Journal of Computational Intelligence in Bioinformatics
ISSN 0973-385X Volume 5, Number 1 (2012) pp. 13-27

http://www.ripublication.com/jcib.htm

Computational Modeling: A Review -Application to Thioredoxin

D. Premalatha[1], P. Ravindra[2], L. Venkateswar Rao[3] and Jackline Paul[4]

[1&3] Department of Microbiology, University College of Science, Osmania University, Hyderabad – 500 007, Andhra Pradesh, India.
[2]School of Engineering and Information Technology, Universiti Malaysia Sabah, Locked Bag No. 2073, 88999, Kota Kinabalu, Sabah, Malaysia
[4]Department of commerce, Osmania University, Hyderabad – 500 007, Andhra Pradesh, India.

Abstract

The revolution in post genomic era has shifted the scenario from the lab to the laptop. The enormously increasing genome sequencing data can be analyzed by translating it into functional genomics through structural genomics. Protein sequence → structure → Function is a well-known paradigm. The bottlenecks of traditional methods to determine 3D-structure demanded structure prediction methods. The structural details at atomic level help to understand and alter the function of proteins to cure human diseases and to produce industrially important products, the ultimate goal of genomic era. Structure prediction methods are reviewed in this paper with special emphasis on comparative modeling. Modeller a reliable software was used to predict the structure of putative thioredoxin and structural details are discussed.

Keywords: 3d-model, Modeller, Procheck, Verify 3d, Thioredoxin, Fold, Cis-proline.

Introduction

Nowadays the analysis of flooding data for genome sequences is a very big challenge for the experimental scientists. Genomic data by itself is insignificant. The information encoded in genome sequences should be coded and utilized for the upliftment of mankind, which is the ultimate goal of genome sequencing projects.

Now the questions arises: How to crack the genomic data? A genome is a repository of 100s of genes and controls the different processes of an organism through proteins, the messengers of life. Functional characterization of the protein is facilitated by 3d-structure which in turn hidden in primary sequence (1, 2).

Traditional methods used to obtain 3d-structure are X-ray crystallography and NMR-spectroscopy(3, 4).They are slow, costly and unable to cope up with the releasing pace of genome sequencing data(5). Hence, there is demand for the development of fast and efficient computational methods.

Computational methods available to predict the 3d structure are classified as Abinitio modeling, Fold recognition/Threading and Comparative modeling(3, 6-10).

Abinitio modeling

The Abinitio modeling is the most challenging method for the prediction of the 3d-structure as the prediction does not rely on any existing experimentally determined structure(11, 12) instead, it is based on the phenomenon of the protein sequence containing the information for the formation of 3d-structure(1). In abinitio modeling 3d-structure is found by searching for conformational space which is low in free energy for the given sequence(11, 13) by simulating the biological processes depending on the principles of physical and chemical sciences like statistical, thermodynamics and quantum mechanics(14).

Different servers/Programs are available to predict the model using abinitio modeling. For ex:I-TASSER (15), ROBETTA (16), BHAGEERATH (9) etc.

Fold Recognition/Threading

In the absence of significant structural homologue, threading methods sheds light to find the 3d structure of target. As the protein folds are limited in nature (2, 17, 18), a newly sequenced protein is likely to fall among the existing folds. Structural fold adopted by target(new) sequence will be identified by fitting the target sequence into structural database and choosing the best fit score(lowest energy)(19-23)). Following different methods can be followed to identify the fold of the target sequence. A) Environment of the residue (24), B) Pair wise interatomic energy of the residue (25) C) Secondary structure prediction (22)Hybrid method(26.)

Following are the some of the servers/programs available to predict the 3d-structure using fold recognition methods. FFAS(27), PROSPECT (28), 3D-PSSM (29), FUGUE (30), GenTHREADER (26), HHPRED (31) etc.

Comparative modeling

Comparative modeling is based on the concept that similar sequences will adopt similar structure (2, 32-34). Since the first structure prediction (35) many articles have been published on this topic(36-43). A theoretical model is predicted by transferring the backbone information of the similar structure/template. If a protein sequence, targeted with 40 % sequence identity to a homologous structure is

available, it can be automatically modeled equal to low resolution X-ray crystallographic structure or medium NMR structures(44-46).

The process of comparative modeling starts with a search for the Template i.e. homologous structure followed by alignment between target and template, Model Building and Model Evaluation. (15, 44, 47-52).

Template search

In comparative modeling, the quality of the predicted structure depends on template structure. Hence identification of the right structure is the most important step. Template can be identified using pair wise sequence alignment methods like BLAST(53), FASTA(54, 55); multiple sequence alignment methods (56, 57) such as PSI-BLAST (57), Hidden Markov Models (58) etc; and sequence structure comparison methods such as 3D-PSSM (30) FUGUE (31), THREADER (26, 58,) etc. Multiple sequence alignment methods will increase the sensitivity of the search. The sequence structure comparison methods are useful for the identification of distant homologues. Any of the database identifies the structure similar to the target sequence through the sequence comparison.

When a set of homologous structures are available, the choice of template can be based on parameters like sequence identity, similarity between secondary structure of target and template and quality of the template structure (49). More than one template can also be used for model building, in fact, it increases the quality of the model (49, 51, 59).

