



Operations Research Society in
Bosnia and Herzegovina

Southeast Europe Journal of Soft Computing
Available online: <http://scjournal.ius.edu.ba>



IUS Soft Computing
Research Group

Assessment of Accuracies of Protein 3-Dimensional Prediction Software

R. Gosto

Faculty of Engineering and Natural Sciences,
International University of Sarajevo, Sarajevo,
Hrasnicka Cesta 15, Ilidža 71210 Sarajevo,
Bosnia and Herzegovina
rijad.gosto1@gmail.com

Article Info

Article history:

Article received on 10 July 2018

Received in revised form 27 August
2018

Keywords:

Protein 3-dimensional structure
prediction; 3D prediction
software; SWISS-MODEL;
Phyre2; RaptorX; i-Tasser;
CPHmodels

Abstract

Protein 3-dimensional structure prediction is determination of the 3-dimensional structure of a protein from its amino acid sequence by using protein structure prediction software. By understanding protein's 3-dimensional structure, we should be able to figure out the function of the said protein. We already have several protein prediction software, but the purpose of this study is to determine how accurate they are, and if the results presented are true and to what extent. To determine how accurate protein 3-dimensional structure prediction software are, we compared x-ray crystallography determined protein structures to software predicted 3-dimensional protein structures. All of the software used showed good accuracy, and according to our results, "i-Tasser" software was the most accurate, closely followed by RaptorX.

1. INTRODUCTION

Protein's 3-dimensional structure prediction is a process of determining the 3-dimensional folded structure of the protein based on its amino acid composition and interaction between those amino acids. As we already know, product of translation process in the cell is a linear chain of amino acids which is not yet functional but it becomes functional only when it assumes its fully folded conformation. This linear chain will have to pass through several folds, to become fully functional. This entire process takes very little time in the living cell. (Alberts, Bruce; A. et al., 2002)

As the technology developed, we got new insights into protein folding and we are able to harness all the processing power of the computers and programming to develop structure prediction software, right after learning more about how amino acids interacted with one another. It is impossible to experimentally figure out the structure of all of the proteins because they are so abundant, and each would take a lot of time and resources. That's where "in silico" structure prediction finds its role. It's much cheaper and faster using computers, and once the prediction becomes reliable enough it may reduce the time spent in the laboratory.

After running the amino acid sequence through structure prediction software, we will usually get 3-dimensional structure of the protein. As stated earlier, protein's function depends on its structure, so getting the most out of those software is of importance, not only for bioinformatics but also for medicine while doing drug design, or biotechnology while designing novel enzymes. How accurate those software are, and comparison of the results obtained and experimental results, will be presented in this work.

The performance of current methods of protein 3-dimensional structure is assessed every two years in CASP experiment (Critical Assessment of protein Structure Prediction). It is a community-wide, worldwide experiment for protein structure prediction taking place every two years since 1994 (Moult, et al., 1995)

The primary method of evaluation is a comparison of the predicted model α -carbon positions with those in the target structure. The comparison is shown visually by cumulative plots of distances between pairs of equivalent α -carbons in the alignment of the model and the structure (a perfect model would stay at zero all the way across), and is assigned a numerical score. (Cozzetto, et al., 2009).

Even though the assessment is done biennially in large scale project, the assessment should be also done by independent researchers to keep the statistics and results as unbiased as possible.

2. METHODOLOGY

As this is "in silico" project, we will mention and describe all the software and bioinformatics tools used during the work. The idea behind the entire project is to compare 3-dimensional structures of proteins obtained experimentally by X-ray crystallography, to the 3-dimensional protein structures predicted by software.

All of the experimental 3-dimensional protein structures were acquired from Protein Data Bank (PDB).

Protein Data Bank is a worldwide archive for large biological molecules including proteins and nucleic acids. It also contains 3-dimensional structures of the proteins determined experimentally, in the laboratory (Berman, et. Al. , 2003). While doing PDB search for proteins, several criteria were used. First one is that all of the protein structures had to be experimentally determined by X-ray crystallography.

