Asian Journal of Pharmacy and Life Science ISSN 2231 – 4423
Vol. 2 (4), Oct-Dec,2012

Stereo chemical and Statistical issues for evaluating the goodness of 30 different models generated by online freely protein 3D structure prediction servers

Sudha Karumuri and Rajib Bandopadhyay*

Department of Biotechnology, Birla Institute of Technology, Mesra, Ranchi-835215, Jharkhand, India

*Corresponding author's E-mail: rajib_bandopadhyay@bitmesra.ac.in Mob: +91-9430378406, Fax: +91-2275401

Abstract

Proteins are known as the building blocks of the living body. For the correct diagnosis of any disease or malfunction of the living body, information about protein provides knowledge about the situation. Therefore protein structure and function plays an important role in assessing the condition of the organism. In silico analysis provides a faster way to detect the structure of proteins through three mechanisms namely homology modelling, Threading or fold recognition and Ab initio method. This can be done by using different server and software freely available online. An effort has been done to identify the ideal three dimensional structure of the protein among the thirty models generated by 30 different structure prediction servers. The evaluation was done taking into account several stereo chemical and statistical factors. The results have shown that the models generated by majority of the homology modelling servers are satisfying the evaluation factors when compared with the models generated by the other two mechanisms namely, Threading and Ab-initio.

Key Words*: Homology modelling, Threading, Ab-initio, Target, Template, Validation*

Introduction

Protein structure prediction plays a major role in knowing the function of that protein. So, there have been several methods available to predict the 3D structure (tertiary structure) of a protein, namely, *Ab initio* method, Homology Modelling and Threading [1]. One of the fastest and accurate methods of protein structure prediction include the *in silico* approach. We can generate the 3D-Structure of a protein from its amino acid sequence using several online available servers, which follow the combination of any of the above mentioned methods. Most of the online available servers follow the homology modelling method. These servers accept the sequence of an amino-acid as input either in 'fasta' format or directly the one-letter code of the amino acids involved. This amino acid sequence of the protein whose tertiary structure is to be determined is called target sequence. From the collected target sequence the server searches for a known protein sequence which already has a proved tertiary structure that is expected to be similar with that of the target sequence. The amino acid sequence of this known protein is collected from protein databases available and is termed as template sequence [2]. The target and template sequences are undergone through sequence alignment to evaluate the accuracy of the target - template similarity [3].

The identification of the suitable template sequence plays a very crucial role in determining the accuracy of the 3D model of the target sequence [4]. More the similar the target and template sequences more will be the correctness of the predicted target 3D structure. The template sequence selection is based on taking into account several factors

that estimate the goodness of the target and template similarity [5]. One of the most used methods to search the template sequences is the BLAST (Basic Local Alignment Search Tool) technique [6]. This tool selects different templates based on the scores of the similarity of the target and template, through local alignments of the query sequence with the sequences in the database. Blastp is the most used type of BLAST techniques for protein sequence similarity search against a protein database.

Some structure prediction servers not only accept the amino acid sequence of the target as input but also demand the template sequence also along with its alignment with the target sequence. In this work, we evaluate the servers which require only target amino acid sequence as input to develop the tertiary structure of the target protein. These servers searched for the template automatically from several databases by their in-built tools.

Need for protein structure validation and analysis: Most of the servers used are freely available online while some of them require login permission from the service provider. Because of this high ease for availability we can easily generate a 3D structure for our protein sequence query. But, there is need to check the quality and accuracy of the model generated, which would promise the effective assessment of functional features of the target protein [7].

The Critical Assessment of protein Structure Prediction (CASP) conducts experiments in every couple of years from CASP1 (1994) to CASP10 (2012). They present the working efficiency of different structure prediction servers basing on some blind predictions. They perform several structural super positions and ranking is done based on the deviation of the Z-Sores and also numerical score GDT-TS (Global Distance Test — Total Score) is taken into consideration [8,9]. In this work, we considered directly the models generated by the prediction servers and evaluated them basing on some factors, which would be very helpful to assess the feasible quality of the output of these public servers.

The validation and analysis of the 3D models generated by the structure prediction servers can be done by taking into account several factors like stereo-chemical factors, statistical and mathematical functions, energy scores, volume parameters, scoring functions of local and global quality estimators, packing quality of the protein, Z-score etc. [10,11]. There are several online available tools and servers to evaluate and analyze the protein model generated by the structure prediction servers, namely 'Procheck'(http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/), 'Promotiff'(http://www.msi.umn.edu/software/promotif/document_2.html),'Whatcheck'(http://swift.cmbi.kun.nl/swift/whatcheck/),'Qmean'(http://swissmodel.expasy.org/qmean/cgi/index.cgi),'Annolea'(http://protein.bio.puc.cl/cardex/servers/anolea/index.html),'Verify3D'(http://nihserver.mbi.ucla.edu/Verify_3D/),'Gromos'(http://www.gromos.net/),'Prove'(http://www.scmbb.ulb.ac.be/SCMBB/PROVE/),'Errat'(http://nihserver.mbi.ucla.edu/ERRAT/), 'Dfire'(http://sparks.informatics.iupui.edu/yueyang/DFIRE/dDFIRE-service) etc.,

Several factors that are used to study the accuracy of the model

The key factors used to assess the goodness of the computationally generated protein three dimensional models are given below [12,13,14,15].

Z-Score: This is generally explained as the "standard deviation from the mean". This gives the score showing the deviation of the model comparing with already known similar structures. The Z-Score should be zero for novel structural prediction. Better than average is given by positive Z-score and worse than average is indicated by negative Z-Scores.

Ramachandran Plot: The 3D structure of the protein is made of different bonds with specific conformations. To prove that the model generated is of good quality, a study of its structural conformations should be done. Ramachandran Plot gives a statistical study of the conformations of the torsional angles by showing them as dots in the allowed and disallowed regions in a map with psi-angles at Y-axis and phi-angles at X-axis. The torsional angles of a protein constitute the rotatable bonds between four atoms, namely phi [N(i-1),C(i),Ca(i),N(i)] and psi [C(i),Ca(i),N(i),C(i+1)] angles. The allowed regions in the ma are considered as low-energy regions and the disallowed regions as the high-energy regions. Therefore, as more are the dots in the allowed region more is the quality of the generated model, with better conformations. [16]

Backbone and Side chain conformation: For a generated 3D model of the protein to be reliable, the back bone conformations should be checked for each residue. The backbone conformation of a protein is judged by the dihedral angles, psi, phi and omega, which determines the entire topology of the protein. The backbone normality check is done by comparing with the structures in the backbone structure database. The normality score like any Z-Score, shows negative for errors in the structure [17].