Alignment between template and target

Alignment between the target and template is the most influential factor for the accurate model. It produces structural equivalences between the target and template residues. When sequence identity between template and target is less, alignment may have gaps and model may have errors (47, 60), but when sequence identity >50%, the model obtained may be good and equal to low resolution X-ray structure(60).

Alignment programs such as CLUSTALX (61), can be used to get alignment. Misalignment of single residue results approximately 4A error in model (44).

Model Building

The methods used to build the Comparative model are:

A) **Rigid body Building** -A model is assembled from the rigid bodies obtained from the aligned protein structures. Protein structures are dissected into structurally conserved regions, loops and side chains and target model is assembled from closely related conserved regions of structures (35, 36).

B) **Modeling by segment matching**-Here a comparative model is constructed based on subset of atomic positions/conserved positions of template structures. These positions are used as guiding positions and all atom segments that fit these conserved

positions are identified and assembled by scanning known protein structures or by conformational search restrained by an energy function(62-64).

C) **Modeling by satisfaction of spatial restraints**-A model is constructed by satisfaction of spatial restraints, which includes Cα-Cα bond length, main chain and side chain dihedral angles, vander waals interactions. The model predicted is with minimized violations of restraints, derived based on Homology and supplemented by stereo chemical restraints and expressed as probability density functions. Restraints or constraints on the structure of sequence are generated from related structures using alignment of corresponding distances between aligned residues of template and target (63, 65).

Finally all above mentioned methods have their own way of determining 3d-model from the given alignment.

Errors

Errors in comparative modeling may be 1) Mistakes in packing of side chains 2) Conformation and shifts of the core segments and loops that will not have an equaling segment from any template structure.3) Improper structure because of wrong template(47).

Comparative models can be built using different softwares /programmes like Modeller (63), SWISSMODEL (66), WHATIF (67), ESYPRED3D (68), CPH(69)etc.

Model evaluation

To evaluate the predicted model quality, two types of evaluations exist, i.e External evaluation and Internal evaluation. The former method identifies the suitability of the template used to build the 3d-model and the latter method identifies the unreliable regions of the model i.e. stereochemistry, bond length, bond angle, dihedral angles, atom-atom overlaps etc(70).Different evaluation programs/servers suchasANOLEA(71), COLORADO3D(72), PROCHECK (73), WHATCHECK(74), PROSAII(75) etc can be used for model evaluation.

The applications of the Homology model depends on the quality of the model. It can be used for A)Identification of ligand binding sites (76, 77), B)Drug designing (78), C) Site directed mutagenesis(52, 79-81).

Thioredoxin

Thioredoxin is a small ubiquitous, redox active protein with vital roles in physiological and biochemical processes (82, 83). It plays an important role in DNA synthesis by providing hydrogen atoms to an essential enzyme ribonucleotide reductase (84).

It is an electron transporter to the enzymes like methionine sulfoxide reductase and sulfate reductase etc(82). It protects, damaged DNA and proteins from Oxidative stress, which causes damage to this molecules, by activating the transcription factors

like NF-kB (85) and AP-1(86). It also activates antioxidant molecules such as glutathione peroxidase and thioredoxin peroxidase to remove H_2O_2 a product of oxidative stress (82) thus playing an important role in limiting the oxidative stress.

Thioredoxin application is not only limited to physiology and biochemical processes but also has its presence in biotechnology and medicine. Allergenic proteins are active in oxidized state. Thioredoxin inactivates the allergenic proteins by reducing their disulfide bonds thus thioredoxin eliminates the allergenicity of foods of plants and animals like soy, milk and wheat Thioredoxin is used for improving the quality of baked products by reducing intramolecular disulphide bonds of flour proteins like glutenins and gliadins and enhancing the formation of intermolecular disulphide bonds. Neurotoxins and other toxic components in venom of snakes, bees and scorpions can be inactivated by thioredoxin (87-89).

In this study, thioredoxin structure was determined from *Streptococcus.pyogenes*, a pathogen, causing different diseases to access the insight into the function of the putative protein.

Identification of the template

Modeling process was initiated with a search for the template using the sequence of putative thioredoxin of *S.pyogenes* mined from Swiss-Prot database (90) as query. Analysis of the results of servers like PDB-BLAST, (a modified version of BLAST)(53); 3D-PSSM, (29) and FUGUE (30), revealed X-ray crystallographic structure of thioredoxin of *E.coli* (PDB identifier 2TRXa) as appropriate template. The atomic co-ordinates of the crystal structure of the thioredoxin of *E.coli* (91) at 1.68Å resolution were accessed from Protein Data Bank (92, 93) files.

Model Building

MODELLER builds the 3D-models based on satisfaction of spatial restraints(63, 65). The input files to the MODELLER are Alignment file in PIR format table.1, prepared using the ClustalX software (61) and co-ordinates of the template.

