

Z-SCORE FOR HOMOLOGY MODELING AND STRUCTURE PREDICTION OF GLOBULAR PROTEIN OF HOMOSAPIENS & PLANT

Vipan Kumar Sohpal¹, Apurba Dey², Amarpal Singh³

¹Department of Chemical & Bio Technology, Beant College of Engineering & Tech, Gurdaspur Punjab, India,

² Department of Bio Technology, National Institute of Technology, Durgapur West Bengal, India,

³ Department of Electronics & Communication Engg, Beant College of Engineering & Tech, Gurdaspur Punjab, India.

ABSTRACT

The Swiss model is a web-based integrated protein structure homology modelling software. The main objective of this paper is to compare the homosapiens and plant globular protein using modeling software. Here we focus on the pairwise energy, solvation energies and C- β interaction among the proteins based on Z score. From the simulation, it observes that Myoglobin and Neuroglobin pseudo-properties on Z scale are similar. The quality of model of proteins also evaluated and its higher values indicate that close relationship among the plant and human proteins. The quality of model for homosapiens hemoglobin and Leg hemoglobin also found around parallel.

Keywords: *Swiss-Model, Homology Modeling, Homosapiens and Plant Globular Protein and Quality of Model*

I INTRODUCTION

The role of computational biology in protein structure prediction is now more prominent, because abundant amounts of protein sequence data and control of experimentally determined protein structures. There are three computational approaches to protein three-dimensional structural modeling and prediction. They are homology modeling, threading, and ab initio prediction. The first two are knowledge-based methods; they predict protein structures based on knowledge of existing protein structural information in databases. Homology modeling builds an atomic model based on an experimentally determined structure that closely relating at the sequence level. Threading identifies proteins that are structurally similar, with or without detectable sequence similarities. The ab initio method is simulation basis and predicts structures based on physicochemical principles governing protein folding without the

use of structural templates. Homology modeling, which is the most accurate in all the prediction approach, derives models from close homologs.

Homology modeling predicts protein structures based on sequence homology virtue of that it is useful as comparative modeling. The concept behind it is that if two proteins share a high sequence similarity, they are likely to have similar 3-D structures. Homology modeling produces an all-atom model based on alignment with template proteins. The overall homology modeling approach consists of six steps. The first step is template selection, which involves identification of homologous sequences in the protein structure database to be used as templates for modeling. The next step is alignment of the target and template sequences. The third step is to create a framework structure for the target protein consisting of main chain atoms. The fourth stage of model building includes the addition and optimization of side chain atoms and loops. The fifth step is to improve and optimize the entire model according to energy criteria. The final step involves evaluating of the overall quality of the model obtained. If necessary, alignment and model building repeat until a satisfactory result, achieve. A number of comprehensive modeling programs is able to perform the complete procedure of homology modeling in an automated fashion. The automation requires assembling a pipeline that includes target selection, alignment, model generation, and model evaluation. Some freely available protein modeling programs and servers is Modeller, Swiss-Model and 3D-Jigsaw. Swiss-Model is protein server that allows a user to submit a sequence and get a structure automatically. The server constructs a model by automatic alignment (First Approach mode) or manual alignment (Optimize mode). In the First Approach mode, the user provides sequence input for modeling. The server performs alignment of the query with sequences in pdb using Blast. After selection of suitable templates, a raw model generates. Refinement of the structure is using GROMOS. Alternatively, the user can specify or upload structures as templates. The final models being achieve to the user for further analysis. In the Optimize mode, the user constructs a sequence alignment in Swiss pdb Viewer and submits it to the server for model construction.

II LITERATURE REVIEW

Currently about 20,000 experimental protein structures are deposited in the Protein Data Bank (PDB) [1]. Therefore, it is a clear need to overcome this structure knowledge gap' and computational methods for protein structure prediction have gained much attention in recent years. They investigated the relationship of an amino acid sequence to its tertiary structure and identify homologous segments, which have homologous native conformations in proteins. They worked on research and prediction of the packing of α -helices against beta-sheet in the tertiary structure of globular proteins [2]. Researcher worked on the distance geometry procedure for computing the tertiary structure of globular proteins [3]. Scientist performed the sequence alignment to predict protein structure and depicted from their evaluation results that non homologous proteins are used for structure prediction without reasonable solutions [4]. Group of bioinformatics persons worked on simulated software Swiss model which is a framework for comparative protein modeling. They provided the Swiss model repository, a database containing

