	X-Ray Diffraction	2.35 Å	A	aa:M1-Y7	ha: E29, K33, R64, L66, R181	aa:D313 – R379; ha: K236, E252
			B	aa:M1-H5	aa:R60-G70, G170-A173; ha: N6, L8, E73, K77, K120, R140	aa:D313 – R379; ha: K236, D301
1RQ2	X-Ray Diffraction	1.86 Å	A	aa:M1-Y7	aa:R64-A69; ha: K33, M177	aa:D313 – R379; ha: K236
			B	aa:M1-Y7	aa:R60-G70, D171-A173; ha: K33, Q45, E73	aa:V314-R379
			A	aa:M1-Y7	aa:R64-A69; ha: K33, L48	aa:D313 – R379; ha: K236, Q255
1RQ7	X-Ray Diffraction	2.60 Å	B	aa:M1-H5	aa:R60-G70, D171-A173; ha: K33, E73, K120, R140, S141, E153	aa:V314-R379; ha: K236, D313
2Q1X	X-Ray Diffraction	2.35 Å	A	aa:M1-Y7	aa:T63-G70	aa:D313 – R379
			B	aa:M1-H5	aa:R60-A71, R140-N142, Q168-A173	aa:D313 – R379
2Q1Y	X-Ray Diffraction	2.30 Å	A	aa:M1-Y7	-	aa:D313 – R379
			B	aa:M1-H5	aa:R60-G70, G170-A173	aa:D313 – R379

aa – lack of respective amino acid residues in the X-ray structure;

ha – absence of heavy atoms (carbon, oxygen or nitrogen) in the side chains of respective amino acid residues.