Table 1: Alignment file in PIR format used to run the MODELLER.

```
>P1; 2trxa
structureX:2trxa: 1:1:108:: A: thioredoxin: Escerichia coli: 1.70:16.5
SDKIIHL--TDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQ—
GKLTVAKLNID-
QNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLA*
>P1;QUERY sequence
--MALEVTDATFVEETKE---GLVLIDFWATWCGPCRMQAPILEQLSQEI-
DEDELKILKMDVDE-NPET
ARQFGIMSIPTLMFKKDGEVVKQVAGVHTKDQLKAIIAELS-*
```

The command " mod trx.top" was used to run the MODELLER(63, 65). Depending on the given option MODELLER generated 20 models which are B00009991.atm to B000099920.atm files.The models were ranked and a model with the lowest score of 362.1932 objective function was selected as the best model Loop building was carried out for the model with SWISS PDB VIEWER(66) for the amino acid residues present in disallowed regions of Ramachandran plot using PROCHECK (73). Final structure was refined by subjecting to limited energy minimization by SYBYL's steepest descent and conjugate gradient algorithms. All computations were carried out on Silicon Graphics workstation using the SYBYL software, operating under IRIX (SYBYL 6.7, Tripos Inc., 1699, South Hanley Rd., St. Louis, Missouri, 63144, USA).

Structural features of Model

Predicted structure uncovered a single polypeptide chain with α-helical and β-sheet secondary structural elements. General 3d structure of thioredoxins i.e trx fold(94) characterized by α β α motif at amino terminal end and β β α motif at carboxy terminal end was observed in predicted structure. Significant difference with the template was not observed in the secondary structural elements. Core of the structure has a single twisted β-sheet with 3 parallel and 2 antiparallel strands flanked by helices. β-sheet is packed by α2, α4 helices on one side and α1, α3 helices on the other side. Content of α-helix is 31.73% and β-strand is 27.88%. Proline 73 is identified as cis proline.

RMSD

The overall root mean square deviation for all C-α atoms 0.52A.confirms the resemblance of generated structure to that of *E.coli* thioredoxin(Template)(90) using SWISS PDBVIEWER software (66).

Stereochemical quality of thioredoxin structure

The ϕ and ψ distributions of Ramachandran plot of non-glycine, non-proline residues of final model are depicted in Fig. No.1 and shown in Tab.No.2. The stereochemical parameters for the backbone angle of amino acids of the model tested by the PROCHECK (73)is identified as excellent due to the presence of 91.4% residues in most favoured and 7.5% in additionally allowed regions..

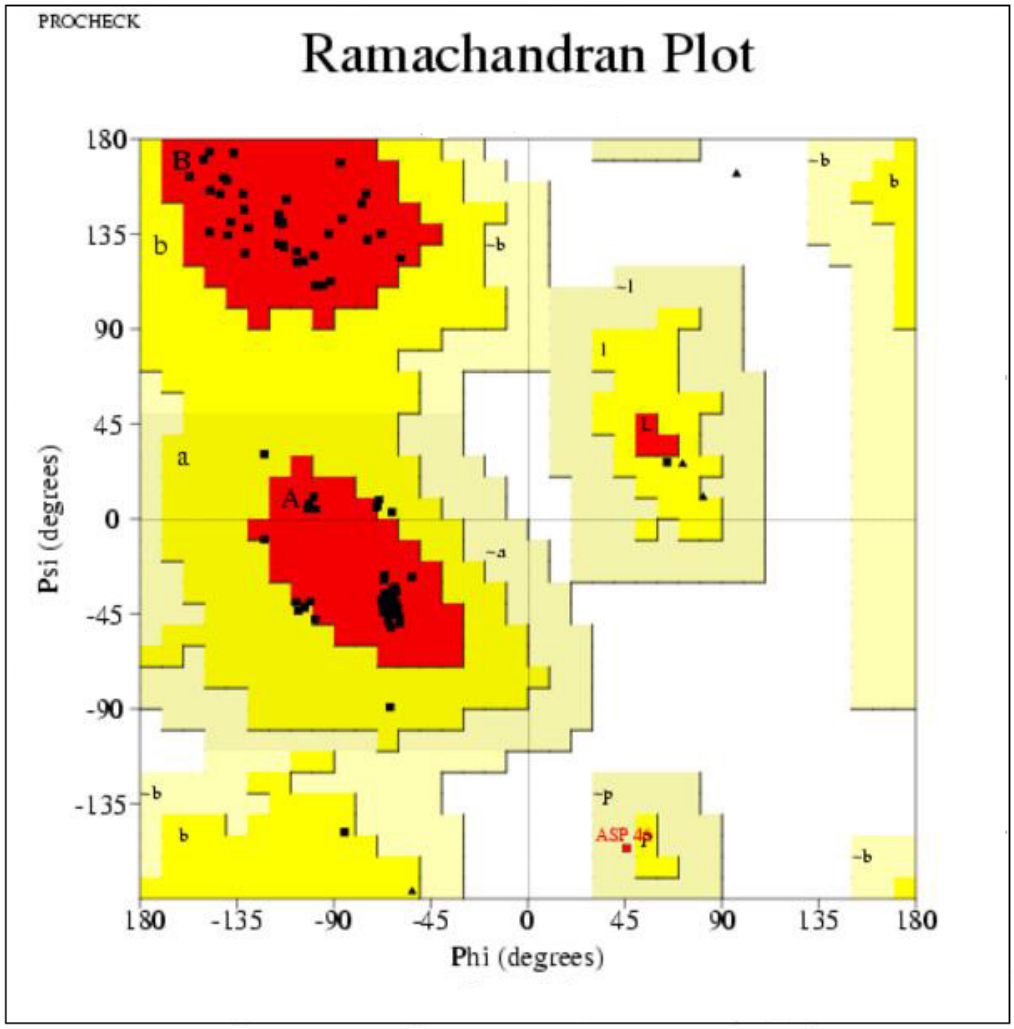

Figure 1: Computed Procheck picture of Ramachandran plot for thioredoxin.