Repulsive steric interactions are involved in determining the protein side chain conformations. A few of this type of interactions depend on chi-1 and chi-2 which are considered as backbone conformation independent interactions. These interactions involve delta Carbons and backbone C or N atoms. Therefore, the chi1- chi2 rotamer normality Z-Score is used in validating the 3D model[18].

Protein folds: The conformation and the folds present in the model should be similar to that in the template used to generate the model. This will assure the homologous relation of the template and the model. Therefore, the conformation and inter atomic contacts should be highly similar. Hence, the 3D-1D comparisons of the model generated and the sequence used plays an important role in the model validation.

In this work we present 30 different models of each of the three proteins obtained from the protein databank, PDB. The quality of each model is assessed by using SAVES server, which include several quality estimation parameters. Comparison of models is done through the quality checking parameters basing on the scores generated.

Materials and Methods

We collected three proteins from the popular database PDB which has experimentally proved 3D structures. The 3 proteins are selected based on the following parameters which establish the novel candidature of their 3D structures.

i) Should have an experimentally proved X-Ray structure

ii) Should have X-Ray resolution scores between 0-2 Angstroms

iii) To avoid complexity in the 3D structure generation only 1 chain in each protein sequence is considered for structure generation

Three proteins with PDB ids 3AZ1, 3UWD and 4A14 are selected as they are satisfying the above conditions. The amino acid sequences of these 3 proteins are collected in the fasta format and are used as target sequences for the

submission to different servers to generate their tertiary structures. 30 different protein structure prediction servers are used in which some of the servers followed 'Homology Modelling' while some of them used 'Threading' principles to generate the 3D structures. Few of these servers used both the 'Homology Modelling' and 'Threading' principles. All these servers are freely available in the internet; some of them require log-in information. The 30 servers which are considered for 3D structure generation with their web links are given in Table I. These servers are supplied with all the three target amino acid sequences each time to generate a new model. Some of the servers accept only the raw amino acid sequence (one-letter code of the amino acid sequence) and some of them accept the amino acid sequence in fasta format. After submitting the target sequence, the servers search for the best suited templates and used them for 3D structure generation. The parameters of each server are allowed to remain as default during the whole process. Thus all the three proteins have thirty 3D models generated by each server.

The quality of the models generated are evaluated using a server of NIH-NLM named SAVES (http://nihserver.mbi.ucla.edu/SAVES/) which has 5 different quality estimating tools mentioned in Table II.

Each of the 30 models generated are submitted to the SAVES server for quality estimation. The pdb coordinate file of each model generated is submitted to SAVES and all the programs included in the SAVES server namely PROCHECK, WHATCHECK, VERIFY 3D, ERRAT and PROVE are made to run. The results obtained from SAVES are re-checked using the new version of SAVES version 4 (http://services.mbi.ucla.edu/SAVES/).

RESULTS

The quality of the 30 models is evaluated using SAVES (Structural Analysis and Verification Server). The factors used to access the goodness of the models are

i.'% of residues in favourable region' predicted by Procheck, showing the Ramachandran Plot, with a scale of % that 90% or more residues in the favourable region correspond to a good model (Fig. Ia).

ii.'Over-all quality score' is the goodness estimation factor predicted by Errat, with a scale of % that above 90% of errat score corresponds to better quality of the model.

iii.Z-Scores of 2nd generation packing quality, Ramachandran plot, Chi 1- chi 2 rotamer normality and Back bone conformation predicted by What Check program, giving the over-all summary of the model on comparison with the proved structures, with a scale that better the quality of the model more will be the positive value of the Z-Scores.

iv.'% of residues with averaged 3D-1D Score> 0.2', is the measurement of compatibility of the tertiary structure of the model generated to its amino acid sequence on a % scale of 100.

v.'Z-Score RMS' is predicted by Prove, giving the Root Mean Square of Z-Value, where the value should be approximately 1.0 (Fig. IIa). Deviations of the value from 1.0 show inaccuracy of the models (Fig. IIb).

The modelling servers are mentioned in the Table III according to their serial number in the Table I. Under each server the values for all the three proteins in response to the quality estimation are noted. I, II and III in the column named 'Protein' indicate the proteins 3AZ1, 3UWD and 4A14 respectively. According to the results from the Table III, the

models generated can be classified based on their deviations from the expected novel scores of the evaluation factors. Six models from the servers namely, Swiss Model, Protmod, AL2TS, HHPred, SamT08, and Phyre2 have good scores of the evaluation factors except from 1-3 deviations from the expected novel scores. The models which show 3-5 deviations from the scores of the evaluation factors are generated by nine servers namely PS2, EsyPred, CPH, BioSerf, M4T, @Tome, ModWeb, Loopp and IntFOLD. The models showing more than 5 deviations from the novel scores are Homer, 3D Jigsaw, Geno3D, COMA, ITasser and Prosp3. Unfortunately some models failed to give results to one or more evaluation factors are the ones developed by the servers Proteus, Fugue, Raptorx, Wurst and Bhageerath-H. The model generated by the server Sparks[x] showed all novel results to the evaluation servers abut failed in showing the Z-Score RMS value. The models generated by the threading servers Lomets, Muster and BioInfoMeta Server showed all satisfactory results but failed in the Ramachandran Plot generation for one or couple of proteins (Fig. Ib).

DISCUSSION

Computational approach for predicting of protein structure from its amino acid sequence started many years ago [19,20, 21, 22, 23] and has been successfully evolved by the development of many structure prediction servers and software [24,25,26]. The two methods used for protein three dimensional structures other than Homology Modelling are, Fold recognition and Ab-*initio* methods.

As discussed above Comparative modelling or Homology modelling [27] builds 3D models basing on the sequence alignment with single or multiple templates [24]. Following this approach model building can be by three ways. One of them is known as rigid-body assembly [27] which uses specific regions from related structures forming structural sketch by averaging of Cα atoms of structurally conserved regions in template structure and modelling is done by selecting known proteins from databases recognizing the best fit protein n structures for the conserved region. The other approach is by segment matching in which the target is dissected into short segments and constructing a framework of the target sequence by taking into account the positions of conserved atoms in templates [28,29] and searching in a database for similar fragments which can better match the constructed framework of the target sequence. The final approach is modelling by gratifying all the spatial restrains found from the target and template structural alignment [30], like distance and dihedral angle restraints obtained from the three dimensional analysis of the alignment. Model building is done in such a way that the resultant model should almost nullify the deviations from input restraints as far as possible. Few minor deviations from the restraints can be considered.