spP64170 M. tuberculosis	▼1	▼7	▼60	▼71	▼142	▼168	▼173
1RUJ_A_FTSZ_MYCTU	MTPPHNYLAVIKVIGIGGGVNAVNRMTIEOGLKGVEFFIAINTDAQALLMSDADVKLVGRDSTRIGIGAGACAPEVGRKAAEDAKDEIEELLRGADMVFVTA						
1RUJ_B_FTSZ_MYCTU	-----LAVIKVIGIGGGVNAVNRMTIEOGLKGVEFFIAINTDAQALLMSDADVKLVGRDSTRIGIGAGACAPEVGRKAAEDAKDEIEELLRGADMVFVTA						
1RQ2_A_FTSZ_MYCTU	-----NYLAVIKVIGIGGGVNAVNRMTIEOGLKGVEFFIAINTDAQALLMSDADVKLVGRDSTRIGIGAGACAPEVGRKAAEDAKDEIEELLRGADMVFVTA						
1RQ2_B_FTSZ_MYCTU	-----LAVIKVIGIGGGVNAVNRMTIEOGLKGVEFFIAINTDAQALLMSDADVKLVGRDSTRIGIGAGACAPEVGRKAAEDAKDEIEELLRGADMVFVTA						
1RQ7_A_FTSZ_MYCTU	-----LAVIKVIGIGGGVNAVNRMTIEOGLKGVEFFIAINTDAQALLMSDADVKLVGRDSTRIGIGAGACAPEVGRKAAEDAKDEIEELLRGADMVFVTA						
1RQ7_B_FTSZ_MYCTU	-----LAVIKVIGIGGGVNAVNRMTIEOGLKGVEFFIAINTDAQALLMSDADVKLVGRDSTRIGIGAGACAPEVGRKAAEDAKDEIEELLRGADMVFVTA						
2Q1X_A_FTSZ_MYCTU	-----LAVIKVIGIGGGVNAVNRMTIEOGLKGVEFFIAINTDAQALLMSDADVKLVGRDSTRIGIGAGACAPEVGRKAAEDAKDEIEELLRGADMVFVTA						
2Q1Y_A_FTSZ_MYCTU	-----LAVIKVIGIGGGVNAVNRMTIEOGLKGVEFFIAINTDAQALLMSDADVKLVGRDSTRIGIGAGACAPEVGRKAAEDAKDEIEELLRGADMVFVTA						
2Q1Y_B_FTSZ_MYCTU	-----NYLAVIKVIGIGGGVNAVNRMTIEOGLKGVEFFIAINTDAQALLMSDADVKLVGRDSTRIGIGAGACAPEVGRKAAEDAKDEIEELLRGADMVFVTA						
spP64170 M. tuberculosis	140▼▼142	168▼	▼173				
1RUJ_A_FTSZ_MYCTU	GEGGGTGCGAPWASTARKLGALTGVVTPRPSFGKERSNOAENGAIALRESCTDIIIVIPNDRLLQMGDAAVSLMDAFRSADDEVLLNGVQGITDIIITT						
1RUJ_B_FTSZ_MYCTU	-----GEGGGTGCGAPWASTARKLGALTGVVTPRPSFGKERSNOAENGAIALRESCTDIIIVIPNDRLLQMGDAAVSLMDAFRSADDEVLLNGVQGITDIIITT						
1RQ2_A_FTSZ_MYCTU	-----GEGGGTGCGAPWASTARKLGALTGVVTPRPSFGKERSNOAENGAIALRESCTDIIIVIPNDRLLQMGDAAVSLMDAFRSADDEVLLNGVQGITDIIITT						
1RQ2_B_FTSZ_MYCTU	-----GEGGGTGCGAPWASTARKLGALTGVVTPRPSFGKERSNOAENGAIALRESCTDIIIVIPNDRLLQMGDAAVSLMDAFRSADDEVLLNGVQGITDIIITT						
1RQ7_A_FTSZ_MYCTU	-----GEGGGTGCGAPWASTARKLGALTGVVTPRPSFGKERSNOAENGAIALRESCTDIIIVIPNDRLLQMGDAAVSLMDAFRSADDEVLLNGVQGITDIIITT						
1RQ7_B_FTSZ_MYCTU	-----GEGGGTGCGAPWASTARKLGALTGVVTPRPSFGKERSNOAENGAIALRESCTDIIIVIPNDRLLQMGDAAVSLMDAFRSADDEVLLNGVQGITDIIITT						
2Q1X_A_FTSZ_MYCTU	-----GEGGGTGCGAPWASTARKLGALTGVVTPRPSFGKERSNOAENGAIALRESCTDIIIVIPNDRLLQMGDAAVSLMDAFRSADDEVLLNGVQGITDIIITT						
2Q1X_B_FTSZ_MYCTU	-----GEGGGTGCGAPWASTARKLGALTGVVTPRPSFGKERSNOAENGAIALRESCTDIIIVIPNDRLLQMGDAAVSLMDAFRSADDEVLLNGVQGITDIIITT						
2Q1Y_A_FTSZ_MYCTU	-----GEGGGTGCGAPWASTARKLGALTGVVTPRPSFGKERSNOAENGAIALRESCTDIIIVIPNDRLLQMGDAAVSLMDAFRSADDEVLLNGVQGITDIIITT						
2Q1Y_B_FTSZ_MYCTU	-----GEGGGTGCGAPWASTARKLGALTGVVTPRPSFGKERSNOAENGAIALRESCTDIIIVIPNDRLLQMGDAAVSLMDAFRSADDEVLLNGVQGITDIIITT						
spP64170 M. tuberculosis	▼313	▼379					
1RUJ_A_FTSZ_MYCTU	PGLINVDFAVDKIMSGAGTALMGIGSARGEGRSLKAAEIAINSPILLEASMEGAQGVLMISAGGSDGLGFIPEINEAASLVQDAAHPDANIIFGTVIDDSLG						
1RUJ_B_FTSZ_MYCTU	-----PGLINVDFAVDKIMSGAGTALMGIGSARGEGRSLKAAEIAINSPILLEASMEGAQGVLMISAGGSDGLGFIPEINEAASLVQDAAHPDANIIFGTVIDDSLG						
1RQ2_A_FTSZ_MYCTU	-----PGLINVDFAVDKIMSGAGTALMGIGSARGEGRSLKAAEIAINSPILLEASMEGAQGVLMISAGGSDGLGFIPEINEAASLVQDAAHPDANIIFGTVIDDSLG						
1RQ2_B_FTSZ_MYCTU	-----PGLINVDFAVDKIMSGAGTALMGIGSARGEGRSLKAAEIAINSPILLEASMEGAQGVLMISAGGSDGLGFIPEINEAASLVQDAAHPDANIIFGTVIDDSLG						
1RQ7_A_FTSZ_MYCTU	-----PGLINVDFAVDKIMSGAGTALMGIGSARGEGRSLKAAEIAINSPILLEASMEGAQGVLMISAGGSDGLGFIPEINEAASLVQDAAHPDANIIFGTVIDDSLG						
1RQ7_B_FTSZ_MYCTU	-----PGLINVDFAVDKIMSGAGTALMGIGSARGEGRSLKAAEIAINSPILLEASMEGAQGVLMISAGGSDGLGFIPEINEAASLVQDAAHPDANIIFGTVIDDSLG						
2Q1X_A_FTSZ_MYCTU	-----PGLINVDFAVDKIMSGAGTALMGIGSARGEGRSLKAAEIAINSPILLEASMEGAQGVLMISAGGSDGLGFIPEINEAASLVQDAAHPDANIIFGTVIDDSLG						
2Q1X_B_FTSZ_MYCTU	-----PGLINVDFAVDKIMSGAGTALMGIGSARGEGRSLKAAEIAINSPILLEASMEGAQGVLMISAGGSDGLGFIPEINEAASLVQDAAHPDANIIFGTVIDDSLG						
2Q1Y_A_FTSZ_MYCTU	-----PGLINVDFAVDKIMSGAGTALMGIGSARGEGRSLKAAEIAINSPILLEASMEGAQGVLMISAGGSDGLGFIPEINEAASLVQDAAHPDANIIFGTVIDDSLG						
2Q1Y_B_FTSZ_MYCTU	-----PGLINVDFAVDKIMSGAGTALMGIGSARGEGRSLKAAEIAINSPILLEASMEGAQGVLMISAGGSDGLGFIPEINEAASLVQDAAHPDANIIFGTVIDDSLG						
spP64170 M. tuberculosis	DEVRTVTIAAGFDVSGPRKPVMGETGGAHRIESAAGKLISITLFEFVDVAVSVPLHTNGATLSIGDDDDVDVPPFMR						
1RUJ_A_FTSZ_MYCTU	-----DEVRTVTIAAGF-----						
1RUJ_B_FTSZ_MYCTU	-----DEVRTVTIAAGF-----						
1RQ2_A_FTSZ_MYCTU	-----DEVRTVTIAAGF-----						
1RQ2_B_FTSZ_MYCTU	-----DEVRTVTIAAGFD-----						
1RQ7_A_FTSZ_MYCTU	-----DEVRTVTIAAGF-----						
1RQ7_B_FTSZ_MYCTU	-----DEVRTVTIAAGFD-----						
2Q1X_A_FTSZ_MYCTU	-----DEVRTVTIAAGF-----						
2Q1X_B_FTSZ_MYCTU	-----DEVRTVTIAAGF-----						
2Q1Y_A_FTSZ_MYCTU	-----DEVRTVTIAAGF-----						
2Q1Y_B_FTSZ_MYCTU	-----DEVRTVTIAAGF-----						

Figure 1: Numerous alignment of the full-length sequence of FtsZ protein (UniProtKB entry: P64170) from *Mycobacterium tuberculosis* and sequences of corresponding X-Ray Protein Data Bank structures: 1RUJ, 1RQ2, 1RQ7, 2Q1X and 2Q1Y.

Slikal: Različice sekvence proteina FtsZ (UniProtKB (P64170)) iz bakterije *Mycobacterium tuberculosis* in sekvence pripadajočih struktur proteinov FtsZ z X žarki: 1RUJ, 1RQ2, 1RQ7, 2Q1X and 2Q1Y.

E. coli FtsZ protein spatial structure prediction

Despite the great interest in mycobacterial FtsZ as the target for antibacterial compounds, majority of commercial analytical kits for *in vitro* binding experiments are more available for FtsZ protein from *E. coli* than its mycobacterial homolog. So, here we have a paradox situation, the presence of well proven three-dimensional structure of *M. tuberculosis* FtsZ protein on the one hand, and at the other hand, the fact that majority of the experimental tools targeted *E. coli* FtsZ, for which there is a clear gap in 3D-structure research. So, now, we have only 1F47 PDB structure presented only by last 17 amino acid residues (Lys367-Asp383), forming a short unstructured element, ending with two α -helix turns in C-end (Mosyak et al. 2000). In order to solve this problem we applied *in silico* homology modelling.