Table 2: Ramachandran plot calculations for thioredoxin using Procheck program after energy minimization

Amino acid residues in most favored regions	91.4%,
Additional allowed regions	7.5%
Generously allowed regions	1.1%
Disallowed regions	0.00
Non-glycine and non-proline residues	100.0%.

Verify 3d

The profile score generated by plotting the verify-3d (24) 0.61 indicating the model as a reasonable good model.

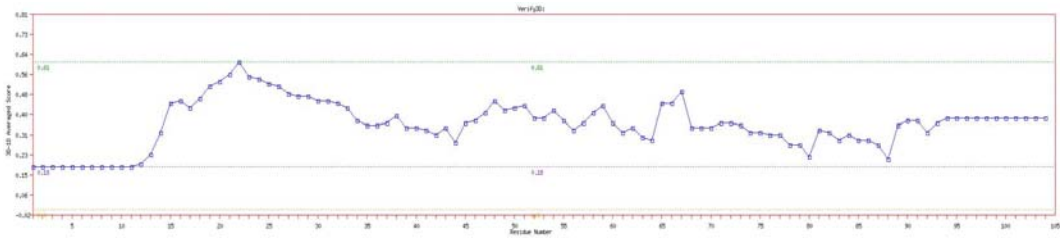

Figure 2: The 3d-profile score generated from coordinates of thioredoxin of *S.pyogenes*.

Site id

Q site finder program(95) was used to identify the binding sites on thioredoxin protein.Results were displayed in table.3.Binding sites are arranged based on their energetic levels i.e favorable sites are arranged in decreasing order.

Table 3

Predicted site	Site Volume	Amino acids present
Site1	207 Cubic angstroms	Asp7, Ala8, Phe10, Val 11, Lys 15, Glu 62, Thr 63, Gln 66, Phe 67, Lys 78, Gly 81.
Site2	164 Cubic angstroms	Met 1, Ala2, Ser 42, Gln 43, Ile 45, Asp 46, Glu 47, Leu 50, Lys 51, Ile 52.
Site3	129 Cubic angstroms	Met 1, Asp22, trp24, Ala 25, Lys 31, , Arg 32, Gln 34, Ala 35, Leu 38, Lys 54.
Site4	106Cubic angstroms	Glu 13, Thr 14, Glu16, Gly17, Leu18, val 19, Glu 47, Asp48, Glu 49, Leu 50, Lys 51, Leu 53.
Site5	76 Cubic angstroms	Phe 67, Gly 68, Ile 69, Met 70, Ser 71, Thr 74, Gln 86, Ala 88.
Site6	63 Cubic angstroms	Met 1, Ala2, Leu 3, Glu 4, Val 5, Trp 24, Lys 54, Met 55, Asp 56.
Site7	84 Cubic angstroms	Trp 27, Val 57, Ala 64, Ile 69, Met 70, Ser 71, Ile 72.
Site8	76 Cubic angstroms	Leu 41, Ile 45, Glu 49, Leu 50, Lys 97, Ile 100. Ala 101, Ser 104.
Site9	70 Cubic angstroms	Lys 15, Glu 16, Gly 17, Leu 18, Lys 79, Asp 80, Gly 81.
Site10	63 Cubic angstroms	Trp 71, Cys 28, Gly 29, Pro 30, Ser 71, Ile 72, Pro 73.

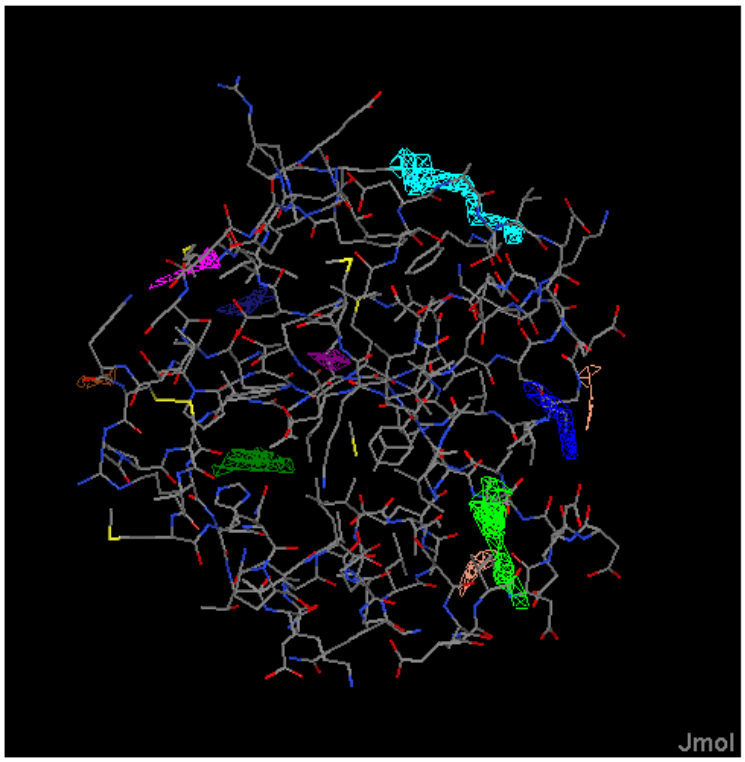

Figure 3: Binding sites of thioredoxin 3D-structure.