X-Ray crystallography is currently the most favored technique for structure determination of proteins and biological macromolecules. The aim of X-ray crystallography is to obtain a three dimensional molecular structure from a crystal. A purified sample at high concentration is crystallized and the crystals are exposed to an X-ray beam. The resulting diffraction patterns can then be processed, in order to yield the information about the

crystal packing symmetry and the size of the repeating unit that forms the crystal. This is obtained from the pattern of the diffraction spots. The intensities of the spots can be used to determine the "structure factors" from which a map of the electron density can be calculated (Carter, and Sweet, 1997).

Second criterion is that all of the proteins have to contain single subunit in their structure. The reason for this is that prediction software often have trouble to accurately predict 3-dimensional structure of more complex proteins which contain many subunits so the accuracy of the predictions is lower than single subunit proteins. Later into the project we did include several multi-subunit proteins to see how accurately their 3-dimensional structure can be predicted and how that software cope with larger structures.

While searching for proteins in PDB, we also did it by size, which is shown as "Chain length" in the PDB advanced search engine. Protein with lowest number of residues contained just around 100 residues while the largest one contained more than 10,000 residues. The reason for size comparison is because we wanted to see if the 3-dimensional prediction accuracy goes down with the increasing number of residues, subunits, and protein complexity, which is expected. All of the PDB data was downloaded and further used with the software.

2.1. Prediction Software

After all of the PDB files were downloaded, the first software we used was SWISS-MODEL. SWISS-MODEL is a structural bioinformatics web-server dedicated to homology modeling of protein 3D structures. (Schwede, et. al., 2003) Homology modeling is currently the most accurate method to generate reliable three-dimensional protein structure models and is routinely used in many practical applications. Homology (or comparative) modeling methods make use of experimental protein structures ("templates") to build models for evolutionary related proteins (Biasini, et al, 2014).

We used SWISS-MODEL to predict of all of the proteins selected, and the results were saved for further analysis. The following prediction software we have used was Phyre2.

Phyre and Phyre2 (Protein Homology/ Analogy Recognition Engine) are web-based services for protein structure prediction that are free for non-commercial use. (Lawrence et. Al., 2011).

Phyre is among the most popular methods for protein structure prediction, having been cited over 1500 times. Like other remote homology recognition techniques, it is able to regularly generate reliable protein models when other widely used methods such as PSI-BLAST cannot. Phyre2 has been designed to ensure a user-friendly interface for users inexperienced in protein structure prediction methods. (Lawrence, and Jefferys, 2011).

The Phyre and Phyre2 servers predict the three-dimensional structure of a protein sequence using the principles and techniques of homology modeling. Typically, the amino acid sequences of a representative set of all known three-dimensional protein structures is compiled, and these sequences are processed by scanning against a large protein sequence database. The result is a database of profiles or HMMs, one for each known 3D structure. A user sequence of interest is similarly processed to form a profile/HMM. This user profile is then scanned against the database of profiles using profile-profile or HMM-HMM alignment techniques. These alignments can also take into account patterns of predicted or known secondary structure elements and can be scored using various statistical models (Bennett-Lovsey, et. Al., 2007). Phyre2 showed very good results but the downside of it is time consumed while processing a single protein. Each protein with more than 2000 residues took at least 10 hours to process.

Third software for protein 3-dimensional structure prediction we used was I-TASSER.

I-TASSER server is an on-line platform that implements the I-TASSER based algorithms for protein structure and function predictions. It allows academic users to automatically generate high-quality model predictions of 3D structure and biological function of protein molecules from their amino acid sequences. When user submits an amino acid sequence, the server first tries to retrieve template proteins of similar folds (or super-secondary structures) from the PDB library by LOMETS, a locally installed meta-threading approach.

In the second step, the continuous fragments excised from the PDB templates are reassembled into full-length models by replica-exchange Monte Carlo simulations with the threading of unaligned regions (mainly loops) built by “ab initio” modeling. In cases where no appropriate template is identified by LOMETS, I-TASSER will build the whole structures by ab initio modeling. The low free-energy states are identified by SPICKER through clustering of the simulation decoys. In the third step, the fragment assembly simulation is performed again starting from the SPICKER cluster centroids, where the spatial restraints collected from both the LOMETS templates and the PDB structures are used to guide the simulations. The final full-atomic models are obtained by REMO which builds the atomic details from the selected I-TASSER decoys through the optimization of the hydrogen-bonding network (Yang, et al., 2010; Zhang, 2008).