Initially the sequence of *E. coli* FtsZ protein (UniProt: P0A9A6) has been sent to the »SWISS-MODEL Workspace« server, for model building (alignment). With completely automatic modelling of *E. coli* FtsZ protein, server selected the chain B of the 1OFU X-ray structure from the Sula-FtsZ complex (2.10 Å, R-value=0.216, R-free=0.255) from *P. aeruginosa* (Cordell et al., 2003) as the template structure. »SWISS-MODEL Workspace« server generated one model for target FtsZ protein of 293 aminoacid residues in length (from Asn24 to Gly316 inclusive). The secondary and tertiary structures of model was completely similar to those in 1OFU except absence of small area in the N-end, covering the first β -fold and significant part of the next α -helix (Fig. 3a). In the structure of the matrix protein these elements are present (Ala11-Gly23 in FtsZ protein from *E. coli*, and Ala12-Gly24 in FtsZ protein from *P. aeruginosa*). When the chain A of 2Q1Y structure was assigned as a matrix (previously selected as the best X-ray structure of *M. tuberculosis* FtsZ protein, see above), »SWISS-MODEL Workspace« server built model of *E. coli* FtsZ protein, which contained in its structure above-mentioned region (Fig. 2b) and was 305 amino acid residues in length (from Ala11 to Ile315 inclusive). Sequences of these structural areas in *E. coli* and *M. tuberculosis* FtsZ proteins have differences in amino acid residues: Ile16-Val14, Val18-Ile16 and G23-V20, respectively. Fitting of these two *E. coli* FtsZ structures,

based on different modelling matrixes (Fig. 2c), demonstrate high level of structural similarity confirmed by root-mean-square deviation of C α -atoms (RMSD=0.862 Å) (Höltje et al. 2008).

E. coli FtsZ protein spatial structure modelling performed by using I-TASSER server

In parallel with using of classical template-based modelling, we have applied 3D-reconstruction using on-line I-TASSER server (<http://zhanglab.ccmb.med.umich.edu>). I-TASSER 3D-models are built based on multiple-threading alignments by LOMETS (Wu and Zhang 2007) and iterative TASSER assembly simulations; function insights are then derived by matching the predicted models with protein function databases. (Roy et al. 2010) Following the sequence-to-structure-to-function paradigm, the I-TASSER procedure (Roy et al. 2010) for structure and function modelling involves four consecutive steps of: (a) template identification by LOMETS (Wu and Zhang 2007); (b) fragment structure reassembly by replica-exchange Monte Carlo simulations (Zhang et al. 2002); (c) atomic level structure refinement using REMO (Li et al. 2009) and FG-MD (Zhang et al. 2011); and (d) structure-based function interpretations using COFACTOR (Roy et al. 2011).

We submitted request of *E. coli* FtsZ-protein sequence (UniProt: P0A9A6) at the I-TASSER server using on default settings without manual assignment of template structure. As seen from Table 3, the server align our query sequence with a range of such template PDB structures as 2VAW (Chain A), 1FSZ (Chain A), 2VAM (Chain A), 2R6R (Chain A), 1W5F (Chain A) and 2RHL (Chain B). As a result, were generated five I-TASSER models of *E. coli* FtsZ protein of full length (from Met1 to Asp383) with different values of C-score (Table 4, Fig. 3). Model 1 has the highest value of C-score, indicating it as optimal model structure among generated by I-TASSER server. Based on server statistics Model 1 was also characterize by TM-score (template modelling score) = 0.56 \pm 0.15 and root-mean-square deviation (RMSD) = 9.6 \pm 4.6Å.

On the Figure 4 we present results of I-TASSER secondary structure and properties prediction for *E. coli* FtsZ protein. Confidence score values for the predicted structures are also indicating a

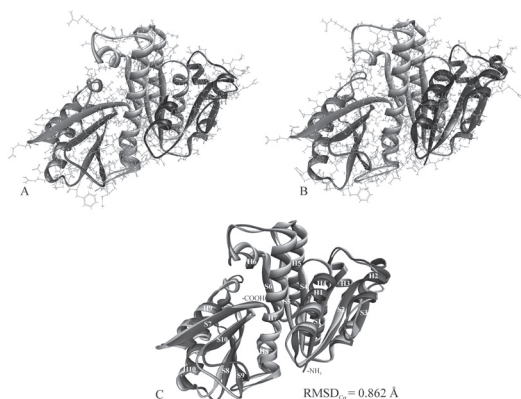


Figure 2: Spatial structure models of FtsZ protein from *Escherichia coli*, built with the protein structure homology-modeling server »SWISS-MODEL Workspace« using X-ray template PDB structures of different origin: A – FtsZ from *E. coli* constructed on template 1OFU (Chain B) from *P. aeruginosa*; B – FtsZ from *E. coli* constructed on template 2Q1Y (Chain A) from *M. tuberculosis*; C – Fitting of constructed 3-D models of FtsZ protein from *E. coli* (light gray – based on 1OFU (Chain B) template, dark gray – based on 2Q1Y (Chain A) template). Protein models presented as solid ribbon diagram with side-chain atoms shown as lines. 3-D models were visualized in Accelrys Discovery Studio Visualizer.

Slika 2: Tri dimenzionalni model FtsZ iz bakterije *Escherichia coli* narejen na serverju »SWISS-MODEL Workspace« ob uporabi matric različnih izvorov: A – FtsZ iz *E. coli* narejen na matrici 1OFU_B (vir – *P. aeruginosa*), B – FtsZ iz *E. coli* narejen na matrici 2Q1Y_A (vir – *M. tuberculosis*), C – Prileganje modela FtsZ narejenega iz *Escherichia coli* (svetlo sivo – matrica 1OFU_B, temno siva – matrica 2Q1Y_A). Model proteina predstavljen kot polni trak s stranskimi verigami atomov (linije). Izdelan v Discovery Studio Visualizer.

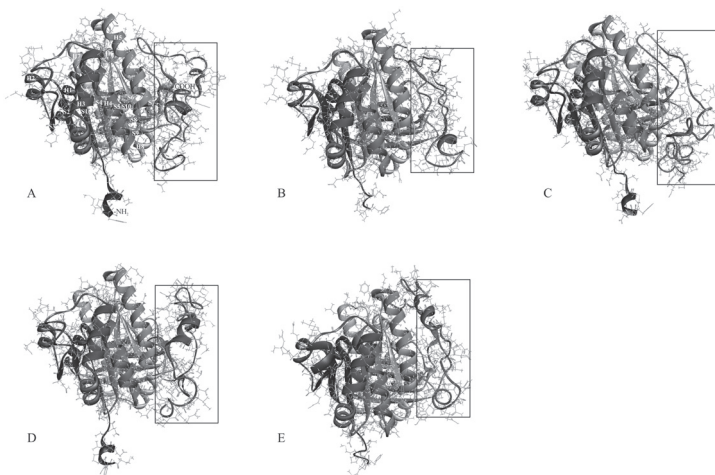


Figure 3: Spatial models of FtsZ *Escherichia coli* built on I-TASSER server.