The fold recognition method can be broadly classified into types basing on the selection of the template namely Sequence-based and Structure -based [31]. The initial method does not take structural information from the templates into consideration. The relationship of sequence similarities are evaluated in the form of a profile or a Hidden Markov Model. In contrast to this, the later type of fold recognition uses the experimentally determined structural information from the template like secondary structural features and local environment issues (like backbone conformation, solvent accessibility) [32].

The *Ab-initio* methods of protein structure prediction are the most promising approaches as the predictions are not governed by classic knowledge-based assumptions but consider force fields and the lowest energy conformation is traced thus promising the most stable structure for the protein [33,34,35,36,37]. The vast number of feasible conformations of the protein space acts as the major hinder for the accuracy of the model prediction through *Ab-initio*

method [37]. De novo method designs the three dimensional protein structure by determining a template evaluating its folds such that the template is more stable than the target sequence [38]. *Ab-initio* method with de novo designing is proved to give better results of protein prediction [39,40].

According to the result it is clearly visible that the models generated by servers following the Homology Modelling/Comparative Modelling approach are showing comparatively better scores with respect to evaluation factors, than the models generated by the servers following Threading/Fold Recognition approach or *Ab-initio* approach. The models generated by the Threading /Fold Recognition approach showed very less deviations in the values of Z-Scores (namely Z-Scores of 2nd generation packing quality, Ramachandran plot, Chi 1- chi 2 rotamer normality and Back bone conformation) when compared to the models genetrated by the servers following other 2 approaches . The servers are classified (in the Results section) according to their deviations from the novel scores by limiting the boundary of the score range for each evaluation factor as i) >85% of residues in favourable region ii)>75% of Over-all quality score iii) Z- Scores of 2nd generation packing quality, Ramachandran plot, Chi 1- chi 2 rotamer normality and Back bone conformation as >-3.00 iv) 90% of residues with averaged 3D-1D Score> 0.2 and v) Z-Score RMS as near as possible to 1. The deviations from these boundaries are used to classify the servers. Some servers failed to give the results for Z-Scores (predicted by What If program), the reason behind this can be explained as the vast and undesired difference in the corresponding factors of the model when compared to that of the proved high-resolution structures. This gives higher probability for the inaccuracy of the model. The failure in obtaining the 'over-all quality score' by ERRAT program indicates the worse resolution of the model. The RMS Z-Score obtained by PROVE program fails because of the increase in the number of the buried outlier protein atoms for which the Z-Scores of their volume cross the upper limit of 3.0 away from the mean score of that atom type.

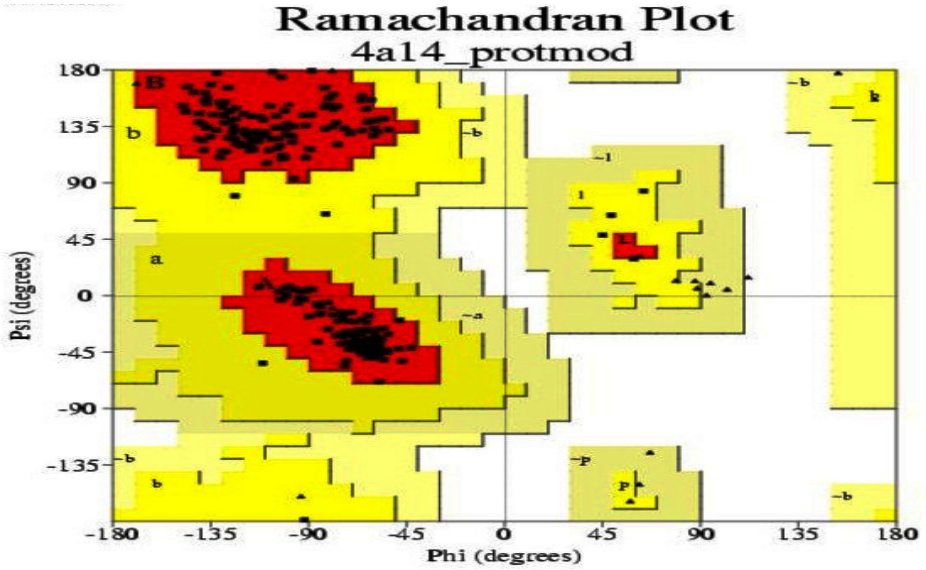

Figure: Ia The Ramachandran Plot for a good model

 ISSN 2231 – 4423

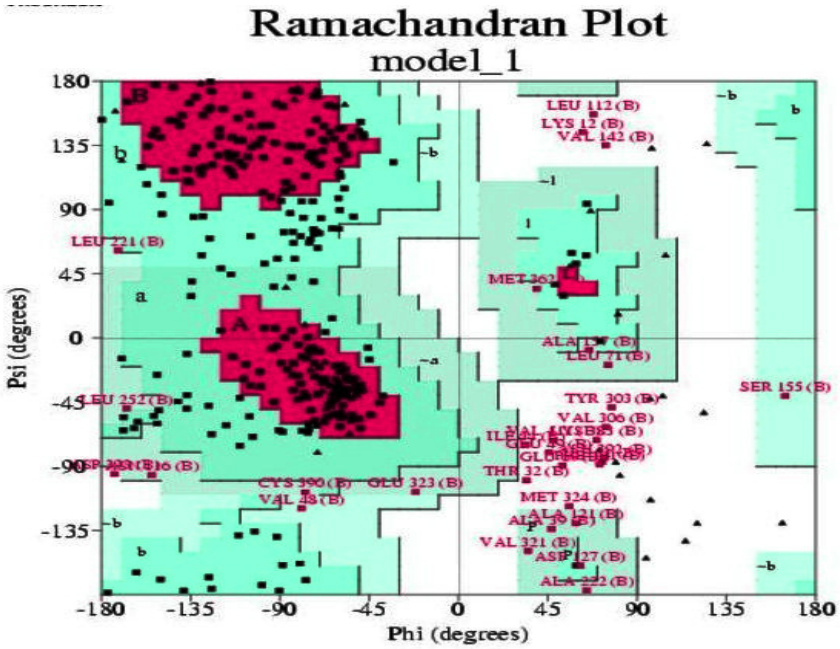

Figure: Ib The Ramachandran Plot for a bad model. The black dots inside the grids show the number of residues in the favorable region, which implies the quality of the model. The more the dots outside the grids the worse is the quality of the model.