Conclusion

The first insight into the properties of this protein has become accessible as the predicted structure accedes with the reference structure used for modeling. Further studies are necessary to elucidate and enhance the role of this protein.

References

[1] Christian, B. Anfinsen, 1973, “Principles that Govern the Folding of Protein Chains, " Science, 181, pp. 223-230.

[2] Cyrus Chothia, and Arthur, M. Lesk, 1986, "The relation between the divergence of sequence and structure in proteins, " The EMBO Journal, 5 No.4, pp. 823-826.

[3] Mathew L. Baker, Wen Jiang, William J. Wedmeyer, Frazer J. Rixon, David Baker, and Wah Chiu, 2006, “Abinitio Modeling of the Herpesvirus VP26 Core Domain Assessed by CryoEM Density, ” Plos Computational Biology, Volume 2 Issue 10, e 14.

[4] Roaa I.Mubark, Hesham A.Keshk, and Mohamed I.Eladawy, 2008, ”Different species and proteins classifiers and protein’s structure predictors systems, ”

International Journal of Biology and Biomedical Engineering, Vol 2, pp. 119-128.

[5] A.Ozlem Tastan Bishop, Tjaart A.P.deBeer, and Fourie Joubert, 2008, "Protein homology modeling and its use in South Africa, " South African Journal of Science, 104, pp. 2-6.

[6] Tom Defay, and Fred E.Cohen, 1995, "Evaluation of current techniques for abinitio protein structure prediction, " Proteins: Str Fun and Genetics, 23, pp. 431-445.

[7] Daisuke Kihara, HuiLu, Andrez Kolinski, and Jeffrey Skolnick, 2001, "TOUCHSTONE; "An ab initio protein structure prediction method that uses threading-based tertiary restraints, " PNAS, Vol. 98, pp. 10125-10130.

[8] Mahesh Shah., Sergei passovets, Dongsup kim, Kyle Ellrott, Liwang, Inna vokler, Philip locascio, Dong xu and Ying xu, 2003, "A computational pipeline for protein structure prediction and analysis at genome scale, " Bioinformatics, Vol. 19, pp. 1985-1996.

[9] Jayaram., Kumkum bhushan., Sandhya R. Shenoy., Pooja narang., Surojit Bose., Praveen Agrawal., Debasish sahu., and Vidhu Pandey., 2006, "Bhageerath: an energy based web enabled computer software suite for limiting the search space of tertiary structures of small globular proteins" Nucl.Acids, V 34 (21), p6195-6204.

[10] Sofia Khan., and Mauno Vihinen., 2009, "Evaluation of accuracy and applicability of protein models:retrospective analysis of biological and biomedical predictions", Insilico Biology 9., 0025.

[11] Bonneau, R., and Baker, D., 2001, "Ab initio structure prediction:progress and Prospects".Annu.Rev.Biophy. Biomol.Struct., 30, pp.173-189.

[12] Sitao, Wu., Jeffrey Skolnick., and Yang Zhang., 2007, " Ab initio modeling of small proteins by iterative TASSER Simulations, " BMC. Biology., 5:17.

[13] Klepeis, J.L., and Floudas, C. A., 2003, "ASTRO-FOLD: A Combinatorial and Global Optimization Framework forAb Initio Prediction of Three-Dimensional Structures of Proteins from the Amino Acid Sequence, " Biophy. J., 85, pp.2119–2146.

[14] Krzysztof Ginalski.., Nick, V. Grishin., Adam Godzik., and Leszek Rychlewski., 2005, "Practical lessons from protein structure prediction, " Nucleic Acids Research., 33(6), pp.1874-1891.

[15] Yang Zhang., 2008, "I-TASSER server for protein 3D structure prediction"BMC Bioinformatics, 9:40.

[16] Kim, D.E., Chivian, D, Baker. D., 2004"Protein structure prediction and analysis using the Robetta server, " Nucleic Acids Res., 32, W526-W531.

[17] Chothia, C., 1992, "One thousand protein families for the molecular biologist".Nature., 357, pp.543-544.

[18] Wang, Z.X., 1996, "How many fold types of protein are there in nature?, "Proteins., 26(2), pp.186-191.

[19] Mario, Albrecht., Daniel, Hanisch., Ralf Zimmer., and Thomas Lengauer., 2002, "Improving fold recognition of protein threading by experimental distance constraints, "In Silico Biology 2, 0030, Bioinformation Systems e.V.