Following protein 3-dimensional structure prediction software we have used was RaptorX. RaptorX is a protein structure prediction server developed by Xu group, excelling at predicting 3D structures for protein sequences without close homologs in the Protein Data Bank (PDB). Given an input sequence, RaptorX predicts its secondary and tertiary structures, contacts, solvent accessibility, disordered regions and binding sites. RaptorX also assigns

some confidence scores to indicate the quality of a predicted 3D model. (Källberg, et., al., (2011).

Final protein 3-dimensional structure prediction software we have used was CPHmodels-3.0. It is a web-server predicting protein 3D-structure by use of single template homology modeling. The server employs a hybrid of the scoring functions of CPHmodels-2.0 and a novel remote homology-modeling algorithm. A query sequence is first attempted modeled using the fast CPHmodels-2.0 profile-profile scoring function suitable for close homology modeling. The new computational costly remote homology-modeling algorithm is only engaged provided that no suitable PDB template is identified in the initial search (Nielsen, et. Al., 2010).

2.2 Accuracy Analysis Software

After all of the proteins were run through the protein 3-dimensional prediction software, we used SuperPose software to analyze how close the predicted 3-dimensional structures were to the experimentally determined structures. SuperPose is a freely available web server designed to perform both pairwise and multiple protein structure superpositions. (Maiti, et. Al., 2004).

Structural superposition can be quantified either in terms of similarity or difference measures. The optimal superposition is the one in which the similarity measure is maximized (the former case) or the difference measure (the latter case) is minimized. The “SuperPose” web server uses “RMSD” or Root-Mean-Square Deviation as a difference measure to find the optimal pairwise or multiple protein structure superpositions. After an initial sequence and secondary structure (in case of low sequence identity) alignment, SuperPose generates a Difference Distance (DD) matrix (Richards, and Kundrot, 1988) from the equivalent C-alpha atoms of two molecules. The sequence/structure alignment and DD matrix analysis information is then fed into a modified algorithm to rapidly perform the structural superposition and calculate the RMSD between aligned regions of two macromolecules. (Kearsley, 1990).

Beginning with an input PDB file or set of files, SuperPose first extracts the sequences of all chains in the file(s). Each sequence pair is then aligned using a Needleman–Wunsch pairwise alignment algorithm (Needleman, and Wunsch, 1970) employing a BLOSUM62 scoring matrix. If the pairwise sequence identity falls below the default threshold (25%), SuperPose determines the secondary structure using VADAR (volume, area, dihedral angle reporter) and performs a secondary structure alignment using a modified Needleman–Wunsch algorithm (Willard, et al 2003). After the sequence or secondary structure alignment is complete, SuperPose then generates a difference distance (DD) matrix (Richards, and Kundrot, 1988) between aligned alpha carbon atoms. A difference distance matrix can be generated by first calculating the distances between all pairs of Ca atoms in one molecule to generate an initial

distance matrix. A second pair wise distance matrix is generated for the second molecule and, for equivalent/aligned Ca atoms, the two matrices are subtracted from one another, yielding the DD matrix. From the DD matrix it is possible to quantitatively assess the structural similarity/dissimilarity between two structures. In fact, the difference distance method is particularly good at detecting domains or hinge motions in proteins.

In further analysis, we have used Verify3D software. The Verify3D (Eisenberg et al. 1997) method assess protein structures using three-dimensional profiles. This program analyzes the compatibility of an atomic model (3D) with its own amino acid sequence (1D). Each residue is assigned a structural class based on its location and environment (alpha, beta, loop, polar, apolar etc). Then a database generated from good structures is used to obtain a score for each of the 20 amino acids in this structural class. The vertical axis in the plot represents the average 3D-1D profile score for each residue in a 21-residue sliding window. The score ranges from -1 (bad score) to +1 (good score).

Furthermore, we used QMEAN score combined with GMQE (Global Model Quality Estimation). GMQE is a quality estimation which combines properties from the target-template alignment and the template search method. The resulting GMQE score is expressed as a number between 0 and 1, reflecting the expected accuracy of a model built with that alignment and template, and the coverage of the target. Higher numbers indicate higher reliability. Once a model is built, the GMQE gets updated for this specific case by also taking into account the QMEAN score of the obtained model in order to increase reliability of the quality estimation. QMEAN is a composite estimator based on different geometrical properties and provides both global (i.e. for the entire structure) and local (i.e. per residue) absolute quality estimates on the basis of one single model (Benkert et al, 1997).

