Published online 7 November 2006 *Nucleic Acids Research, 2006, Vol. 34, No. 21* ***6195–6204***
doi:10.1093/nar/gkl789

Bhageerath: an energy based web enabled computer software suite for limiting the search space of tertiary structures of small globular proteins

B. Jayaram*, Kumkum Bhushan, Sandhya R. Shenoy, Pooja Narang, Surojit Bose, Praveen Agrawal, Debashish Sahu and Vidhu Pandey

Department of Chemistry and Supercomputing Facility for Bioinformatics and Computational Biology, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110 016, India

Received April 5, 2006; Revised September 27, 2006; Accepted October 2, 2006

ABSTRACT

We describe here an energy based computer software suite for narrowing down the search space of tertiary structures of small globular proteins. The protocol comprises eight different computational modules that form an automated pipeline. It combines physics based potentials with biophysical filters to arrive at 10 plausible candidate structures starting from sequence and secondary structure information. The methodology has been validated here on 50 small globular proteins consisting of 2–3 helices and strands with known tertiary structures. For each of these proteins, a structure within 3–6 Å RMSD (root mean square deviation) of the native has been obtained in the 10 lowest energy structures. The protocol has been web enabled and is accessible at http://www.scfbio-iitd.res.in/bhageerath.

INTRODUCTION

The tertiary structure prediction of a protein using amino acid sequence information alone is one of the fundamental unsolved problems in computational biology/molecular biophysics (1). The folding of protein molecules with a large number of degrees of freedom spontaneously into a unique three-dimensional (3-D) structure is of scientific interest intrinsically and due to its application in structure based drug design endeavors. The cost and time factors involved in experimental techniques urge for an early *in silico* solution to protein folding problem (2). The ultimate goal is to use computer algorithms to identify amino acid sequences that not only adopt particular 3-D structures but also perform specific functions i.e. to propose designer proteins (3).

Contemporary approaches for protein structure prediction can be broadly classified under two categories viz. (i) comparative modeling, which includes homology modeling and threading (4–7) and (ii) *de novo* folding (8–12). The first category of methods utilizes the structures of already solved proteins as templates (either locally or globally, at the sequence level or at the sub-structure level). With large amounts of genome and proteome data accumulating via sequencing projects, comparative modeling has become the method of choice to characterize sequences where related representatives of a family exist in structural databases (13–18). There are several web servers based on comparative modeling approaches such as Swiss Model (4), CPHmodels (19), FAMS (20) and ModWeb (21). The assessors for comparative modeling at CASP6 (Critical Assessment of protein Structure Prediction methods) have noted small improvements in model quality despite increase in the available structures but marginal improvement in alignment accuracy when compared to CASP5 (22). A natural limit for these approaches is the quantity of information available in the structural databases. This highlights the importance of *de novo* techniques for protein folding.

Significant progress has been made in recent years towards physics-based computation of protein structure, from a knowledge of the amino acid sequence. This approach, commonly referred to as an *ab initio* method (23–25) is based on the *thermodynamic hypothesis* formulated by Anfinsen (1973), according to which the native structure of a protein corresponds to the global minimum of its free energy under given conditions (26). Protein structure prediction using *ab initio* method is accomplished by a search for a conformation corresponding to the global-minimum of an appropriate potential energy function without the use of secondary structure prediction, homology modeling, threading etc. (27). In contrast, methods characterized as *de novo* use the *ab initio* strategies partly as well as database information directly or indirectly. Table 1 summarizes different known web servers/groups for protein structure prediction and the function(s) therein. The tertiary structure prediction of protein starting from its sequence has been successfully demonstrated on protein sequences <85 residues in length by Baker's group (28,29) using a fragment assembly methodology. The ProtInfo

*To whom correspondence should be addressed. Tel: +91 11 2659 1505; Fax: +91 11 2658 2037; Email: bjayaram@chemistry.iitd.ac.in

Table 1. Some *de novo*/*ab initio* servers for protein folding

Sl. No.	Name of the Web Server/Group	Description
1.	ROBETTA (28,29) (http://robetta.bakerlab.org)	*De novo* Automated structure prediction analysis tool used to infer protein structural information from protein sequence data
2.	PROTINFO (30) (http://protinfo.compbio.washington.edu)	*De novo* protein structure prediction web server utilizing simulated annealing for generation and different scoring functions for selection of final five conformers
3.	SCRATCH (31) (http://www.igb.uci.edu/servers/psss.html)	Protein structure and structural features prediction server which utilizes recursive neural networks, evolutionary information, fragment libraries and energy
4.	ASTRO-FOLD (32)	Astro-fold: first principles tertiary structure prediction based on overall deterministic framework coupled with mixed integer optimization
5.	ROKKY (33) (http://www.proteinsilico.org/rokky/rokky-p/)	*De novo* structure prediction by the simfold energy function with the multi-canonical ensemble fragment assembly
6.	BHAGEERATH (http://www.scfbio-iitd.res.in/bhageerath)	Energy based methodology for narrowing down the search space of small globular proteins

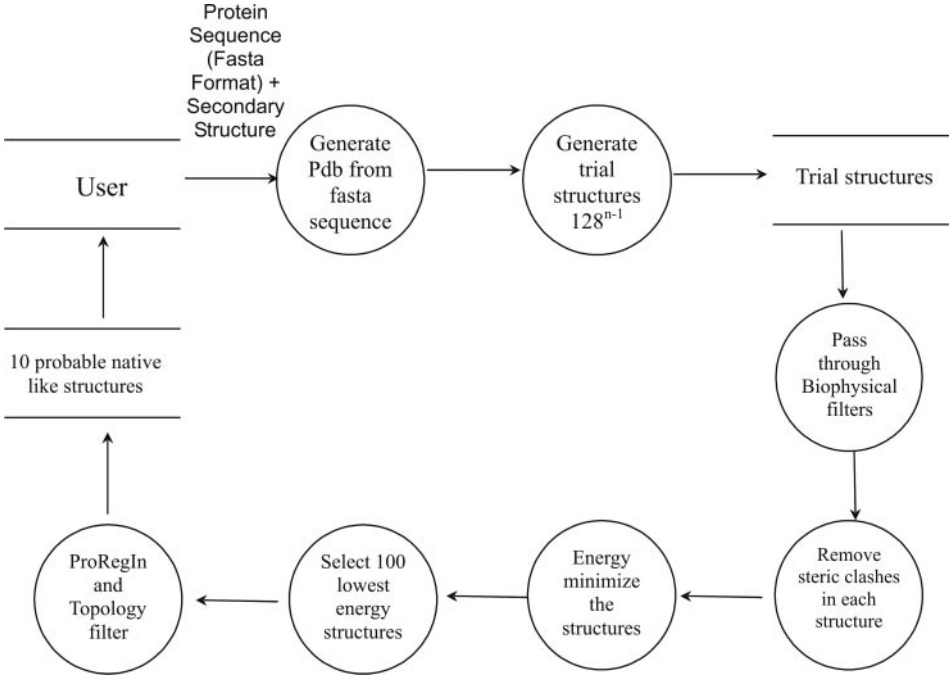

Figure 1. The flow of information in *Bhageerath* web server, starting with the input from the user to the final 10 predictions made available to the user.

web server by Samudrala *et al*. (30) predicts protein tertiary structure for sequences <100 amino acids using *de novo* methodology, where by structures are generated using simulated annealing search phase which minimizes a target scoring function. Scratch web server by Baldi *et al*. (31) predicts the protein tertiary structure as well as structural features starting from the sequence information alone. Astro-fold (32) an *ab initio* structure prediction framework by Klepeis and Floudas employs local interactions and hydrophobicity for the identification of helices and beta-sheets respectively followed by global optimization, stochastic optimization and torsion angle dynamics. *De novo* structure prediction by simfold energy function with the multi-canonical ensemble fragment assembly has been developed by Fujitsuka *et al*. (33). The function has been tested on 38 proteins along with the fragment assembly simulations and predicts structures within 6.5 Å RMSD (root mean square deviation) of the native in 12 of the cases. Arriving at structures between 3 and 6 Å RMSD of the native expeditiously using *ab initio* or *de novo* methodologies remains a formidable challenge.

