Research Article

Homology Modeling Studies of Human Genome Receptor Using Modeller, Swiss-Model Server and Esypred-3D Tools

A. Ranganadha Reddy*[1], T.C. Venkateswarulu[1], D. John Babu[1], M.Indira[1]
*[1&1]School of Biotechnology, VFSTR University, Guntur, Andhra Pradesh, India.
***Corresponding author's E-mail:** rangaaluri@gmail.com

Accepted on: 26-08-2014; **Finalized on:** 31-12-2014.

ABSTRACT

The importance of homology modeling is that, for a set of given proteins that are hypothesized to be homologous, their three-dimensional structures are conserved to a greater extent than their primary structures. This observation has been used to generate protein models from homologues with very low sequence similarities. Thus, we attempted to model human genome receptor protein using MODELLER, SWISS-MODEL SERVER and ESYPRED-3D TOOLS. Comparison of the three Insilco tools for modeling the same protein showed that MODELLER outperformed SWISS-MODEL SERVER and ESYPRED-3D.

Keywords: Homology modeling, Human genome receptor, Modeller, Swiss-Model Server, Esypred-3D

INTRODUCTION

Homology modeling prognosticates the three-dimensional (3D) structure of a given protein of unknown structure (target) dependent primarily on its alignment to one or more proteins of known structure (templates).[1-5] Two prerequisites must be met to build a useful model. First, the congruity between the target sequence and the template structure must be detectable. Second, a considerably correct alignment between the target sequence and the template structures must be calculated. Homology modeling is possible because small changes in the protein sequence usually result in small changes in its 3D structure.[6] Although substantial progress has been made in ab initio protein structure prediction[7], Homology modeling remains the most reliable prediction method. The overall accuracy of homology models spans a wide range, from low resolution models with only a correct fold to more accurate reliable models comparable to medium resolution structures determined by crystallography or nuclear magnetic resonance (NMR) spectroscopy[5]. Even low resolution models can be useful in biology because some attributes of function can sometimes be predicted only from the coarse structural features characteristics of a model.

The 3D structures of proteins in a family are more conserved than their sequences[8]. Therefore, if homogeneity between two proteins is recognized at the sequence level, structural similarity can usually be understood. Moreover, even proteins that have non detectable sequence similarity can have similar structures. It has been estimated that approximately one third of all sequences are distinguishably related to at least one known protein structure.[9-13] Because the number of known protein sequences is approximately 500,000[14], homology modeling could in principle be applied to more than 150,000 proteins. This number can be compared to approximately 10,000 protein structures determined by experiment.[15,16] The usefulness of homology modeling is steadily increasing because the number of unique structural folds that proteins adopt is restricted[17] and because the number of experimentally determined new structures are increasing exponentially[18]. It is possible that in less than 10 years at least one example of most structural folds will be known, making homology modeling applicable to most protein sequences.[18] The beta-1 adrenergic receptor (β1 adrenoreceptor), also known as ADRB1, is a beta adrenergic receptor, and also denotes the human gene encoding it[19]. It is a G-protein coupled receptor associated with the Gs heterotrimeric G-protein and is expressed predominantly in cardiac tissue.

All current homology modeling methods consist of four sequential steps (Figure 1): fold assignment and template selection, template–target alignment, model building, and model evaluation. If the model is not satisfactory, template selection, alignment, and model building can be repeated until a satisfactory model is obtained. For each of the steps in the modeling process, there are many different methods, programs, and World Wide Web servers (Table 1).

Table 1: Programs and World Wide Web servers useful in comparative modeling

Program	Website address	Program description
Geno3d	http://pbil.ibcp.fr/	Automatic modeling of protein three-dimensional structure
Swiss Model	http://www.expasy.org/swissmod/SWISS-MODEL.html	An automated knowledge-based protein modeling server; first approach and optimize
CPHmodels	Http://www.cbs.dtu.dk/services/CPHmodels/	Automated neural-network based protein modeling server
Modeller	http://salilab.org/	A program for automated protein Homology Modeling
Amber	http://amber.scripps.edu/	Similar package as CHARMm. Developed by Kollaman's group at UCSF
Homology	http://www.accelrys.com/	Automatic Homology Modeling module. The software suite also has Modeller, SeqFold modules, Quanta
Wloop	http://psb00.snv.jussieu.fr/wloop/	The Loop Homology Modeling Server
What-If Server	http://www.cmbi.kun.nl/gv/servers/WIWWWI/	V.Friend's What-IF Homology Modeling Server
Esypred3D	http://www.fundp.ac.be/sciences/biologie/urbm/bioinfo/esypred/	ESyPred3D is a new automated homology modeling program.
SDSC1	http://cl.sdsc.edu/hm.html	SDSC Protein Structure Homology Modeling Server