3. RESULTS

The results obtained were very close to the expected results before starting this project. As the size of the protein and their complexity increased, the accuracy of prediction lowered. In several cases, the accuracy results were lower than expected, and that might be due to several reasons. First one is that the protein consists of many loops and software had hard time predicting the overall conformation of the protein, or the software does not do well when the protein is more complex. In the following table analysis results using SuperPose software are presented.

SuperPose results are given as RMSD value or Root Mean Square Deviation, the measure of the average distance between the atoms (usually the backbone atoms) of superimposed proteins. An RMSD value is expressed in

length units. The most commonly used unit in structural biology is the Ångström (Å) which is equal to 10⁻¹⁰ m. It gives us quite good representation of the 3-dimensional structure accuracy (Maiti, et. Al., 2004).

Table 1. Results of SuperPose software analysis expressed in Ångström (Å)

Code	N	S-M	P2	RX	i-T	CPH
5ODY	138	0,86	0.70	0.61	0.66	0.60
5MV1	407	0.52	0.71	0.65	0.70	0.67
5O2L	785	8.15	1.12	0.70	0.69	0.64
5V2S	841	0.34	1.18	0.72	0.72	0.52
2I37	1047	0.97	0.76	0.72	0.80	0.79
5NJY	1635	0.99	0.98	0.80	0.97	0.84
6F0K	2759	1.45	1.32	1.01	1.28	0.98
6ETI	2180	1.43	0.84	0.86	1.17	1.12
3AK5	3900	3.18	1.35	1.21	1.19	1.17
5XYM	6880	1.66	1.55	1.75	1.84	-
5LUQ	8644	2.21	2.55	2.00	2.01	-
2PFF	11277	4.22	3.01	2.66	2.17	-
3HMJ	11814	4.57	2.97	2.81	2.40	-
Mean		2.35	1.46	1.27	1.27	0.81
STDV		2.12	0.81	0.84	0.59	0.21

In the following table (table 2) the results obtained by "Verify3D" software are represented. It determines the compatibility of an atomic model (3D) with its own amino acid sequence (1D) by assigning a structural class based on its location and environment.

Verify3D results are given on the scale from -1 to 1, with -1 being the bad score and 1 being the good score. (Eisenberg, et al. (1997)

Table 2. Results of Verify3D analysis

Code	N	SM	P2	RX	i-T	CPH
5ODY	138	0,87	0.90	0.91	0.90	0.97
5MV1	407	0.91	0.89	0.92	0.88	0.87
5O2L	785	-0.49	0.31	0.75	0.75	0.90
5V2S	841	0.88	0.88	0.89	0.80	0.73
2I37	1047	0.85	0.86	0.86	0.83	0.70
5NJY	1635	0.50	0.53	0.70	0.71	0.63
6F0K	2759	0.62	0.53	0.60	0.59	0.56
6ETI	2180	0.71	0.74	0.76	0.77	0.92
3AK5	3900	0.58	0.55	0.72	0.59	0.60
5XYM	6880	0.73	0.63	0.69	0.69	-
5LUQ	8644	0.66	0.22	0.32	0.49	-
2PFF	11277	-0.27	0.21	0.31	0.44	-
3HMJ	11814	-0.31	0.06	0.41	0.28	-
Mean		0.48	0.52	0.71	0.67	0.76
STDV		0.42	0.27	0.20	0.17	0.14

In table 3, we shows the GMQE (Global Model Quality Estimation) scores. It is expressed as a number between 0 and 1 reflecting the expected accuracy of a model built

with that alignment and template and the coverage of the target. [Benkert, et. Al., 2008)

Table 3. GMQE scores of each 3-D prediction software

Code	N	SM	P2	RX	i-T	CP H
5ODY	138	0.87	0.90	0.91	0.90	0.97
5MV1	407	0.91	0.89	0.92	0.88	0.87
5O2L	785	-0.49	0.31	0.75	0.75	0.90
5V2S	841	0.88	0.88	0.89	0.80	0.73
2I37	1047	0.85	0.86	0.86	0.83	0.70
5NJY	1635	0.50	0.53	0.70	0.71	0.63
6F0K	2759	0.62	0.53	0.60	0.59	0.56
6ETI	2180	0.71	0.74	0.76	0.77	0.92
3AK5	3900	0.58	0.55	0.72	0.59	0.76
5XYM	6880	0.73	0.63	0.69	0.69	-
5LUQ	8644	0.66	0.22	0.32	0.49	-
2PFF	11277	-0.27	0.27	0.31	0.54	-
3HMJ	11814	-0.31	0.25	0.41	0.51	-
Mean		0.48	0.58	0.71	0.69	0.79
STDV		0.42	0.24	0.20	0.13	0.13

