

RESEARCH

Open Access

Bhageerath-H: A homology/*ab initio* hybrid server for predicting tertiary structures of monomeric soluble proteins

B Jayaram^{1,2,3*}, Priyanka Dhingra^{1,2}, Avinash Mishra^{2,3}, Rahul Kaushik^{2,3}, Goutam Mukherjee^{1,2}, Ankita Singh², Shashank Shekhar²

From Asia Pacific Bioinformatics Network (APBioNet) Thirteenth International Conference on Bioinformatics (InCoB2014)

Sydney, Australia. 31 July - 2 August 2014

Abstract

Background: The advent of human genome sequencing project has led to a spurt in the number of protein sequences in the databanks. Success of structure based drug discovery severely hinges on the availability of structures. Despite significant progresses in the area of experimental protein structure determination, the sequence-structure gap is continually widening. Data driven homology based computational methods have proved successful in predicting tertiary structures for sequences sharing medium to high sequence similarities. With dwindling similarities of query sequences, advanced homology/ *ab initio* hybrid approaches are being explored to solve structure prediction problem. Here we describe *Bhageerath*-H, a homology/ *ab initio* hybrid software/server for predicting protein tertiary structures with advancing drug design attempts as one of the goals.

Results: *Bhageerath*-H web-server was validated on 75 CASP10 targets which showed TM-scores ≥ 0.5 in 91% of the cases and $C\alpha$ RMSDs $\leq 5\text{\AA}$ from the native in 58% of the targets, which is well above the CASP10 water mark. Comparison with some leading servers demonstrated the uniqueness of the hybrid methodology in effectively sampling conformational space, scoring best decoys and refining low resolution models to high and medium resolution.

Conclusion: *Bhageerath*-H methodology is web enabled for the scientific community as a freely accessible web server. The methodology is fielded in the on-going CASP11 experiment.

Background

“The native conformation of a protein is determined by the totality of interatomic interactions and hence, by the amino acid sequence, in a given environment” (Nobel Lecture, Christian B. Anfinsen, December 11, 1972). According to Anfinsen’s protein folding hypothesis, a protein’s native structure is determined by its amino acid sequence which drives protein into its minimum Gibbs energy state [1]. This hypothesis evolved as a basic tenet for protein structure prediction algorithms (PSPAs). However limited

understanding of net balance of forces involved in protein folding creates deficiencies in various proposed PSPAs. One of the early efforts in solving protein folding problem was driven by thermodynamic calculations, which incorporate searching algorithms to investigate a conformation that corresponds to minimum free energy [2]. Here the large number of degrees of freedom of a protein gives rise to innumerable conformations, an enumeration of which is practically impossible. This despite, proteins fold rapidly into their native structure in milliseconds to seconds time scales implying that a brute force enumeration of all possible conformations may not be required as implicit in Levinthal’s Paradox [3]. The fact that sequence introduces local structural bias, narrows down the accessible

* Correspondence: bjayaram@chemistry.iitd.ac.in

¹Department of Chemistry, Indian Institute of Technology, Hauz Khas, New Delhi-110016, India

Full list of author information is available at the end of the article

conformational space and introduces local as well as long range interactions, suggesting a halfway solution to the paradox [4-7]. As a result, PSPAs need two key components: (a) a rapid computational algorithm for protein conformational search and (b) an accurate scoring function to capture the best available conformation. The first component involves use of different physics based as well as knowledge based approaches for extensive sampling of the vast conformational space [8,9]. Physics based sampling methods include use of Monte Carlo (MC) methods [10-15], Genetic algorithms [16], molecular dynamics simulations (MD) [17,18], simulated annealing [19,20], replica-exchange MC or MD and local enhanced sampling [21-23]. Knowledge based methods use information from the solved protein structures and knowledge based potentials for sampling protein conformational space [24].

Homology modeling [25-30] and fold recognition/threading methods [31-35] are knowledge based approaches, which are routinely used to generate reliable models for proteins with overall fold topology similar to an available template in the protein databases. Query protein with no sequence and structural similarity are modeled from scratch using physics based/ *ab initio* approaches. The success of *ab initio* or physics-based sampling methods is limited by lack of accurate energy functions [36,37], heavy computational requirements, force field errors [38-40] and protein size, while knowledge based approaches are limited by sequence similarity and evolutionary relationships [41-43]. A popular trend in protein conformational sampling is the fragment assembly method, which uses parts of known protein or protein fragments to generate a structure of the target. After conformational sampling, the next immediate concern is to capture the best available structure by means of a scoring function [44-56]. These functions combine chemical, physical, geometrical and energetic constraints to capture native or near native models [57,58].

A thorough literature survey reveals that the available protein structure prediction algorithms are based on methods such as (a) homology modeling, (b) fold recognition, (c) *ab initio* and (d) hybrid [59,60]. Different software/tools are available in the public domain based on these computational approaches and are evaluated every two years during the Critical Assessment of techniques for protein structure prediction (CASP experiments) [61]. Recent CASP experiments have shown significant progress by hybrid approaches, which combine homology, *ab initio* along with atomic level model refinements for protein structure prediction [62]. This article describes *Bhageerath-H*, a homology/ *ab initio* hybrid software for predicting tertiary structure of monomeric proteins. *Bhageerath-H* makes use of *Bhageerath-H* Strgen algorithm [63] for extensive sampling

of the protein fold space and generates a large basket of decoys containing near-native protein conformations, which are further supplemented by a chemical logic based alignment scheme and then clustered to eliminate non-unique redundant structures. These are then screened by a physico-chemical scoring metric (pcSM) and assessed for their quality. The selected models are refined via a unique and effective quantum mechanics based loop bond angle optimization method, which drives the selected models further close to the native topology. *Bhageerath-H* automated pipeline is freely available to the scientific community across the world via http://www.scfbio-iitd.res.in/bhageerath/bhageerath_h.jsp.

Methodology

Bhageerath-H software suite for protein tertiary structure prediction narrows down the conformational search space and predicts five probable near native candidate structures for an input amino acid sequence. The software comprises seven computational modules which work in conduit and together form an automated pipeline. Figure 1 shows a diagrammatic representation of *Bhageerath-H* software suite. Following sections discuss each module of the automated pipeline.

(A) *Bhageerath-H* Strgen for candidate structures

The first step in the pipeline involves generation of a large pool of full length decoys. In the proposed protein structure prediction pipeline, *Bhageerath-H* Strgen algorithm for protein conformational sampling [63] is the first module. The module takes as input protein amino acid sequence and provides as output a large pool of decoys. A revised and improved version of structure generation algorithm is incorporated in the *Bhageerath-H* software suite. *Bhageerath-H* Strgen makes use of the current sequence and structural database knowledge

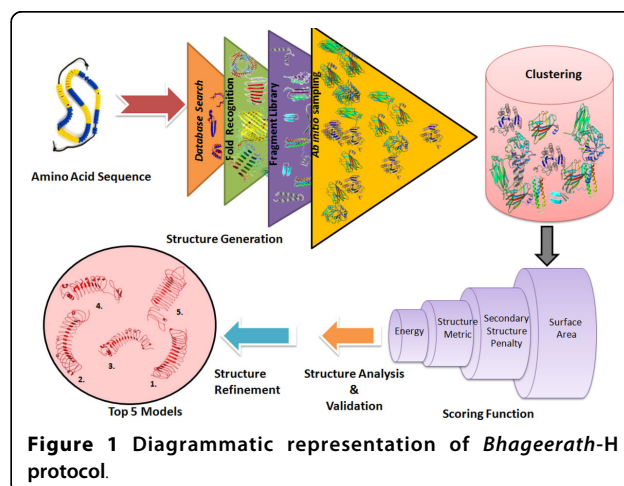


Figure 1 Diagrammatic representation of *Bhageerath-H* protocol.

along with *Bhageerath ab initio* folding [64,65] in order to effectively search the fold space for an input protein sequence. It starts with amino acid sequence, followed by secondary structure prediction and BLAST [66] search for sequence based homologs. In addition, it also searches for distant analogs and structural homologs using tools such as pGenthrader [67,68], ffas [69,70], spark-x [71] and HHSearch [72]. A new addition to this methodology is a chemical logic based [73] procedure for template selection followed by alignment generation. It utilizes amino acid chemical properties such as hydrogen bond donor, conformational flexibility, shape and size of side chains for generating an amino acid substitution scoring matrix. This scoring matrix is used for template selection as well as template-target alignment generation. The matrix helps in selecting distant homologs, which are generally missed during a normal database search. The templates and template-target alignments are used for modeling fragments of varying length via Modeller [74,75]. Modeled fragments are then screened for missing links with no available templates. These missing stretches are generated using *Bhageerath ab initio* modeling method [65,76,77]. All the incomplete protein fragments are patched in order to generate full-length models, which are energy scored and top 5 lowest energy decoys are sent for *Bhageerath ab initio* loop sampling. The newly sampled structures are added to the growing pool of full length protein decoys. The output of the first step is a large pool of protein decoys. The average size of the decoy pool is on the order of 10^4 - 10^5 structures.