We have developed a computationally viable *de novo* strategy for tertiary structure prediction, processing and evaluation. The web server christened *Bhageerath* takes as input the amino acid sequence and secondary structure information for a query protein and returns 10 candidate structures for the native. In this article, we report the validation and testing of the protein structure prediction web suite *Bhageerath* with application to 50 small globular proteins. The programs are written in standard C++, with a total of more than ~8000 lines of code and are easily portable on any POSIX (UNIX, LINUX, IRIX and AIX) compliant system.

MATERIALS AND METHODS

Bhageerath (www.scfbio-iitd.res.in/bhageerath) software suite for protein tertiary structure prediction narrows down the search space to generate probable candidate structures for the native. The flow chart diagram of *Bhageerath* is depicted in Figure 1.

The first module involves the formation of a 3-D structure from the amino acid sequence with the secondary structural elements in place. The second module involves generation of a large number of trial structures with a systematic sampling of the conformational space of loop dihedrals. The number of trial structures generated is $128^{(n-1)}$ where n is the number of secondary structural elements. These structures are generated by choosing seven dihedrals from each of the loops (three at both ends and one dihedral from the middle of the loop) and sampling two conformations for each dihedral. The values assigned for dihedrals Φ, Ψ to each amino acid during structure generation are given in supplementary information (Supplementary Table S1). The trial structures generated via dihedral sampling are screened in the third module through persistence length and radius of gyration filters (34), developed for the purpose of reducing the number of improbable candidates. The resultant structures are refined in the fourth module by a Monte Carlo sampling in dihedral space to remove steric clashes and overlaps involving atoms of main chain and side chains. In module five, the structures are energy minimized to further optimize the side chains. The energy minimization is carried out in vacuum with distance dependent dielectric for 200 steps (75 steps steepest descent + 125 steps conjugate gradient). Module six involves ranking of structures using an all atom energy based empirical scoring function (35) followed by selection of the 100 lowest energy structures. Module seven reduces the probable candidates based on the protein regularity index of the Φ and Ψ dihedral values based on the threshold value of 1.5 for Φ and 4.0 for Ψ (Thukral *et al*., manuscript accepted in *J. Biosci*.). Module eight further reduces the structures selected in the previous module to 10 using topological equivalence criterion and the accessible surface area [calculated using NACCESS (36)]. The above eight modules are configured to work in a conduit.

Overview of the organization of the suite

Bhageerath is a fully automated web enabled protein structure prediction software suite that is made available through a convenient user interface which returns 10 predictions for a given protein query sequence. A click on the *Bhageerath* server opens into a window wherein a user can paste a query protein sequence in FASTA format. The current version supports continuous sequences up to 100 amino acids. The user is prompted for amino acid range as secondary structural input. Upon submission the user receives an unique job id for his/her sequence. User has the option to provide an email ID to receive an output link which contains 10 lowest energy candidate structures.

RESULTS

We present here a performance appraisal of the protein tertiary structure prediction software suite on 50 globular proteins with known structures. All the proteins have been extracted from the Protein Data Bank (PDB) (37) and are functionally diverse. We have extracted ~8000 unique proteins from the PDB at 50% sequence similarity or less. From these, ~8000 unique proteins, we obtained 329 proteins satisfying the criterion that the number of residues is <100 and the number of secondary structural elements varies between two and three. We have selected our test set of 50 proteins randomly from these 329 proteins. The length of the polypeptide chain varies from 17 to 70 and the total number of helices and strands ranges between two and three.

The results obtained for the 50 globular proteins with the web server are shown in Table 2. The table gives the PDB ID, the number of amino acids in the sequence as well as the number and type of secondary structural elements present in each protein in columns (i)–(iii). The number of structures obtained after the persistence length and radius of gyration filters are given in column (iv) of Table 2. The lowest RMSD obtained in the 100 structures along with its energy rank are provided in the next two columns, (v) and (vi). This is followed by the number of structures selected by ProRegIn filter in column (vii). The number in parenthesis in column (vii) indicates the number of structures with RMSD < 6 Å in the selected structures. The lowest RMSD and the corresponding energy rank after selection with ProRegIn filter are reported in column (viii) and (ix). The structures selected after the Topology filter are reported in column (x) and the number in parenthesis indicates the number of structures with RMSD <6 Å in the final 10 structures. The last two columns of Table 2 [column (xi) and (xii)] show the lowest RMSD with respect to the native obtained from amongst the 10 predicted structures along with the energy rank of the structure. For all the 50 test proteins, irrespective of the nature of secondary structural elements and the length of intervening loops, it may be noted that a few topologically correct structures within an RMSD of 3–6 Å from the native structure are obtained in the final 10 predicted structures. Thus, the 'needle in a haystack' problem can be reduced to finding a solution in the best 10 structures at least for small proteins.

Figure 2 shows a superimposition of the lowest RMSD structure with the respective native structures for all the 50 globular test proteins.

A comparison of the structures obtained with the protein structure prediction web server presented here was carried out with six freely available homology modeling servers: CPHmodels (19), Swiss Model (4), EsyPred3D (38), ModWeb (21), Geno3D (39) and 3Djigsaw (40). While SwissModel, EsyPred3D, Geno3D and 3Djigsaw provide an option for template selection the other two servers are automatic. For the 50 test proteins validated, we have first carried out sequence alignment using PSI BLAST (41) and the templates were selected such that the sequence similarity of the template is >30% and the template is not from the same family. For most of the proteins there was very less sequence similarity with proteins of other families and the templates were restricted to the same family. In such cases the quality of model built is quite high and the RMSD with respect to the native is <1 Å in few cases. The proteins where the templates are selected from different families result in RMSDs comparable to those obtained with *Bhageerath* web server. Table 3 shows the RMSD of the structures obtained by homology modeling from the respective web servers for all the 50 globular proteins. The template ID, percentage sequence similarity and alignment of the target-template sequence for each method and each structure therein is provided in

Table 2. A performance appraisal of *Bhageerath* web server for 50 small globular proteins