As we can conclude from Tables 1, 2 and 3, all of the software's showed good prediction when it comes to less complex proteins. While the number of residues was low, CHPmodels software showed the best results and fastest processing time but it is limited by number of amino acid residues it can process which is 4000. As the complexity was getting higher i-Tasser showed the best results, closely followed by Phyre2 and RaptorX. Downside of i-Tasser and Phyre2 is that they take very long time to process the results, especially if the protein have several thousand residues in its structure. Some of the fields in the table are left blank and the reason is that protein 3-dimensional structure prediction software's required more than 10 days to complete, and due to the limited time, their prediction was not completed. In order to complete the prediction of all the remaining protein 3-dimensional structure, would require several months, without counting the analysis.

CHPmodels and I-Tasser software showed lowest RMSD values.

Important thing to notice here is that results obtained in this graduation project match with the data found in worldwide experiment CASP (Critical Assessment of protein Structure Prediction) database Based on their report, the CHPmodels protein 3-dimensional prediction software showed best results, with it being limited to 4000 residues, closely followed by i-Tasser and Phyre2.

4. DISCUSSION

According the obtained results, the protein 3-dimensional prediction accuracy got lower with increasing number of amino acid residues. As the number of residues and protein

complexity was getting higher, the 3-dimensional structure prediction was getting lower. Out of the software's used, the "CHPmodels" software showed the best results regarding its accuracy, as proven by RMSD results (see table 1), followed by i-Tasser and Phyre2, meaning that it's deviation from X-ray crystallography determined structure was lowest. This confirms the results where the CHPmodels software proved to be best predictor in a community CASP (Critical Assessment of protein Structure Prediction) project, (Källberg, and Wang 2012). It is important to note that CHP models software is limited to 4000 amino acid residue count, and it cannot process higher complexity proteins.

In couple of instances RMSD values were higher than expected, which could be tied to protein complexity. Even though in those instances proteins had less than 2000 residues, their structure might have been quite complex for 3-dimensional prediction engine to process due to all of the interactions in the structure.

Due to lack of time we were not able to use all of the protein 3-dimensional prediction software's as planned at the beginning of the project. Some of the software used here required over 10 days finishing their prediction, and some had queue of over 1200 projects

5. CONCLUSION

As expected before starting this graduation project, the software showed lower accuracy of 3-dimensional protein prediction as the number of residues and protein complexity increased. In several cases, the accuracy was lower than expected, but it happened not only in one software but two or more which leads me to think that the protein structure was too complex for a 3-dimensional prediction software to process.

Software which showed us the best results was CHPmodels followed by I-Tasser, Phyre2 and RaptorX.

Some of the software showed poor results in SuperPose software analysis but when using Verify3D they would show better results, which might lead to the conclusion that some of proteins selected in this study might have several possible conformations, depending on their role in the organism.

For two 3-dimensional prediction software, we did not complete the prediction, since they require minimum 10 days to finish their predictions. According to the obtained data, we have filtered out the best 3D predictor, among the most common used 3D predictors. In this regard we can conclude that there is definitely a room for improvement in 3-dimensional protein prediction accuracy for all software, where in this study the CHP models software showed the highest accuracy of protein 3-dimensional structure prediction.

