

Computational Methods for Protein Structure Prediction and Its Application in Drug Design

Nishant Toomula^{1*}, D Sathish Kumar² and VVL Pavan Kumar A³

¹Department of Biotechnology, GITAM Institute of Technology, GITAM University, Visakhapatnam, India

²Department of Biotechnology, University of Hyderabad, Hyderabad, India

³Department of Bioinformatics, Bharath University, Chennai, India

Abstract

Drug design and drug discovery are of critical importance in human health care. Computational approaches have become a major part of structure based drug design. Structure-based drug design utilizes the three dimensional structure of a protein target to design candidate drugs that are predicted to bind with high affinity and selectivity to the target. In this review computational methods for prediction of the protein structure are described and their use towards the drug design is discussed.

Keywords: Protein structure prediction; SOPMA; TMHMM; Comparative modeling; MODELLER; Drug design; Docking

Introduction

Most modern drug discovery projects start with protein target identification and verification to obtain a verified drug target. For structure-based drug design the three-dimensional structure of the protein needs to be determined experimentally by using either x-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy [1]. While both methods are increasingly being applied in a high-throughput manner, structure determination is not yet a straightforward process. X-ray crystallography is limited by the difficulty of getting some proteins to form crystals, and NMR can only be applied to relatively small protein molecules.

Proteins are essential to biological processes. They are responsible for catalyzing and regulating biochemical reactions, transporting molecules, the chemistry of vision and of the photosynthetic conversion of light to growth, and they form the basis of structures such as skin, hair, and tendon. Protein function can be understood in terms of its structure. Indeed, the three-dimensional structure of a protein is closely related to its biological function. Proteins that perform similar functions tend to show a significant degree of structural homology [2].

The amino acid sequence of a protein is known as its primary structure, while local conformations in this sequence, namely alpha-helices, beta sheets, and random coils are known as secondary structures. The angles between adjacent amino acids, called the torsion angles [3], determine the twists and turns in the sequences which result in these secondary structures. The three-dimensional configuration of the primary structure is defined as the tertiary structure, describing the fold of the protein.

Each amino acid consists of a rigid plane formed by single nitrogen, carbon, alpha-carbon (Ca), oxygen, and hydrogen atoms, and a distinguishing side chain. The individual amino acids are distinguished from each other by a number of physical chemical properties that give rise to the three dimensional structure [4].

ExPASy [5,6] is a proteomics server operated by the Swiss Institute of Bioinformatics, it is used to analyze protein sequences and structures and two-dimensional gel electrophoresis (2-D Page electrophoresis) [7-9].

Primary structure analysis

Amino acid sequence analysis [10] provides important insight into the structure of proteins, which in turn greatly facilitates the understanding of its biochemical and cellular function. Efforts to use computational methods in predicting protein structure based only on sequence information started 30 years ago [11]. However, only during the last decade, has the introduction of new computational techniques such as protein fold recognition and the growth of sequence and structure databases due to modern high-throughput technologies led to an increase in the success rate of prediction methods.

Sequence retrieval database searches

Sequence similarity searching is a crucial step in analyzing newly determined protein sequences. Typically, large sequence databases such as the non-redundant (nr) database at the NCBI [12] (synthesis of Gen-Bank, EMBL and DDBJ databases) or genome sequences are scanned for DNA or amino acid sequences that are similar to a target sequence. Alignments of the target sequence are constructed for each database entry, typically using dynamic programming algorithms [13]. Scores derived from these alignments are used to identify statistically significant matches.

Traditionally, searches were carried out using programs for pairwise sequence comparisons like FASTA [14] or BLAST [15-17]. However, the relationship between sequences of homologous proteins can be recognized by pairwise sequence comparisons. The most sensitive methods available today use the initial search for homologues to construct a multiple sequence alignment (MSA) [18], which provide insight into the positional constraints of the amino acid composition,

***Corresponding author:** Nishant Toomula, Department of Biotechnology, GITAM Institute of Technology, GITAM University, Visakhapatnam, India, E-mail: nishanththumula@gmail.com

Received October 25, 2011; **Accepted** December 03, 2011; **Published** December 08, 2011

Citation: Nishant T, Sathish Kumar D, VVL Pavan Kumar A (2011) Computational Methods for Protein Structure Prediction and Its Application in Drug Design. J Proteomics Bioinform 4: 289-293. doi:10.4172/jpb.1000203