Sl. No.	PDB ID (i)	Number of amino acids (ii)	Number of secondary structure elements (iii)	Number of structures accepted after Persistence length and Radius of gyration filters (iv)	Lowest RMSD in the final 100 structures (Å) (v)	Energy Rank of the lowest RMSD structure in 100 structures (vi)	After ProRegIn filter			After topology and accessible surface area filter		
							Number of structures selected (Number of structures <6 Å) (vii)	Lowest RMSD (Å) (viii)	Energy Rank of the lowest RMSD structure in 100 structures (ix)	Number of structures selected (Number of structures <6 Å) (x)	Lowest RMSD (Å) (xi)	Energy Rank of the lowest RMSD structure in 10 structures (xii)
1	1E0Q	17	2E	128	2.5	2	100 (29)	2.5	2	10 (10)	2.5	2
2	1B03	18	2E	64	4.4	2	64 (5)	4.4	2	10 (5)	4.4	2
3	1WQC	26	2H	128	2.5	6	100 (53)	2.5	6	10 (10)	2.5	3
4	1RJU	36	2H	64	4.6	48	64 (3)	4.6	48	10 (2)	5.9	6
5	1EDM	39	2E	128	2.9	100	100 (59)	2.9	100	10 (10)	3.5	2
6	1AB1	46	2H	128	2.4	10	100 (82)	2.4	10	10 (10)	2.9	6
7	1BX7	51	2E	128	2.2	71	100 (85)	2.2	71	10 (10)	3.1	8
8	1B6Q	56	2H	128	3.1	27	100 (8)	3.1	27	10 (5)	3.1	10
9	1ROP	56	2H	128	4.3	2	100 (6)	4.3	2	10 (2)	4.3	2
10	1NKD	59	2H	128	3.8	8	100 (4)	3.8	8	10 (4)	3.8	6
11	1RPO	61	2H	128	3.8	2	100 (6)	3.8	2	10 (4)	3.8	2
12	1QR8	68	2H	128	4.4	80	100 (3)	4.4	80	10 (2)	4.4	10
13	1FME	28	1H,2E	15 592	2.9	52	100 (90)	2.9	52	10 (8)	3.7	5
14	1ACW	29	1H,2E	15 726	3.9	97	100 (45)	3.9	97	10 (5)	5.1	8
15	1DFN	30	3E	13 174	4.4	77	98 (11)	4.4	77	10 (4)	5.0	1
16	1Q2K	31	1H,2E	16 020	4.2	46	100 (20)	4.2	46	10 (4)	4.2	9
17	1SCY	31	1H,2E	15 423	3.1	10	100 (40)	3.1	10	10 (4)	3.1	5
18	1XRX	34	1E,2H	14 630	3.9	28	100 (19)	3.9	28	10 (1)	5.6	1
19	1ROO	35	3H	1071	2.5	14	100 (100)	2.5	14	10 (10)	2.8	5
20	1YRF	35	3H	15 180	3.8	16	100 (62)	3.8	16	10 (9)	4.8	4
21	1YRI	35	3H	15 180	2.8	81	100 (70)	2.8	81	10 (8)	3.8	6
22	1VII	36	3H	16 380	3.7	7	100 (50)	3.7	7	10 (6)	3.7	2
23	1BGK	37	3H	14 139	3.8	33	100 (56)	3.8	33	10 (8)	4.1	3
24	1BHI	38	1H,2E	14 923	5.3	2	100 (5)	5.3	2	10 (2)	5.3	2
25	1OVX	38	1H,2E	12 074	3.2	8	100 (76)	3.2	8	10 (5)	4.0	1
26	1I6C	39	3E	2927	4.1	31	100 (32)	4.1	31	10 (3)	5.1	2
27	2ERL	40	3H	16 268	3.1	18	100 (32)	3.1	18	10 (2)	3.2	6
28	1RES	43	3H	16 135	4.0	30	100 (40)	4.0	30	10 (7)	4.2	2
29	2CPG	43	1E,2H	10 905	3.6	20	100 (18)	3.6	20	10 (1)	5.3	2
30	1DV0	45	3H	14 488	4.0	20	100 (21)	4.0	20	10 (1)	5.1	4
31	1IRQ	48	1E,2H	11 592	3.5	74	100 (18)	3.5	74	10 (1)	5.3	9
32	1GUU	50	3H	13 410	4.5	74	100 (42)	4.5	74	10 (7)	4.6	6
33	1GV5	52	3H	11 109	3.5	33	99 (24)	3.5	33	10 (5)	4.1	2
34	1GVD	52	3H	10 626	3.8	18	100 (35)	3.8	18	10 (6)	4.9	9
35	1MBH	52	3H	10 632	3.8	48	100 (24)	3.8	48	10 (5)	4.0	4
36	1GAB	53	3H	14 495	3.6	16	100 (12)	3.6	16	10 (3)	3.6	6
37	1MOF	53	3H	16 384	2.4	57	100 (96)	2.4	57	10 (10)	2.9	5
38	1ENH	54	3H	13 622	3.2	12	100 (23)	3.2	12	10 (3)	4.6	3
39	1IDY	54	3H	11 133	3.3	84	100 (52)	3.3	84	10 (8)	3.5	6
40	1PRV	56	3H	5468	4.4	55	99 (25)	4.4	55	10 (7)	4.9	9
41	1HDD	57	3H	12 849	3.2	74	100 (22)	3.2	74	10 (2)	4.8	8
42	1BDC	60	3H	11 255	4.2	44	100 (19)	4.2	44	10 (2)	4.8	5
43	1I5X	61	3H	16 384	2.6	29	99 (54)	2.6	29	10 (10)	2.6	6
44	1I5Y	61	3H	16 384	2.6	20	100 (48)	2.6	20	10 (10)	2.6	7
45	1KU3	61	3H	5701	4.9	68	100 (14)	4.9	68	10 (3)	5.5	4
46	1YIB	61	3H	16 384	2.9	7	100 (75)	2.9	7	10 (9)	3.5	5
47	1AHO	64	1H,2E	2429	4.7	58	100 (15)	4.7	58	10 (1)	6.0	6
48	1DF5	68	3H	16 384	3.1	10	100 (41)	3.1	10	10 (6)	3.1	8
49	1QR9	68	3H	16 384	2.9	49	100 (33)	2.9	49	10 (9)	3.8	2
50	1AIL	70	3H	16 384	4.2	42	100 (5)	4.2	42	10 (3)	4.2	7

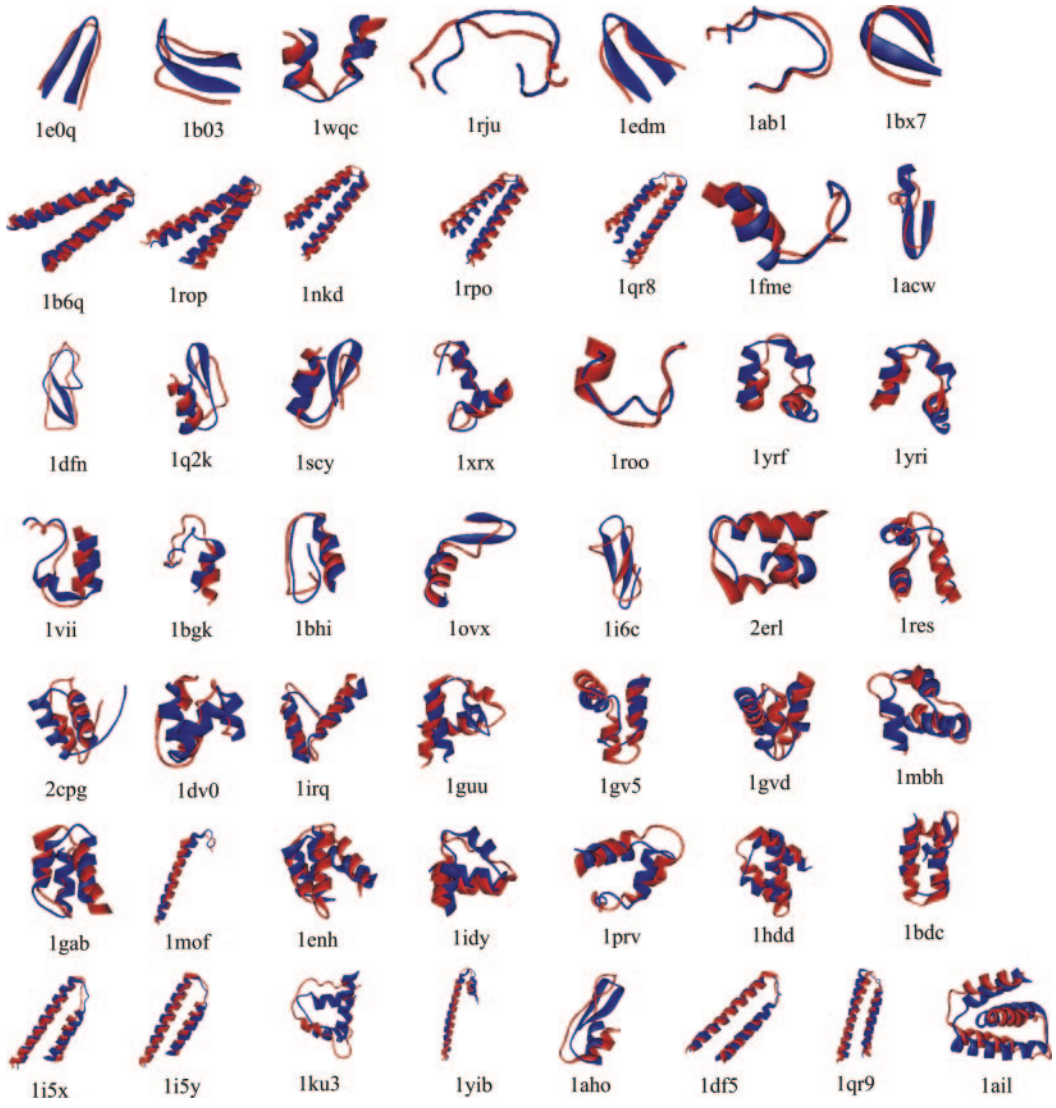

Figure 2. The superimposed lowest RMSD structures for the 50 small globular test proteins used for the validation of *Bhageerath* web server. The PDB ID's are shown underneath each structure. The predicted structure is shown in red color and the native in blue.

supplementary information (Supplementary Tables S2–S7). Thus, for new sequences with no known sequence homologues, the *Bhageerath* web server has the potential to predict a structure to within 3–6 Å RMSD of the native structure with accuracies comparable to the homology modeling servers.