Bhageerath-H Strgen module includes locally installed copies of Psipred, BLAST, PFAM [78], SCOP [79,80], nr [81], pdb database [82] <http://www.pdb.org/pdb/home/home.do>, HHSearch, Spark-X, pGenthrader, ffas and modeller. The scalable *Bhageerath-H* Strgen algorithm is currently configured to utilize 64 processors of Linux Cluster. Programs are written in C++, MPI language and involve use of linux shell scripting. Average time taken for *Bhageerath-H* Strgen run is 1-2hrs. This first module of the *Bhageerath-H* pipeline generates a large pool of decoys which needs to be further filtered, processed and refined. We would like to note that *Bhageerath-H* software is not just limited to *Bhageerath-H* Strgen an already published algorithm. *Bhageerath-H* Strgen is a protein decoy generation program which is the first module here. After protein decoy generation, protein decoy selection and refinement are the other two very important steps in protein structure prediction pipeline. In *Bhageerath-H* software modules 2-5 are dedicated for decoy clustering, selection and refinement, which are not included in *Bhageerath-H* Strgen. Output from this module is submitted for clustering in the next step.

(B) Clustering

Recurring structural models sampled in the previous step are clustered using K-means clustering algorithm. The main aim of this step is to retain a single representative structure of each unique topology. MMTSB [83] toolkit's *k-clust* is used to perform clustering. The tool requires list of protein decoys to cluster. Following command was executed:

```
kclust -modermsd -cdist -heavy -lsqfit -radius 1.0 -maxerr 1 pdblist > cluster_file
```

This command gives as an output a cluster file, which contains the centroid in the pdb format along with the members of each centroid and the root mean square deviation (rmsd) distance of each member from the centroid. The centroids themselves are mathematical constructs and convey no information, but utilizing rmsd information one lowest rmsd member from each cluster is picked [83]. To overcome the time limitation, clustering is performed in a parallel mode. The output of K-mean clustering is a set of decoys, which are unique, non-recurring and contain near-native structural models. This set of decoys containing near-native models is submitted for physico-chemical scoring in the third step.

(C) Scoring based on a physico-chemical metric

The third step in the *Bhageerath-H* pathway involves the use of a robust metric that combines chemical, physical, geometrical and energetic constraints known to show universalities among native protein structures. The physico-chemical scoring metric (pcSM) consists of different parameters, which include (a) **P**: Secondary structure penalty, (b) **M**: Euclidean distance, (c) **A1-A4**: Surface areas and (d) **E**: Empirical potential energy functions. The scoring function calculates a final cumulative score (CS), which comprises each of these parameters.

$$CS = c_{A1}A1 + c_{A2}A2 + c_{A3}A3 + c_{A4}A4 + c_{pmax}(P_H, P_S) + c_{M1}M1$$

where A1 is the fractional area of exposed non-polar residues, A2 is the fractional area of exposed non polar part of residues, A3 is the weighted exposed area, A4 is the total surface area, PH and Ps are secondary structure penalties for helix and sheet respectively, M1 is Euclidean distance. The prefix "c" for each parameter in the above equation refers to its optimized coefficient. $c_{A1} = 10$, $c_{A2} = 0.1$, $c_{A3} = 0.00001$, $c_{A4} = 0.001$, $c_{M1} = 0.001$, $c_p = 0.15$ (P_H) and 0.21 (P_S).

In order to get the top 10 structures, each of the seven parameters are evaluated for all the clustered decoys and a short energy minimization is performed to remove steric clashes. For the given input decoy pool, pcSM gives as an output top 10 ranked native-like candidates structures. pcSM algorithm runs in parallel mode and utilizes 64 processors. On an average, time

taken for scoring varies from 2 to 3 hours. The top 10 pcSM ranked models are submitted for protein structure analysis and validation in the next step.

(D) Protein Structure Analysis and Validation (PROTSAV) based ranking

PROTSAV is a protein structure quality assessment meta-server (manuscript under preparation). Currently, it comprises six tools namely Procheck [84], Verify-3D [85], ERRAT [86], Naccess [87], PROSA [88] and dDFIRE [89], for quality assessment of protein structures. PROTSAV generates an overall protein quality score, which is a summation of scores predicted by individual modules. High PROTSAV values reflect poor structure quality of query protein and low values close to zero represent good quality of query protein structure. Run time for this module is 40-45 seconds. In this step, pcSM selected top 10 protein models are analysed and ranked. The top ranked model is submitted for QM based loop bond angle refinement in the next step.

(E) Quantum mechanics (PM6) based loop bond angle optimization

Quantum mechanics (PM6) based loop bond angle optimization (manuscript in preparation) takes topmost PROTSAV selected model as an input, optimizes loop bond angles and performs *ab initio* loop sampling [66]. The small pool of decoys generated in the process is side chain optimized using Scwrl4. Scwrl4 is a program for prediction of protein side chain conformation [90]. Scwrl4 uses latest backbone-dependent library to provide rotamer frequency, dihedral angles and variances. The side chain optimized decoys are further energy minimized (SD = 500, CG = 500) using sander module of AMBER10 software [91].

These optimized and energy minimized refinement generated decoys are scored using pcSM and the top 10 ranked QM refined models are passed to next step.

(F) Final ranking

Input to this step is top 10 pcSM ranked QM refined models from step (E) and top 5 PROTSAV ranked models from the step (D). PROTSAV ranked models are side chain optimized and energy minimized before final ranking. The selected 15 models are re-ranked using pcSM and the top 5 are given to the user as an output.

The *Bhageerath-H* protocol is a careful combination of different algorithms which are configured to work in conduit. Starting from *Bhageerath-H* Strgen followed by clustering, pcSM scoring, PROTSAV and QM refinement each module has its own importance and role in providing the user, near-native candidate structures as final output. The software takes protein amino acid sequence as input and provides a user as output five native-like candidate

structures. Figure 2 shows the flow chart of *Bhageerath-H* software suite.