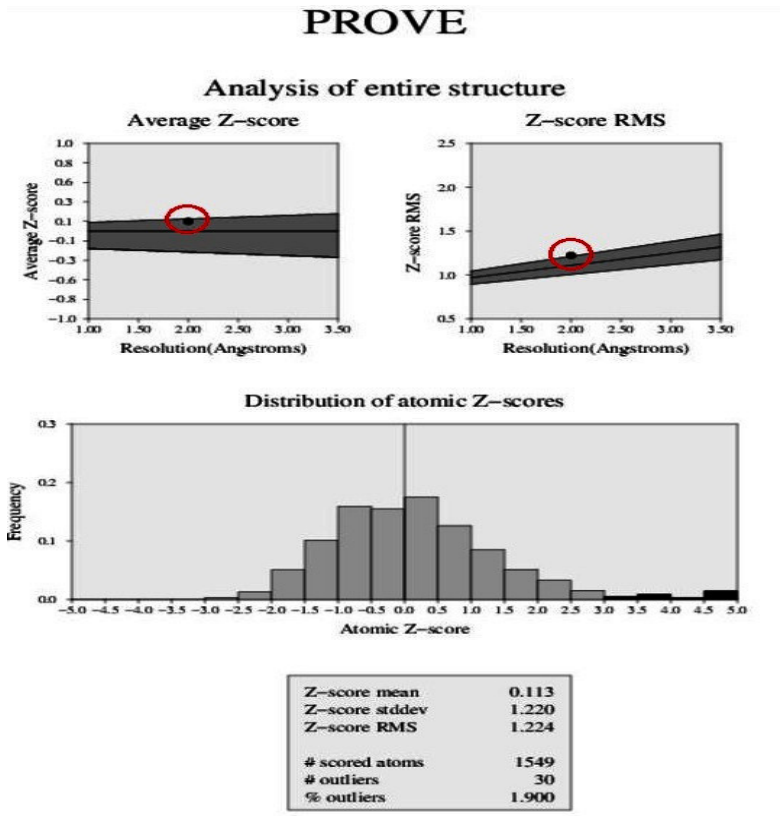

Figure: IIa The Prove Plot of a model with good Z-Score

 ISSN 2231 – 4423

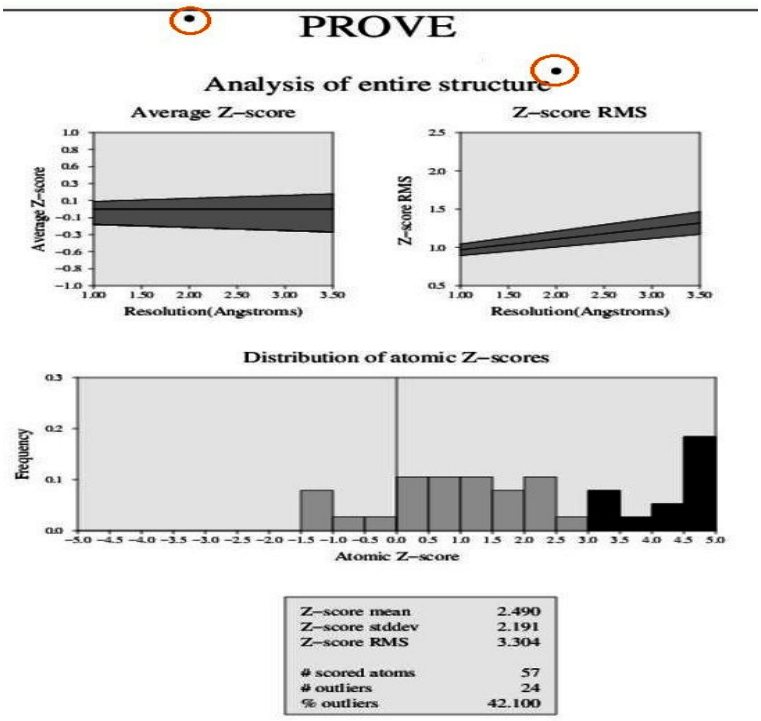

Figure: IIb The Prove Plot of a model with bad Z-Score. The black dot rounded up corresponds to the position of the outliers. The nearer the black dot (outliers) to the mean, the more will be the accuracy of the model.

Table I: The thirty online available servers used for comparison with their mechanism followed and websites.

Sl.No	Server name	Web link	Method followed
1.	Swiss Model	http://swissmodel.expasy.org/	Homology Modelling/Comparative Modelling
2.	Homer	http://protein.bio.unipd.it/homer/auto.html	
3.	PS2	http://ps2.life.nctu.edu.tw/index.php	
4.	EsyPred	http://www.fundp.ac.be/sciences/biologie/urbm/bioinfo/esypred/	
5.	CPH	http://ffas.burnham.org/protmod-cgi/checkLogin.pl	
6.	Protmod	http://bioinf.cs.ucl.ac.uk/bio_serf/public_job/	
7.	BioSerf	http://bmm.cancerresearchuk.org/~3djigsaw/	
8.	3D Jigsaw	http://wks16338.biology.ualberta.ca/proteus2/	
9.	Proteus	http://geno3d-pbil.ibcp.fr/cgi-bin/geno3d_automat.pl?page=/GENO3D/geno3d_home.html	
10	Geno 3D	http://tardis.nibio.go.jp/fugue/prfsearch.html	
11	Fugue	http://proteinmodel.org/AS2TS/AS2TS/as2ts.html	
12	AL2TS	http://manaslu.aecom.yu.edu/M4T/	

13	M4T	http://atome.cbs.cnrs.fr/AT2/meta.html	
14	@TOME	http://modbase.compbio.ucsf.edu/ModWeb20-html/modweb.html	
15	Modweb	http://bioinformatics.ibt.lt:8085/coma/	
16	COMA	http://toolkit.tuebingen.mpg.de/hhpred	
17	HHPred	http://compbio.soe.ucsc.edu/SAM_T08/T08-query.html	Threading/Fold Recognition
18	Sam T08	http://zhanglab.ccmb.med.umich.edu/I-TASSER/	
19	I Tasser	http://zhanglab.ccmb.med.umich.edu/LOMETS/	
20	Lomets	http://zhanglab.ccmb.med.umich.edu/MUSTER/	
21	Muster	http://meta.bioinfo.pl/submit_wizard.pl	
22	BioInfoBankMeta Server	http://sparks.informatics.iupui.edu/yueyang/sparks-x/	
23	Sparks[x]	http://raptorx.uchicago.edu/predict/	
24	Raptorx	http://neuropa.zbh.uni-hamburg.de/wurst/index.php	
25	Wurst	http://clsb.ices.utexas.edu/loopp/web/	