Further comparison of the 10 structures obtained from *Bhageerath* was carried out with the five candidate structures obtained from the ProtInfo web server (30) and 10 structures obtained with ROBETTA software (28) configured locally. The results shown in Table 4 indicate that the server described here is able to predict structures with RMSDs comparable to those obtained by ProtInfo web server and ROBETTA software. Supplementary Table S8 in the supplementary information provides the comparison of the GDT_TS scores obtained using LGA server (42) for structures obtained with *Bhageerath* and ProtInfo web servers and ROBETTA software. The GDT_TS scores are also found to be comparable for structures obtained from these three different structure prediction methodologies.

DISCUSSION

We describe here an energy based computational web server *Bhageerath*, for an automated candidate tertiary structure prediction. The web server permits predictive folding with moderate computational resources. The validation of the computational protocol on 50 globular proteins has shown that the web server selects one or more candidate structures within an RMSD of 3–6 Å with respect to the native in the 10 lowest energy structures. The results presented are for proteins having 2–3 secondary elements with α, β and α/β structures and are obtained solely from the amino acid sequence and secondary structure information (without the aid of multiple sequence alignment, or fold recognition). The results provide a benchmark as to the level of model accuracy one can expect from this web server.

All of the eight modules are currently being executed on a cluster with 32 dedicated UltraSparc III 900 MHz processors. In contrast to typical short return times (ranging from 1 to 10 min) for receiving results from comparative modeling

Table 3. A comparison of protein tertiary structure prediction accuracies with different homology modeling servers available in public domain

Sl. No.	PDB ID	CPHModels (19) RMSD (Å)	SwissModel (4) RMSD (Å)	EsyPred3D (38) RMSD (Å)	ModWeb (21) RMSD (Å)	Geno3D (39) RMSD (Å)	3DJigSaw (40) RMSD (Å)	*Bhageerath* RMSD (Å)
1	1E0Q(1–17)	—	—	1.7 (1–17)	—	—	1.5 (1–16)	2.5
2	1B03(1–18)	—	—	3.5 (2–18)	—	—	—	4.4
3	1WQC(1–26)	0.5 (1–26)	0.4 (1–26)	—	—	—	—	2.5
4	1RJU(1–36)	2.0 (1–36)	1.7 (1–36)	2.1 (1–36)	—	—	—	5.9
5	1EDM(1–39)	1.5 (1–39)	1.4 (1–39)	0.8 (2–38)	0.5 (1–39)	—	1.8 (1–39)	3.5
6	1AB1(1–46)	0.6 (1–46)	2.8 (1–46)	0.4 (1–46)	0.4 (1–46)	0.7 (1–46) 0.7 (1–46)	—	2.9
7	1BX7(1–51)	0.6 (1–51)	0.8 (1–51)	2.2 (3–50)	0.6 (1–51)	2.6 (4–51)	2.2 (3–50)	3.1
8	1B6Q(1–56)	4.7 (1–56)	5.0 (1–56)	2.7 (3–56)	5.1 (1–56)	0.7 (1–56) 1.4 (1–56)	4.8 (1–56)	3.1
9	1ROP(1–56)	1.3 (1–56)	0.6 (1–56)	4.7 (3–56)	0.7 (1–56)	0.7 (1–56) 0.8 (1–56)	1.3 (1–56)	4.3
10	1NKD(1–59)	0.5 (1–59)	7.7 (1–50)	0.6 (1–59)	1.9 (1–59) 0.7 (1–59)	1.3 (1–59) 1.2 (1–59)	0.4 (1–59)	3.8
11	1RPO(1–61)	0.5 (1–61)	7.7 (1–50)	0.5 (1–59)	1.9 (1–61) 0.7 (1–59)	0.8 (1–61) 0.9 (1–61)	0.4 (1–61)	3.8
12	1QR8(1–68)	0.5 (1–68)	0.5 (1–68)	1.1 (2–66)	1.6 (1–68)	0.9 (1–68) 1.2 (1–68)	0.5 (1–68)	4.4
13	1FME(1–28)	0.7 (1–28)	0.9 (1–28)	—	—	—	—	3.7
14	1ACW(1–29)	0.7 (1–29)	0.4 (1–29)	—	—	—	—	5.1
15	1DFN(1–30)	0.8 (1–30)	0.4 (1–30)	1.3 (2–30)	—	—	—	5.0
16	1Q2K(1–31)	0.9 (1–31)	0.5 (1–31)	—	—	—	—	4.2
17	1SCY(1–31)	0.6 (1–31)	0.7 (1–31)	—	—	—	—	3.1
18	1XRX(1–34)	0.5 (1–34)	0.3 (1–34)	—	0.7 (1–34)	—	3.1 (1–31)	5.6
19	1ROO(1–35)	0.8 (1–35)	0.7 (1–35)	—	—	—	—	2.8
20	1YRF(1–35)	1.6 (1–35)	0.5 (1–35)	1.2 (1–35)	1.3 (1–35)	—	—	4.8
21	1YRI(1–35)	1.7 (1–35)	0.7 (1–35)	1.4 (1–35)	1.5 (1–35)	—	—	3.8
22	1VII(1–36)	2.4 (2–36)	0.9 (1–36)	2.2 (2–36)	2.0 (2–36)	—	—	3.7
23	1BGK(1–37)	0.8 (1–37)	0.5 (1–37)	–	0.7 (1–37)	—	—	4.1
24	1BHI(1–38)	0.8 (1–38)	0.4 (1–38)	1.0 (1–38)	1.1 (1–38)	—	—	5.3
25	1OVX(1–38)	0.9 (1–38)	0.3 (1–38)	1.0 (1–38)	0.6 (1–38)	—	0.3 (1–38)	4.0
26	1I6C(1–39)	4.2 (1–39)	4.4 (1–39)	4.5 (1–39)	0.8 (1–39)	—	3.1 (1–34)	5.1
27	2ERL(1–40)	1.3 (1–40)	0.9 (1–40)	0.4 (1–40)	0.4 (1–40)	1.2 (1–40) 1.1 (1–40)	—	3.2
28	1RES(1–43)	4.2 (1–43)	4.1 (1–43)	4.2 (1–43)	0.8 (1–43)	1.2 (1–43) 1.1 (1–43)	—	4.2
29	2CPG(1–43)	0.8 (1–43)	0.6 (1–43)	1.1 (1–43)	0.9 (1–43)	0.9 (1–43)	0.6 (1–43)	5.3
30	1DV0(1–45)	4.2 (1–45)	10.5 (1–35)	2.0 (1–42)	0.7 (1–45) 2.4 (1–44)	0.9 (1–45) 1.1 (1–45)	0.6 (1–45)	5.1