[20] Jones, D.T., 1997, “Progress in protein structure prediction, ” Curr.Opin.Struct.Biol.., 7, pp.37-387.
[21] Zhang, B., Jaroszewski., L., Rychlewski., L., and Godzik, A., 1997, ”Similarities and differences between non homologous proteins with similar folds:Evaluation of threading strategies, ”Fold Des., 2, pp.307-317.
[22] Rost, B., Schneider, R.., and Sander, C., 1997, ”Protein fold recognition by prediction based threading, ”.J.Mol.Biol., 270, pp.471-480.
[23] Sternberg, M.J.E., Bates, P.A., Kelley, L.A.., and Maccallum, R..M., 1999, ”Progress in Protein structure prediction: assessment of CASP3, ” Curr.Opin.Struct.Biol., 9, pp.368-373.
[24] Bowie, J.U., Luthy, R., Eisenberg, D., 1991, ”A method to identify protein sequences that fold into a known three-dimensional structure, ”Science., 253, pp, 164-170.
[25] Jones, D.T., Taylor, W.R.., and Thornton, J.M.., 1992, ” A new approach to protein fold recognition, “ Nature., 358, pp.86-89.
[26] David, T. Jones., 1999, “GenTHREADER: An Efficient and Reliable Protein Fold Recognition Method for Genomic Sequences, ” J. Mol. Biol., 287, pp.797±815.
[27] Rychlewski. L., Jaroszewski, L., Li, W., and Godzik, A., 2000, ” Comparision of sequence profiles.Strategies for structural predictions using sequence information, ”Protein Sci., 9, pp. 232-241.
[28] Xu, Y., and Xu, D., 2000, ” Protein threading using PROSPECT:design and evaluation.” Protein., 40 (3), pp.343-354.
[29] Kelley, Lawrence.A., MacCallum.R.M., and. Sternberg, M.J., 2000., “Enhanced genome annotation using structural profiles in the program 3D-PSSM, ” J. Mol. Biol., 299, pp.499-520.
[30] Jiye Shi., Tom L. Blundell., and Kenji Mizuguchi., 2001, ” Fugue: Sequence structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties, ” J. Mol. Biol., 310, pp.243-257.
[31] Soding, J., Biegert, A., and Lupas, A.N., 2005, ”The HHpred interactive server for protein homology detection and structure prediction, ” Nucleic Acids Research., 33, pp.W244—W248.
[32] Hubbard, T.J.P., &Blundell, T.L., 1987, ”Comparision of solvent inaccessible cores of homologous proteins:definitions useful for protein modeling, ”.Protein Eng., 1, pp.159-171.
[33] Andrej, Sali., and John overington., 1994, ”Derivation of rules for comparative protein modeling from a database of protein structure alignments, ”Protein science., 3, pp.1582-1596.
[34] Krzysztof, Ginalski., 2006, ”Comparative modeling for protein structure prediction, ”Current Opinion in Structural Biology., 16, pp.172-177.
[35] Browne, W. J., North, A. C. T., Phillips, D. C., Brew, K., Vanaman, T. C., & Hill, R. C., 1969, ”A possible three-dimensional structure of bovine -lactalbumin based on that of hen’s egg-white lysozyme, ” J.Mol. Biol., 42, pp.65–86.

[36] Greer, J., 1981, " Comparative model-building of the mammalian serine proteases, " J.Mol.Biol., 153, pp.1027-1042.

[37] Chyh-Chong Chuang, , Shih-Hsiung Wu, , Shyh-Horng Chiou., and Gu-Gang Chang., 1999, "Homology Modeling of Cephalopod Lens S-Crystallin: A Natural Mutantof Sigma-Class Glutathione Transferase with DiminishedEndogenous Activity, ". Biophysical Journal.., 76, pp.679–690.

[38] Johnson, M. S., Srinivasan, N., Sowdhamini, R., Blundell, T.L., 1994, "Knowledge based protein modeling, ".CRC.Crit.Rev. Biochem.Mol.Biol., 29, .pp.1-68.

[39] Floppe, N., Ferrand, M., Breton, J., & Smith, J.C., 1995, "Structural model of the Photosynthetic reaction centre of Rhodobacter capsulatus, "Proteins., 22(3)., pp.226-244.

[40] Larry Cosenza., Andrew Rosenbach., James, V.White., John R.Murphy and Temple Smith., 2000, " Comparative model building of interleukin-7 using interleukin-4 as a template: a structural hypothesis that displays a typical surface chemistry in helix D important for receptor activation, "Protein science., 9, pp.916-926.

[41] Choubey Jyotsna., Patel Ashish., Khatri Sheetal., Dewangal Ruby., Shailendra Kumar Gupta., Verma., M.K, 2009, "Homology modeling of Hypoxanthine-guanine Phosphoribosyl transferase, enzyme involved in salvage pathway of purine metabolism, " J.of Comp. Sci &Syst.Bioology, 2(5), pp.259-261.

[42] Raj, K. Prasada., Rajesh Sharma., and G. L. Prajapati., 2010, " Homology modeling and evaluation of human TEK tyrosine kinase using SWISS-MODEL Workspace, "Chem. Pharm. Res., 2(2), pp.440-451

[43] VinodRishi., AshishPatel., JyotsnaChoubey, M.K., Verma, A.K.Awasthi., 2010, "Construction of 3d model for cyclobutane pyrimidine(CPD) phytolase from Chlorella pyrenoidosa, "International Journal of Engineering Science and Technology., .2 (8), pp.3816-3823.

[44] Roberto Sanchez., Sali, A., 1997, "Comparative protein structure modeling as an optimization problem, "Jrnl of Molecular structure(Theochem)., 398-399, pp. 489-496.

[45] Bino John., and Sali, A., 2003, "Comparative protein structure modeling by iterative alignment, model building and model assessment, " Nucl.Acids. Res., 31, pp.3982-3992.

[46] Zhexin Xiang., 2006, "Advances in Homology Protein Structure Modeling, " Current Protein Pept. Sci.., 7(3), pp.217-227.

[47] Marti-Renom, M.A., Stuart, A., Fiser, A., Sanchez, R.., Melo, F., and Sali, A., 2000 "Comparative protein structure modeling of genes and genomes, "Annu Rev Biophys. Biomol.Struct., 29, pp.291-325.

