



Online Servers and Offline Tools for Protein Modelling, Optimization and Validation: A Review

Lalit R. Samant*¹, Vikrant C. Sangar², Abhay Chowdhary^{1,2}

¹ Systems Biomedicine Division, Haffkine Institute for Training, Research and Testing, Mumbai, India.

² Department of Virology & Immunology, Haffkine Institute for Training, Research and Testing, Mumbai, India.

*Corresponding author's E-mail: samantlalit@gmail.com

Accepted on: 06-07-2014; Finalized on: 31-08-2014.

ABSTRACT

Protein modelling plays a major role in the drug discovery process. Often lack of proper structure is a major problem to use receptors as targets. The ultimate goal of protein modelling is to predict a structure from its sequence with an accuracy that is comparable to the best results achieved experimentally. This would allow users to safely use rapidly generated *in silico* protein models in all the contexts where today only experimental structures provide a solid basis: structure-based drug design, analysis of protein function, ligand-receptor and protein-protein interactions, antigenic behaviour, and rational design of proteins with increased stability or novel functions. The servers which are freely available for protein modelling and optimization are mentioned in this review briefly which can be accessed and used for academic purpose.

Keywords: Protein modelling, Protein model optimization, Protein model validation, Phyre2, Chiron, UCSF Chimera.

INTRODUCTION

Proteins are large, complex molecules which are required for the structure, function, and regulation of the body's tissues and organs. Proteins are made up of hundreds or thousands of smaller units called amino acids, which attach to one another in long chains. There are 20 different types of amino acids that can be combined to make a protein. The sequence of amino acids determines each protein's unique 3-dimensional structure and its specific functions.¹

Protein-protein interactions play a central role in various aspects of the structural and functional organization of the cell and their elucidation is crucial for a better understanding of processes such as metabolic control, signal transduction and gene regulation. Genome-wide proteomics studies like yeast two-hybrid assays provide an increasing list of interacting proteins but only a small fraction of the potential complexes are amenable to direct experimental analysis. Thus, it is important to develop docking methods that can elucidate the details of specific interactions at the atomic level. The current problem in the developing country like INDIA for research scholars and Ph.D. students in the field of bioinformatics is lack of awareness about available tools, software or servers. In this review, we tried to cover all possible ways of protein modelling, structure optimization and *in silico* model validation by freely available tools, software or servers.

Protein Modelling

Online Servers

Multiple Mapping Method with Multiple Templates (M4T) Server ver. 3.0.

Multiple Mapping Method with Multiple Templates (M4T) is a fully automated comparative protein structure modelling server. The novelty of M4T resides in two of its major modules, Multiple Mapping Method (MMM) and Multiple Templates (MT).² The MT module of M4T selects and optimally combines the sequences of multiple template structures through an iterative clustering approach that takes into account the 'unique' contribution of each template, its sequence similarity to other template sequences and to the target sequences and the quality of its experimental resolution. MMM module is a sequence-to-structure alignment method which aims at improving the alignment accuracy, especially at lower sequence identity levels. The current implementation of MMM takes inputs from three profile-to-profile-based alignment methods and iteratively compares and ranks alternatively aligned regions according to their fit in the structural environment of the template structure. The performance of M4T was benchmarked on CASP6 comparative modelling target sequences and on a larger independent test set. Then it showed a favourable performance to current state-of-the-art methods. Comparative Modelling using a combination of multiple templates and iterative optimization of alternative alignments. The advantage of this server is our job can be retrieved using email address submitted for the job while submitting the query.^{3,4}

RaptorX

RaptorX is a protein structure prediction server developed by Xu group, This RaptorX is useful at predicting 3D structures for protein sequences without close homologs in the Protein Data Bank (PDB).^{5,6} After providing an input sequence, RaptorX predicts its secondary and tertiary structures as well as solvent accessibility and disordered regions. It also assigns the



following confidence scores to indicate the quality of a predicted 3D model: P-value for the relative global quality, GDT (global distance test) and uGDT (un-normalized GDT) for the absolute global quality and RMSD for the absolute local quality of each residue in the model. RaptorX-Binding is a web server that predicts the binding sites of a protein sequence, based upon the predicted 3D model by RaptorX.^{5,6,7} RaptorX excels the alignment of hard targets, which have less than 30% sequence identity with solved structures in PDB. Till now, it is tested on the 50 hardest CASP9 template-based modelling targets and RaptorX outperforms all the CASP9 participating servers including those using consensus and refinement methods. The advantage of this server is submitted jobs can be deleted from the server 6 months after completion.^{8,9,10}