26	Loopp	http://www.reading.ac.uk/bioinf/IntFOLD/IntFOLD_form.html	
27	IntFOLD	http://cssb.biology.gatech.edu/skolnick/webservice/pro-sp3-TASSER/index.html	
28	Prosp3 Tasser	http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index	
29	Phyre2	http://www.scfbio-iitd.res.in/bhageerath/bhageerath_h.jsp	Ab-initio
30	Bhageerath-H		

Table II The five tools under the SAVES package used to evaluate the structure prediction result of the thirty servers.

Sl.No	Name of the tool	Function	Web link
1.	PROCHECK	Checks the Stereo-Chemical factors of the model, including Rama Chandran Plot analysis and psi-phi plots.	http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/
2.	WHATCHECK	Checks the Stereo Chemical properties of the residues in a broad scope.	http://swift.cmbi.kun.nl/swift/whatcheck/
3.	VERIFY 3D	Checks for compatibility between the 3D model generated and its amino acid sequence, basing on the study of valid structures.	http://nihserver.mbi.ucla.edu/Verify_3D/
4.	ERRAT	Analyzes the non-bonded interactions between the atom types and also shows the error % of the model by comparing it with highly refined structures.	http://nihserver.mbi.ucla.edu/ERRAT/
5.	PROVE	Calculates the Z-Score deviations of the model from highly resolved structures.	http://www.scmbb.ulb.ac.be/SCMBB/PROVE/

Table III Comparison of the 3D models generated for three proteins by 30 servers basing on their scores obtained in the quality estimation test.

Server	Protein	% of residues in favourable region	Over-all quality score	2nd generation packing quality Z-Score	Ramachandran plot Z-Score	Chi 1-chi 2 rotamer normality Z-Score	Back bone conformation Z-Score	% of residues with averaged 3D-1D Score> 0.2	Z-Score RMS
ORIGINAL	I	92.4%	98.760%	-0.061	1.032	0.186	-2.453	88.05%	1.057
	II	93.4%	93.750%	-1.1	-0.5	-0.3	-3.8	96.85%	39.679
	III	96.9%	97.826%	-0.675	0.598	0.330	-1.722	97.31%	44.932
Swiss Model	I	92.1%	98.776%	0.469	0.970	2. 526	-3.062	94.49%	1.189
	II	91.4%	94.301%	-0.277	-0.135	0.865	-5.633	97.97%	25.671
	III	87.7%	90.826%	-0.875	-0.290	1.747	-5.158	91.39%	1.339
Homer	I	82.8%	79.821%	-1.866	-5.700	4.881	-6.226	65.52%	1.928
	II	94.1%	93.247%	-0.339	-0.556	1.639	-5.141	95.19%	1.402
	III	94.1%	60.526%	-0.398	-0.211	3.576	-3.421	92.97%	1.585
PS2	I	94.7%	98.776%	-0.313	2.665	1.654	-2.868	93.70%	1.462
	II	95.6%	93.264%	-0.971	1.177	0.518	-4.961	93.92%	25.988
	III	90.6%	57.585%	-1.548	0.751	-0.976	-6.391	90.72%	1.671
EsyPred	I	95.2%	95.918%	-0.211	2.711	1.737	-2.856	95.28%	1.430
	II	93.4%	80.211%	-1.733	0.239	-0.732	-6.250	94.85%	38.481
	III	92.5%	73.125%	-1.433	1.296	0.119	-5.464	91.07%	1.574
CPH	I	90.7%	98.367%	0.221	-6.419	1.858	-2. 718	96.46%	1.192
	II	87.8%	88.571%	-0.762	-5.318	0.560	-6.686	95.43%	1.224
	III	84.0%	73.700%	-1.560	-4.711	0.140	-6.105	95.25%	1.404
Protmod	I	91.6%	94.672%	0.157	0.837	7.177	-2.884	95.65%	1.339
	II	93.1%	86.423%	-0.207	0.038	7.067	-4.839	95.17%	26.121
	III	96.9%	79.298%	-0.331	0.592	7.203	-1.550	94.61%	1.496
BioSerf	I	95.6%	98.367%	-0.566	2.668	0.243	-2.792	94.09%	1.469
	II	94.4%	91.451%	-1.160	1.040	0.629	-5.004	93.42%	25.964
	III	91.9%	72.289%	-1.543	1.287	0.498	-8.916	91.59%	1.529
Proteus	I	92.5%	------	0.296	0.785	0.972	-2.500	94.49%	1.552
	II	92.9%	59.221%	-1.248	-1.050	-0.310	-7.968	95.19%	26.532
	III	Homology Modelling failed							
3D Jigsaw	I	91.6%	98.776%	-0.488	-1.507	-1.669	-2.959	91.73%	1.488
	II	75.1%	58.549%	-2.635	-5.707	-3.397	-14.602	73.92%	27.932
	III	75.8%	54.103%	-3.142	-4.716	-0.051	-9.403	84.06%	1.877