Results and Discussion

Validation of *Bhageerath-H* software suite

Bhageerath-H automated pipeline was thoroughly tested and validated on the benchmark CASP10 dataset. Each CASP experiment reveals the state of the art in the field of protein structure prediction. About 75 CASP10 targets of varying size and complexity were considered here for the analysis. To begin with the assessment, CASP-like conditions were mimicked, which means the native and near-native homologs were excluded during structure prediction. Any template released later than the first CASP10 server target i.e. fifth May of 2012 was not considered. For structure assessment an automated pipeline was developed. For each CASP10 target, sequence was extracted from the native structure. Then predicted structure sequence and the native sequence were aligned using ClustalW [92]. Residues with missing coordinates were removed from the predictions in order to make the sequence of the two structures match exactly. The native and the *Bhageerath-H* generated final five models were compared based on the widely used criteria of C α root mean square deviation (C α RMSD) and Template modeling score (TM-score). C α RMSD is a global indicator of structural identity, while TM-score identifies local substructures and evaluates local identity. TM-score refers to template modeling score. TM-score is considered as a quantitative measure for classification of

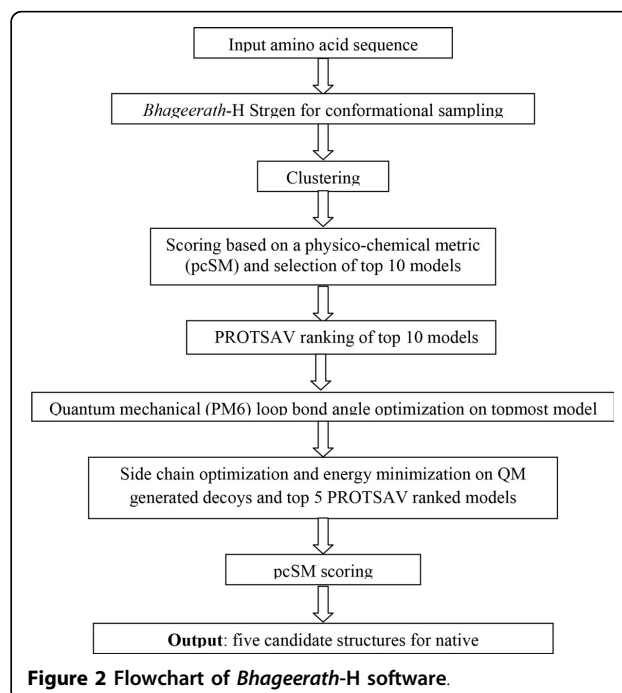


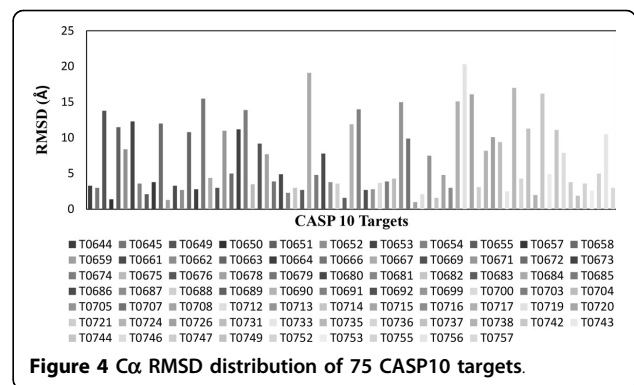
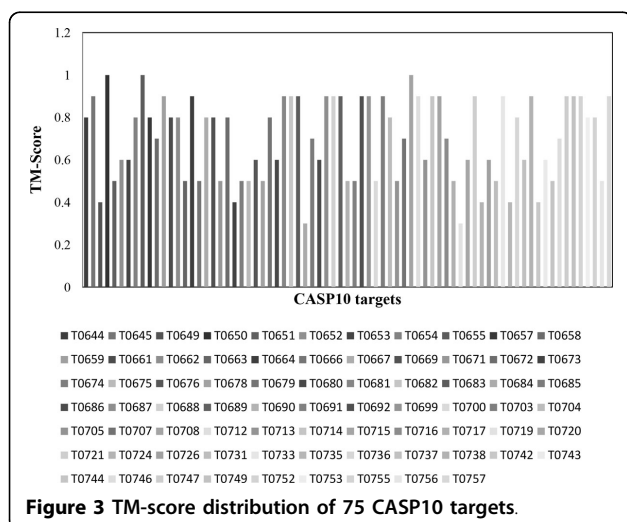
Figure 2 Flowchart of *Bhageerath-H* software.

protein topology. A TM-score > 0.5 signifies that protein pairs share same fold whereas a TM-score < 0.5 are mostly not of the same fold and a TM-score of 0.17 indicates random prediction [93,94].

(A) Bhageerath-H performance on 75 CASP10 targets

Bhageerath-H was validated on 75 CASP10 targets. $C\alpha$ RMSDs and TM-scores of final five *Bhageerath-H* predictions from the native were calculated. In 68 out of 75 systems i.e. in 91% of the cases *Bhageerath-H* predicted model has a TM-score ≥ 0.5 , while in 44 targets i.e. in 59% of the cases *Bhageerath-H* was able to predict a model in top 5 having a $C\alpha$ RMSD from the native $\leq 5.0\text{\AA}$ (Additional File 1). Figure 3 shows the TM-score distribution and Figure 4 shows the $C\alpha$ RMSD distribution of all the 75 targets.

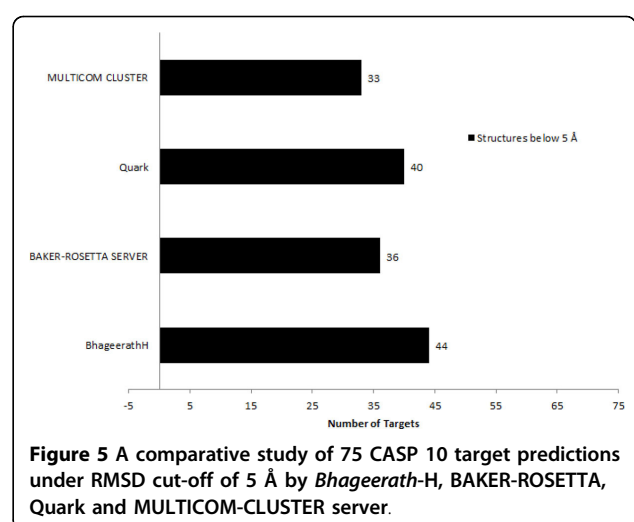
Comparison of Bhageerath-H performance with BAKER-ROSETTA, Quark and MULTICOM-CLUSTER For comparative analyses, we considered three state-of-the-art servers for protein tertiary structure prediction. Predictions submitted by BAKER-ROSETTA [95], Quark [96] and MULTICOM-CLUSTER [97] during CASP10 [62] experiment were used. Their submitted five predictions were downloaded from the CASP10 website <http://www.predictioncenter.org/casp10/index.cgi> and analyzed using the automated evaluation pipeline described above. The minimum RMSD obtained among the five submitted models was considered. In 36 cases, BAKER-ROSETTA server submitted a model among five predictions having $C\alpha$ RMSD from the native $\leq 5.0\text{\AA}$. Quark submitted 40 predictions among 75 under the $C\alpha$ RMSD cutoff of 5.0\AA , whereas MULTICOM-CLUSTER succeeded in 33 cases. In comparison to these three servers, *Bhageerath-H* server was successful in 44 cases i.e. in 59% of the cases, this server was able to propose a model in top 5 having a $C\alpha$ RMSD from the native $\leq 5.0\text{\AA}$ (Figure 5).



CASP organizers assign a unique target id to each protein fielded in the CASP experiment. While validating and comparing performance of *Bhageerath-H* software on 75 CASP10 targets, we have closely analyzed some of the CASP10 target proteins in which *Bhageerath-H* outperformed other three servers under consideration. A brief description of the biological role of the targets T0655, T0672, T0675, T0700, T0716, T0736, T0747, T0755, T0669, T0713, T0686, T0724 is given in Additional File 2.

For targets T0655, T0672, T0675, T0700, T0716, T0736, T0747, T0755 *Bhageerath-H* outperformed BAKER-ROSETTA server. It predicted a structure in top 5 within the defined $C\alpha$ RMSD cutoff ($\leq 5.0\text{\AA}$). In case of Quark, *Bhageerath-H* exceeded in 6 cases T0669, T0672, T0675, T0685, T0716, T0747, while *Bhageerath-H* was successful in 11 cases when compared with MULTICOM-CLUSTER. For targets T0655, T0672, T0675, T0716, T0747, *Bhageerath-H* achieved high prediction accuracy than all the three servers (Figure 6).