I-TASSER

I-TASSER server is an internet service for protein structure and function predictions. It allows academic users to automatically generate high-quality predictions of 3D structure and biological function of protein molecules from their amino acid sequences (<1,500 residues, in FASTA format). This server is also used to make structure of the target. I-TASSER server is able to store job data for 3 months.^{13,14,15}

Critical Assessment of Techniques for Protein Structure Prediction (CASP) is a community-wide experiment for testing the state-of-the-art of protein structure predictions which takes place every two years since 1994. The experiment is often referred as a competition in international bioinformatics world which is strictly blind because the structures of testing proteins are unknown to the predictors. I-TASSER server (as "Zhang-Server") participated in the Server Section of 7th (2006), 8th (2008), 9th (2010), and 10th CASPs (2012), and was ranked as the No 1 server in CASP7 and CASP8. In CASP9 and CASP10, I-TASSER server and QUARK were ranked as No 1 and No 2 servers respectively.¹⁶

Integrated Protein Structure and Function Prediction Server (IntFOLD)(Version 2.0)

Integrated Protein Structure and Function Prediction (IntFOLD) Server allows to predict tertiary structures, assess the quality of 3D models, detect disordered regions, predict the boundaries for structural domains and predict likely ligand binding site residues for a submitted amino acid sequence.¹⁷

MODELLER

MODELLER is used for homology or comparative modelling of protein three-dimensional structures. In this server, the user need to provide an alignment of a sequence to be modelled with known related structures and MODELLER automatically calculates a model containing all non-hydrogen atoms. MODELLER implements comparative protein structure modelling by satisfaction of spatial restraints and can perform many

additional tasks, including de novo modelling of loops in protein structures, optimization of various models of protein structure with respect to a flexibly defined objective function, multiple alignment of protein sequences and/or structures, clustering, searching of sequence databases, comparison of protein structures, etc.^{18,19}

Rosetta™

Rosetta™ is a molecular modelling software package for understanding protein structures, protein design, protein docking, protein-DNA and protein-protein interactions. The Rosetta software contains multiple functional modules like Rosetta Ab initio, Rosetta Design, Rosetta Dock, Rosetta Antibody, Rosetta Fragments, Rosetta NMR, Rosetta DNA, Rosetta RNA, Rosetta Ligand, Rosetta Symmetry.²⁰ This is freely available for academic purpose. It builds model on the basis of Ab initio method. PyRosetta is a version of Rosetta implemented with a Python interface and developed primarily by Jeffrey Gray's Laboratory at Johns Hopkins University.²¹ The software is freely available for academic purpose and licences for academic and commercial purpose is available on its website.

Protein Homology/analogy Recognition engine 2

Protein Homology/analogy Recognition engine 2 (PHYRE2) is a free online homology modelling server.^{22,23} Phyre2 uses the alignment of hidden Markov models via HHsearch to significantly improve accuracy of alignment and detection rate. This incorporates a new *ab-initio* folding simulation called "Poing" to model those regions of proteins in question which have no detectable homology to known structures.²⁴

Swiss-Model 8.05

SWISS-MODEL is a fully automated protein structure homology-modeling server, accessible via the ExPASy web server, or from the program DeepView (Swiss Pdb-Viewer). The purpose of this server is to make protein modelling accessible to all biochemists and molecular biologists in the world.²⁵ A personal working environment is provided for each user where several modelling projects can be carried out in parallel. Tools for template selection, model building and structure quality evaluation can be invoked from within the workspace. This server is capable to store user job for 14 days only.²⁶

The other servers for protein modelling are CPH models-3.0, Modweb, (PS)2, (PS)V2, Homer.²⁷⁻³⁵ The efficacy of online servers with respect to modelling of protein target depending on length of target sequence is explored and explained elaborately. The commercial software's which are available for protein modelling by comparative approach, threading or fold recognition approach or ab initio approach. Accelrys Discovery Studio 4.0 and Schrodiners biologics suit are the two widely used software's for academic and industry.³¹