Copyright: © 2011 Nishant T, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

and allow the identification of conserved and variable regions in the family, comprising the target and its presumed homologues

Protein domain identification

After protein discovery, there are many questions that are associated with protein's overall identity, putative function and biologically significant sites identification [19]. To answer these questions, a number of databases and tools have been customized. Most proteins are composed from a finite number of evolutionarily conserved modules or domains. Protein domains are distinct units of three-dimensional protein structures, which often carry a discrete molecular function, such as the binding of a specific type of molecule. These domains vary in length from between about 25 amino acids up to 500 amino acids. The direct functional and structural determination of all the proteins in an organism is prohibitively costly and time consuming because of the relative scarcity of 3D structural information therefore primary sequence analysis is preferred to identify majority of protein domain families [20].

A few thousand conserved domains, which cover more than two thirds of known protein sequences have been identified and described in literature. PFAM [21,22] and SMART [23] databases are the largest collections of the manually curated protein domains of information. Each deposited domain family is extensively annotated in the form of textual descriptions, as well as cross-links to other resources and literature references.

Secondary structure prediction

The important concepts in secondary structure prediction [24] are identified as: residue conformational propensities, sequence edge effects, moments of hydrophobicity, position of insertions and Deletions in aligned homologous sequence, moments of conservation, auto-correlation, residue ratios, secondary structure feedback effects, and filtering [25].

The early methods of secondary structure prediction are suffered from a lack of data and Predictions were performed on single sequences rather than families of homologous sequences, and there were relatively few known 3D structures from which to derive parameters. The most famous early methods are those of Chou & Fasman, Garnier, Osguthorpe & Robson (GOR) and Lim. Although the authors originally claimed quite high accuracies (70-80 %), under careful examination, the methods were shown to be only between 56 and 60% accurate [26].

Recent improvements

The availability of large families of homologous sequences revolutionized secondary structure prediction. Traditional methods, when applied to a family of proteins rather than a single sequence proved much more accurate at identifying core secondary structure elements. The combination of sequence data with sophisticated computing techniques such as neural networks [27] has led to accuracies well in excess of 70 %. Though this seems a small percentage increase, these predictions are actually much more useful than those for single sequence, since they tend to predict the core accurately. Moreover, the limit of 70-80% may be a function of secondary structure variation within homologous proteins.

SOPMA (Self Optimized Prediction Method from Alignment) [28] was employed for prediction of secondary structure features like alpha helix, extended strand, beta turn and random coils in terms of percentage for all the sequences. These features were considered as

input parameters for self organizing maps for further analysis [29]. SOPMA accurately predicts 69.5% of amino acid for the three states describing the secondary structure (α -helix, β -beta sheet and coil). This tool works on the basis of neural network method (PHD) [30]. The PHD algorithm first performs a database search for possible homologous proteins, then aligns and filters the sequences to decide on the most likely homologues, and finally feeds the sequences and alignment profile to a feed-forward neural network for secondary structure prediction [31].

Transmembrane region prediction

Different servers TMHMM, SOSUI [32], HMMTOP and TMpred servers were accessed to validate the TM region [33-34]. TMHMM, a new membrane protein topology prediction method, is based on a hidden Markov model [35].

Transmembrane topology predictions

In a study conducted by Kumar et al. the transmembrane topology of AHA1 was predicted from the amino acid sequence [36] by averaging the results of four different predictive algorithms: DAS [37], HMMTOP [38], TMHMM [39] and TMPRED [40]. The accuracy of the prediction was assessed by using the same algorithms on 3b8c protein and the results were compared with the topology defined in the 3.6A° structures (3b8c).

Tertiary structure

Knowing a protein's 3-dimensional structure (Tertiary Structure) helps us to understand its functionality and provides means for planning experiments and drug design. The Brookhaven Protein Data Bank (PDB) is the repository for those structures. Files including atom coordinates which are suited for visualization by graphical molecule viewers like rasmol can be obtained at this site. PDB is also searchable with a sequence as a query, e.g. with the BLAST service located at NCBI with a polypeptide as a query.

Tertiary structure of a protein is build by packing of its secondary structure elements to form discrete domains or autonomous folding units [41]. Two main approaches in determination of protein 3D structure are: Ab initio prediction and comparative modeling.