[48] Eswar, N., John., B., Mirkovic., N., Fiser, A., Ilyin, VA., Pieper, U., Stuart, AC., Marti-Renom, MA., . Madhusudhan, M. S., Yerkovich, B., and Sali, A., 2003, " Tools for comparative protein structure modeling and analysis, " Nucleic Acids Research., 31(13)., pp. 3375–3380.

[49] Centeno, NB., , Planas-Iglesias, J., and Oliva, B., 2005, "Comparative modeling of protein structure and its impact on microbial cell factories, " Microbial Cell Factories., 4:20 doi:10.1186/1475-2859-4-20.

[50] Ursula Pieper., Narayanan Eswar., Ashley, C stuart., Valentin A.IIyin., and Sali, A., 2002, "ModBase, a database of annotated comparative protein structure models, " Nucl.Acids. Res., 30(1), pp.255-259.

[51] Fernandez –Fuentes., N., .Rai., B.K., .Madrid-Aliste., C.J., J Fajardo., J.E., and Fiser, A., 2007, "Comparative protein structure modeling by combining multiple templates and optimizing sequence –to-structure alignments, " Bioinformatics V., 23(19), pp.2558-2565.

[52] Bordoli., L., Kiefer., F., Arnold., K., Benkert, .P., , Battey, J, ., Schwede., J., . 2009, "Protein structure homology modeling using SWISS-MODEL workspace" Nature Protocols., 4, (1), pp.1 – (13).

[53] Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. 1990. "Basic local alignment search tool, ".J Mol. Biol., 215(3), pp.403–410.

[54] Pearson, W. R., Lipman, D.J., 1988, "Improved tools for biological sequence comparison, " Proc. Natl.Acad.Sci. U S A., 85, pp.2444-2448.

[55] Pearson, W.R., 2000, "Flexible sequence similarity searching with the FASTA3 program package, ". Methods.Mol. Biol., 132, pp.185-219.

[56] Henikoff, J.G., Pietrovski, S., .McCallum, C.M., Henikoff, S., 2000, "Bolcks based methods for detecting protein Homology, " Electrophoresis., 21(19), pp. 1700-1706.

[57] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D.J., 1997, "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, "Nucl.Acids.Res., .25, pp.3389-3402.

[58] Eddy, S.R.., 1996, "Hidden Markov Models, "Curr.Opin.Struct.Biol., 6, pp.361-65.

[59] Srinivasan, N., Blundell, T.L., 1993, "An evaluation of the performance of an automated procedure for comparative modeling of protein tertiary structure, ".Protein Eng., 6, pp.501-512.

[60] Kopp, J., Schwede, T., 2004., "Automated protein structure homology modeling:A progress report, " Pharmacogenomics. J., 5(4):, pp.405-416.

[61] Thomson, J.D., .. Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G, 1997, "The CLUSTAL_X windows interface:flexible strategies for multiple sequence alignment aided by quality analysis tools"". Nucleic Acids Res, 25(24) pp.4876-82.

[62] Jones, T.A., Thirup, .S., 1986, "Using known substructures in protein model building and crystallography, " EMBO Journal., 4, pp.819-822.

[63] Sali, A., & Blundell, T. L. 1993, " Comparative protein modeling by satisfaction of spatial restraints, ". J.Mol. Biol., 234, pp.779-815.

[64] Chou.K.C., 2005, "Modeling the tertiary structure of human cathepsin-E, " Biochemical and Biophysical Research Communications., 331(1), pp.56-60.

[65] Andrej Sali., Liz Potterton., Feng Yuan., Herman van Vlijmen, and Martin Karplus., . 1995, . "Evaluation of Comparative Protein Modelling by

MODELLER, ". PROTEINS: Structure, Functions, and Bioinformatics., 23, pp.318-326.

[66] Guex., N., and Peitsch, M.C., 1997., ".SWISS-MODEL and the SWISSPDBVIEWER "an environment for comparative protein modeling, " Electrophoresis., 18, pp.2714-2723.

[67] Vriend, G., 1990, "WHAT IF: a molecular modeling and drug design program, " J.Mol. Graph., 8(1), pp. 52-56.

[68] Lambert, C., Leonard N, , Debolle X, ., Depiereux E., 2002, "Esypred 3D:Prediction of proteins 3Dstructures, "Bioinformatics., 18(9), pp.1250-1256.

[69] Nielsen., Lundegard, C., Lund O., Petersen, T.N, 2010" CPH Models.Remote homology modeling using structure guided sequence profile"Nucl.Acids Res. 38, W576-W581

[70] Sippl, M.J., 1993, "Recognition of errors in three dimensional structure of proteins, " Proteins, 17(4), pp.355-62.

[71] Melo, F., devos, D., Depiereux, E., Feytmans., E., 1997, "ANOLEA: a www server to assess protein structures, " Proc.Int.Conf. Intell.syst.Mol Biol, 5, pp.1870-90.

[72] Sasin, J.M., and Bujnicki, J.M., 2004, "COLORADO3D, a web server for the visual analysis of protein structures, "Nucleic Acids Res., 32, W586-589.

[73] Laskowski, R.A., Mac Arthur, M.W., Moss, D.S., Thornton, J.M., 1993, "PROCHECK: a program to check the stereochemical quality of protein structures., "J.Appl. Crystallogr. 26, pp.283-291.