A close inspection of the reason for better performance of *Bhageerath-H* revealed that for targets such as



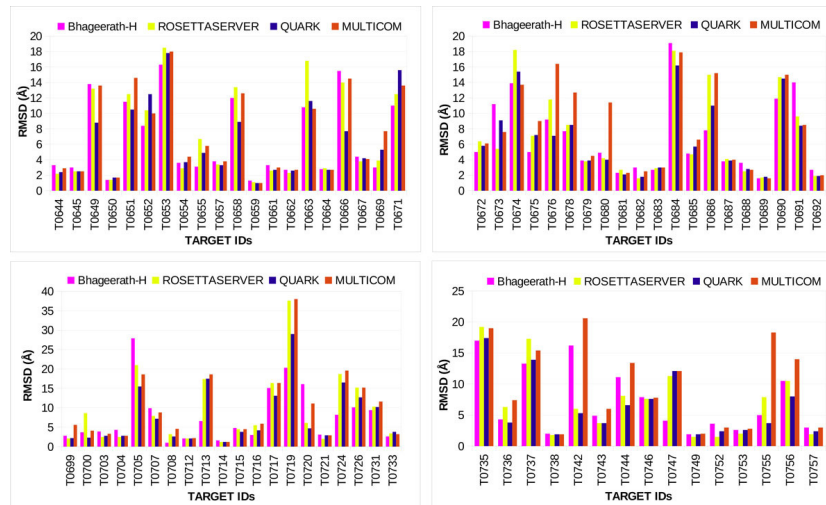


Figure 6 Comparison of *Bhageerath-H* software suite with BAKER-ROSETTA server, Quark server and MULTICOM-CLUSTER server for 75 CASP10 targets.

T0675, T0672, T0669, T0716, T0736, T0700 it was *Bhageerath-H* Strgen patching module as well as *ab initio* loop sampling which generated a low RMSD near-native structure. In systems T0655, T0747, the low RMSD sampled structure is due to the amino acid chemical logic based scoring matrix. The amino acid substitution scoring matrix is a new addition to *Bhageerath-H* Strgen methodology and performs a very thorough search of the database for homologs based on amino acids chemical properties. This matrix helped in template search and alignment generation especially in targets T0655 and T0747, where most other servers failed to predict a low RMSD structure. It identified correct templates and generated better target-template alignments, which resulted in high quality near-native structural models for proteins with low sequence similarity. In cases where a full length template is unavailable, the matrix helped in generating high quality alignments for short sequence fragments. Other than amino acid chemical logic based scoring matrix the major contributor for better performance of *Bhageerath-H* software is *abinitio* loop sampling. Loops are the most flexible parts of a protein structure involved in molecular recognition. Correct modeling of loops has always been a challenge. *Ab initio* loop sampling module helped in systematic and thorough sampling of the loop conformation space and generated low RMSD models. CASP 10 targets where *Bhageerath-H* outperformed other participating servers were mainly modeled through chemical logic and *ab-initio* loop sampling.

Other than above specified targets, *Bhageerath-H*'s performance is noteworthy for targets T0713, T0686 and T0724 when compared to the other three servers under consideration. Though high quality *Bhageerath-H*

models were not predicted, these targets need special attention and discussion. These three targets are described below as case studies for illustration of *Bhageerath-H* performance.

(i) **Target T0713:** This target is a hypothetical protein from *Eubacterium ventriosum* having PDB: 4H09 and 739 amino acid residues. It has four leucine rich repeats domains which take solenoid shape in protein structure. These domains help protein to interact with its complementary protein partner. *Bhageerath-H* sampled a lowest RMSD structure of 8.91Å in pool of trial structures. After clustering and pcSM decoy selection the lowest RMSD model in top 10 was 9.80Å. The topmost PROTSAV selected model was given to QM based structure refinement. QM refined the input model and generated a decoy in the small pool having 6.61Å C α RMSD from the native. It is due to the bond angle optimization which assisted in a better conformational sampling and a lower RMSD decoy, which was picked by pcSM during final five ranking. *Bhageerath-H* successfully modeled and picked a structure in the top five having leucine repeat domain similar to the native structure. The domain form horseshoe shape reflects its biological activity.

(ii) **Target T0686:** This target is a sporozite surface protein of *plasmodium vivax*, one of the causative agents for malarial disease. It is also called TRAP (thrombospondin repeat anonymous protein) which mediates the invasion of mosquitoes and vertebrates host cells in malaria. TRAP protein has two functional domains (i) TSP (thrombospondin type I) and (ii) VWA (von willebrand factor type A) that are responsible for cell adhesion. *Bhageerath-H* Strgen generated a 7.41Å RMSD structure which was retained post clustering.

pcSM and PROTSAV picked an 8.13Å structure which was submitted for QM based refinement. The final lowest RMSD model in top 5 is 7.75Å, which is a much better prediction in comparison to other server predictions. Model structure closely superimposes with VWA domain of native crystal structure (PDB: 4QHO) protein while there are a few anomalies in TSP domain. VWA domain is mainly responsible for protein's biological activity and covers a stretch of ~180 amino acids. TSP is a shorter domain (~40 amino acids). The final ranked *Bhageerath-H* modelled structure missed an extended β-sheet, which resulted in a high RMSD of the prediction from the native.

(iii) **Target T0724:** This target is a hypothetical uncharacterized protein from *bacteroides vulgates* having PDB: 4FMR. It has only one characterized functional domain i.e DNA binding. QM based structure refinement assisted in better conformational sampling and in generating a near-native decoy. A brief biological description of the studied targets is given in the Additional File 2.

In a nut shell, major reasons behind the ability of *Bhageerath-H* to predict lower RMSD near-native models are firstly exhaustive sampling technique. *Bhageerath-H* Strgen and the newly developed amino acid chemical logic based scoring matrix help in a thorough search of template and protein conformational space, ensuring generation of near-native models in maximum instances. Secondly, it is the pcSM scoring function which cherry picks these native-like candidates with 93% accuracy. Apart from these two major modules, it is the PROTSAV structure analysis which ranks models accordingly and submits for QM refinement. Finally, QM based refinement protocol facilitates in going one step ahead and improves prediction accuracy.

(B) Assessment of individual modules of *Bhageerath-H* pipeline

To comprehend the potential of individual modules of *Bhageerath-H* automated pipeline, we further analyzed 7 targets where *Bhageerath-H* outperformed all the three

servers. Table 1 details the output of individual modules of *Bhageerath-H* i.e *Bhageerath-H* Strgen, clustering, pcSM scoring, PROTSAV ranking and final output. Table 1 column 3 contains the result of module 1, *Bhageerath-H* Strgen. It shows the lowest Cα RMSD sampled in the decoy pool. Column 4 shows the size of the decoy pool. Column 5 has Cα RMSD result for module 2, clustering. It contains information of the lowest RMSD structure in the decoy pool after clustering. Column 6 represents the size of the decoy pool post clustering. Column 7 contains the result for module 3, pcSM scoring. It shows the lowest Cα RMSD among the top 10 pcSM ranked decoys. Column 8 has results of module 4, PROTSAV ranking. The Cα RMSD of top-most PROTSAV ranked model. The last column has the final prediction results of *Bhageerath-H* pipeline, the lowest Cα RMSD among final five *Bhageerath-H* predictions for the given target.

As discussed earlier the backbone of any protein tertiary structure prediction software/tool is its protein conformational sampling module. Unless a near-native decoy is sampled/generated, it is impossible to attain high prediction accuracy. In Table 1 for all the 7 targets near-native decoys (Cα RMSD ≤5.0Å) were present in *Bhageerath-H* Strgen sampled decoy pool. These decoys were retained post K-mean clustering. While filtering bad decoys from good ones, it is extremely important to retain the sampled near-native decoys in the smaller basket. As can be seen from Table 1 clustering was able to reduce the basket size while retaining good structures. Second major module of prediction pipeline is scoring. pcSM scoring function has successfully picked the best decoys in top10 except in the case of T0655. PROTSAV has further assisted in ranking the best model (lowest Cα RMSD) as topmost model in 5 cases. In 2 cases we missed out the lowest RMSD sampled decoy in final ranking but successfully selected a ≤5 Å in final predicted output. The last column shows the final prediction results of *Bhageerath-H* pipeline. Figure 7(a1-a7) shows superimposition of lowest

Table 1 Assessment of individual modules of *Bhageerath-H* pipeline for 7 CASP10 targets.