Comparative modeling

Homology or comparative protein structure modeling constructs a three-dimensional model of a given protein sequence based on its similarity to one or more known structures [42]. It is carried out in four sequential steps: finding known structures (templates) related to the sequence to be modeled (target), aligning the target sequence with the templates, building the model, and assessing the model [43]. Therefore, comparative modeling is only applicable when the target sequence is detectably related to a known protein structure.

3D structure generation by using MODELLER

Modeller is a computer program for comparative modeling of protein three-dimensional structures. Alignment of a sequence to be modeled is provided with known related structures and modeller automatically calculates a model containing all non-hydrogen atoms. Modeller [44] implements comparative protein structure modeling by satisfaction of spatial restraints. The homology modeling requires sequences of known 3D structure and the target having above 35% of similarity.

Template identification and sequence alignment

Template identification is an important step. It lays the foundation by identifying appropriate homologues of known protein structure, called template, which are sufficiently similar to the target sequence to be modeled. Template sequence were selected by a simple search submits the target sequence to programs BLASTP search along with default parameters was performed against the Brook Heaven Protein Data Bank (PDB). Based on the high identity, lowest e-value and low gaps the high resolution having sequence was selected as a template. To ensure the high accuracy of the structure, the target and the template sequence can be aligned.

Model building and refinement

Although the theory behind building a protein homology model is complicated, using available programs is relatively easy. Several modeling programs [45] are available, using different methods to construct the 3D structures. In segment matching methods, the target is divided into short segments, and alignment is done over segments rather than over the entire protein. Satisfying spatial restraints is the most common method. It uses either distances or optimization techniques to satisfy the spatial restraints. The method is implemented using the popular program, Modeller [46] and which includes the CHARMM [47] energy terms that ensure valid stereochemistry is combined with spatial restraints [48].

Validation

The best validation combines common sense, biological knowledge and results from analytical tools. Most refinement involves adjusting the alignment. PROCHECK [49,50] is used to calculate the main-chain torsion angles, i.e. the Ramachandran plot [51] for our predicted structures. Three models were predicted using different templates among those the one that shows the good resolution factor and R-factor was used as a template and evaluated by Procheck performing full geometric analysis with a resolution of 1.5 Å. The validation for structure models obtained from the three software tools was performed by using PROCHECK [52].

Comparative modeling of rat cathepsin L

In a research work done by sunil kumar, Priya ranjan and supakar the amino acid sequence of rat cathepsin L was retrieved from the sequence database of NCBI. It was ascertained that the three-dimensional structure of the protein was not available in Protein Data Bank; hence BLAST search was performed against Brookhaven Protein Data Bank (PDB) with the default parameters to find suitable templates for homology modeling. Sequences were aligned and the one that showed the maximum identity with high score and lower e-value and 73% sequence identity was used as a reference structure to build a 3D model for rat cathepsin L. The rat cathepsin L structure was modeled by means of comparative modeling procedure using the 1CS8 as the template. The rat cathepsin L sequence was submitted to Genesilico protein fold-recognition metaserver [53].

Fold-recognition server Fugue and 3D PSSM reported 1CS8 as the best template with highly significant score. The academic version of MODELLER 9v2 [54] was used for model building. Backbone of the core regions of the protein were transferred directly from the corresponding coordinates of 1CS8. Side chains confirmation for backbone residues was generated automatically by homology. Out of 20 models generated by MODELLER, the one with the best G-score

of PROCHECK and with the best VERIFY3D [55,56] profile was subjected to energy minimization.

Comparative modeling of Viral Protein R (VpR)

In a study conducted by Seenivasagan et al. the protein sequence of VpR [57] was retrieved form from the NCBI database (POC1P5) which has 96 amino acids. The target sequence was searched for similar sequence using the BLAST, against Protein Database (PDB). The BLAST results yielded NMR structure of HIV-1 regulatory protein R VpR with 85% similarity to our target protein. The theoretical structure of VpR is generated using Modeller-9v1 for comparative modeling of protein structure prediction [58]. It implements comparative structural modeling by conforming special restraints [59].

Quaternary structure

In the case of complexes of two or more proteins, where the structures of the proteins are known or can be predicted with high accuracy, protein-protein docking methods can be used to predict the structure of the complex.

Methods for the prediction of protein interactions

Ramon Aragues and his co workers used four different methods for predictions of protein-protein interactions [60]: (i) Gene fusion, in which two proteins are predicted to interact if their corresponding genes appear fused in another genome [61].

(ii) Phylogenetic profiles, in which similarity of phylogenetic profiles is interpreted as being indicative of two proteins need to be simultaneously present to perform a given function together [62].