6. REFERENCES

- Benkert, P., Tosatto, S.C.E. and Schomburg, D. (2008). "QMEAN: A comprehensive scoring function for model quality assessment." *Proteins: Structure, Function, and Bioinformatics*, 71(1):261-277
- Bennett-Lovsey, R. M.; Herbert, A. D.; Sternberg, M. J. E.; Kelley, L. A. (2007). "Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre". *Proteins: Structure, Function, and Bioinformatics*. 70
- Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, M., Bordoli, L., Schwede, T. (2014) "SWISS-MODEL: modeling protein tertiary and quaternary structure using evolutionary information". *Nucleic Acids Research*. 42 (W1): 195–201
- Berman, H.M., Henrick, K. Nakamura, H. (2003) Announcing the worldwide Protein Data Bank *Nature Structural Biology* 10 (12): 980.
- Bruce, A., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walters, P. (2002) "The Shape and Structure of Proteins". *Molecular Biology of the Cell; Fourth Edition*. New York and London: Garland Science
- Carter, C.W., Sweet, R.M., eds. *Macromolecular crystallography, part A. Methods Enzymol* 1997;276
- Carter, C.W., Sweet, R.M., eds. *Macromolecular Crystallography, part B. Methods Enzymol* 1997;277.
- Cozzetto, D., Kryshchak, A., Fidelis K., Moulton, J., Burkhardt, R., Tramontano, A. (2009) "Evaluation of template-based models in CASP8 with standard measures". *Proteins*. 77 (Suppl 9): 18–28
- Diamond, R. (1992) "On the multiple simultaneous superposition of molecular structures by rigid body transformations". *Protein Sci.*, 1, 1279–1287
- Eisenberg, D. Luthy, R., Bowie, J.U., (1997). "VERIFY3D: assessment of protein models with three-dimensional profiles." *Methods Enzymol* 277: 396-404.
- Kabsch, W.A. (1978) Discussion of solution for best rotation of two vectors. *Acta Crystallogr. A.*,34, 827–828.
- Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., and Xu, J. (2012) "Template-based protein structure modeling using the RaptorX web server". *Nature Protocols* 7, 1511–1522.
- Kearsley, S.K. (1990). "An algorithm for the simultaneous superposition of a structural series". *J. Comput. Chem.* 11 (10): 1187–1192
- Kelley, K., Bennett-Lovsey, R., Herbert, A., and Fleming, K., "Phyre: Protein Homology/analogy Recognition Engine". Structural Bioinformatics Group, Imperial College, London. Retrieved 22 April 2011.
- MacLachlan, A.D. (1982) Rapid comparison of protein structures. *Acta Crystallogr. A.*,38, 871–873.
- Maiti, R., Van Domselaar, G. H., Zhang, H., and Wishart, D. S. (2004) "SuperPose: a simple server for sophisticated structural superposition" *Nucleic Acids Res.* 32 (Web Server issue): W590W594.
- Moulton, J., Pedersen J.T., Judson R., and Fidelis K. (1995) "A large-scale experiment to assess protein structure prediction methods". *Proteins*. 23 (3): ii–iv.
- Needleman, S.B. and Wunsch, C.D. (1970) "A general method applicable to the search for similarities in the amino acid sequence of two proteins". *J. Mol. Biol.*,48, 443–453.
- Nielsen M., Lundegaard C., Lund O., Petersen TN., (2010) "CPHmodels-3.0 - Remote homology modeling using structure guided sequence profiles", In Press - NAR web issue 2010
- Rajarshi, M., Van Domselaar, G.H., Zhang, H., Wishart, D.S. (2004) "SuperPose: a simple server for sophisticated structural superposition". *Nucleic Acids Res.* 32 (Web Server issue)
- Richards, F.M.; Kundrot, C.E. (1988). "Identification of structural motifs from protein coordinate data: secondary structure and first-level supersecondary structure". *Proteins*. 3 (2): 71–84.
- Roy, A., Kucukural, A., Zhang, Y. "I-TASSER: a unified platform for automated protein structure and function prediction". *Nature Protocols*, 5: 725-738 (2010)
- Schwede T, Kopp J, Guex N, Peitsch MC (2003). "SWISS-MODEL: an automated protein homology-modeling server". *Nucleic Acids Research*. 31 (13): 3381–3385.
- Willard, L., Ranjan, A., Zhang, H., Monzavi, H., Boyko, R.F., Sykes, B.D. and Wishart, D.S. (2003) "VADAR: a web server for quantitative evaluation of protein structure quality". *Nucleic Acids Res.*, 31, 3316–3319.
- Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, J., "The I-TASSER Suite: Protein structure and function prediction". *Nature Methods*, 12: 7-8 (2015)
- Zhang, Y., (2008) "I-TASSER server for protein 3D structure prediction". *BMC Bioinformatics*, vol 9, 40