Protein Optimization

Online Servers

Chiron

Chiron is a protein energy minimization server.³⁶ It helps in rapid energy minimization of protein molecules using discrete molecular dynamics with an all-atom representation for each residue in the protein.³⁷ Gaia, is a tool to estimate the nature and quality of a given protein structure. Gaia compares a given protein structure against high resolution crystal structures for certain parameters including but not limited to unphysical atomic overlaps, unsatisfied hydrogen bonds and packing artifacts and reports the standing of the input structure with respect to high resolution crystal structure.³⁸

3D^{refine} server

3D^{refine} server is a web service for consistent and computationally efficient protein structure refinement.³⁹ The protocol is based on two steps of refinement process: First step is based on optimization of hydrogen bonding (HB) network. The second step applies atomic-level energy minimization on the optimized model using a composite physics and knowledge-based force fields. The goal of 3D^{refine} server is simultaneous improvement in both global and local structural qualities of the initial models to bring it closer to the native state in a computationally inexpensive manner. It hardly takes only few minutes (less than 5 minutes) to refine a protein structure of typical length (300 residues). But the time is directly proportional to the size of the protein submitted for refinement and a smaller protein takes short time than a larger protein.⁴⁰

YASARA server

This server performs an energy minimization using the YASARA force field.⁴¹ To use this server, we need to simply enter user email address then we need to upload our protein model in PDB format and afterwards we need to click the 'Submit' button. The server provides energy minimized structure within a day on email id submitted. YASARA offline tool is also available to work on but it consumes more amount of RAM.⁴²

Offline tools and Software:

University of California San Francisco Chimera:

University of California San Francisco Chimera (UCSF) is a highly extensible program for interactive visualization and analysis of molecular structures and related data, including density maps, supramolecular assemblies, sequence alignments, docking results, trajectories, and conformational ensembles. High-quality images and animations can be generated. UCSF is also used to minimize the energy and optimize the structure.⁴³⁻⁴⁹

Swiss-PdbViewer (aka DeepView):

Energy minimization is good to release local constraints, "make room" for a residue, but it will not pass through high energy barriers and stops in a local minima. Swiss-

PdbViewer includes a version of the GROMOS 43B1 force field [W.F. van Gunsteren et al. (1996) in Biomolecular simulation: the GROMOS96 manual and user guide. VdfHochschulverlag ETHZ]. This force field allows to evaluate the energy of a structure as well as repair distorted geometries through energy minimization. In this implementation, all computations are done in vacuum, without reaction field.⁵⁰

Protein validation Servers

Structural Analysis and Verification Server (SAVES)

In SAVES, user can upload modelled protein in.pdb format and it will check for PROVE, ERRAT, VERIFY3D.^{51,52}

Harmony Server

HARMONY server a server which is used to assess the compatibility of an amino acid sequence with a proposed three-dimensional structure. Structural descriptors such as backbone conformation, solvent accessibility and hydrogen bonding are used to characterise the structural environment of each residue position. Propensity and Substitution values are used together to predict the occurrence of an amino acid at each position in the sequence on the basis of the local structural environment. We demonstrate that the information from amino acid substitutions among homologous sequences (in the form of environment-dependent amino acid substitution tables) is a powerful tool for identifying errors that may be present in the protein structure.⁵²

CONCLUSION

These are freely available tools and servers which can be exploited further to the maximum potential to get insight of protein in 3D and to gather in depth information about protein using in silico approach.

REFERENCES

1. Genetics Home Reference Your Guide to Understanding Genetic Conditions, 2014 <http://ghr.nlm.nih.gov/handbook.pdf> (Last accessed on 28/05/2014).
2. <http://www.fiserlab.org/servers/m4t> (Last accessed on 28/05/2014).
3. Fernandez-Fuentes N, Madrid-Aliste CJ, Rai BK, Fajardo JE, Fiser A, M4T: a comparative protein structure modeling server, *Nucleic acids research*, 35, 2007, 363-368.
4. Fernandez-Fuentes N, Rai BK, Madrid-Aliste CJ, Fajardo JE, Fiser A, Comparative protein structure modeling by combining multiple templates and optimizing sequence-to-structure alignments, *Bioinformatics*, 23(19), 2007, 2558-65.
5. Rykunov D, Steinberger E, Madrid-Aliste CJ, Fiser A, Improved scoring function for comparative modeling using the M4T method, *Journal of structural and functional genomics*, 10(1), 2009, 95-99.
6. Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, Template-based protein structure modeling using the RaptorX web server, *Nature protocols*, 7(8), 2012, 1511-1522.