A – model 1, B – model 2, C – model 3, D – model 4 and E – model 5. Protein models presented as solid ribbon diagram with side-chain atoms shown as lines. Boxes mark the C-terminal region that is absent in all crystal structures of FtsZ proteins and was overbuilt by the server based on sequence similarity to the regions of other classes proteins. Generated in Discovery Studio Visualizer.

Slika 3: Prostorski model FtsZ iz *Escherichia coli* narejen na I-TASSER serverju:

A – model 1, B – model 2, C – model 3, D – model 4 in E – model 5. Modeli proteinov so predstavljeni kot polni trak s stranskimi verigami atomov (linije). Okvir označuje C terminalno regijo, ki je odsotna v vseh strukturah kristala proteina FtsZ in je dograjena na serverju na osnovni podobnosti zaporedja v regijah pri ostalih proteinih. Izdelan v Discovery Studio Visualizer.

rather high quality of all built models. All this along with a variety of conformational representations of C-terminal region provided by five server-built models (Fig. 4) can not regard any from these structures as the quite probable and suitable for further bioinformatic research.

Quality evaluation of *E. coli* FtsZ protein models

All seven models of *E. coli* FtsZ were examined using the online MolProbity server and »Protein Structure & Model Assessment Tools« of »SWISS-MODEL Workspace« server.

Table 2: Results of evaluation of structure quality of chain 2Q1Y_A FtsZ *Mycobacterium tuberculosis* on a MolProbity server.

Tabela 2: Rezultati vrednotenja kakovosti strukture verige 2Q1Y_A FtsZ iz bakterije *Mycobacterium tuberculosis* na serverju MolProbity.

All-AtomContacts	Clashscore for all atoms:	4.13	96 th percentile*
	Poor rotamers	0.00%	Goal: < 1%
	Ramachandran outliers	0.00%	Goal: < 0.2%
	Ramachandran favored	100.00%	Goal: > 98%
	C β deviations > 0.25 Å	0	Goal: 0
	MolProbity score	1.20	99 th percentile*
	Residues with bad bonds:	0.00%	Goal: 0%
	Residues with bad angles:	0.00%	Goal: < 0.1%
	ProteinGeometry		

Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.

* 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst.

^ MolProbity score is defined as the following: $0.42574 * \log(1 + \text{clashscore}) + 0.32996 * \log(1 + \max(0, \text{pctRotOut} - 1)) + 0.24979 * \log(1 + \max(0, 100 - \text{pctRamaFavored} - 2)) + 0.5$

Table 3: Top 10 template X-Ray structures selected by I-TASSER server for homology modelling of *Escherichia coli* FtsZ protein

Tabela 3: Zgornjih 10 matric struktur z X žarki uporabljenih na I-TASSER-ju za modeliranje homolognosti FtsZ proteina bakterije *Escherichia coli*.

Rank	PDB Hit	Ident1	Ident2	Cov.	Norm. Z-score
1	2VAW (Chain A)	0.67	0.55	0.82	3.46
2	2VAW (Chain A)	0.67	0.55	0.82	5.88
3	1FSZ (Chain A)	0.44	0.40	0.86	6.38
4	2VAM (Chain A)	0.54	0.43	0.79	4.68
5	2R6R (Chain A)	0.46	0.39	0.84	6.20
6	2VAW (Chain A)	0.67	0.55	0.82	0.00
7	2VAW (Chain A)	0.67	0.55	0.82	13.50
8	1FSZ (Chain A)	0.48	0.26	0.53	5.77
9	1W5F (Chain A)	0.46	0.38	0.81	6.87
10	2RHL (Chain A)	0.52	0.43	0.82	5.00

(a) Rank of templates represents the top ten threading templates used by I-TASSER.

(b) Ident1 is the percentage sequence identity of the templates in the threading aligned region with the query sequence.

(c) Ident2 is the percentage sequence identity of the whole template chains with query sequence.

(d) Cov. represents the coverage of the threading alignment and is equal to the number of aligned residues divided by the length of query protein.

(e) Norm. Z-score is the normalized Z-score of the threading alignments. Alignment with a Normalized Z-score >1 mean a good alignment and vice versa.

Table 4: MolProbity server and »Protein Structure and Model Assessment Tools« of »SWISSMODEL Workspace« server structure quality evaluation for 3-D models of *Escherichia coli* FtsZ protein.

Tabela 4: Rezultati vrednotenja kakovosti strukture 3-D modela FtsZ iz bakterije *Escherichia coli* na serverju MolProbity in »Protein Structure and Model Assessment Tools« na serverju »SWISS-MODEL Workspace«.

Verification Type	Models from I-Tasser			Template based models from Swiss-Model			Goal
	Model	Model	Model	Model	Model	Model	
	1-1.239	2-1.726	3-2.056	4-2.638	5-2.396	IOFU_B 2Q1Y_A	
All-Atom contacts	121.27(0 th percentile*)	177.41(0 th percentile*)	122.68(0 th percentile*)	141.22(0 th percentile*)	131.33(0 th percentile*)	38.41(9 th percentile*)	48.84(4 th percentile*)
Poor rotamers	1.36%	2.38%	2.38%	3.06%	2.72%	1.83%	0.00%
Ramachandran outliers	3.15%	2.89%	3.41%	3.94%	1.84%	0.34%	0.33%
Ramachandran favored	92.39%	89.76%	89.50%	89.76%	92.65%	95.53%	98.68%
Protein geometry	9	13	5	9	3	2	0
C β deviations >0.25 Å	3.12(19 th percentile*)	3.55(8 th percentile*)	3.40(11 th percentile*)	3.54(8 th percentile*)	3.37(11 th percentile*)	2.58(43 rd percentile*)	2.16(67 th percentile*)
MolProbity score [^]	0.00%	0.52%	0.00%	0.00%	0.00%	0.00%	0.00%
Residues with bad bonds:	5.74%	6.27%	5.74%	4.18%	4.70%	0.00%	0.00%
Residues with bad angles:	-494.67	-455.72	-471.66	-461.74	-471.66	-377.70	-412.32
DFire energy	0.614	0.62	0.628	0.595	0.592	0.684	0.747
QMEAN Raw score 6	-1.79	-1.72	-1.62	-2.01	-2.04	-0.96	-0.31
QMEAN Z-score 6							

Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.