31	1IRQ(1–48)	0.6 (1–48)	0.8 (1–48)	1.3 (2–48)	0.7 (1–48)	1.2 (1–48)	0.9 (1–48)		5.3
32	1GUU(1–50)	2.5 (1–50)	2.6 (1–50)	2.3 (38–50)	5.7 (1–50) 0.5 (1–50) 2.2 (1–49)	1.5 (1–48) 1.6 (1–48)	1.6 (1–42)		4.6
33	1GV5(1–52)	1.4 (1–52)	0.6 (1–52)	1.3 (1–52)	0.68 (1–52) 2.2 (3–52)	2.1 (3–46) 2.0 (3–46)	1.8 (3–45)		4.1
34	1GVD(1–52)	1.4 (1–52)	4.2 (1–51)	1.3 (1–52)	5.5 (1–52)	6.6 (1–44) 9.8 (1–44)	6.4 (1–43)		4.9
35	1MBH(1–52)	1.8 (1–52)	3.3 (1–51)	1.8 (1–52)	1.9 (1–52)	1.6 (1–52) 2.1 (1–52)	1.1 (6–45)		4.0
36	1GAB(1–53)	0.6 (1–53)	1.6 (1–53)	3.3 (1–53)	3.3 (1–53)	2.2 (1–53) 2.7 (1–53)	0.5 (1–53)		3.6
37	1MOF(1–53)	0.6 (1–53)	1.8 (1–53)	1.9 (1–53)	1.7 (1–53) 2.3 (3–51)	3.4 (1–53) 3.4 (1–53)	1.7 (1–53)		2.9
38	1ENH(1–54)	0.5 (1–54)	0.8 (1–54)	0.9 (3–53)	1.7 (1–53) 1.0 (5–53) 2.5 (1–54) 3.0 (1–51)	1.7 (1–54) 1.7 (1–54)	0.5 (1–54)		4.6
39	1IDY(1–54)	4.0 (2–54)	10.8 (1–50)	3.8 (2–52)	0.9 (1–54)	10.0 (5–46) 10.0 (5–46)		0.3 (1–54)	3.5
40	1PRV(1–56)	5.7 (2–56)	2.1 (1–56)	5.6 (3–56)	1.6 (1–56)	5.7 (2–56) 5.4 (2–56)		5.6 (2–56)	4.9
41	1HDD(1–57)	13.2 (1–57)	13.3 (1–57)	1.2 (1–56)	2.7 (1–57) 3.3 (1–51) 3.6 (1–57) 1.5 (9–55) 1.3 (1–56)	2.1 (1–56) 2.7 (1–56)		0.3 (1–57)	4.8
42	1BDC(1–60)	3.4 (1–60)	2.7 (6–39) 3.1 (1–38)	2.7 (6–39)	2.7(5–59)	2.1 (1–60) 1.8 (1–60)		2.6 (5–37)	4.8
43	1I5X(1–61)	0.7 (1–61)	1.1 (1–61)	1.6 (1–61)	1.8 (1–61)	1.5 (1–61) 1.4 (1–61)		0.9 (1–61)	2.6
44	1I5Y(1–61)	0.7 (1–61)	1.1 (1–61)	1.6 (1–61)	1.4 (1–61)	1.7 (1–61) 1.1 (1–61)		0.9 (1–61)	2.6
45	1KU3(1–61)	1.3 (1–61)	2.8 (4–61)	1.5 (1–61)	1.5 (1–61)	1.9 (1–61) 1.7 (1–61)		0.4 (1–61)	5.5
46	1YIB(1–61)	1.8 (2–61)	1.7 (1–61)	3.4 (1–61)	2.8 (2–60)	1.5 (1–61) 1.6 (1–61)		1.9 (2–61)	3.5
47	1AHO(1–64)	0.6 (1–64)	0.5 (1–64)	1.3 (1–64)	0.4 (1–64)	1.8 (1–64)		0.3 (1–64)	6.0
48	1DF5(1–68)	0.6 (1–68)	1.5 (1–68)	1.8 (2–66)	2.2 (1–68)	1.8 (1–68) 1.8 (1–68)		1.6 (1–68)	3.1
49	1QR9(1–68)	0.5 (1–68)	0.7 (1–68)	1.4 (2–66)	1.8 (1–68)	1.7 (1–68) 1.8 (1–68)		0.6 (1–68)	3.8
50	1AIL(1–70)	0.87 (1–70)	0.73 (1–70)	0.46 (1–70)	0.6 (1–70)	0.88 (1–70) 0.97 (1–70)		0.9 (1–70)	4.2

The numbers in parenthesis indicate the length of the protein model obtained. Supplementary Tables S2–S7 in the supplementary information contain the template ID, % sequence identity and alignment for each method and structure shown above.

Table 4. A comparison of protein tertiary structure prediction accuracy with ProtInfo web server and ROBETTA software available in the public domain for 50 test proteins