S. No	Target	Lowest Cα RMSD structure in <i>Bhageerath-H</i> Strgen sampled decoy pool	Number of decoys generated	Lowest Cα RMSD structure post K-mean clustering	Number of filtered decoy	Lowest Cα RMSD structure among top10 pcSM ranked decoys	Cα RMSD of the topmost SAVPRO ranked model	Lowest Cα RMSD among final five <i>Bhageerath-H</i> predictions
1.	T0655	2.98	77826	3.13	19742	3.13	3.13	3.13
2	T0669	2.85	20002	2.85	7634	2.85	2.85	2.85
3	T0675	4.90	108131	4.95	31104	5.05	5.1	4.93
4	T0716	2.97	109056	2.97	11708	2.97	3.06	2.97
5	T0733	2.54	105124	2.54	14460	2.54	2.65	2.59
6	T0747	4.11	60584	4.11	17874	4.11	4.11	4.11
7	T0672	5.00	116786	5.00	26806	5.00	5.09	5.00

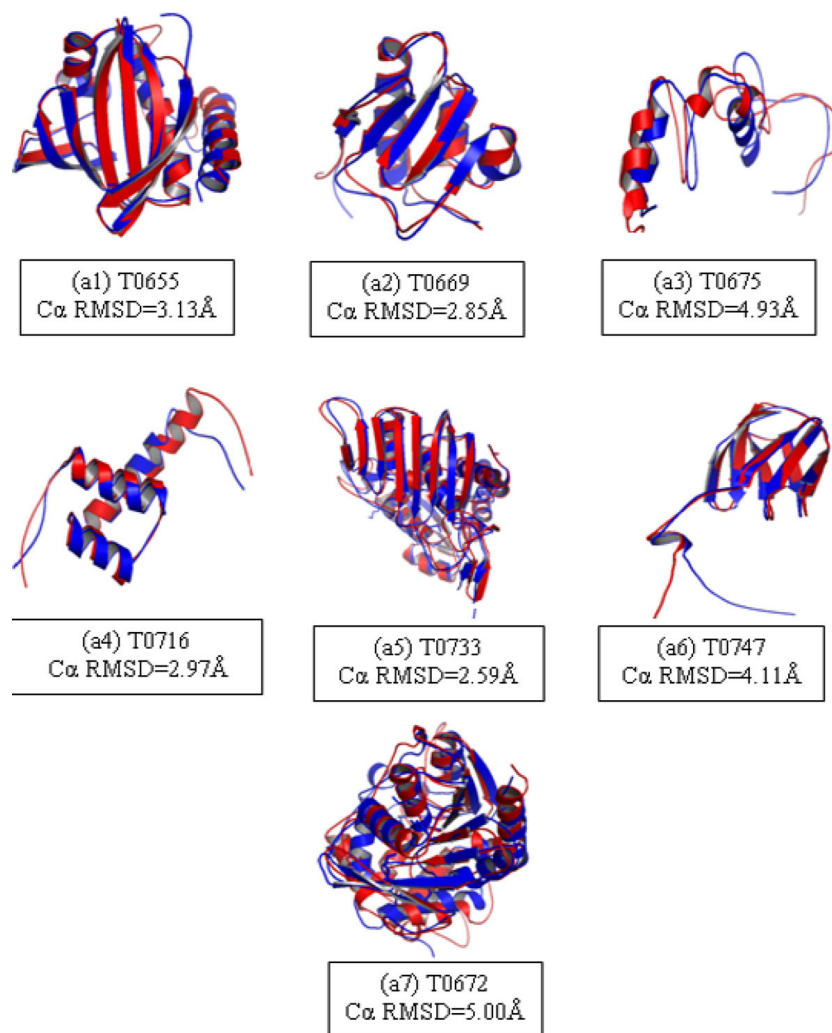


Figure 7 (a1-a7): A superimposition of best *Bhageerath-H* predicted model with native for 7 targets. *Bhageerath-H* predicted model is in red and native is in blue.

C α RMSD *Bhageerath-H* predicted models with the corresponding natives.

(D) Quality assessment of *Bhageerath-H* predictions

Finally, the quality of *Bhageerath-H* predictions was assessed based on Molprobity score [98]. Molprobity score evaluates the stereochemistry of input structure. Online Molprobity server <http://molprobity.biochem.duke.edu> was used for score calculation. Additional File 3 shows the Molprobity score of the best *Bhageerath-H* predictions. Best refers to the lowest C α RMSD in the final five *Bhageerath-H* predictions. The average Molprobity score is 1.94 for 75 predictions.

Bhageerath-H web server

Bhageerath-H automated pipeline is available for the scientific community as a freely accessible web server at url http://www.scfbio-iitd.res.in/bhageerath/bhageerath_h.jsp.

The web server takes as input amino acid sequence of the query protein. The processed results are sent to the users at the email id provided by them. Each submitted job is provided with a unique Jobid, which can be used to check job status. The server provides an option for specifying templates. A user can either opt for automatic template searching option or user defined template option. In automatic template searching option software itself searches for the best templates and uses hybrid approach to predict tertiary structure. In user defined template option, user is required to input template information i.e. template's pdb-id and chain id. Structures based on the defined templates will be given to the user as output. Complete *Bhageerath-H* run takes approximately 5-6 hours depending on the size of the protein. The software runs on a 35 node Quad-Core AMD Opteron(tm) Processor 2380 based cluster on CentOS platform over an Infini-band QDR backbone.

Bhageerath-H receives at least 10-20 jobs every day from all across the world. *Bhageerath*-H is participating in CASP11 competition (1st May 2014 - 16th July 2014). Figure 8 show a screenshot of *Bhageerath*-H webserver.

Conclusions

We have developed *Bhageerath*-H, an automated pipeline for protein tertiary structure prediction and made it into a freely accessible web server http://www.scfbio-iitd.res.in/bhageerath/bhageerath_h.jsp. The pipeline comprise six different modules which are *Bhageerath*-H Strgen for decoy generation, K-mean clustering, pcSM for decoy selection, PROTSAV for structure validation, QM (PM6) based loop refinement and final ranking. Together each module assists in pushing the prediction accuracy to higher limits. *Bhageerath*-H server was validated on 75 CASP10 targets and results show that the methodology is effective in predicting good structures for proteins with varying sequence and structural similarities. Comparison with some of the existing softwares demonstrated the uniqueness of the hybrid methodology in effectively sampling conformational space, scoring best decoys and refining low resolution models to high and medium resolution. A critical analysis of the targets

where *Bhageerath*-H was unsuccessful in predicting low RMSD structures highlights the areas of improvement. These include better secondary structure prediction, better alignment strategies, improvement in *ab initio* modeling for sampling new folds and refinement strategies. We are currently working on these areas especially for targets with very low sequence similarity. The current version of *Bhageerath*-H has already taken the structure prediction field beyond CASP10. This improved methodology is fielded in the ongoing CASP11 experiment.

Several proteins exhibit partial or complete instability in their structures. These proteins are classified as intrinsically disordered proteins (IDPs). *Bhageerath*-H is a homology and *abinitio* hybrid method for modeling structures of monomeric proteins. The current web-enabled version of the protocol is not specifically programmed to model structures of IDPs. Rather, the *ab initio* loop modeling section of the first module as well as QM(PM6) method for loop bond angle refinement attempt to sample conformation space of long loop stretches/disordered regions.