(iii) Distant conservation of sequence patterns and structure relationships, in which structural similarities among domains of known interacting proteins and conservation of pairs of sequence patches involved in protein-protein interfaces are used to predict putative protein interaction pairs [63] and

(iv) Structural interologs, in which interactions are transferred between proteins with the same structural domains [64].

Drug development based on protein structure

The object of drug design is to find or develop a, mostly small, drug molecule that tightly binds to the target protein, moderating its function or competing with natural substrates of the protein. Such a drug can be best found on the basis of knowledge of the protein structure. If the spatial shape of the site of the protein is known, to which the drug is supposed to bind, then docking methods can be applied to select suitable lead compounds that have the potential of being refined to drugs.

Docking

Docking is a method which predicts the preferred orientation of one molecule to a record when bound to each other to form stable complex knowledge of the preferred orientations in turn may be used to predict the binding strength of association or binding affinity between two molecules [65]. Docking is frequently used to predict the binding orientations of small molecules drug candidates to protein targets in order to in turn predict the affinity and activity of the small molecule [66].

The development and implementation of a range of molecular docking algorithms [67] based on different search methods [68] were

observed in the last few years. This approach has had several recent successes in drug discovery [69].

A number of powerful software programs, e.g. AutoDock [70,71], HEX [72,73], GOLD [74,75], FlexX, DOCK, Glide, Surflex, LigandFit, have been developed over the past several decades to carry out docking calculations, and good success in both binding mode and binding affinity prediction has often been achieved in selected test cases [76].

Conclusion

Computational methods for protein structure prediction are still in the stage of development and methods like homology-based prediction become especially helpful in an environment where the methods can be used in concert with experimental techniques for structure and function determination of protein. The use of computers and computational methods permeates all aspects of drug discovery today and forms the core of structure-based drug design. Availability of protein 3D structures, high-performance computing, data management software and internet are facilitating the access of huge amount of data generated and transforming the massive complex biological data into workable knowledge in modern day drug discovery process. Computational tools offer the advantage of delivering new drug candidates more quickly and at a lower cost.