more than 3500 automatically generated protein [5]. They predicted the unknown structures of five globular proteins on the basis of global optimization of a potential energy function [6]. Biotechnologist evaluated the accuracy of predicted models for structure prediction using MaxSub integrated with CASP3 [7]. They evaluated myoglobin using automated VIDAS assessment and found that primary myocardial infection caused by myoglobin [8]. They performed wet lab experiment for cloning and sequence analysis of porcine myoglobin c-DNA [9]. They showed that the changes in α -helix content were influenced by the initial helical content and the dipole moment [10]. Group of resource persons provided automated tools for protein structure prediction and analysis, which used for comparative modeling or structure prediction methods [11]. They analyzed the solvent model to stabilize the native structure of a single protein based on free energy force field [12]. They observed the diffusion of small ligand through the matrix and found that local cavities tracing a pathway for the diffusion [13]. Group of Indian people analyzed the geometrical models for computing the tertiary structure of globular from the primary structure [14]. They worked on three different types of hemoglobin occur in plants and found that globin is porous, providing tunnels that may assist in ligand binding [15]. Swiss scientist developed SWISS-MODEL and database of 3D protein structure models generated by the Swiss model homology-modelling pipeline [16]. They introduced a standalone pipeline, allows users to assign their own computing assets to process a potentially infinite number of queries [17]. Swiss person estimated of protein model quality by combining a composite scoring function with structural density data. They improved the model selection using a composite scoring function operating on single models. It improves the quality models which are subsequently used to calculate the structural consensus. They also worked on the estimation of the absolute quality of individual protein and extent of nativness. [18] and [19]. They focused on Swiss model workspace that is web-based integrated service dedicated to protein structure homology modeling [20].

III MATERIALS AND METHODS

The primary sequences of the globular proteins retrieve from the National center for Biotechnology information, public domain for protein and DNA database. The proteins sequences retrieve ([http:// www.ncbi.nlm.nih.gov/fasta](http://www.ncbi.nlm.nih.gov/fasta)) in fasta format for further analysis. Table 1 depicts, the sequences use for work in fasta format.

```
>sp|Q9NPG2|NGB_HUMAN Neuroglobin OS=Homo
sapiens GN=NGB PE=1 SV=1
MERPEPELIRQSWRAVSRSPLEHGTVLFARLFALEP
DLLPLFQYNCRQFSSPEDCLSSPEFLDHIRKVMLVI
DAVTNVEDLSSLEEYLASLGRKHRAVGKLSST
VGESLLYMLEKCLGPAFTPATRAAWSQLYGAVVQ
AMSRGWDGE
>gi|4504349|ref|NP_000509.1| hemoglobin subunit beta
[Homo sapiens]
MVHLTPEEKSAVTALWGKVVNDEVGGEALGRLL
```

```

VVYPWTQRFESFGDLSTPDAVMGNPKVKAHGKK
VLGAFSDGLAHLNLDNLKGTFFATLSELHCDKLHVDPE
NFRLLGNVLVCVLAHHFGKEFTPPVQAAAYQKVVA
GVANALAHKYH
>sp|P02144|MYG_HUMAN Myoglobin OS=Homo
sapiens GN=MB PE=1 SV=2
MGLSDGEWQLVLNVWGKVEADIPGHGQEVLRIRLF
KGHPETLEKFDKFKHLKSEDEMKASEDLKKHGAT
VLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVK
YLEFISECIIQVLQSKHPGDFGADAQGAMNKALELF
RKDMASNYKELGFQG
>sp|P02235|LGB1_SOYBN Leg hemoglobin C1
OS=Glycine max PE=2 SV=2
MGAFTEKQEALVSSSFEAFKANIPQYSVVFYNSILE
KAPAAKDLFSFLANGVDPTNPKLTGHAEKLFALVR
DSAGQLKTNGTVVADAALVSIHAQKAVTDPQFVV
VKEALLKTIKEAVGGNWSDELSSAWEVAYDELAA
AIKKA
>gi|13096683|pdb|1D8U|A Chain A, Crystal Structure Of
Non-Symbiotic Plant Hemoglobin From Rice
MALVEDNNAVAVSFSEEQEALVLKSWAILKKDSA
NIALRFFLKIFEVAPSASQMFSFLRNSDVPLEKNPK
LKTHAMSVFVMTCEAAAQLRKAGKVTVRDITLK
RLGATHLKYGVGDAHFEVVKFALLDTIKEEVPAD
MWSPAMKSAWSEAYDHLVAAIKQEMKPAE

```

TABLE 1: Fasta format of Neuroglobin, Hemoglobin and Myoglobin protein sequences of homosapiens, Leg hemoglobin (Soybean) and Plant Hemoglobin (rice).

IV RESULTS AND DISCUSSION

The pseudo-energies of the C_{beta} Interaction, Energy All-atom pairwise Energy Solvation Energy Torsion Angle Energy are below together with their Z-scores in Table 2 and Table 3. Neuroglobin and Myoglobin have Z-score -0.31 while the counterpart of that hemoglobin has Z score comparatively less (-0.08). Contrary to human globular protein plant globular protein maintain different Z score. Rice hemoglobin has -1.19 and leg hemoglobin show 0.03 Z score.