Thus to summarize, in the recent years, data driven homology based computational methods have proved successful in predicting tertiary structures for sequences

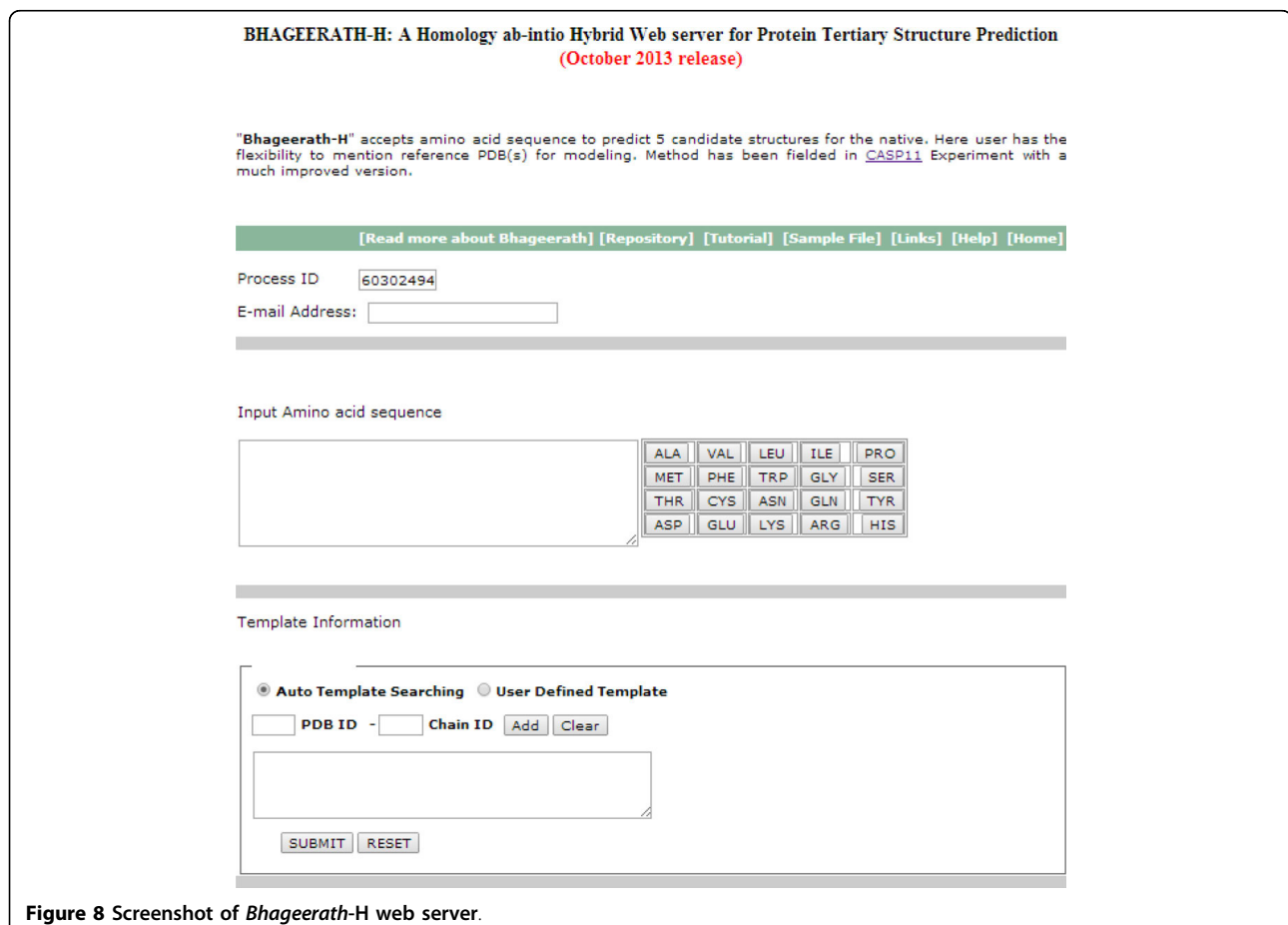


Figure 8 Screenshot of *Bhageerath*-H web server.

with high sequence similarity. With the dwindling similarities of query sequences, advanced homology/ *ab initio* hybrid approaches are being explored to solve structure prediction problem. Overcoming these limitations while pushing the frontiers of protein structure prediction, we have proposed *Bhageerath-H* algorithm. The proposed algorithm finds applications in the field of protein structure/function prediction, active-site directed drug design, in studying protein-protein interactions, and in protein design and engineering. In the absence of experimental protein structure, the availability of computational protein tertiary structural models helps to probe biological functions of proteins.

Additional material

Additional File 1: C α RMSD and TM-Score of best *Bhageerath-H* prediction. Best refers to lowest C α RMSD predicted model in final five *Bhageerath-H* predictions.

Additional File 2: Description of biological and structural relevance of CASP10 Targets (T0655, T0672, T0675, T0700, T0716, T0736, T0747, T0755, T0669, T0713, T0686, T0724).

Additional File 3: Molprobrity score of the best *Bhageerath-H* prediction.

List of abbreviations

CASP: Critical Assessment of Protein Tertiary Structure Prediction.
RMSD: Root mean square deviation.
C α RMSD: C-alpha root mean square deviation
TM-Score: Template modeling score.
PSPA: Protein structure prediction algorithms.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

PD developed *Bhageerath-H* Strgen, AM developed pcSM, RK and AS developed PROTSAV, PD and GM developed quantum mechanics (PM6) based loop bond angle refinement. BJ supervised the above projects. PD, AM and RK analyzed and validated the server. BJ and PD wrote the manuscript. PD and SS web-enabled the software.

Acknowledgements

Programme support to the Supercomputing Facility for Bioinformatics & Computational Biology (SCFBio), IIT Delhi from the Department of Biotechnology Govt. of India and Indian Council of Medical Research is gratefully acknowledged. RK is a recipient of Senior Research Fellowship from Council of Scientific and Industrial Research (CSIR), India

Declarations

The publication charges of this article were funded by Professional Development Fund, IIT Delhi, India.
This article has been published as part of *BMC Bioinformatics* Volume 15 Supplement 16, 2014: Thirteenth International Conference on Bioinformatics (InCoB2014): Bioinformatics. The full contents of the supplement are available online at <http://www.biomedcentral.com/bmcbioinformatics/supplements/15/S16>.

Authors' details

¹Department of Chemistry, Indian Institute of Technology, Hauz Khas, New Delhi-110016, India. ²Supercomputing Facility for Bioinformatics & Computational Biology, Indian Institute of Technology, Hauz Khas, New Delhi-110016, India. ³Kusuma School of Biological Sciences, Indian Institute of Technology, Hauz Khas, New Delhi-110016, India.