Geno 3D	I	63.5%	75.720%	-3.972	-6.229	-4.391	-13.734	79.53%	2.103
	II	60.2%	81.937%	-3.896	-5.662	-3.804	-12.761	93.67%	41.837
	III	53.8%	76.147%	-4.156	-6.066	-4.131	-13.098	78.87%	32.193
Fugue	I	90.3%	0.000%	-10.801	-2.061	----	-25.456	0.00%	2.481
	II	94.2%	0.000%	-10.788	-0.468	----	-32.483	13.88%	3.261
	III	88.5%	1.389%	-10.697	-1.251	----	-23.551	6.56%	3.304
AL2TS	I	92.4%	100.00%	0.242	-0.161	1.436	-2.386	89.24%	1.143
	II	90.5%	83.333%	-0.870	-1.930	2.526	-5.702	99.75%	37.351
	III	96.1%	89.841%	-0.546	0.572	1.830	-3.651	95.44%	1.377
M4T	I	94.3%	98.776%	-0.215	2.816	1.856	-2.979	88.98%	1.425
	II	94.7%	94.560%	-1.152	1.158	-0.080	-6.635	92.91%	1.561
	III	91.2%	68.293%	-1.411	1.568	-0.427	-7.442	87.32%	30.068
@TOME	I	92.1%	100.00%	0.288	0.856	0.975	-2.599	95.67%	32.97
	II	95.9%	94.560%	-0.982	1.394	0.254	-4.914	99.75%	1.563
	III	92.8%	66.234%	-1.556	0.986	-0.409	-10.040	87.70%	1.665
Modweb	I	94.3%	99.592%	-0.382	2.883	0.501	-3.163	94.49%	1.486
	II	95.0%	94.819%	-0.732	0.929	-0.810	-5.636	95.95%	25.397
	III	91.8%	55.919%	-1.754	0.503	-1.375	-7.588	86.98%	1.708
COMA	I	84.0%	66.667%	-1.835	0.578	6.420	-12.347	53.97%	36.641
	II	93.1%	69.974%	-0.986	0.268	6.760	-6.054	90.82%	26.496
	III	90.3%	67.296%	-1.513	0.927	6.571	-7.573	84.64%	1.627
HHPred	I	94.7%	98.367%	-0.221	2.707	1.302	-2.886	87.80%	1.413
	II	95.6%	95.337%	-1.017	1.093	0.065	-5.437	95.44%	36.658
	III	93.6%	78.882%	-1.711	1.346	-0.341	-5.455	87.54%	1.549
Sam T08	I	91.2%	98.776%	0.243	1.200	3.961	-3.598	94.49%	32.751
	II	92.3%	90.674%	-0.034	0.189	5.767	-4.655	98.99%	1.316
	III	95.9%	77.679%	-0.658	1.241	4.324	-2.882	97.97%	1.617
I Tasser	I	94.7%	99.592%	-0.422	2.667	-0.192	-2.772	91.73%	1.461
	II	91.4%	45.078%	-1.900	-0.118	-2.954	-6.349	96.96%	27.639
	III	88.6%	41.493%	-2.022	-0.435	-3.301	-6.812	90.72%	42.952
Lomets	I	95.6%	97.551%	-0.711	2.793	0.668	-3.067	91.73%	1.425
	II	0.0%	91.710%	-1.045	1.236	0.506	-6.551	93.42%	1.597
	III	71.4%	67.365%	-2.207	0.819	-0.459	-9.165	82.32%	30.693
Muster	I	0.0%	97.551%	-0.322	2.816	1.047	-2.765	84.65%	1.456
	II	0.0%	91.710%	-1.045	1.236	0.506	-6.551	93.42%	1.597
	III	80.0%	76.994%	-1.450	1.101	0.958	-5.535	91.01%	1.461
BioInfoBank	I	0.0%	94.694%	-0.387	2.998	1.698	-2.896	96.46%	1.510
Meta Server	II	0.0%	92.228%	-1.145	0.984	-0.540	-5.167	93.16%	25.937
	III	91.8%	55.919%	-1.754	0.503	-1.375	-7.588	86.98%	1.708
Sparks[x]	I	95.6%	97.959%	-0.233	2.776	1.268	-2.821	91.73%	-----

	II	94.1%	77.865%	-1.772	0.601	-1.089	-6.529	96.46%	----
	III	88.3%	61.905%	-1.951	0.286	0.144	-9.454	88.70%	----
Raptorx	I	96.0%	-----	-0.582	-2.379	0.862	1.392	83.46%	96.533
	II	------	-----	-0.835	-1.870	0.476	-0.217	80.00%	-------
	III	88.9%	-----	-2.431	-2.756	-1.725	-3.334	84.64%	--------
Wurst	I	92.5%	98.086%	-7.189	1.490	------	-2.323	0.00%	2.403
	II	94.1%	98.760%	-7.204	-0.556	-------	-5.139	23.29%	2.416
	III	90.2%	99.419%	-7.195	-2.491	-------	-2.036	7.67%	2.298
Loopp	I	96.0%	94.286	-0.856	2.587	0.518	-3.560	97.64%	1.475
	II	95.0%	90.155	-1.196	1.042	0.517	-4.972	94.18%	25.816
	III	90.8%	57.187%	-1.747	1.152	-1.299	-6.136	92.08%	1.527
IntFOLD	I	95.2%	98.367%	-0.140	2.773	0.606	-2.798	92.13%	1.393
	II	95.6%	95.337%	-1.017	1.093	0.065	-5.437	95.44%	36.658
	III	91.9%	72.508%	-1.319	1.117	-0.158	-5.534	95.36%	1.656
Pro sp3 Tasser	I	86.8%	70.204%	-0.367	-1.124	-0.072	-3.709	79.92%	60.414
	II	84.0%	43.782%	-1.506	-2.257	-0.561	-7.079	95.19%	54.399
	III	76.8%	50.000%	-1.750	-3.473	-0.343	-10.509	88.12%	2.249
Bhageerath-H	I	85.5%	-----	------	-3.204	-2.888	-------	0.00%	51.822
	II	90.2%	-----	------	-2.646	-1.979	------	0.00%	37.354
	III	79.5%	-----	------	-3.922	-2.897	------	0.00%	30.929
Phyre2	I	92.5%	100.00%	0.296	1.043	0.970	-2.807	88.19%	1.232
	II	94.1%	92.208%	-0.364	-0.556	1.472	-5.139	94.43%	1.420
	III	91.1%	71.166%	-0.942	-1.271	3.994	-6.478	89.85%	1.752

CONCLUSION

Because of large number of software and servers coming into scene for the prediction of the 3 Dimensional structures of proteins, our study provides an easy way to analyse and appreciate the models generated. The above mentioned tools to access the quality of the protein models are not only easy to be performed but are also faster and freely available.

ACKNOWLEDGEMENT

We gratefully acknowledge BTISNet SubDIC (BT/BI/065/2004) for providing internet facilities and the Government of Jharkhand, Dept. of Agriculture for providing infrastructure development fund. RB is thankful for getting Cumulative Professional Development Grant (CPDG) from BIT, Mesra during preparation of this manuscript.