7. Ma J, Wang S, Zhao F, Xu J, Protein threading using context-specific alignment potential, *Bioinformatics*, 29(13), 2013, 257-265.
8. Ma J, Peng J, Wang S, Xu J, A conditional neural fields model for protein threading, *Bioinformatics*, 28(12), 2012, 59-66.
9. Peng J, Xu J, Boosting Protein Threading Accuracy, Research in computational molecular biology: Annual International Conference, RECOMB: proceedings International Conference on Research in Computational Molecular Biology, 5541, 2009, 31-45.
10. Peng J, Xu J, Low-homology protein threading, *Bioinformatics*, 26(12), 2010, 294-300.
11. Peng J, Xu J, RaptorX: exploiting structure information for protein alignment by statistical inference, *Proteins*, 79(10), 2011, 161-171.
12. Peng J, Xu J, A multiple-template approach to protein threading, *Proteins*, 79(6), 2011, 1930-1939.
13. Roy A, Kucukural A, Zhang Y, I-TASSER: a unified platform for automated protein structure and function prediction, *Nat Protoc*, 5(4), 2010, 725-38.
14. Roy A, Yang J, Zhang Y, COFACTOR: an accurate comparative algorithm for structure-based protein function annotation, *Nucleic acids research*, 40, 2012, 471-477.
15. Zhang Y, I-TASSER server for protein 3D structure prediction, *BMC bioinformatics*, 9(40), 2008.
16. <http://predictioncenter.org/> (Last accessed on 28/05/2014).
17. Roche DB, Buenavista MT, Tetchner SJ, McGuffin LJ, The IntFOLD server: an integrated web resource for protein fold recognition, 3D model quality assessment, intrinsic disorder prediction, domain prediction and ligand binding site prediction, *Nucleic acids research*, 39, 2011, 171-176.
18. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A, Comparative protein structure modeling of genes and genomes. Annual review of biophysics and biomolecular structure, 29, 2000, 291-325.
19. Sali A, Blundell TL, Comparative protein modelling by satisfaction of spatial restraints, *Journal of molecular biology*, 234(3), 1993, 779-815.
20. Raman S, Vernon R, Thompson J, Tyka M, Sadreyev R, Pei J, Structure prediction for CASP8 with all-atom refinement using Rosetta, *Proteins: Structure, Function, and Bioinformatics*, 77(S9), 2009, 89-99.
21. http://c4c.uw4c.com/express_license_technologies/rosetta (Last accessed on 27th may 2014).
22. Kelley LA, Sternberg MJ, Protein structure prediction on the Web: a case study using the Phyre server, *Nat Protoc*, 4(3), 2009, 363-371.
23. Soding J, Protein homology detection by HMM-HMM comparison, *Bioinformatics*, 21(7), 2005, 951-960.
24. Malik IA, Sharif S, Malik F, Hakimali A, Khan WA, Badruddin SH, Nutritional aspects of mammary carcinogenesis: a case-control study, *The Journal of the Pakistan Medical Association*, 43(6), 1993, 118-120.
25. <http://swissmodel.expasy.org/workspace> (Last accessed on 27th may 2014).
26. Guex N, Peitsch MC, SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling, *Electrophoresis*, 18(15), 1997, 2714-2723.
27. Nielsen M, Lundegaard C, Lund O, Petersen TN, CPHmodels-3.0--remote homology modeling using structure-guided sequence profiles, *Nucleic acids research*, 38, 2010, 576-581.
28. Eramian D, Eswar N, Shen MY, Sali A, How well can the accuracy of comparative protein structure models be predicted?, *Protein science : a publication of the Protein Society*, 17(11), 2008, 1881-1893.
29. Eswar N, John B, Mirkovic N, Fiser A, Ilyin VA, Pieper U, Tools for comparative protein structure modeling and analysis, *Nucleic acids research*, 31(13), 2003, 3375-3380.
30. Chen CC, Hwang JK, Yang JM, (PS)2: protein structure prediction server, *Nucleic acids research*, 34, 2006, 152-157.
31. Nema V, Pal SK, Exploration of freely available web-interfaces for comparative homology modelling of microbial proteins, *Bioinformation*, 9(15), 2013, 796-801.
32. Notredame C, Higgins DG, Heringa J, T-Coffee: A novel method for fast and accurate multiple sequence alignment, *Journal of molecular biology*, 302(1), 2000, 205-217.
33. Schaffer AA, Aravind L, Madden TL, Shavirin S, Spouge JL, Wolf YI, Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements, *Nucleic acids research*, 29(14), 2001, 2994-3005.
34. Schaffer AA, Wolf YI, Ponting CP, Koonin EV, Aravind L, Altschul SF, IMPALA: matching a protein sequence against a collection of PSI-BLAST-constructed position-specific score matrices, *Bioinformatics*, 15(12), 1999, 1000-1011.
35. Wallner B, Elofsson A, Identification of correct regions in protein models using structural, alignment, and consensus information, *Protein science : a publication of the Protein Society*, 15(4), 2006, 900-13.
36. Ramachandran S, Kota P, Ding F, Dokholyan NV, Automated minimization of steric clashes in protein structures, *Proteins: Structure, Function, and Bioinformatics*, 79(1), 2011, 261-270.
37. <http://troll.med.unc.edu/chiron/login.php> (Last accessed on 27th May 2014).
38. Kota P, Ding F, Ramachandran S, Dokholyan NV, Gaia: automated quality assessment of protein structure models, *Bioinformatics*, 27(16), 2011, 2209-2215.
39. <http://sysbio.rnet.missouri.edu/3Drefine> (Last accessed on 27th may 2014).
40. Bhattacharya D, Cheng J, 3Drefine: consistent protein structure refinement by optimizing hydrogen bonding network and atomic-level energy minimization, *Proteins*, 81(1), 2013, 119-1131.
41. <http://yasara.org/servers.htm> (Last accessed on 27th may 2014).
42. Krieger E, Joo K, Lee J, Lee J, Raman S, Thompson J, Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: four approaches that performed well in CASP8, *Proteins: Structure, Function, and Bioinformatics*, 77(9), 2009, 114-122.