* 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst.

[^] MolProbity score is defined as the following: $0.42574 * \log(1 + \text{clashscore}) + 0.32996 * \log(1 + \text{max}(0, \text{petRotOut} - 1)) + 0.24979 * \log(1 + \text{max}(0, 100 - \text{petRamaFavored} - 2)) + 0.5$.

DFire is an all-atom statistical potential based on a distance-scaled finite ideal-gas reference state. It's used to assess non-bonded atomic interactions in the protein model. A lower energy indicates that a model is closer to the native conformation.

QMEAN6 scoring function is a linear combination of six structural descriptors using statistical potentials: The local geometry is analysed by a torsion angle potential over three consecutive amino acids. Two distance-dependent interaction potentials are used to assess long-range interactions: the first is a residue-level implementation based on C-beta atoms only and the second an all-atom potential which is able to capture more details of the model. A solvation potential investigates the burial status of the residues. Two additional terms describing the agreement of the predicted (from sequence) and the calculated secondary structure and solvent accessibility of the model.

Evaluation of all functions demonstrates a significantly higher quality of models, built with SWISS-MODEL server (Table 4). Evaluation of all-atom contacts, namely the number of serious steric overlaps of all atoms, allows us to determine model based on PDB matrix structure 1OFU (Chain B) as the best among all *E. coli* FtsZ protein models. At the same time, evaluation indexes of protein geometry, listed in Table 4, characterize another SWISS-MODEL structure, based on PDB template 2Q1Y (Chain A), as the most accurate.

According to all-atom statistical potential DFire index, among I-TASSER models, the undisputed leader are the Model 1 and an outsider – Model 4. At the same time, among SWISS-MODEL reconstructed models, structure, based on 2Q1Y (Chain A), demonstrate substantially superior quality in comparison with model based on *Pseudomonas* FtsZ structure. Evaluation indices »Raw score« and »Z-score« of QMEAN6, demonstrate undeniable superiority of *E. coli* FtsZ model based on 2Q1Y (Chain A) template.

Discussion

Choice of FtsZ M. tuberculosis model

As shown in Figure 1, polypeptide chains of all FtsZ structures deposited in the database are lacking the first 5-7 a.a. residues of N-terminal region and 65-66 residues of C-terminal region (in fact, the last C-terminal region is absent at all). Also, eight chains, which are representing all five of the studied PDB-structures of FtsZ *M. tuberculosis* (in 1RLU and 2Q1Y – only Chains B), contain 5-10 residues gap within the loop, which corresponds to the loop between sheet S3 and helix H3 from 1FSZ – crystal structure of FtsZ protein from *M. jannaschii* (Oliva et al. 2004). B-chains of each structure are characterized by the absence of 3-6 residues in the loop that extend to helix H7. Additionally, the chain B of 2Q1X has a short gap (3 a.a. residues) within the α -helix that corresponds to helix H5 of 1FSZ. Relative to the artifacts of residues, only 2Q1X and 2Q1Y crystals have all heavy atoms in all amino acid residues of structure (Table 1). Both of this two structures of FtsZ *M. tuberculosis* are

characterized by complete composition of amino acid side chains, whereas the X-Ray structure obtained by Leung et al. 2004 (PDB: 1RQ7) are suffer from lack of some heavy atoms in at least three residues in each chain.

Thus, Chain A from the 2Q1Y structure, which does not possess the seven initial (N-end) and sixty six C-end amino acid residues but in contrast to the Chain A from 1RLU has full atom composition of all available amino acid residues, in our opinion is a good three-dimensional model of FtsZ *M. tuberculosis*. This assumption was confirmed by the scores of all MolProbity quality evaluating functions for this structure (model) (Table 2).

Construction and verification of FtsZ E. coli models

The »SWISS-MODEL Workspace« is a web-based integrated service dedicated to protein structure homology modelling. It assists in building protein homology models at different levels of complexity. Successful model building requires at least one experimentally determined 3D structure (template) that shows significant amino acid sequence similarity with the target sequence. Building a homology model comprises four main steps: identification of structural template(s), alignment of target sequence and template structure(s), model building, and model quality evaluation. These steps can be repeated until a satisfying modelling result is achieved. Each of the four steps requires specialized software and access to up-to-date protein sequence and structure databases. In full automatic mode server performs all four steps itself and gives the complete three-dimensional model of the target protein with specified matrix structure. There is also a possibility to specify the necessary structure as the matrix manually (Arnold et al. 2006).

Thus, in full automatic modelling of FtsZ *E. coli* server used Chain B from 1OFU structure as a matrix. It should be noted that selection of this object as a matrix was based on the origin of these two bacterial species. Both of them are belonging to the class γ -proteobacteria from phylum Proteobacteria. Although orders of these species are different, an affiliation to a common class is important enough to explain/suppose a higher level of similarity between sequence of

FtsZ from *E. coli* and *P. aeruginosa* (Vaughan et al. 2004, Demchuk and Blume 2005), than with sequences from other bacterial species from different phylums for which crystal structures are also presented in PDB database.

However, such relatively close relationship between *E. coli* and *P. aeruginosa* was not sufficient to build a complete model of FtsZ *E. coli*. Incomplete N-terminal end of this model starting with the Asn24, which indicates the absence of first beta-sheet and significant part of the next alpha-helix in the model structure (Fig. 2a), although in the structure of the matrix this elements are present (Ala11-Gly23 in FtsZ from *E. coli* and Ala12-Gly24 in FtsZ from *P. aeruginosa*). This artifact is surprising because of the fact that primary sequence of missing part of FtsZ *E. coli* is completely similar to that area in the template structure – FtsZ protein from *P. aeruginosa*.

We can only assume some failure in the algorithm of »SWISS-MODEL Workspace« server taking note the fact that the using a chain 2Q1Y_A as specific template allowed server to generate FtsZ *E. coli* model, which contains the mentioned area in its structure (Fig. 2b) and that is why it was stretched from the Ala11 to Ile315 inclusive.