REFERENCES

1. Floudas C.A., Fung H.K., McAllister S.R., Monnigmann M., Rajgaria R. Advances in protein structure prediction and de novo protein design: A review. Chemical Engineering Science.2006; 61:966 – 988.
2. Nuria B.C., Joan P.I., Baldomero O. Comparative modelling of protein structure and its impact on microbial cell factories. Microbial Cell Factories.2005 4:20, doi:10.1186/1475-2859-4-20.

3. Marc A.M.R., Ashley C.S., Andras F.R.S., Francisco M., Andrej S. Comparative Protein Structure Modeling of Genes and Genomes. Annu Rev Biophys Biomol Struct.2009. 29:291–325.
4. Bruno C.M., Paul W.F. , Paul A.B.. Insilico Protein Recombination: Enhancing Template and Sequence Alignment Selection for Comparative Protein Modelling. J Mol Biol.2003; 328:593–608.
5. Donald P., Zhexin X., Christopher L.T., Lei X., Marina G., Therese M., Cinque S.S., Sharon G.F., Andrew K., Avner S., Ingrid Y.Y.K., Emil A., Barry H.. Using Multiple Structure Alignments, Fast Model Building, and Energetic Analysis in Fold Recognition and Homology Modelling. PROTEINS: Structure, Function, and Genetics.2003; 53:430–435.
6. Johannes S.. Protein homology detection by HMM–HMM comparison. Bioinformatics.2005; 21:951–960.
7. Matthew J., Andrej S.. Comparative Protein Structure Modelling and its Applications to Drug Discovery. Annual Reports In Medicinal Chemistry.2004; 39:259-273.
8. Andriy K., Oleh K., Pawel D., Zinovii D., Krzysztof F.. Protein structure prediction center in CASP8. Proteins: Structure, Function, and Bioinformatics.2009; 77:5-9.
9. Domenico C., Andriy K., Krzysztof F., John M., Burkhard R., Anna T. Evaluation of template-based models in CASP8 with standard measures. Proteins: Structure, Function, and Bioinformatics.2009; 77:18-28.
10. Leonid A.M., Eugene I.S.. Protein Structure Prediction by Threading. Why it Works and Why it Does Not. J Mol Biol.1998; 283:507-526.
11. Chris A.E.M.S., Sander B.N., Elmar K., Gert V., Geerten W.V. Validation of protein structures derived by NMR spectroscopy. Progress in Nuclear Magnetic Resonance Spectroscopy.2004; 45:315–337.
12. Rolando R., Glay C., Nelia L., Tirso P., Gert V.. Homology modeling, model and software evaluation: three related resources. Bioinformatics.1998; 14:523-528.
13. Donald P., Barry H.. Free energy determinants of tertiary structure and the evaluation of protein models. Protein Science.2000; 9:2181–2191.
14. Paolo M., Maria L.G., Elena P., Piercarlo F., Luca D.G. Validation of protein models by a neural network approach. BMC Bioinformatics.2008; 9:66.
15. Nagano K.. Logical analysis of the mechanism of protein folding. I. Predictions of helices, loops and beta-structures from primary structure. J Mol Biol.1973; 75:401–420.
16. Ramakrishnan C.. Ramachandran and his map. Resonance.2001; 48-56.
17. Spronk C.A.E.M., Nabuurs S.B., Krieger E., Vriend G., Vuister G.W.. Validation of protein structures derived by NMR spectroscopy. Progress in Nuclear Magnetic Resonance Spectroscopy.2004; 45:315–337.
18. Dunbrack R.L.Jr., Karplus M.. Conformational analysis of the backbone-dependent rotamer preferences of protein sidechains. Nature Structural Biology.1994; 1:334-340.
19. Chou P.Y., Fasman G.D.. Prediction of protein conformation. Biochemistry.1974; 13:222–245.
20. Floudas C.A.. Computational Methods in Protein Structure Prediction. Biotechnology and Bioengineering.2007; 97:207-213.
21. Daiya T., Peter A.J.. The CpG Island Searcher: A New WWW Resource. In Silico Biology.2003; 3:235-240.
22. Soumita D., Keya C.. Identification of a Unique IAHP (IcmF Associated Homologous Proteins) Cluster in Vibrio cholerae and other Proteobacteria through In Silico Analysis. In Silico Biology.2002; 3:287-300.
23. Niels G.. AliBaba2: Context Specific Identification of Transcription Factor Binding Sites. In Silico Biology.2002; 2: S1-S15.

24. Andrzej K., Dominik G.. Comparative modeling without implicit sequence alignments. Bioinformatics.2007; 23:2522–2527.
25. Michael Y.G., Eugene V.K.. Sources of Systematic Error in Functional Annotation of Genomes: Domain Rearrangement, Non-Orthologous Gene Displacement and Operon Disruption. In Silico Biology.1998; 1:55-67.
26. Kishore R.S., Meena K.S., Vincent T.K.C.. A Novel Genomics Approach for the Identification of Drug Targets in Pathogens, with Special Reference to Pseudomonas Aeruginosa. In Silico Biology.2004; 4:355-360.
27. Blundell T.L., Sibanda B.L., Sternberg M.J., Thornton J.M.. Knowledge based prediction of protein structures and the design of novel Molecules. Nature.1987; 326:347-352.
28. Jones D.T., Taylor W.R., Thornton J.M.. A new approach to protein fold recognition. Nature.1992; 358:86-89.
29. Kelley L.A., Mac C.R.M., Sternberg M.J.. Enhanced genome annotation using structural profiles in the program 3D-PSSM. J Mol Biol.2000; 299:499-520.
30. Havel T.F., Snow M.E.. A new method for building protein conformations from sequence alignments with homologues of known structure. J Mol Biol.1991; 217:1-7.
31. James U.B., Roland L., David E.. A method to identify protein sequences that fold into a known three dimensional structure. Science.1991; 253:164-170.
32. Cymerman I., Feder M., PawLowski M., Kurowski M.A., Bujnicki J.M.. Computational Methods for Protein Structure Prediction and Fold Recognition. Nucleic Acids and Molecular Biology.2004; 15:1- 34.
33. Klepeis J.L., Floudas C.A.. Ab initio tertiary structure prediction of proteins. J Global Optim.2003; 25:113–140.
34. Klepeis J.L., Floudas C.A.. ASTRO-FOLD: a combinatorial and global optimization framework for ab initio prediction of three dimensional structures of proteins from the amino-acid sequence. Biophysical J.2003; 85:2119–2143.
35. John L.K., Yinan W., Michael H.H., Christodoulos A.F.. Ab initio Prediction of the Three-Dimensional Structure of a De novo Designed Protein: A Double-Blind Case Study. PROTEINS: Structure, Function, and Bioinformatics.2005; 58:560–570.
36. Liwo A., Lee J., Ripoll D., Pillardy J., Scheraga H.. Surmounting the multiple minima problem in protein folding. Proc Natl Acad Sci.1999; 96:5482–5485.
37. Pillardy J., Czaplewski C., Liwo A., Wedemeyer W., Lee J., Ripoll D., Arlukowicz P., Oldziej S., Arnautova E., Scheraga H.. Development of physics-based energy functions that predict medium resolution structure for proteins of α, β and α/β structural classes. J. Phys. Chem B.2001; 105:7299-7311.
38. Fung H.K., Taylor M.S., Floudas C.A.. Novel Formulations for the Sequence Selection Problem in De novo Protein Design with Flexible Templates. Optimization Methods and Software.2006; 00:1-27.
39. Bin Q., Srivatsan R., Rhiju D., Philip B., Airlie J.M.C., Randy J.R., David B.. High-resolution structure prediction and the crystallographic phase problem. Nature.2007; 450:259-264.
40. Rojan S., Francois B., Kam Y.J.Z.. Accelerating ab initio phasing with de novo models. Acta Cryst.2011; D67:804–812.