Sl. No.	PDB ID	RMSD without end loops (Å) (*Bhageerath*)	RMSD without end loops (Å) (ProtInfo)[a] (30)	RMSD without end loops (Å) (ROBETTA)[a] (28)
1	1E0Q	4.5, 2.5, 3.0, 5.0, 3.4, 3.3, 3.2, 3.3, 5.9, 3.3	4.0, 4.1, 3.7, 3.9, 4.2	1.1[b]
2	1B03	10.3, 4.4, 5.9, 5.5, 6.7, 5.4, 4.5, 6.1, 6.9, 7.5	4.0, 4.7, 4.1, 4.5, 4.4	2.7, 3.0
3	1WQC	4.0, 4.5, 2.5, 3.8, 2.9, 5.1, 4.2, 5.7, 3.8, 4.7	2.1, 1.8, 1.8, 2.0, 2.1	2.3, 3.4
4	1RJU	6.1, 6.3, 6.6, 5.9, 6.6, 5.9, 6.6, 7.0, 6.7, 7.4	3.4, 4.9, 3.3, 4.8, 6.0	3.4, 4.0, 2.5, 3.2, 3.0, 3.6, 4.8, 2.9, 3.0, 3.1
5	1EDM	3.9, 3.5, 3.8, 4.0, 3.6, 5.2, 5.4, 4.1, 3.9, 4.7	3.4, 4.0, 3.7, 3.3, 3.1	0.4, 0.5, 0.4, 0.5, 0.6, 0.4, 0.7, 0.7, 1.1, 0.4
6	1AB1	4.8, 4.5, 4.3, 5.2, 4.2, 2.9, 4.5, 3.8, 5.8, 3.3	3.3, 5.1, 6.3, 3.6, 4.9	2.2, 2.8, 2.9, 2.4, 2.9, 2.7, 3.7, 3.5, 2.2, 3.3
7	1BX7	3.3, 4.0, 5.0, 3.2, 4.5, 3.8, 4.8, 3.1, 4.0, 3.5	2.6, 4.2, 3.7, 4.5, 2.1	0.9, 1.5, 1.0, 1.6, 1.5, 1.6, 1.4, 1.0, 2.0, 1.5
8	1B6Q	6.1, 8.4, 4.0, 4.4, 3.8, 10.1, 5.3, 9.7, 10.7, 3.1	10.2, 10.0, 10.0, 10.4, 10.5	10.0, 9.6, 8.5, 7.6, 12.0, 8.3, 8.2, 7.0, 10.2, 9.0
9	1ROP	5.3, 4.3, 9.2, 7.3, 7.5, 11.0, 14.2, 11.5, 8.7, 6.2	10.8, 11.5, 11.5, 10.1, 12.4	5.8, 10.3, 10.0, 11.7, 8.6, 7.0, 8.3, 7.7, 11.2, 13.6
10	1NKD	3.9, 16.2, 10.1, 7.0, 10.6, 3.8, 4.8, 4.9, 7.9, 14.7	13.5, 13.5, 13.3, 13.4, 11.7	8.9, 8.9, 10.6, 11.0, 12.6, 10.7, 12.2, 10.1, 11.0, 9.1
11	1RPO	9.9, 3.8, 4.0, 7.5, 14.4, 4.8, 6.0, 13.5, 3.8, 7.5	10.8, 10.4, 10.4, 10.9, 11.2	10.3, 8.7, 6.9, 6.0, 12.4, 7.7, 10.1, 7.2, 10.0, 7.7
12	1QR8	9.0, 11.1, 8.2, 7.1, 9.7, 14.0, 8.1, 10.9, 5.4, 4.4	10.1, 9.5, 10.0, 10.4, 12.2	11.3, 9.3, 9.0, 7.6, 9.5, 12.2, 10.5, 7.1, 11.3, 8.5
13	1FME	4.9, 5.0, 4.8, 6.5, 3.7, 4.5, 4.2, 6.2, 4.3, 4.1	2.2, 2.3, 2.5, 2.7, 1.6	3.8, 2.8, 3.3, 4.5, 3.6, 3.1, 2.7, 3.9, 4.4, 3.7
14	1ACW	5.5, 7.0, 5.3, 6.0, 7.4, 5.7, 7.0, 5.1, 7.2, 5.6	5.8, 5.8, 6.0, 6.2, 7.1	1.3, 1.7
15	1DFN	5.0, 5.9, 6.5, 5.8, 6.8, 6.0, 7.1, 6.1, 6.5, 7.4	5.6, 6.8, 6.4, 6.6, 6.4	1.7, 5.3, 6.0, 5.5, 4.0, 6.3, 5.2, 6.5, 5.2, 6.6
16	1Q2K	7.4, 7.4, 7.2, 4.8, 5.8, 6.5, 5.7, 6.2, 4.2, 7.3	5.9, 6.0, 5.8, 6.4, 9.1	1.7, 3.0, 3.3, 1.6, 4.7
17	1SCY	6.1, 4.8, 6.6, 7.2, 3.1, 5.0, 6.5, 6.9, 7.2, 5.6	5.5, 5.6, 6.5, 6.4, 6.2	2.2, 2.7, 3.3
18	1XRX	5.6, 8.8, 7.6, 7.7, 9.6, 8.4, 9.0, 6.2, 8.4, 8.2	8.6, 8.8, 7.8, 8.8, 4.0	5.2, 9.1, 7.1, 6.2, 4.4, 9.4, 6.6, 4.5, 5.4, 8.2
19	1ROO	3.9, 3.4, 3.3, 3.8, 2.8, 4.1, 3.5, 3.2, 3.2, 3.3	2.8, 2.7, 2.7, 3.0, 2.7	1.8, 2.1, 1.9, 2.9, 2.5, 1.2, 2.5, 1.9, 2.8, 2.2
20	1YRF	5.9, 5.7, 5.7, 4.8, 4.9, 5.0, 4.9, 5.0, 6.2, 5.8	4.3, 4.1, 3.3, 3.3, 4.3	1.7, 3.1, 4.3
21	1YRI	5.9, 5.5, 4.6, 6.0, 5.5, 3.8, 5.5, 5.4, 5.5, 6.1	4.2, 4.0, 3.2, 3.2, 4.2	1.7, 3.9, 2.8
22	1VII	5.5, 3.7, 6.6, 5.9, 6.1, 5.7, 5.6, 6.0, 6.3, 5.7	4.4, 4.7, 4.5, 4.3, 3.7	2.4, 3.3, 1.8, 5.6, 4.3, 3.0, 3.2, 3.7, 4.8, 1.9
23	1BGK	5.8, 5.9, 4.1, 6.1, 5.8, 5.5, 5.5, 4.9, 5.2, 6.1	6.2, 6.0, 6.4, 6.4, 6.2	6.5, 4.1, 4.6, 2.5, 3.8, 5.9, 3.5, 3.3, 3.5, 6.1
24	1BHI	7.9, 5.3, 6.7, 7.2, 5.4, 8.9, 6.3, 6.6, 6.2, 7.1	3.7, 3.8, 4.5, 4.5, 5.0	2.4, 2.4, 1.7, 1.1, 2.8, 1.7, 2.6, 2.2, 2.3, 1.9
25	1OVX	4.0, 6.4, 6.3, 4.3, 6.1, 5.4, 5.3, 5.9, 7.7, 6.1	4.6, 4.9, 4.4, 5.6, 5.2	3.2, 1.5, 3.1, 2.6, 4.2, 4.4, 2.3, 2.6, 1.9, 5.0
26	1I6C	7.5, 5.1, 5.4, 6.2, 5.4, 6.2, 8.0, 6.2, 6.7, 7.6	5.6, 5.7, 5.6, 7.3, 6.9	3.0, 3.0, 2.2, 3.2, 2.1
27	2ERL	6.7, 8.6, 7.1, 8.4, 7.2, 3.2, 4.1, 6.2, 6.8, 8.1	7.0, 7.4, 7.1, 7.2, 8.3	1.3, 7.1
28	1RES	6.1, 4.2, 5.2, 7.7, 4.8, 4.8, 4.3, 7.0, 5.6, 5.5	7.6, 7.1, 7.0, 7.3, 5.1	3.5, 3.0, 2.8, 4.3, 4.2, 2.3, 2.0
29	2CPG	10.1, 5.3, 10.0, 8.5, 9.4, 10.6, 7.8, 9.4, 7.4, 7.5	4.2, 4.5, 5.3, 5.1, 11.0	8.0, 4.3, 8.5, 8.4, 6.5, 10.0, 4.8, 8.6, 5.5, 7.6
30	1DV0	7.7, 7.1, 8.0, 5.1, 8.3, 6.0, 7.8, 8.7, 8.4, 8.5	3.2, 4.4, 4.0, 2.8, 6.2	1.6, 1.5, 1.6, 2.0, 1.5, 4.5, 2.4, 2.0, 2.3, 4.2
31	1IRQ	6.8, 6.9, 6.4, 6.7, 10.2, 8.4, 9.8, 9.0, 5.3, 8.2	8.2, 8.9, 9.1, 9.0, 8.5	6.1, 4.3, 6.0, 5.0, 6.6, 6.0, 7.4, 5.2, 6.4, 7.5
32	1GUU	5.5, 5.3, 7.7, 4.6, 5.0, 4.6, 5.1, 5.7, 8.9, 9.1	10.1, 10.1, 9.8, 9.3, 10.1	2.9, 4.2, 2.9, 7.0, 3.2, 3.7, 2.4, 6.5, 5.6
33	1GV5	4.9, 4.1, 4.8, 4.8, 9.0, 9.4, 4.6, 9.2, 9.3, 8.9	9.4, 9.1, 9.5, 8.9, 3.3	8.5, 3.7, 9.1, 4.5, 4.7, 5.3, 4.2, 9.1, 3.1, 3.5
34	1GVD	5.7, 6.4, 8.0, 5.1, 6.0, 4.9, 4.9, 6.9, 4.9, 5.5	9.4, 9.4, 8.8, 9.1, 3.9	8.5, 3.5, 2.7, 3.0, 4.7, 4.4, 4.3, 2.3, 6.7, 8.9
35	1MBH	9.1, 9.2, 9.2, 4.0, 9.5, 8.4, 5.5, 5.5, 5.0, 5.3	4.3, 4.1, 5.7, 3.5, 9.5	8.3, 8.1, 4.2, 2.8, 8.9, 2.4, 7.9, 3.5, 7.7, 7.7
36	1GAB	4.9, 9.2, 6.2, 6.0, 6.8, 3.6, 8.5, 9.7, 8.8, 6.3	5.5, 5.6, 6.4, 5.4, 5.9	2.3, 8.8, 2.7, 7.9, 2.8, 8.1, 2.7, 2.3, 2.2, 7.7
37	1MOF	5.7, 3.7, 3.9, 4.2, 2.9, 4.0, 4.9, 4.3, 4.0, 4.9	12.7, 13.6, 12.5, 12.7, 13.5	13.7, 11.8, 11.2, 12.6, 12.6, 12.0, 12.2, 12.9, 12.8, 11.2
38	1ENH	6.3, 9.9, 4.6, 9.1, 9.7, 5.8, 5.7, 9.5, 6.2, 6.4	5.0, 4.6, 4.3, 8.7, 4.2	2.2, 1.7, 1.8, 5.1, 2.3, 4.6, 3.0, 5.2, 3.1, 3.2
39	1IDY	4.6, 4.9, 8.7, 4.0, 3.6, 3.5, 5.3, 3.7, 6.0, 9.3	8.7, 8.3, 8.3, 8.8, 4.6	2.7, 2.5, 3.0, 8.5, 2.1, 2.0, 2.1, 6.8, 2.6, 2.9
40	1PRV	6.9, 5.1, 6.9, 5.8, 5.0, 5.6, 5.6, 9.5, 4.9, 4.9	2.3, 2.6, 3.0, 3.2, 5.4	2.5, 2.1, 3.4, 2.9, 3.7, 4.9, 2.9, 2.4, 4.2, 6.8
41	1HDD	10.2, 6.3, 10.2, 5.5, 11.1, 6.2, 9.8, 4.8, 7.0, 6.7	4.4, 4.7, 5.8, 4.6, 9.7	2.3, 2.5, 2.2, 3.3, 3.6, 4.4, 3.4, 3.0, 4.2, 4.2
42	1BDC	7.7, 6.1, 6.6, 8.3, 4.8, 7.0, 7.5, 5.0, 6.7, 6.6	3.1, 3.0, 3.5, 2.8, 5.1	2.5, 2.5, 3.7, 3.2, 7.7, 4.0, 3.7, 7.9, 2.6, 7.8
43	1I5X	5.5, 5.9, 3.6, 5.4, 5.8, 2.6, 4.3, 6.0, 3.9, 5.1	11.4, 11.0, 11.0, 11.5, 9.2	10.8, 6.8, 8.6, 12.5, 4.5, 9.8, 7.1, 13.1, 9.0, 7.0
44	1I5Y	5.8, 5.1, 4.3, 4.3, 3.4, 4.9, 2.6, 3.7, 3.2, 4.0	9.8, 8.9, 8.4, 11.8, 9.1	9.6, 7.8, 10.2, 9.1, 8.2, 5.0, 12.5, 11.3, 8.4, 8.1
45	1KU3	6.6, 7.4, 6.4, 5.5, 7.2, 5.6, 6.3, 6.2, 5.6, 8.3	5.6, 5.4, 4.9, 5.4, 9.6	4.7, 4.4, 5.8, 4.5, 5.3, 5.3, 5.5, 6.2, 4.7, 2.9
46	1YIB	6.7, 5.3, 5.5, 5.8, 3.5, 4.8, 5.1, 4.5, 5.2, 4.6	17.5, 17.6, 18.3, 17.3, 17.4	17.8, 17.5, 17.1, 17.1, 17.3, 17.5, 18.5, 16.3
47	1DF5	3.4, 5.3, 6.0, 6.1, 7.0, 3.8, 3.4, 3.1, 8.1, 3.4	9.3, 10.3, 8.7, 9.3, 11.7	9.9, 8.2, 5.7, 5.6, 9.9, 8.5, 8.6, 11.1, 6.3, 7.0
48	1AHO	7.8, 7.6, 9.1, 8.7, 6.6, 6.0, 7.2, 7.7, 9.2, 7.7	8.1, 6.6, 4.1, 5.2, 6.0	0.6, 1.1, 0.6, 1.2, 1.0, 0.4, 0.8, 1.4, 1.2, 0.8
49	1QR9	4.3, 3.8, 4.9, 5.1, 10.9, 6.0, 4.0, 4.0, 4.2, 4.6	11.0, 11.1, 9.6, 11.2, 12.9	6.3, 8.5, 4.3, 9.9, 8.6, 6.5, 8.7, 11.7, 12.1, 10.7
50	1AIL	10.8, 6.6, 4.4, 6.4, 7.2, 8.9, 4.2, 8.5, 6.0, 4.2	9.0, 8.9, 8.4, 7.6, 10.3	3.2, 4.4, 4.5, 5.3, 7.2, 5.4, 6.4