The QMEAN-Z composite scoring function of C_{beta} interaction energy All-atom pairwise Energy, Solvation Energy, Torsion Angle Energy. The combination of these terms helped to develop model selection. C-beta interaction is the basis on optimization of a simplified energy function of a peptide backbone. The energy is the backbone of interaction between peptide groups and function of the local interactions within all amino acid residues. Hemoglobin has positive C-β interaction while other homosapiens and plant globular proteins have

negative interaction energy. It signifies that hemoglobin demonstrate the capacity of local interaction with their residual. Table 2 and 3 reflects that human globular C- β interaction is more rule than plant globular protein. The hierarchy of C- β interaction is as hemoglobin> Neuroglobin>Myoglobin>Leg hemoglobin>Rice hemoglobin. Rice hemoglobin has least C- β interaction energy with their residual and consequently the stability in respect of other proteins is less.

TABLE 2: The pseudo-energies of the contributing terms of homosapiens proteins using Swiss model

Protein	Z-score	C_beta Interaction Energy	All-atom pairwise Energy	Solvation Energy	Torsion Angle Energy
Hemoglobin	-0.08	0.59	0.23	0.82	-0.66
Neuroglobin	-0.31	-0.10	0.09	1.21	-1.14
Myoglobin	-0.31	-0.10	0.09	1.21	-1.14

TABLE 3: The pseudo-energies of the contributing terms of plants proteins using Swiss model

Protein	Z-score	C_beta Interaction Energy	All-atom pairwise Energy	Solvation Energy	Torsion Angle Energy
Leg Hemoglobin	0.03	-0.35	0.32	-0.15	-0.22
Rice Hemoglobin	-	-0.69	0.09	-0.99	-1.05
Hemoglobin	1.19				

Inter residue interactions control the structural stability of proteins. Amino acid pairing frequencies produce higher interaction between proteins arises and hence more stable proteins formation takes place. However, the summation of energies is not suitable for intrinsically unstructured proteins. To overcome these limitations, Swiss model present a novel method for estimating the total pair wise interaction energy. Myoglobin, Neuroglobin and rice hemoglobin, have similar pair wise interaction energy 0.09, while leg hemoglobin has 0.32 and hemoglobin resides on 0.23. On the basis of Swiss result, it observes that plant proteins have higher inter residue interaction as compare to human proteins. Hemoglobin protein also shows the strong support for inter residue interaction succeeding the stability is higher.

The stability of protein structures in solvent is the function of their atomic coordinates. The solvation free energy is the product of the atomic solvation parameter and of the atom accessibility to solvent. The solvation energy of the peptide backbone is calculated using solvation parameters computed from the side chain solvation energies due to lack of experimental data. Solvation energies of Neuroglobin and Myoglobin are also quantitatively indistinguishable 1.21, but hemoglobin is lagging behind other homosapiens globular proteins. The solvation energies of plants are in negative domain state plants globular proteins have less side chains as compare to homosapiens globular proteins. Higher solvation of homosapiens globular proteins means their side chains are more hydrophilic and amphiphilic in nature while plants proteins side chains are water repellent, virtue of that the stability of Myoglobin and hemoglobin reduces.

The torsion angle potential over three consecutive amino acids is the useful tool for the analysis of the local geometry of a model. The amino acid is representing by three points (di-hydril angle) and a direction. The points are the geometric center CZ, the CB (β -C-Atom), and O (oxygen from the carbonyl group of the backbone) of the amino acid. The direction is defining using the vector CZCB, and a plane creates in CZ, CB, and O. In Swiss server 45 degree for the centre residue, 90 degree for the 2 adjacent residues is considering for evaluation of torsion angle energy. Again the torsion energy of Myoglobin and Neuroglobin are similar -1.14 as solvation energies. The negative energy indicates that a staggered conformation due to angle reduce to 60 degree. Minimum energy also signifies the more stability. Globular protein of rice is more stable than soybean protein.

TABLE 4: The Q means score of homosapiens and plants globular using Swiss model for quality estimation of model.

Protein	Hemoglobin	Neuroglobin	Myoglobin	Leg Hemoglobin in (Soybean)	Rice Hemoglobin
Q-mean Score	0.786	0.765	0.765	0.799	0.705

The raw Q mean score values in the range from 0 to 1 and higher QMEAN means better agreement with predicted features and lower mean force to potential energy. We have applied QMEAN Z-scores to sequence of homosapiens from the NCBI database Figure1 and show the Mean score analysis (Neuroglobin, Hemoglobin and Myoglobin) Leg hemoglobin (soybean) and Rice hemoglobin shown in table 4. The Q mean score Neuroglobin, and Myoglobin is alike and highest in leg hemoglobin 0.799 predict the Model quality. Hemoglobin has the highest degree of nativness that shows the Q mean score value 0.786.