Published: 8 December 2014

References

1. Anfinsen CB: Principles that govern the folding of protein chains. *Science* 1973, **181**:223-230.
2. Guo JT, Ellrott K, Xu Y: A historical perspective of template-based protein structure prediction. *Methods Mol Biol* 2008, **413**:3-42.
3. Levinthal C: Are there pathways for protein folding? *Journal de Chimie Physique et de Physico-Chimie Biologique* 1968, **65**:44-45.
4. Srinivasan R, Rose GD: A physical basis for protein secondary structure. *Proc Natl Acad Sci USA* 1999, **96**:14258-14263.
5. Street AG, Mayo SL: Intrinsic β -sheet propensities result from van der Waals interactions between side chains and the local backbone. *Proc Natl Acad Sci USA* 1999, **96**:9074-9076.
6. Honig B: Protein folding: From the Levinthal paradox to structure prediction. *J Mol Biol* 1999, **293**:283-293.
7. Chikenji G, Fujitsuka Y, Takada S: Shaping up the protein folding funnel by local interaction: Lesson from a structure prediction study. *Proc Natl Acad Sci USA* 2006, **103**:3141-3146.
8. Liwo A, Czaplowski C, Oldziej S, Scheraga HA: Computational techniques for efficient conformational sampling of proteins. *Curr Opin Struct Biol* 2008, **18**:134-139.
9. Baldwin RL, Rose GD: Is protein folding hierarchic? I. Local structure and peptide folding. *Trends Biochem Sci* 1999, **24**:26-33.
10. Scheraga HA, Lee J, Pillardy J, Ye YJ, Liwo A, Ripoll D: Surmounting the Multiple-Minima Problem in Protein Folding. *Journal of Global Optimization* 1999, **15**:235-260.
11. Wales DJ, Scheraga HA: Global Optimization of Clusters, Crystals, and Biomolecules. *Science* 1999, **285**:1368-1372.
12. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E: Equations of State Calculations by Fast Computing Machines. *Journal of Chemical Physics* 1953, **21**:1087-1092.
13. Da Silva RA, Degreve L, Caliri A: LMProt: An Efficient Algorithm for Monte Carlo Sampling of Protein Conformational Space. *Biophys J* 2004, **87**:1567-1577.
14. Tang K, Zhang J, Liang J: Fast Protein Loop Sampling and Structure Prediction Using Distance-Guided Sequential Chain-Growth Monte Carlo Method. *PLoS Comput Biol* 2014, **10**:e1003539.
15. Zhang J, Lin M, Chen R, Liang J, Liu JS: Monte Carlo sampling of near-native structures of proteins with applications. *Proteins* 2007, **66**:61-68.
16. Lee J, Scheraga HA, Rackovsky S: New optimization method for conformational energy calculations on polypeptides: Conformational space annealing. *J Comput Chem* 1997, **18**:1222-1232.
17. Cavas LS, Evanseck JD, Karplus M: Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. *Protein Sci* 1998, **7**:649-666.
18. Abrams CF, Vanden-Eijnden E: Large-scale conformational sampling of proteins using temperature-accelerated molecular dynamics. *Proc Natl Acad Sci USA* 2010, **107**:4961-4966.
19. Chou KC, Caracci L: Simulated annealing approach to the study of protein structures. *Protein Eng* 1991, **4**:661-667.
20. Kannan S, Zacharias M: Simulated annealing coupled replica exchange molecular dynamics—an efficient conformational sampling method. *J Struct Biol* 2009, **166**:288-294.
21. Hansmann UHE: Parallel Tempering Algorithm for Conformational Studies of Biological Molecules. *Chem Phys Lett* 1997, **281**:140.
22. Zhou R: Methods Replica exchange molecular dynamics method for protein folding simulation. *Mol Bio* 2007, **350**:205-223.
23. Zhang W, Chen J: Efficiency of adaptive temperature-based replica exchange for sampling large-scale protein conformational transitions. *J Chem Theory Comp* 2013, **9**:2849-2856.
24. Zhou H T, Skolnick : Ab initio protein structure prediction using chunk-TASSER. *J Biophys J* 2007, **93**:1510-1518.
25. 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.
26. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N: ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* 2010, **38**:W529-W533.
27. Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R: A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res* 2010, **38**(suppl 2):W695-704.

28. Jayaram B, Dhingra Priyanka: **Towards creating complete proteomic structural databases of whole organisms.** *Current Bioinformatics* 2012, **7**:424-435.
29. Kopp J, Schwede T: **Automated protein structure homology modeling: a progress report.** *Pharmacogenomics* 2004, **5**:405-416.
30. Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, Schwede T: **Protein structure homology modelling using SWISS-MODEL Workspace.** *Nature Protocols* 2009, **4**:1-13.
31. Norel P, Petrey D, Honig B: **PUDGE: a flexible, interactive server for protein structure prediction.** *Nucleic Acid Research* 2010, **38**:W550-554.
32. Rost B, Schneider R, Sander C: **Protein Fold Recognition by Prediction-based threading.** *J Mol Biol* 1997, **270**:471-80.
33. Godzik A: **Fold recognition methods.** *Methods Biochem Anal* 2003, **44**:525-46.
34. Taylor RWilliam, Jonassen Inge: **A structural pattern-based method for protein fold recognition.** *Proteins Structure Function and Bioinformatics* 2004, **56**:222-34.
35. Jinbo X, Feng J, Libo Y: **Protein structure prediction using threading.** *In Protein Structure prediction Methods in Molecular biology* 2008, **413**:91-121.
36. Richard B, David B: **Ab initio protein STRUCTURE PREDICTION: Progress and Prospects.** *Annual Review of Biophysics and Biomolecular Structure* 2001, **30**:173-189.
37. Themis L, Martin K: **Effective energy functions for protein structure prediction.** *Current Opinion in Structural Biology* 2000, **10**:139-145.
38. Fan H, Periole X, Mark AE: **Mimicking the action of folding chaperones by Hamiltonian replica-exchange molecular dynamics simulations: Application in the refinement of de novo models.** *Proteins* 2012, **80**:1744-1754.
39. Lin MS, Gordon TH: **Reliable protein structure refinement using a physical function.** *J Comp Chem* 2011, **32**:709-717.
40. Zhu J, Fan H, Periole X, Honig B, Mark AE: **Refining homology models by combining replica-exchange molecular dynamics and statistical potentials.** *Proteins* 2008, **72**:1171-1188.
41. Margelevicius M, Venclovas C: **Re-searcher: a system for recurrent detection of homologous protein sequences.** *BMC Bioinformatics* 2010, **11**:89-102.
42. Wernisch L, Hunting M, Wodak SJ: **Identification of Structural Domains in Proteins by a Graph Heuristic.** *Proteins Struct Funct Genet* 1999, **35**:338-352.
43. Liu T, Guerquin M, Samudrala R: **Improving the accuracy of template-based predictions by mixing and matching between initial models.** *BMC Structural Biology* 2008, **8**:24.
44. Rykunov D, Fiser A: **New statistical potential for quality assessment of protein models and a survey of energy functions.** *BMC Bioinformatics* 2010, **11**:28.
45. Fang Q, Shortle D: **Protein refolding in silico with atom-based statistical potentials and conformational search using a simple genetic algorithm.** *J Mol Biol* 2006, **359**:1456.
46. McConkey BJ, Sobolev V, Edelman M: **Discrimination of native protein structures using atom-atom contact scoring.** *Proc Natl Acad Sci USA* 2003, **100**:3215-3220.
47. Benkert P, Kunzli M, Schwede T: **QMEAN server for protein model quality estimation.** *Nucleic Acids Research* 2009, **37**:W510-W514.
48. Benkert P, Tosatto SC, Schomburg D: **QMEAN: A comprehensive scoring function for model quality assessment.** *Proteins* 2008, **71**:261-277.
49. McConkey BJ, Sobolev V, Edelman M: **Discrimination of native protein structures using atom-atom contact scoring.** *Proc Natl Acad Sci USA* 2003, **100**:3215.
50. Zhang J, Chen R, Liang J: **Empirical potential function for simplified protein models: Combining contact and local sequence-structure descriptors.** *Proteins: Structure Function and Bioinformatics* 2006, **63**:949-960.
51. Lu M, Dousis AD, Ma J: **OPUS-PSP: An Orientation-dependent Statistical All-atom Potential Derived from Side-chain Packing.** *Journal of Molecular Biology* 2008, **376**:288-301.
52. Rykunov D, Fiser A: **Effects of amino acid composition, finite size of proteins, and sparse statistics on distance-dependent statistical pair potentials.** *Proteins Structure, Function and Bioinformatics* 2007, **67**:559-568.
53. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K: **Scalable molecular dynamics with NAMD.** *Journal of Computational Chemistry* 2005, **26**:1781-1802.
54. Fang Q, Shortle D: **A consistent set of statistical potentials for quantifying local side-chain and backbone interactions.** *Proteins* 2005, **60**:90.
55. Shen MY, Sali A: **Statistical potential for assessment and prediction of protein structures.** *Protein Sci* 2006, **15**:2507-2524.
56. Sippl MJ: **Recognition of errors in three-dimensional structures of proteins.** *Proteins* 1993, **17**:355.
57. Mishra A, Rao S, Mittal A, Jayaram B: **Capturing Native/Native like Structures with a Physico-Chemical Metric (pcSM) in Protein Folding.** *BBA - Proteins and Proteomics* 2013, **1834**:1520-31.
58. Mishra A, Rana PS, Mittal A, Jayaram B: **D2N: Distance to native.** *BBA - Proteins and Proteomics* 2014, **1844**:1798-1807.
59. Morea V, Tramontano A: **Assessment of homology-based predictions in CASP5.** *Proteins Structure Function and Bioinformatics* 2003, **53**:352-368.
60. Tress M, Tai CH, Wang G, Ezkurdia I, López G, Valencia A, Lee B, Dunbrack RL: **Domain definition and target classification for CASP6.** *Proteins Structure Function and Bioinformatics* 2005, **61**:8-18.
61. Moulton J, Fidelis K, Krysztafowicz A, Schwede T, Tramontano A: **Critical assessment of methods of protein structure prediction (CASP) – round x.** *Proteins Structure Function and Bioinformatics* 2013, **82**:1-6.
62. Zhang Y: **Progress and challenges in protein structure prediction.** *Curr Opin Struct Biol* 2008, **18**:342-348.
63. Dhingra P, Jayaram B: **A homology/ab initio hybrid algorithm for sampling near-native protein conformations.** *J Comput Chem* 2013, **34**:1925-1936.
64. Jayaram B, Bhushan K, Shenoy RS, Narang P, Bose S, Agarwal P, Sahu D, Pandey V: **Bhageerath: An Energy Based Web Enabled Computer Software Suite for Limiting the Search Space of Tertiary Structures of Small Globular Proteins.** *Nucl Acids Res* 2006, **34**:6195-6204.
65. 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.
66. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-402.
67. Lobley A, Sadowski MI, Jones DT: **pGenTHREADER and pDomTHREADER: new methods for improved protein fold recognition and superfamily discrimination.** *Bioinformatics* 2009, **25**:1761-1767.
68. McGuffin LJ, Jones DT: **Improvement of the GenTHREADER method for genomic fold recognition.** *Bioinformatics* 2003, **19**:874-881.
69. Jaroszewski L, Rychlewski L, Li Z, Li W, Godzik A: **FFAS03: a server for profile-profile sequence alignments.** *Nucleic Acids Res* 2005, **33**:W284-288.
70. Jaroszewski L, Li Z, Cai XH, Weber C, Godzik A: **FFAS server: novel features and applications.** *Nucleic Acids Res* 2011, **39**:W38-W44.
71. Yang Y, Faraggi E, Zhao H, Zhou Y: **Improving protein fold recognition and template-based modeling by employing probabilistic-based matching between predicted one-dimensional structural properties of query and corresponding native properties of templates.** *Bioinformatics* 2011, **27**:2076-2082.
72. Söding J: **Protein homology detection by HMM-HMM comparison.** *Bioinformatics* 2005, **21**:951-960.
73. Jayaram B: **Decoding the design principles of amino acids and the chemical logic of protein sequences.** *Nature Precedings* 2008 [<http://precedings.nature.com/documents/2135/version/1>].
74. Sali A, Blundell TL: **Comparative protein modelling by satisfaction of spatial restraints.** *J Mol Biol* 1993, **234**:779-815.
75. Sali A, Potterton L, Yuan F, Vlijmen HV, Karplus M: **Evaluation of comparative protein modeling by MODELLER.** *Proteins* 1995, **23**:318-326.
76. Narang P, Bhushan K, Bose S, Jayaram B: **A computational pathway for bracketing native-like structures for small alpha helical globular proteins.** *Phys Chem Chem Phys* 2005, **7**:2364-2375.
77. Narang P, Bhushan K, Bose S, Jayaram B: **Protein structure evaluation using an all-atom energy based empirical scoring function.** *J Biomol Str Dyn* 2006, **23**:385-406.
78. Finn RD, Tate J, Mistry J, Coghill PC, Sammut SJ, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer EL, Bateman A: **The Pfam protein families database.** *Nucleic Acid Res* 2010, **38**:D211-222.
79. Murzin A, Brenner SE, Hubbard T, Chothia C: **SCOP: a structural classification of proteins database.** *Journal of Molecular Biology* 1995, **247**:536-540.
80. Holm L, Sander C: **Protein folds and families: sequence and structure alignments.** *Nucleic Acids Res* 1999, **27**:244-247.
81. Pruitt KD, Tatusova T, Maglott DR: **NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins.** *Nucleic Acids Res* 2007, **35**:D61-65.

82. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE: **The Protein Data Bank**. *Nucleic Acids Research* 2000, **28**:235-242.
83. Feig M, Karanicolas J, Brooks CL 3rd: **MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology**. *J Mol Graph Model* 2004, **22**(5):377-395.
84. Laskowski RA, MacArthur MW, Moss DS, Thornton JM: **PROCHECK - a program to check the stereo chemical quality of protein structures**. *J App Cryst* 1993, **26**:283-291.
85. Luthy R, Bowie JU, Eisenberg D: **Assessment of protein models with three-dimensional profiles**. *Nature* 1992, **356**:83-85.
86. Colovos C, Yeates TO: **Verification of protein structures: Patterns of nonbonded atomic interactions**. *Protein Science Cambridge University Press* 1993, **2**:1511-1519.
87. Lee B, Richards FM: **The interpretation of protein structures: estimation of static accessibility**. *J Mol Biol* 1971, **55**:379-400.
88. Wiederstein M, Sippl MJ: **ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins**. *Nucleic Acids Research* 2007, **35**:W407-W410.
89. Yang Y, Zhou Y: **Ab initio folding of terminal segments with secondary structures reveals the fine difference between two closely related all-atom statistical energy functions**. *Protein Science* 2008, **17**:1212-1219.
90. Krivov G, Shapovalov MV, Dunbrack RL: **Improved prediction of protein side-chain conformations with SCWRL4**. *Proteins* 2009, **77**:778-795.
91. Case DA, Darden TA, Simmerling CL, Wang J, Duke RE, Luo R, Crowley M, Walker RC, Zhang W, Merz KM, Wang B, Hayik S, Roitberg A, Seabra G, Kolossváry I, Wong KF, Paesani F, Vanicek J, Wu X, Brozell SR, Steinbrecher T, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Mathews DH, Seetin MG, Sagui C, Babin V, Kollman PA: **Amber 10**. University of California, San Francisco; 2008.
92. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG: **Clustal W and Clustal x version 2.0**. *Bioinformatics* 2007, **23**:2947-2948.
93. Xu J, Zhang Y: **How significant is a protein structure similarity with TM-score = 0.5?** *Bioinformatics* 2010, **26**:889-895.
94. Zhang Y, Skolnick J: **TM-align: A protein structure alignment algorithm based on the TM-score**. *Nucleic Acids Res* 2005, **33**:2303-2309.
95. Rohl CA, Strauss CE, Misura KM, Baker D: **Protein structure prediction using Rosetta**. *Methods Enzymol* 2004, **383**:66-93.
96. Xu D, Zhang Y: **Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field**. *Proteins* 2012, **80**:1715-1735.
97. Wang Z, Eickholt J, Cheng J: **MULTICOM: a multi-level combination approach to protein structure prediction and its assessments in CASP8**. *Bioinformatics* 2010, **26**:882-888.
98. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC: **MolProbity: all-atom structure validation for macromolecular crystallography**. *Acta Crystallogr D Biol Crystallogr* 2010, **66**:12-21.

doi:10.1186/1471-2105-15-S16-S7

Cite this article as: Jayaram et al.: *Bhageerath-H: A homology/ab initio hybrid server for predicting tertiary structures of monomeric soluble proteins*. *BMC Bioinformatics* 2014 **15**(Suppl 16):S7.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