Modeller is used for homology or comparative modeling of protein three-dimensional structures .The user provides an alignment of a sequence to be modeled with known related structures and Modeller automatically calculates a model containing all non-hydrogen atoms. Modeller models 3D structure of proteins. It is built in FORTRAN. Modeller is most frequently used for homology or comparative protein structure modeling. Modeller models protein 3D structure keeping in the constraints of spatial restraints. **Swiss Model Server** is a fully automated protein structure Homology Modeling server. Swiss Model however does not accept the sequences for homology modelling when similarity is less than 25%. **ESyPred3D** is a new automated homology modeling program. The method gets benefit of the increased alignment performances of a new alignment strategy using neural networks. Alignments are obtained by combining, weighting and screening the results of several multiple alignment programs.

COMPUTATIONAL TOOLS

All calculations were carried out in Maestro v9.2 installed in Cadd-WS3 machine under 64-bit centos operating system placed in CADD department, Institute of Life Sciences. The machine was built up with:

A) 4 cores and 8 processers with Intel Xenon CPU E5620 @ 2.40GHZ

B) 16 GB RAM

C) NVidia Qudvo FX3800 Graphical Process Unit (GPU)

D) The PROCHECK analysis provides an idea of the stereo chemical quality of all protein chains in a given PDB structure. They highlight regions of the proteins which appear to have unusual geometry and provide an overall assessment of the structure as a whole.

E) Other Servers

1) Primary sequence of the beta-1 adrenergic receptor Protein (Homo sapiens) was retrieved from Swiss Prot (accession number **NP_000675** and GI number **4557265**) from the ExPASy (Expert Protein Analysis System) proteomics serves of the Swiss Institute of Bioinformatics.

2) Homology search for Ras protein was carried out using BLAST software.

F) The crystal structure for Ras protein (PDB ID: **2vt4**) was obtained from PDB database RCSB.

DATABASES

Database

A database is a collection of information that is organized so that it can easily be accessed, managed, and updated.

Data mining

Data mining[20] or knowledge discovery is the computer-oriented process of digging and analyzing large volumes of data and finally extracting the meaning of the data. Applications of data mining to bioinformatics include gene finding, protein function domain detection, function motif detection, protein function inference, disease diagnosis, disease prognosis, disease treatment optimization, protein and gene interaction network reconstruction, data cleansing, and protein sub-cellular location prediction.

Swiss-Prot

Swiss-Prot was created in 1986 by Amos Bairoch and developed by the Swiss Institute of Bioinformatics and subsequently developed by Rolf Apweiler at the European Bioinformatics Institute.[21-23] Swiss-Prot aimed to provide reliable protein sequences associated with a high level of annotation a minimal level of redundancy and high level of integration with other databases.

Protein Data Bank

PDB consists of 3D (Three Dimensional) data of experimentally determined structures of proteins and nucleic acids[24] established at Brookhaven National Laboratory.[25] The archive is managed by the Worldwide Protein Data Bank organization (wwPDB), whose mission is to ensure that a single, global PDB data archive is and will remain freely and publicly available.[26]

TOOLS

Basic Local Alignment Search Tool (BLAST)

Nowadays Similarity searching, including sequence comparison, is one of the principal techniques used by computational biologists and has found widespread use among biologists in general. The most popular tool for this purpose is BLAST[27] (**Basic Local Alignment Search Tool)** which performs comparisons between pairs of sequences, searching for regions of local similarity. NCBI BLAST is available from the NCBI[28] **(National Center for Biotechnology Information**).

METHODOLOGY

Homology modeling was performed by using three tools: Modeller, Swiss Model Server, and Esypred3D.

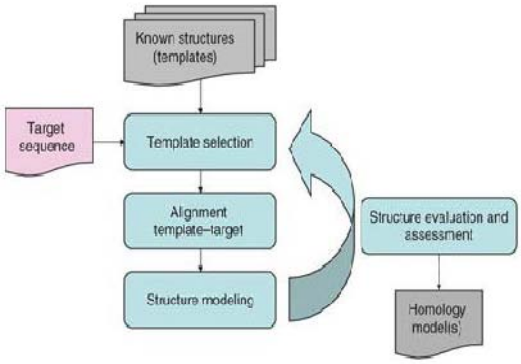


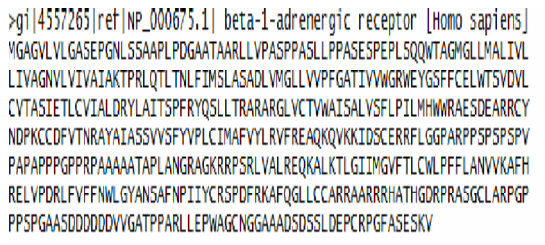
Figure 1: Four main steps of homology modeling of protein i.e. template selection, target–template Alignment, model building and model quality evaluation[29]