[a]The secondary structure information was utilized from the native structure along with the sequence information for both *Bhageerath* and ROBETTA (Rosetta++ software suite was obtained from UW TechTransfer Digital Ventures). We have generated 10000 decoys starting from sequence and secondary structure information. The top 2000 scoring decoys were selected and top 10 cluster centers were extracted. The ProtInfo (http://protinfo.compbio.washington.edu) predictions were obtained from the sequence information alone.
[b]For the system 1e0q it took ~12 days on a dedicated processor to generate 1000 decoys.

servers, the expected prediction time with *Bhageerath* web server for two helix systems is 4–5 min while for three helix systems it is ~2–3 h. However, this depends on the length of the sequence, number of secondary structure elements and the number of structures accepted after the biophysical filters for processing the energetics of each trial structure at the atomic level. It is currently able to process ~4–5 normally sized jobs per day on 32 processors.

The current version of the web server elicits secondary structure information from the user. For new sequences where secondary structure information is not available, web based secondary structure prediction tools can be employed. We have characterized the results obtained from five different freely available secondary structure prediction servers (43–47) available on the web for the 50 test proteins. The predictions are provided in the supplementary information

Table 5. A list of modules of *Bhageerath* converted to independent web utilities with their respective URL's

Sl. No.	Name of the utility	Description
1	Persistence length filter (http://www.scfbio-iitd.res.in/software/proteomics/perlen.jsp)	A filter based on the maximum uninterrupted length of the polypeptide chain persisting in a particular direction
2	Radius of gyration filter (http://www.scfbio-iitd.res.in/software/proteomics/rg.jsp)	A filter based on the radius of the molecule and defined as the root mean square distance of the collection of atoms from their common centre of gravity
3	Hydrophobicity ratio filter (http://www.scfbio-iitd.res.in/software/proteomics/hyphb.jsp)	A filter based on hydrophobicity ratio, which is defined as the ratio of loss in accessible surface area (ASA) per atom of non-polar atoms to the loss in accessible surface area per atom of the polar atoms
4	Packing fraction filter (http://www.scfbio-iitd.res.in/software/proteomics/pf.jsp)	A filter based on packing density, which utilizes observation that proteins are known to exhibit packing fractions ~0.7
5	Protein structure optimizer (http://scfbio-iitd.res.in/software/proteomics/promin.jsp)	A utility that minimizes the energy of the protein structure using a combination of steepest descent and conjugate gradient minimization algorithms
6	Scoring function for protein structure evaluation (http://www.scfbio-iitd.res.in/utility/proteomics/energy.jsp)	An all-atom empirical energy based scoring function which combines second generation force field parameters with a hydrophobicity function
7	Protein regularity index (http://www.scfbio-iitd.res.in/software/proregin.jsp)	A utility based on the regularity seen in the main chain loop dihedral angles of proteins

(Supplementary Table S9). We envisage the introduction of a secondary structure predictor in module one shortly. For larger systems, i.e. those containing more than 100 amino acid residues and those with more than three secondary structural elements, we conceive the introduction of loop filters to control the combinatorial explosion in the number of trial structures. We have utilized two biophysical filters presently in module three for trial structure selection and plan to utilize a few more such as hydrophobicity and packing fraction at later stages. Also one could profitably employ constraints on strands for sheet formation, constraints on metal ions to cluster residues and disulphide bridges as filters for reducing the number of trial structures. The all atom empirical energy function utilized in module six was tested previously and was seen to separate native from the decoy structures in 67 of the 69 protein sequences from among 61 640 decoys studied (35). The scoring function calculates the non-bonded energy of each trial structure as a sum of the electrostatics, van der Waals and hydrophobicity. There is scope for improvement in the scoring function particularly in describing the hydrophobicity component. Work on the above mentioned lines as also on a Flexible Monte Carlo simulation strategy to bring down the RMSD < 3 Å of the native is in progress.

The individual modules of *Bhageerath* are web enabled for free access. These include the four biophysical filters (persistence length, radius of gyration, hydrophobicity ratio and packing fraction), a protein structure optimizer, an all-atom empirical energy based scoring function and ProRegIn utility. These are listed in Table 5 along with their corresponding URL's.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENTS

Funding from the Department of Biotechnology is gratefully acknowledged. Ms Kumkum Bhushan is a recipient of the Senior Research Fellow award from the Council of Scientific & Industrial Research (CSIR), India. Help received from Ms Lipi Thukral and Mr Shailesh Tripathi is gratefully acknowledged. The Open Access publication charges for this article were waived by Oxford University Press.

Conflict of interest statement. None declared.