References

1. Afergan E, Najjireh Y, Gutman D, Epstein H, Elmalak O, et al. (2010) 31P-NMR and Differential Scanning Calorimetry Studies for Determining Vesicle's Drug Physical State and Fraction in Alendronate Liposomes. *J Bioanal Biomed* 2: 125-131.
2. Chan HS, Dill K (1993) The protein folding problem. *Physics Today* 24–32.
3. Jung S, Bae SE, Son HS (2011) Validity of Protein Structure Alignment Method Based on Backbone Torsion Angles. *J Proteomics Bioinform* 4: 218-226.
4. Wilcox GL, Poliac M, Liebman MN (1990) Neural network analysis of protein tertiary structure. *Tetrahedron Computer Methodology* 3: 191–211.
5. Chen B (2011) Reconstruction of Local Biochemical Reaction Network Based on Human Chromosome 9 Sequence Data. *J Proteomics Bioinform* 4: 087-090.
6. Asthana M, Singh VK, Kumar R, Chandra R (2011) Isolation, Cloning and In silico Study of Hexon Gene of Fowl Adenovirus 4 (FAV4) Isolates Associated with Hydro Pericardium Syndrome in Domestic Fowl. *J Proteomics Bioinform* 4: 190-195.
7. Mazzara S, Cerutti S, Iannaccone S, Conti A, Olivieri S, et al. (2011) Application of Multivariate Data Analysis for the Classification of Two Dimensional Gel Images in Neuroproteomics. *J Proteomics Bioinform* 4: 016-021.
8. Kanamoto T, Souchelnytskyi N, Toda R, Rimayanti U, Kiuch Y (2011) Proteomic Analyses of Proteins Differentially Expressed in Recurrent and Primary Pterygia. *J Proteomics Bioinform* 4: 058-061.
9. Rajkumar R, Ilayaraja R, Alagendran S, Archunan G, Muralidharan AR, et al. (2011) Characterization of Rat Odorant Binding Protein Variants and its Post-Translational Modifications (PTMs): LC-MS/MS analyses of Protein Eluted from 2D-Polyacrylamide Gel Electrophoresis. *J Proteomics Bioinform* 4: 210-217.
10. Lu CH, Ge F, Liu Z, Li R, Xiao CL, et al. (2010) Detection of Abundant Proteins in Multiple Myeloma Cells by Proteomics. *J Proteomics Bioinform* 3: 005-009.
11. Nagano K (1973) Logical analysis of the mechanism of protein folding. I. Predictions of helices, loops and beta-structures from primary structure. *J Mol Biol* 75: 401–420.
12. Amit KB, Neelima A, Varakantham P, Murty USN (2008) Exploring the Interplay of Sequence and Structural Features in Determining the Flexibility of AGC Kinase Protein Family: A Bioinformatics Approach. *J Proteomics Bioinform* 1: 077-089.
13. Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* 48: 443–453.
14. Pearson WR, Lipman DJ (1988) Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A* 85: 2444–2448.
15. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403–410.
16. Neha S, Vrat BS, Kumud J, Thakur PD, Rajinder K, et al. (2011) Comparative In silico Analysis of Partial Coat Protein Gene Sequence of Zucchini Yellow Mosaic Virus Infecting Summer Squash (*Cucurbita pepo* L.) Isolated From India. *J Proteomics Bioinform* 4: 068-073.
17. Neha G, Sachin P, Anil P, Anil K (2008) Primer Designing for Dreb1A, A Cold Induced Gene. *J Proteomics Bioinform* 1: 028-035.
18. Dangre DM, Rathod DP, Gade AK, Rai MK (2009) An in Silico Molecular Evolutionary Analysis of Selected Species of Phoma : A Comparative Approach. *J Proteomics Bioinform* 2: 295-309.
19. Varsale AR, Wadnerkar AS, Mandage RH, Jadhavrao PK (2010) Cheminformatics. *J Proteomics Bioinform* 3: 253-259.
20. Sonnhammer EL, von Heijne G, Krogh A (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* 6: 175–182.
21. Asthana M, Singh VK, Kumar R, Chandra R (2011) Isolation, Cloning and In silico Study of Hexon Gene of Fowl Adenovirus 4 (FAV4) Isolates Associated with Hydro Pericardium Syndrome in Domestic Fowl. *J Proteomics Bioinform* 4: 190-195.
22. Mishra A, Pandey D, Goel A, Kumar A (2010) Molecular Cloning and In silico Analysis of Functional Homologues of Hypersensitive Response Gene(s) Induced During Pathogenesis of Alternaria Blight in Two Genotypes of Brassica. *J Proteomics Bioinform* 3: 244-248.
23. Allam AR, Sridhar GR, Suresh BM, Vamsidhar E, Gunna K (2008) Computational Analysis of Mutations in Colon Cancer Genes Reveals a Possible Role of Micro Satellite in Mutagenesis. *J Proteomics Bioinform* S1: S041-S045.
24. Gomase VS, Kale KV, Shyamkumar K (2008) Prediction of MHC Binding Peptides and Epitopes from Groundnut Bud Necrosis Virus (GBNV). *J Proteomics Bioinform* 1: 188-205.
25. Garnier J, Osguthorpe DJ, Robson B (1978) Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J Mol Biol* 120: 97-120.
26. Kabsch W, Sander C (1983) How good are predictions of protein secondary structure. *FEBS Letters* 155: 179-182.
27. Yadav BS, Pokhariyal M, Ratta B, Rai G, Saxena M, et al. (2010) Predicting Secondary Structure of Oxidoreductase Protein Family Using Bayesian Regularization Feed-forward Backpropagation ANN Technique. *J Proteomics Bioinform* 3: 179-182.
28. Geourjon C, Deléage G (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput Appl Biosci* 11: 681-684.
29. Murty USN, Amit KB, Neelima A (2009) An In Silico Approach to Cluster CAM Kinase Protein Sequences. *J Proteomics Bioinform* 2: 097-107.
30. Selvaraj D, Loganathan A, Sathishkumar R (2010) Molecular Characterization and Phylogenetic Analysis of BZIP Protein in Plants. *J Proteomics Bioinform* 3: 230-233.
31. Liang K, Wang X (2011) Protein Secondary Structure Prediction using Deterministic Sequential Sampling. *J Data Mining in Genom Proteomics* 2:107.
32. Vaseeharan B, Valli SJ (2011) In silico Homology Modeling of Prophenoxydase activating factor Serine Proteinase Gene from the Haemocytes of *Fenneropenaeus indicus*. *J Proteomics Bioinform* 4: 053-057.
33. Krogh A, Larsson B, Von HG, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hid-den Markov model: application to complete genomes. *J Mol Biol* 305: 567-80.
34. Hirokawa T, Boon CS, Mitaku S (1998) SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* 14: 378-379.
35. Ganesh CS, Manas RD, Mukta R, Pradeep D (2009) Homology Modeling and Functional Analysis of LPG2 Protein of Leishmania Strains. *J Proteomics Bioinform* 0: 032-050.
36. Kumar S, Sahu BB, Tripathy NK, Shaw BP (2009) In Silico Identification of