C α -atom RMSD of obtained FtsZ *E. coli* models, which were built on different matrices (Fig. 2c), is 0.862 Å. This indicates a significant structural similarity of the obtained models (Gu and Bourne 2009). And it can become an additional demonstration of the structure conservatism of this class of proteins (Erickson, 1998) and significant modelling accuracy while taking proteins from bacteria phylums class as a template. Minor differences observed in the turn of main chain (Fig. 2c) are confined to the unstructured elements between β -sheets and alpha-helices and fully capable of leveling through the lability of the secondary structure elements in the time (Nyporko and Blume 2001).

In parallel with the modelling on »SWISS-MODEL Workspace« server we have performed modelling of FtsZ *E. coli* on I-TASSER server, an Internet service for protein structure and function predictions. It built 3-D models based on multiple-threading alignments by LOMETS and iterative TASSER assembly simulations; function insights are then derived by matching the predicted models with protein function databases. I-TASSER (as

‘Zhang-Server’) was ranked as the No 1 server for protein structure prediction in recent CASP7, CASP8 and CASP9 experiments (<http://predictioncenter.org/>). It was also ranked as the best for function prediction in CASP9. CASP (or Critical Assessment of Techniques for Protein Structure Prediction) is a community-wide experiment for testing the state-of-the-art of protein structure predictions which takes place every two years since 1994. The experiment (often referred as a competition) is strictly blind because the structures of testing proteins are unknown to the predictors (Roy et al. 2010, <http://zhanglab.ccmb.med.umich.edu/I-TASSER/about.html>).

As a result, server generated five potential full length models of FtsZ proteins from *E. coli* (Fig. 3) with different values of C-score (see Table 4). C-scores a confidence score for estimating the quality of predicted models by I-TASSER. It is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations. C-score is typically in the range of [-5,2], where a C-score of higher value signifies a model with a high confidence and vice-versa (Roy et al. 2010). As it is shown in the Table 4 the best value of this parameter belongs to the model 1.

I-TASSER server estimated accuracy for this model: 0.56 ± 0.15 TM-score and 9.6 ± 4.6 Å RMSD. TM-score and RMSD are known standards of structural similarity between two structures which are usually used as measure of model accuracy when the native structure is known. In case where the native structure is not known, it becomes necessary to predict the quality of the modelling prediction, i.e. what is the distance between the predicted model and the native structures. TM-score is a recently proposed scale for measuring the structural similarity between two structures (Zhang and Skolnick 2004). The purpose of proposing TM-score is to solve the problem of RMSD which is sensitive to the local error. Because RMSD is an average distance of all residue pairs in two structures, a local error (e.g. a misorientation of the tail) may cause a significant RMSD value although the global topology may be correct. In TM-score, however, the small distance is weighted stronger than the big distance which makes the score insensitive to the local modelling error. A TM-score >0.5 indicates a model of

correct topology and a TM-score <0.17 means a random similarity. This cutoff does not depend on the protein length (Roy et al. 2010). Thus, the quality of prediction of model 1 is quite satisfactory for these scores.

On Fig. 3 we presented a predicted secondary structures and properties of FtsZ protein from *E. coli* models generated by I-TASSER. It should be noted that the distribution of secondary structure elements (shown as H for alpha helixes, S for beta sheets and C for coils) matches with that for crystal structures of FtsZ (see figure 2 in Oliva et al., 2007). Confidence score values for the proposed structure also indicate a rather high quality of all predicted models. However, it should be noted that the value of this estimated parameter are at a slightly lower level for the C-tail region (which is not represented in any of the crystal structures of FtsZ proteins) in comparison with the globular N- and C-domains. We can speculate the possibility of predicted by I-TASSER server unstructured conformation of short N-terminal region in the structure of FtsZ *E. coli* because analogs are observed in structures 2VAW and 1FSZ. But predicted secondary structure of long and not structured C-terminal region between B10-sheet and final α -helix, provided in the FtsZ *E. coli* (Mosyak et al. 2000), does not inspire confidence in us.

Low values of Confidence score and some crystal structures of FtsZ testified against coiled coil structure of this C-terminal region provided by I-TASSER server. Thus, in contrast to several crystal FtsZ structures that are starting with β -sheet B1 and breaking immediately after the β -sheet B10 (1OFU, 1RQ7, 1RLU, 1RQ2, 2Q1X and 2Q1Y), FtsZ structures from 2R6R and 2R75 as well as 2VAP and 1FSZ have two additional β -sheets B11 and B12. 1W5X structure has also a short α -helix between B10 and B11, which is terminated element of C-end region in structures 1W5F, 2RHH and 2RHJ. All together with a variety of conformational representations of this C-terminal region in five models provided by I-TASSER server (Fig. 4) can not regard any of these models as the most likely and suitable for further bioinformatics researches.

For the final determination of the best potential model of FtsZ *E. coli* among all constructed we have verified all seven candidates with the online server MolProbity and »Protein Structure

& Model Assessment Tools« of »SWISS-MODEL Workspace«.

MolProbity allows to evaluate the quality both of all atoms contact and so the protein geometry of any three-dimensional biopolymer molecule. Evaluation of all-atom contacts, specifically the number of serious steric overlaps of all atoms, allows to choose model, built basing on template 1OFU_B, as a favourable. In this case second model from »SWISS-MODEL Workspace« loses much less for this parameter in compare to models of I-TASSER. On the other hand, estimation of all parameters of protein geometry (Table 4), allow us to consider the »SWISS-MODEL Workspace« model built based on template structure 2Q1Y (chain A), as the best one. However, on such parameters as »poor rotamers« and »Ramachandran favoured« it's not much better than the second model from this server and all models from I-TASSER. A significant benefit of both models from »SWISS-MODEL Workspace« over models from I-TASSER is observed in rates of »Ramachandran outliers«, »C β deviations«, »MolProbity score«. Significant differences are also observed in »Residues with bad angles« parameter. While two models from »SWISS-MODEL Workspace« have no residues with bad angles, in structures from I-TASSER models the quantity of these residues reaches 4-6 %. It should be noted that residues with bad bonds observed only in the case of model 2 among all seven analyzed models.

Benefits of models built on the I-TASSER in all-atom statistical potential »DFire energy« due to the bigger size of this models that are completely full unlike to the incomplete models from SWISS-MODEL server. Therefore, this parameter from »Protein structure & model assessment tools« we used to compare structures built only on the same server.