43. Couch GS, Hendrix DK, Ferrin TE, Nucleic acid visualization with UCSF Chimera, *Nucleic acids research*, 34(4), 2006, 29-31.
44. Goddard TD, Huang CC, Ferrin TE, Software extensions to UCSF chimera for interactive visualization of large molecular assemblies, *Structure*, 13(3), 2005, 473-482.
45. Goddard TD, Huang CC, Ferrin TE, Visualizing density maps with UCSF Chimera, *Journal of structural biology*, 157(1), 2007, 281-287.
46. Morris JH, Huang CC, Babbitt PC, Ferrin TE, StructureViz: linking Cytoscape and UCSF Chimera, *Bioinformatics*, 23(17), 2007, 2345-2347.
47. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, UCSF Chimera-a visualization system for exploratory research and analysis, *Journal of computational chemistry*, 25(13), 2004, 1605-1612.
48. Pintilie GD, Zhang J, Goddard TD, Chiu W, Gossard DC, Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions, *Journal of structural biology*, 170(3), 2010, 427-438.
49. Yang Z, Lasker K, Schneidman-Duhovny D, Webb B, Huang CC, Pettersen EF, UCSF Chimera, MODELLER, and IMP: an integrated modelling system, *Journal of structural biology*, 179(3), 2012, 269-278.
50. Guex N, Peitsch MC, Schwede T, Automated comparative protein structure modelling with SWISS-MODEL and Swiss-Pdb Viewer: A historical perspective, *Electrophoresis*, 30(1), 2009, 162-173.
51. Colovos C, Yeates TO, Verification of protein structures: patterns of nonbonded atomic interactions, *Protein science: a publication of the Protein Society*, 2(9), 1993, 1511-1519.
52. Bowie JU, Luthy R, Eisenberg D, A method to identify protein sequences that fold into a known three-dimensional structure, *Science*, 253(5016), 1991, 164-170.
53. <http://caps.ncbs.res.in/harmony/>(Last accessed on 27th May 2014).

Source of Support: Nil, **Conflict of Interest:** None.