REFERENCES

1. Liwo,A., Khalili,M. and Scheraga,H.A. (2005) *Ab initio* simulation of protein-folding pathways by molecular dynamics with united residue model of polypeptide chains. *Proc. Natl Acad. Sci. USA*, **102**, 2362–2367.
2. Baker,D. (2000) A surprising simplicity to protein folding. *Nature*, **405**, 39–42.
3. Klepeis,J.L. and Floudas,C.A. (2004) *In silico* protein design: a combinatorial and global optimization approach. *SIAM News*, **37**, 1.
4. Guex,N. and Peitsch,M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, **18**, 2714–2723.
5. Sánchez,R. and Šali,A. (1997) Evaluation of comparative protein structure modeling by MODELLER-3. *Proteins*, **29**, 50–58.
6. Panchenko,A.R., Marcbr-Bauer,A.E. and Bryant,S.H. (2000) Combination of threading potentials and sequence profiles improves fold recognition. *J. Mol. Biol.*, **296**, 1319–1331.
7. Skolnick,J.E. and Kihara,D. (2001) Defrosting the frozen approximation: PROSPECTOR-a new approach to threading. *Proteins*, **42**, 319–331.
8. Aszodi,A., Gradwell,M.J. and Taylor,W.R. (1995) Global fold determination from a small number of distance restrains. *J. Mol. Biol.*, **251**, 308–326.
9. Kolinski,A., Jaroszewski,L., Rotkiewicz,P. and Skolnick,J. (1998) An efficient Monte Carlo model of protein chains. Modeling the short-range correlations between side group centers of mass. *J. Phys Chem*, **102**, 4628–4637.

10. Ortiz,A.R., Kolinski,A. and Skolnick,J. (1998) Fold assembly of small proteins using Monte Carlo simulations driven by restraints derived from multiple sequence alignments. *J. Mol. Biol.*, **277**, 419–448.
11. Huang,E.S., Samudrala,R. and Ponder,J.W. (1999) *Ab initio* fold prediction of small helical proteins using distance geometry and knowledge-based scoring functions. *J. Mol. Biol.*, **290**, 267–281.
12. Simons,K.T., Strauss,C. and Baker,D. (2001) Prospects for *ab initio* protein structural genomics. *J. Mol. Biol.*, **306**, 1191–1199.
13. Rost,B. and Sander,C. (1996) Bridging the protein sequence-structure gap by structure predictions. *Annu. Rev. Biophys. Biomol. Struct.*, **25**, 113–136.
14. Guex,N., Diemand,A. and Peitsch,M.C. (1999) Protein modeling for all. *Trends Biochem. Sci.*, **24**, 364–367.
15. Moult,J. (1999) Predicting protein three-dimensional structure. *Curr. Opin. Biotechnol.*, **10**, 583–588.
16. Al-Lazikani,B., Jung,J., Xiang,Z. and Honig,B. (2001) Protein structure prediction. *Curr. Opin. Struct. Biol.*, **5**, 51–56.
17. Venclovas,C. (2001) Comparative modeling of CASP4 target proteins: Combining results of sequence search with three-dimensional structure assessment. *Proteins*, **45**, 47–54.
18. Tramontanoa,A. and Morea,V. (2003) Assessment of homology based predictions in CASP5, *Proteins*, **53**, 352–368.
19. Lund,O., Nielsen,M., Lundegaard,C. and Worning,P. (2002) X3M a computer program to extract 3D models. Abstract at the CASP5 conference, A102.
20. Ogata,K. and Umeyama,H. (2000) An automatic homology modeling method consisting of database searches and simulated annealing. *J. Mol. Graph Model*, **18**, 258–272, 305–306.
21. Sali,A. and Blundell,T. (1993) Comparative protein modeling by satisfaction of spatial restraints. *J. Mol. Biol.*, **234**, 779–815.
22. Tress,M., Ezkurdia,I., Graña,O., Lopez,G. and Valencia,A. (2005) Assessment of predictions submitted for the CASP6 comparative modeling category. *Proteins*, **61**, 27–45.
23. Scheraga,H.A. (1992) Some approaches to the multiple-minima problem in the calculation of polypeptide and protein structures. *Int. J. Quantum Chem.*, **42**, 1529–1536.
24. Scheraga,H.A. (1996) Recent developments in the theory of protein folding: searching for the global energy minimum. *Biophys. Chem.*, **59**, 329–339.
25. Vasquez,M., Nemethy,G. and Scheraga,H.A. (1994) Conformational energy calculations on polypeptides and proteins. *Chem. Rev.*, **94**, 2183.
26. Anfinsen,C.B. (1973) Principles that govern the folding of protein chains. *Science*, **181**, 223.
27. Pillardy,J. (2001) Recent improvements in prediction of protein structure by global optimization of a potential energy function. *Proc. Natl Acad. Sci. USA*, **98**, 2329–2333.
28. Kim,D.E., Chivian,D. and Baker,D. (2004) Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.*, **32**, W526–W531.
29. Bradley,P., Misura,K.M.S. and Baker,D. (2005) Towards high-resolution de novo structure prediction for small proteins. *Science*, **309**, 1868–1871.
30. Hung,L.-H., Ngan,S.-C., Liu,T. and Samudrala,R. (2005) PROTINFO: new algorithms for enhanced protein structure predictions. *Nucleic Acids Res.*, **33**, W77–W80.
31. Cheng,J., Randall,A.Z., Sweredoski,M.J. and Baldi,P. (2005) SCRATCH: a protein structure and structural feature prediction server. *Nucleic Acids Res.*, **33**, W72–W76.
32. Klepeis,J.L. and Floudas,C.A. (2003) ASTRO_FOLD: A combinatorial and global optimization framework for *ab initio* prediction of three-dimensional structures of proteins from the amino acid sequence. *Biophys. J.*, **85**, 2119–2146.
33. Fujitsuka,Y., Chikenji,G. and Takada,S. (2005) SimFold energy function for *de novo* protein structure prediction: consensus with Rosetta. *Proteins*, **62**, 381–398.
34. Narang,P., Bhushan,K., Bose,S. and Jayaram,B. (2005) A computational pathway for bracketing native-like structures for small alpha helical globular proteins. *Phys. Chem. Chem. Phys.*, **7**, 2364–2375.
35. Narang,P., Bhushan,K., Bose,S. and Jayaram,B. (2006) Protein structure evaluation using an all-atom energy based empirical scoring function. *J. Biomol. Struct. Dyn.*, **23**, 385–406.
36. Hubbard,S.J. and Thornton,J.M. (1993) *'NACCESS', Computer Program,* Department of Biochemistry and Molecular Biology, University College London, UK.
37. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
38. Lambert,C., Leonard,N., De Bolle,X. and Depiereux,E. (2002) EsyPred3D: Prediction of proteins 3D structures. *Bioinformatics*, **18**, 1250–1256.
39. Combet,C., Jambon,M., Deleage,G. and Geourjon,C. (2002) Geno3D: Automatic comparative molecular modeling of protein. *Bioinformatics*, **18**, 213–214.
40. Bates,P.A., Kelley,L.A., MacCallum,R.M. and Sternberg,M.J.E. (2001) Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. *Proteins*, **45**, 39–46.
41. Altschul,S.F., Madden,T.L., Schäffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
42. Zemla,A. (2003) LGA - a method for finding 3D similarities in protein structures. *Nucleic Acids Res.*, **31**, 3370–3374.
43. Bryson,K., McGuffin,L.J., Marsden,R.L., Ward,J.J., Sodhi,J.S. and Jones,D.T. (2005) Protein structure prediction servers at University College London. *Nucleic Acids Res.*, **33**, W36–W38.
44. Rost,B., Yachdav,G. and Liu,J. (2003) The PredictProtein server. *Nucleic Acids Res.*, **32**, W321–W326.
45. Cuff,J.A., Clamp,M.E., Siddiqui,A.S., Finlay,M. and Barton,G.J. (1998) Jpred: a consensus secondary structure prediction server. *Bioinformatics*, **14**, 892–893.
46. Sen,T.Z., Jernigan,R.L., Garnier,J. and Kloczkowski,A. (2005) GOR V server for protein secondary structure prediction. *Bioinformatics*, **21**, 2787–2788.
47. Frishman,D. and Argos,P. (1996) Incorporation of non-local interactions in protein secondary structure prediction from the amino acid sequence. *Protein Eng.*, **9**, 133–142.