- Putative Proton Binding Sites of a Plasma Membrane H⁺-ATPase Isoform of *Arabidopsis Thaliana*, AHA1. *J Proteomics Bioinform* 2: 349-359.
37. Cserzo M, Wallin E, Simon I, von Heijne G, Elofsson A (1997) Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein Eng* 10: 673-676.
 38. Tusnady GE, Simon I (1998) Principles governing amino acid composition of integral membrane proteins: application to topology prediction. *J Mol Biol* 283: 489-506.
 39. Min XJ (2010) Evaluation of Computational Methods for Secreted Protein Prediction in Different Eukaryotes. *J Proteomics Bioinform* 3: 143-147.
 40. Hofmann K, Stoffel W (1993) TMbase - A database of membrane spanning protein segments. *Biol Chem Hoppe-Seyler* 374: 166-171.
 41. Subramanian R, Muthurajan R, Ayyanar M (2008) Comparative Modeling and Analysis of 3-D Structure of EMV2, a Late Embryogenesis Abundant Protein of *Vigna Radiata* (Wilczek). *J Proteomics Bioinform* 1: 401-407.
 42. Wheeler KE, Zemla A, Jiao Y, Aliaga Goltsman DS, Singer SW, et al. (2010) Functional Insights from Computational Modeling of Orphan Proteins Expressed in a Microbial Community. *J Proteomics Bioinform* 3: 266-274.
 43. Sahay A, Shakya M (2010) In silico Analysis and Homology Modelling of Antioxidant Proteins of Spinach. *J Proteomics Bioinform* 3: 148-154.
 44. Sali A, Blundell TL (1993) Comparative protein modeling by satisfaction of spatial restraints. *J Mol Biol* 234: 779-815.
 45. Shakil S, Khan AU (2010) Interaction of 2009 CTX - M Variants with Drugs and Inhibitors: a Molecular Modeling and Docking Study. *J Proteomics Bioinform* 3: 130-134.
 46. Marti-Renom MA, Stuart A, Fiser A, Sánchez R, Melo F, et al. (2000) Comparative protein structure modeling of genes and genomes. *Annu Rev Biochem Biomol Struct* 29: 291-325.
 47. Suresh N, Vasanthi NS (2010) Pharmacophore Modeling and Virtual Screening Studies to Design Potential Protein Tyrosine Phosphatase 1B Inhibitors as New Leads. *J Proteomics Bioinform* 3: 020-028.
 48. Brooks BR, Brucoleri RE, Olafson BD, States DJ, Swaminathan S, et al. (1993) CHARMM: a program for macromolecular energy minimization and dynamics calculations. *J Comput Chem* 4: 187-217.
 49. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereo chemical quality of protein structures. *J Appl Crystallography* 26: 283-291.
 50. Nath M, Goel A, Taj G, Kumar A (2010) Molecular Cloning and Comparative In silico Analysis of Calmodulin Genes from Cereals and Millets for Understanding the Mechanism of Differential Calcium Accumulation. *J Proteomics Bioinform* 3: 294-301.
 51. Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) Stereochemistry of polypeptide chain configurations. *J Mol Biol* 7: 95-99.
 52. Laskowski RA, Rullmann JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 8: 477-486.
 53. Sunil K, Priya RD, Prakash CS (2008) Prediction of 3-Dimensional Structure of Cathepsin L Protein of *Rattus Norvegicus*. *J Proteomics Bioinform* 1: 307-314.
 54. Luthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. *Nature* 356: 83-85.
 55. Seenivasagan R, Kasimani R, Marimuthu P, Kalidoss R, Shanmughavel P (2008) Comparative Modeling of Viral Protein R (VpR) from Human Immunodeficiency Virus 1 (HIV 1). *J Proteomics Bioinform* 1: 073-076.
 56. Seetharaman B, Krishnan VM (2008) In Silico Analysis of Alkaline Shock Proteins in Enterobacteria. *J Proteomics Bioinform* S1: S021-S037.
 57. Eswar N, Marti-Renom MA, Webb B, Madhusudhan MS, Eramian D, et al. (2006) Comparative Protein Structure Modeling using MODELLER. *Curr Protoc Bioinformatics*. 561: 5630.