V CONCLUSION

This work is the application of Swiss model for homology prediction in homosapiens and plants globular proteins. It evaluates the core structural features of proteins and absolute quality estimate score from fasta. Neuroglobin and Myoglobin have cumulative properties are similar and virtue of that quality of the estimated model (Q mean score) also same. The model quality of homosapiens and plant hemoglobin also reveals that close relationship. Rice hemoglobin has least quality estimation among five proteins due to the core structure.

REFERENCES

- [1] Y. Kubota, Ken Nishikawa, Sho Takahashi, Tatsuo Ooi, “*Correspondence of homologies in amino acid sequence and tertiary structure of protein molecules.*” *Journal of Protein Structure and Molecular Enzymology*, Volume 701, Issue 2, pp 242–252, 1982
- [2] FE Cohen, MJE Sternberg - “*Analysis and prediction of the packing of [alpha]-helices against a [beta]-sheet in the tertiary structure of globular proteins.*” *Journal of molecular biology*, Volume 156, Issue 4, pp 821–862, 1982,
- [3] P Cariani and Narendra S. Goel, “*On the computation of the tertiary structure of globular proteins—IV. Use of secondary structure information.*” *Bulletin of mathematical biology*, Volume 47, Number 3, pp 367-407, 1985
- [4] J. Garnier, “*Protein structure prediction.*” *Biochimie* Volume 72, Issue 8, pp 513–524, 1990
- [5] N. Guex, Manuel C. Peitsch, “*SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling.*” *Electrophoresis* Volume 18, Issue 15, pp 2714–2723, 1997
- [6] A Liwo, J Lee, DR Ripoll, “*Protein structure prediction by global optimization of a potential energy function.*” *Biophysics* Vol. 96, pp. 5482–5485, 1999
- [7] N Siew, A Elofsson, L Rychlewski, “*MaxSub: an automated measure for the assessment of protein structure prediction quality.*” *Bioinformatics*, Volume 16 (9), pp 776-785. (2000)
- [8] F Le Moigne, MC Beauvieux, P Derache, “*Determination of myoglobin: comparative evaluation of the new automated VIDAS assay with two other immunoassays.*” *Clinical Biochemistry*, Volume 35, Issue 4, pp 255–262, 2002
- [9] A Eiko, “*Cloning of the human myoglobin gene.*” *Gene* Volume 33, Issue 3, pp 241–249, 1985
- [10] TM Przybycien, JP Dunn, P Valax, “*Secondary structure characterization of β -lactamase inclusion bodies.*” *Protein Eng.* Volume 7 (1), pp 131-136, 1994
- [11] DE Kim, D Chivian, “*Protein structure prediction and analysis using the Robetta server.*” *Nucl. Acids Res.* Volume 32 (suppl 2), pp W526-W531. 2004
- [12] T Herges, W Wenzel, “*An All-Atom Force Field for Tertiary Structure Prediction of Helical Proteins.*” *Biophysical Journal* Volume 87, Issue 5, pp 3100–3109, 2004

- [13] M Brunori, D Bourgeois, "*The structural dynamics of myoglobin,*" Journal of Structural Biology ,Volume 147, Issue 3, pp 223–234, 2002
- [14] NS Goel, "*On the computation of the tertiary structure of globular proteins II ,*" Journal of Theoretical Biology ,Volume 77, Issue 3, pp 253–305, 1979
- [15] JA Hoy, "*The structure and function of plant hemoglobins, Plant Physiology and Biochemistry ,*"Volume 46, Issue 3, pp 371–379, 2008
- [16] F Kiefer, K Arnold, M Künzli, L Bordoli, "*The SWISS-MODEL Repository and associated resources,*" Nucl. Acids Res. Volume 37 (suppl 1), pp D387-D392.2009
- [17] Lee MS, Bondugula R, Desai V, Zavaljevski N, Yeh I-C, et al. "*PSPP: A Protein Structure Prediction Pipeline for Computing Clusters,*" PLoS , Volume4(7), pp e6254, 2009
- [18] P Benkert, M Künzli, "*QMEAN server for protein model quality estimation,*" Nucleic Acids Res. (2009) 37 (suppl 2), pp W510-W514,2009
- [19] P Benkert, M Biasini, "*Toward the estimation of the absolute quality of individual protein structure models,*" Bioinformatics (2011) 27 (3), pp 343-350, 2011
- [20] K Arnold, L Bordoli, J Kopp, "*The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling,*" Bioinformatics Volume 22 (2), pp 195-201.2006