REFERENCES FOR MODELLING SERVERS

(The modelling servers are mentioned according to their serial number in the Table I)

1. Arnold K., Bordoli L., Kopp J., Schwede T.. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. Bioinformatics.2006; 22:195-201.
2. http://protein.bio.unipd.it/homer/auto.html
3. Chen C.C., Hwang J.K., Yang J.M.. (PS)2: protein structure prediction server. NAR.2006; 34:W152-W157.
4. Lambert C., Leonard N., De B.X., Depiereux E.. ESyPred3D: Prediction of proteins 3D structures. Bioinformatics.2002; 18:1250-1256.
5. Nielsen M., Lundegaard C., Lund O., Petersen T.N.. CPHmodels-3.0 - Remote homology modeling using structure guided sequence profiles. Nucleic Acids Research.2010; 38:6. http://ffas.burnham.org/protmod-cgi/checkLogin.pl
6. http://bioinf.cs.ucl.ac.uk/bio_serf/public_job/
7. Bates P.A., Kelley L.A., Mac C.R.M., Sternberg M.J.E.. Enhancement of Protein Modelling by Human Intervention in Applying the Automatic Programs 3D-JIGSAW and 3D-PSSM. Proteins: Structure, Function and Genetics.2001; 5:39-46.
8. http://wks16338.biology.ualberta.ca/proteus2/contact.jsp.
9. Combet C., Jambon M., Deleage G., Geourjon C.. Geno3D: Automatic comparative molecular modelling of protein. Bioinformatics.2002; 18:213-214.
10. 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; 310:243-57.
11. Altschul S.F.. Amino acid substitution matrices from an information theoretic perspective. J. Mol. Biol.1991; 219:555-565.
12. http://manaslu.aecom.yu.edu/M4T/
13. Pons L.J., Labesse G., TOME-2: a new pipeline for comparative modeling of protein–ligand complexes. Nucleic Acids Research.2009; 37:W485-W491.
14. http://modbase.compbio.ucsf.edu/ModWeb20-html/modweb.html
15. Margelevicius M., Venclovas C.. Detection of distant evolutionary relationships between protein families using theory of sequence profile-profile comparison. BMC Bioinformatics.2010; 11:89.
16. Biegert A., Mayer C., Remmert M., Soding J., Lupas A.. The MPI Toolkit for protein sequence analysis. Nucleic Acids Res.2006; 34:W335-339.
17. George S., Kevin K.. Contact Prediction using Mutual Information and Neural Nets. Proteins: Structure, Function, and Bioinformatics.2007; 69:159-164.

18. Ambrish R., Alper K., Yang Z.. I-TASSER: a unified platform for automated protein structure and function prediction. Nature Protocols.2010; 5:725-738.

19. Wu S., Zhang Y.. LOMETS: A local meta-threading-server for protein structure prediction. Nucleic Acids Research.2007; 35: 3375-3382.

20. Wu S., Zhang Y.. MUSTER: Improving protein sequence profile-profile alignments by using multiple sources of structure information. Proteins: Structure, Function, and Bioinformatics.2008; 72:547-556.

21. Ginalski K., Elofsson A., Fischer D., Rychlewski L.. 3D-Jury: a simple approach to improve protein structure predictions. Bioinformatics.2003; 19:1015-1018.

22. Yuedong Y., Eshel F., Huiying Z., Yaoqi Z.. Improving protein fold recognition and template-based modeling by employing probabilistic-based matching between predicted one-dimensional structural properties of the query and corresponding native properties of templates. Bioinformatics.2011; 27:2076-2082.

23. Feng Z., Jian P., Joe D.B., Karl F.F., Tobin R.S., Jinbo X.. A probabilistic and continuous model of protein conformational space for template-free modelling. J Comput Biol.2010; 17:783-798.

24. Torda A.E., Procter J.B., Huber T.. Wurst: A protein threading server with a structural scoring function, sequence profiles and optimized substitution matrices. Nucleic Acids Res.2004; 32:W532-W535.

25. Brinda K.V., Jaroslaw P., Peter M., Jaroslaw M., Thomas B., Bao Q.C., Ron E.. Building and assessing atomic models of proteins from structural templates: Learning and benchmarks. Proteins: Structure, Function, and Bioinformatics.2009; 76:930-945.

26. Roche D.B., Buenavista M.T., Tetchner S.J., McGuffin L.J.. The IntFOLD server: an integrated web resource for protein fold recognition, 3D model quality assessment, intrinsic disorder prediction, domain prediction and ligand binding site prediction. Nucleic Acids Res.2011; 39:W171-W176.

27. Zhou H., Pandit S.B., Skolnick J.. Performance of the Pro-sp3-TASSER Server in CASP8, Proteins.2009; 77:123-127.

28. Kelley L.A., Sternberg M.J.E.. Protein structure prediction on the web: a case study using the Phyre server. Nature Protocols.2009; 4:363 – 371.

29. Jayaram B., Dhingra P., Lakhani B., Shekhar S.. Bhageerath - Targeting the Near Impossible: Pushing the Frontiers of Atomic Models for Protein Tertiary Structure Prediction. Journal of Chemical Sciences.2012; 124:83-91.