 58. Fiser A, Do RK, Sali A (2000) Modeling of loops in protein structures. *Protein Sci* 9: 1753-73.
 59. Ramón A, Javier GG, Baldo O (2008) Integration and Prediction of PPI Using Multiple Resources from Public Databases. *J Proteomics Bioinform* 1: 166-187.
 60. Enright AJ, Iliopoulos I, Kyripides NC, Ouzounis CA (1999) Protein interaction maps for complete genomes based on gene fusion events. *Nature* 402: 86-90.
 61. Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO (1999) Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proceedings of the National Academy of Sciences of the United States of America* 96: 4285-4288.
 62. Espadaler J, Romero-Isart O, Jackson RM, Oliva B (2005b) Prediction of protein-protein interactions using distant conservation of sequence patterns and structure relationships. *Bioinformatics* 21: 3360-3368.
 63. Aragues R, Jaeggi D, Oliva B (2006) PIANA: protein interactions and network analysis. *Bioinformatics* 22: 1015-1017.
 64. Ramanathan K, Karthick H, Arun N (2010) Structure Based Drug Designing for Diabetes Mellitus. *J Proteomics Bioinform* 3: 310-313.
 65. Abdelouahab C, Abderrahmane B (2008) Comparative Study of the Efficiency of Three Protein-Ligand Docking Programs. *J Proteomics Bioinform* 1: 161-165.
 66. Girija CR, Karunakar P, Poojari CS, Begum NS, Syed AA (2010) Molecular Docking Studies of Curcumin Derivatives with Multiple Protein Targets for Procarcinogen Activating Enzyme Inhibition. *J Proteomics Bioinform* 3: 200-203.
 67. Taylor RD, Jewsbury PJ, Essex JW (2002) A review of protein-small molecule docking methods. *J Comput Aided Mol Des* 16: 151-166.
 68. Halperin I, Ma B, Wolfson H, Nussinov R (2002) Principles of docking: an overview of search algorithms and a guide to scoring functions. *Proteins* 47: 409-443.
 69. Sechi M, Sannia L, Carta F, Palomba M, Dallochio R, et al. (2005) Design of novel bioisosteres of beta-diketo acid inhibitors of HIV-1 integrase. *Antivir Chem Chemother* 16: 41-46.
 70. Kaushik SH, Chandregowda V, Venkateswara RG, Anil K, Chandrasekara RG (2009) In-silico Interaction Studies of Quinazoline Derivatives for their Inhibitory Action on Both Wild and Mutant EGFRs. *J Proteomics Bioinform* 2: 126-130.
 71. Chikhi A, Bensegueni A (2010) In Silico Study of the Selective Inhibition of Bacterial Peptide Deformylases by Several Drugs. *J Proteomics Bioinform* 3: 061-065.
 72. Bagchi P, Mahesh M, Somashekar R (2009) Pharmaco-Informatics: Homology Modeling of the Target Protein (GP1, 2) for Ebola Hemorrhagic Fever and Predicting an Ayurvedic Remediation of the Disease. *J Proteomics Bioinform* 2: 287-294.
 73. Prakash N, Patel S, Faldu NJ, Ranjan R, Sudheer DVN (2010) Molecular Docking Studies of Antimalarial Drugs for Malaria. *J Comput Sci Syst Biol* 3: 070-073.
 74. Kumar PM, Raj KK, Ramachandran D, Kumar PMNS, Vaddavalli R, et al. (2010) An Insight into Structural and Functional Characteristics of Plasmodium falciparum Farnesyltransferase (PfFT) 3d7: Comparative Modeling and Docking Studies. *J Proteomics Bioinform* 3: 305-309.
 75. Vamsidhar E, Swamy GV, Chittu S, Babu PA, Venkatasatyanarayana G, et al. (2010) Screening and Docking Studies of 266 Compounds from 7 Plant Sources as Antihypertensive Agents. *J Comput Sci Syst Biol* 3: 016-020.
 76. Daisy P, Sasikala R, Ambika A (2009) Analysis of Binding Properties of Phosphoinositide 3-kinase Through In silico Molecular Docking. *J Proteomics Bioinform* 2: 274-284.