Finally, for evaluating quality of three-dimensional structures, it was implemented QMEAN6 score (<http://swissmodel.expasy.org/qmean/cgi/index.cgi>). QMEAN6 is a reliability score for the whole model which can be used in order to compare and rank alternative models of the same target. The quality estimate ranges between 0 and 1 with higher values for better models. Additionally, the pseudo energies of the four contributing statistical potential terms are provided as well as the percentage agreement

between predicted and measured features from the sequence and model, respectively. The comparison of the differences of the terms among the models may help to understand the reason for the differences in the estimated model quality. In addition to the »Raw scores«, »Z-scores« of the QMEAN composite score as well as all terms are provided relating the quality estimates to scores obtained for high-resolution reference structures solved experimentally by X-ray crystallography (Benkert et al. 2011). The QMEAN »Z-score« represents an measure of the absolute quality of a model by providing an estimate of the »degree of nativeness« of the structural features observed in a model and by describing the likelihood that a given model is of comparable quality to experimental structures. Models of low quality are expected to have strongly negative QMEAN Z-scores (i.e. the model's QMEAN score is several standard deviations lower than expected for experimental structures of similar size). Evaluation QMEAN6 scores, especially »Raw score« and »Z-score«, puts everything in its place, clearly demonstrating the undeniable advantage of *E. coli* FtsZ protein model based on template 2Q1Y (chain A) over all other constructed models.

Among the variety of crystallographic structures of FtsZ proteins from tuberculous pathogen *M. tuberculosis*, deposited in »Protein Data Bank«, we have chosen chain 2Q1Y (chain A) as a reasonable model for further bioinformatics

researches. It lacks first seven and last 66 residues, but it is characterized by full atom composition of all available residues and excellent estimated values of parameters from MolProbity.

Also, among the seven potential FtsZ *E. coli* models, built by I-TASSER server and »SWISS-MODEL Workspace«, we have selected a model from the last server, based on matrix 2Q1Y (chain A). This model prevailed for qualitative evaluation parameters not only full atom models of I-TASSER, but model, built on a matrix 1OFU (chain B), which represents FtsZ structure from *P. aeruginosa* – bacterial species systematically much closer to *E. coli* (the same class of γ -proteobacteria), than *M. tuberculosis*.

Thus, we have successfully reconstructed 3-D models of the FtsZ proteins of *E. coli* and *M. tuberculosis* is of sufficient quality for further *in silico* studies such as molecular docking, molecular dynamics simulations and computational drug discovery.

Acknowledgments

This work was supported by Scientific and Technologic Centre of Ukraine (STCU) grant #4939. Ukrainian participants were also supported by The Ukrainian National Grid (UNG)- <http://grid.nas.gov.ua/>.

References

- Arnold, K., Bordoli, L., Kopp, J., Schwede, T., 2006. The »SWISS-MODEL Workspace«: A web-based environment for protein structure homology modelling. *Bioinformatics*, 22, 195–201.
- Benkert, P., Biasini, M., Schwede, T., 2011. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, 27, (3), 343–350.
- Benkert, P., Schwede, T., Tosatto, S.C.E., 2009. QMEANclust: Estimation of protein model quality by combining a composite scoring function with structural density information. *BMC Struct. Biol.*, 9 (35), doi:10.1186/1472-6807-9-35.
- Berman, H.M., Henrick, K., Nakamura, H., 2003. Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.*, 10 (12), p. 980.
- Chen, V.B., Arendall, W.B. III, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., Richardson, D.C., 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D: Biol. Crystallogr.*, D66 (1), 12–21.
- Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.M., Harris, D.E., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E. III, Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R.M., Devlin, K., Barrell, B.G., 1998. Deciphering the biology of *M. tuberculosis* from the complete genome sequence, *Nature*, 393, 537–544.

- Cordell, S.C., Robinson, E.J.H., Löwe, J., 2003. Crystal structure of the SOS cell division inhibitor SulA and in complex with FtsZ. *Proc. Natl. Acad. Sci. USA*, 100 (13), 7889–7894.
- Demchuk, O.N., Blume, Ya.B., 2005. Phylogenetic tree of bacterial and eucaryotic FtsZ-proteins created according to the homology of their primary sequences, *Cytol. Genetics*, 39 (4), 3–12.
- Erickson, H.P., 1998. Atomic structures of tubulin and FtsZ, *Trends Cell Biol.*, 8, 133–137.
- Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J.D., DeBoy, R.T., Dodson, R.J., Gwinn, M.L., Haft, D.H., Hickey, E.K., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.D., Salzberg, S.L., Delcher, A., Utterback, T.R., Fraser, C.M., 2002. Whole-genome comparison of *M. tuberculosis* clinical and laboratory strains. *J. Bacteriol.*, 184, 5479–5490.
- Gu, J., Bourne, P.E., 2009. *Structural Bioinformatics*, 2nd ed. John Wiley and Sons, New Jersey, 1035 pp.
- Guex, N., Peitsch, M.C., 1997. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modelling. *Electrophoresis*, 18 (15), 2714–2723.
- Haydon, D.J., Stokes, N.R., Ure, R., Galbraith, G., Bennett, J.M., Brown, D.R., Baker, P.J., Barynin, V.V., Rice, D.W., Sedelnikova, S.E., Heal, J.R., Sheridan, J.M., Aiwale, S.T., Chauhan, P.K., Srivastava, A., Taneja, A., Collins, I., Errington, J., Czaplowski, L.G., 2008. An inhibitor of FtsZ with potent and selective anti-staphylococcal activity. *Science*, 321, 1673–1675.
- Höltje, H.-D., Sippl, W., Rognan, D., Folkers, G., 2008. *Molecular modeling. Basic Principles and Applications*, 3rd Ed., Wiley-VCH, p. 320.
- Huang, Q., Kirikae, F., Kirikae, T., Pepe, A., Slayden, R.A., Tonge, P.J., Ojima, I., 2006. Targeting FtsZ for anti-tuberculosis drug discovery: non-cytotoxic taxanes as novel anti-tuberculosis agents. *J. Med. Chem.*, 49 (2), 463–466.
- Jaiswal, R., Beuria, T.K., Mohan, R., Mahajan, S.K., Panda, D., 2007. Tatarol inhibits bacterial cytokinesis by perturbing the assembly dynamics of FtsZ. *Biochem.*, 46 (14), 4211–4220.
- Kumar, K., Awasthi, D., Berger, W.T., Tonge, P.J., Slayden, R.A., Ojima, I., 2010. Discovery of anti-TB agents that target the cell-division protein FtsZ. *Future Med Chem.*, 2 (8), 1305–1323.
- Kumar, K., Awasthi, D., Lee, S.-Y., Zanardi, I., Ruzsicska, B., Knudson, S., Tonge, P.J., Slayden, R.A., Ojima, I., 2011. Novel trisubstituted benzimidazoles, targeting Mtb FtsZ, as a new class of antitubercular agents. *J. Med. Chem.*, 54, 374–381.
- Läppchen, T., Pinas, V.A., Hartog, A.F., Koomen, G.J., Schaffner-Barbero, C., Andreu, J.M., Tram-baiolo, D., Löwe, J., Juhem, A., Popov, A.V., den Blaauwen, T., 2008. Probing FtsZ and tubulin with C8-substituted GTP analogs reveals differences in their nucleotide binding sites. *Chem. Biol.*, 15, 189–199.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*, 23, 2947–2948.
- Leung, A.K., Lucile White, E., Ross, L.J., Reynolds, R.C., DeVito, J.A., Borhani, D.W., 2004. Structure of *M. tuberculosis* FtsZ reveals unexpected, G protein-like conformational switches. *J. Mol. Biol.*, 342 (3), 953–970.
- Li, Y., Zhang, Y., 2009. REMO: A new protocol to refine full atomic protein models from C-alpha traces by optimizing hydrogen-bonding networks. *Proteins*, 76, 665–676.
- Löwe, J., Amos, L.A., 1998. Crystal structure of the bacterial cell-division protein FtsZ. *Nature*, 391, 203–206.
- Margalit, D.N., Romberg, L., Mets, R.B., Hebert, A.M., Mitchison, T.J., Kirschner, M.W., Chaudhuri, D.R., 2004. Targeting cell division: small-molecule inhibitors of FtsZ GTPase perturb cytokinetic ring assembly and induce bacterial lethality. *PNAS*, 101, 11821–11826.
- Mosyak, L., Zhang, Y., Glasfeld, E., Haney, S., Stahl, M., Seehra, J., Somers, W.S., 2000. The bacterial cell-division protein ZipA and its interaction with an FtsZ fragment revealed by X-ray crystallography. *EMBO J.*, 19 (13), 3179–3191.
- Nyporko, A.Yu., Blume, Ya.B., 2001. Comparative analysis of the tubulin secondary structure. *Biopolym. Cell*, 17, (1), 61–69, in Russian.