[74] Hooft, R.W., Vriend, G., Sander, C., Abola, E. E., 1996, "Errors in protein structures, "Nature, 381, pp.272-272.

[75] Markus Wiederstein., and Manfred J.Sippl., 2007, " PROSA-web:Interactive web service for the recognition of errors in three dimensional structures of proteins, " Nucleic Acids res., 35, pp.W407-W410.

[76] Kanthi, B.K., Liou, K, Sohng, J.K., 2010, "Homology modeling, binding site identification and docking in flavone hydroxylase CYP105P2 in streptomyces peucetius ATCC 27952." Comput.boil.Chem, 34(4), pp. 226-31.

[77] Evers, A., Gohlke, H and Klebe, G., 2003, "Ligand supported homology modeling of protein binding sites using knowledge based potentials" J.Mol.Biol., 334(2), pp.327-345.

[78] Petray, D., and Honig, B., 2005. "Protein structure prediction:In roads to Biology".Mol cell., 20(6), pp.811-819,

[79] Chmiel, A.A., Radlinska, M.., .Pawlak, SD., Krowarsch, D., Bujnicki, J.M., and Skowronek, K.J., 2005, " a theoretical model of restriction endonuclease NlaIV in complex with DNA, predicted by fold recognition and validated by site directed mutagenesis and circular dichroism spectroscopy" Protein Eng Des and Sel., 18 (4), pp.181-189.

[80] Wells, G.A., Birkholtz, L.M., Joubert, F., Walter, R.D., Louw, A.I., 2006 "Novel properties of malarial s-adenosylmethionine decarboxylase as revealed by structural modeling, "J.Mol.Graph Model., 24(4), pp.307-318.

[81] Nakonieczna, J., Kaczorowski, T., Obarska Kosinka, A., and Bunicki, J.M., 2009, "Functional analysis of MmeI from methanol utilizer Methylophilus

Methylotrophus, a subtype IIC restriction-modification enzyme Related to type I enzymes, ”Appl and Env.Microbiol., 75(1), pp.212-223.

[82] Holmgren, A., 1985 “ Thioredoxin., ”. Annu. Rev. Biochem., 54, pp.237-271.

[83] Arner, E.S., and Holmgren, A., 2000, “Physiological functions of thioredoxin and thioredoxin reductase, ”.Eur.J.Biochem. 267(20), pp.6102-6109.

[84] Holmgren, A., 1989, ”Thioredoxin and glutaredoxinsystems, ”.J.Biol.Chem., 264, pp.13963-13966.

[85] Hirota, K., Murata, M., Sachi, Y., Nakamura, H., Takeuchi, J., Mori, K., and Yadoi, J., 1985, “Distinct Roles of Thioredoxins in the cytoplasm and in the Nucleus”. J. Biol. Chem., 274(39), pp.27891-7.

[86] Schenk, H., Klein, M., Erdbrugger, W., Droge, W., & Schulze-Osthoff, .K., 1994, ”Distinct effects of thioredoxin and antioxidants on the activation of transcription factors NF-kappa B and AP-1, ” PNAS.USA 91(5), pp.1672-6

[87] Buchanan, B.B., Schurman, P., Decottignies, P., and Lozano, R.M., 1994, Perspectives in Biochemistry and Biophysics- “Thioredoxin: A Multifunctional regulatory Protein with a Bright Future in Technology and Medicine”. Arch. Biochem. Biophys., 314(2), pp. 257-260.

[88] Lozano R. M., . Yee, B. C., and Buchanan, B.B., 1994, “Thioredoxin linked reductive inactivation of venom neurotoxins”. Arch. Biochem.Biophys. 309(2), pp.356-362.

[89] Besse, Isabelle and Buchanan, B.B, 1997, “Thioredoxin linked plant and animal processes: the new generation”. Bot. Bull. Acad. Sin., 38, pp.1-11.

[90] Bairoch, A., and Apweiler, R., 1997, “The SWISSPROT Protein sequence data bank and its supplement TREMBL, ”. Nucl. Acids. Res., 25(1), pp.31-36.

[91] Katti, S.K., LeMaster, D.M., Eklund, H., 1990, ”Crystal structure of thioredoxin from Escherichia coli at 1.68 A resolution, ” J.Mol.Biol., 212(1): pp.167-184.

[92] Bernstein, F C., Koetzle, T. F., Williams, G. J., Meyer, E. F.Jr., Brice, M. D, Rogers, J.R., Kennard, O., Shimanouchi, T & Tasumi, M., 1977, The Protein Data Bank: a computer based archive file for macromolecular structures, ” J.Mol. Biol., 112(3), pp. 535-542.

[93] Berman, H.M., Westbrook, J., Feng ZGilliland, G., Bhat, T.N., Weissiq, H., Shindyaloy, N., Bourne, P.E., .2000, ”The Protein Data Bank, “.Nucl.Acids. Res.28(1), pp.235-242.

[94] Martin, J..L., 1995, ”Thioredoxin a fold for all reasons, “ Structure, 3 pp.245-250.

[95] Laurie, A.T., Jackson, R.M., 2005., ”Q-site finder: an energy-based method for the prediction of protein –ligand binding site, ” Bioinformatics, 21(9), pp.1908-16.