- Ohashi, Y., Chijiwa, Y., Suzuki, K., Takahashi, K., Nanamiya, H., Sato, T., Hosoya, Y., Ochi, K., Kawamura, F., 1999. The lethal effect of a benzamide derivative, 3-methoxybenzamide, can be suppressed by mutations within a cell division gene, *ftsZ*, in *Bacillus subtilis*. *J. Bacteriol.*, 181 (4), 1348–1351.
- Oliva, M.A., Cordell, S.C., Löwe, J., 2004. Structural insights into FtsZ protofilament formation, *Nat. Struct. Mol. Biol.*, 11 (12), 1243–1250.
- Oliva, M.A., Trambaiolo, D., Löwe, J., 2007. Structural insights into the conformational variability of FtsZ. *J. Mol. Biol.*, 373 (5), 1229–1242.
- Perna, N.T., Plunkett 3rd, G., Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Pósfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grotbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E.T., Potamousis, K.D., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A., Blattner, F.R., 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7, *Nature*, 409 (6819), 529–533.
- Raymond, A., Lovell, S., Lorimer, D., Walchli, J., Mixon, M., Wallace, E., Thompkins, K., Archer, K., Burgin, A., Stewart, L., 2009. Combined protein construct and synthetic gene engineering for heterologous protein expression and crystallization using Gene Composer, *BMC Biotechnol.*, 9 (37), doi: 10.1186/1472-6750-9-37.
- Raviglione, M.C., 2000. Issues facing TB control (7). Multiple drug-resistant tuberculosis, *Scott. Med. J.*, 45, (5), 52–55.
- Respicio, L., Nair, P.A., Huang, Q., Anil, B., Tracz, S., Truglio, J.J., Kisker, C., Raleigh, D.P., Ojima, I., Knudson, D.L., Identification of FtsZ polymerization regulatory elements using a *M. tuberculosis* FtsZ temperature sensitive mutant. To be published, DOI:10.2210/pdb2q1y/pdb
- Roy, A., Kucukural, A., Zhang, Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.*, 5 (4), 725–738.
- Roy, A., Zhang, Y., 2011 COFACTOR: protein-ligand binding site predictions by global structure similarity match and local geometry refinement, – in print.
- The UniProt Consortium, 2008. The Universal Protein Resource (UniProt). *Nucl Acids Res.*, 36, 190–195.
- Vaughan, S., Wickstead, B., Gull, K., Addinall, S.G., 2004. Molecular evolution of FtsZ protein sequences encoded within the genomes of archaea, bacteria, and eukaryote. *J. Mol. Evol.*, 58 (1), 19–39.
- White, E.L., Ross, L.J., Reynolds, R.C., Seitz, L.E., Moore, G.D., Borhani, D.W., 2000. Slow polymerization of *M. tuberculosis* FtsZ. *J. Bacteriol.*, 182 (14), 4028–4034.
- White, E.L., Suling, W.J., Ross, L.J., Seitz, L.E., Reynolds, R.C., 2002. 2-alkoxycarbonylaminopyridines: inhibitors of *M. tuberculosis* FtsZ. *J. Antimicrob Chemother.*, 50 (1), 111–114.
- Wu, S., Zhang, Y., 2007. LOMETS: a local meta-threading-server for protein structure prediction. *Nucleic Acids Res.*, 35, 3375–3382.
- Yu, X.-C., Margolin, W., 1998. Inhibition of assembly of bacterial cell division protein FtsZ by the hydrophobic dye 5,5*-Bis-(8-anilino-1-aphthalenesulfonate). *J. Biol. Chem.*, 273 (17), 10216–10222.
- Zhang, J., Zhang, Y., 2011. High-resolution protein structure refinement using fragment guided molecular dynamics simulations. *Structure*, in press.
- Zhang, Y., Kihara, D., Skolnick, J., 2002. Local energy landscape flattening: parallel hyperbolic Monte Carlo sampling of protein folding. *Proteins*, 48, 192–201.
- Zhang, Y., Skolnick, J., 2004. Scoring function for automated assessment of protein structure template quality. *Proteins*, 57 (4), 702–710.
- Zhou, H., Zhou, Y., 2002. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein Sci.*, 11, 2714–2726.