Homology modeling

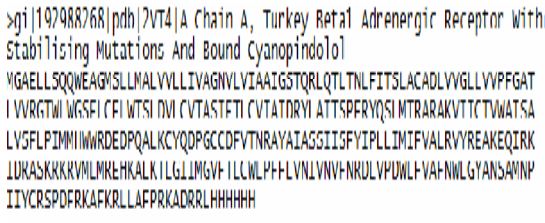
The steps to creating a homology model are as follows:

- Identify homologous proteins and determine the extent of their sequence similarity with one another and the unknown.
- Align the sequences.
- Identify structurally conserved and structurally variable regions
- Generate coordinates for core (structurally conserved) residues of the unknown structure from those of the known Structure(s).
- Generate conformations for the loops (structurally variable) in the unknown structure.
- Build the side-chain conformations.
- Refine and evaluate the unknown structure.

RESULTS AND DISCUSSION



Sequence of target molecule in FASTA format



Sequence of template molecule in FASTA format

Table 2: Sequential steps of modeling in Modeller, Swiss-Model Server and Esypred-3D

MODELLER	SWISS-MODEL SERVER	ESYPRED-3D
Retrieve the amino acid sequence of the target and template in FASTA format from a suitable database.	Open the home page of Swiss-Model server	Open the home page of ESYPRED-3D Server
Retrieve the pdb file of the template from PDB.	Open the first approach mode in Swiss-Model server	Give the e-mail id and a title for the work
Prepare the alignment file of the target and the sequence in PIR format and name it as Alignment.ali.	Give the e-mail id and a title for the work	Paste the sequence of the target protein
Open the script files named model-def ault.py and do necessary corrections and save as model.py.	Paste the sequence of the target protein	Submit model request to the server
Open mod9v6 (Command prompt). Change the dir.	Submit modeling request to the server	The model was collected from mail which was sent by the server
Run the MODELLER (by typing mod9v6 model.p y).	The model was obtained from server	The obtained model was energy minimized to get stable conformation
Choose the best model based on molecule pdf. The obtained model was energy minimized to get stable conformation.	The obtained model was energy minimized to get stable conformation	The energy minimized model was validated by analyzing Ramachandran plot using SAVES server
The energy minimized model was validated by analyzing Ramachandran plot using SAVES server.	The energy minimized model was validated by analyzing Ramachandran plot using AVES server	

Table 4: Comparative studies of Ramachandran plot analysis for Modeller, Swiss model server and Esypred 3D tools.

Residues of protein in various regions	MODELLER	SWISS-MODEL SERVER	ESYPRED-3D
Residues in most favoured region (A, B, L)	95.2%	92.5%	90.7%
Residues in additional allowed region (a, b, l, p)	14.0%	63%	6.7%
Residues in generously allowed region (~a. ~b, ~l, ~p)	0.3%	0.7%	21%
Residues in disallowed region	0.5%	0.4%	0.5%
Number of non-glycine and non-proline residues	100.0%	100.0%	100.0%
Number of end residues (excl Gly and Pro)	2	2	7
Number of glycine residues (shown as triangles)	31	11	10
Number of proline residues	47	10	7
Total number of residues	476	291	218

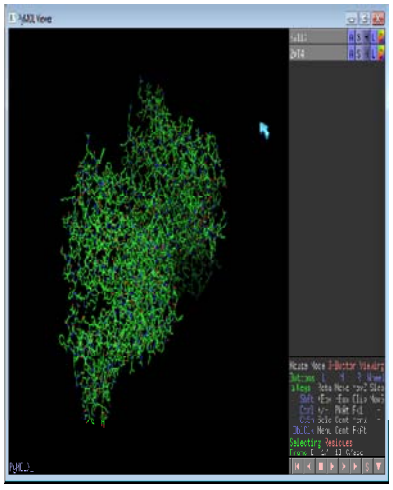


Figure 2(a): Three dimensional structure model of template using PyMol Viewer

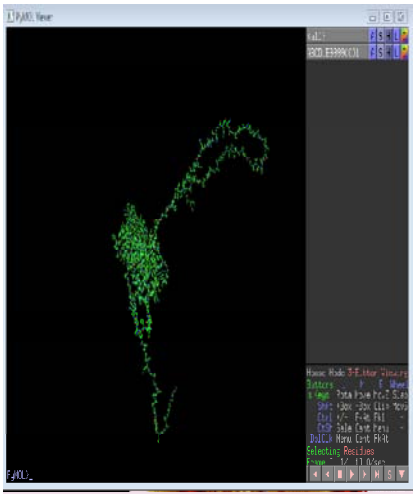


Figure 2(b): Three-dimensional structure of model with lowest model probability density

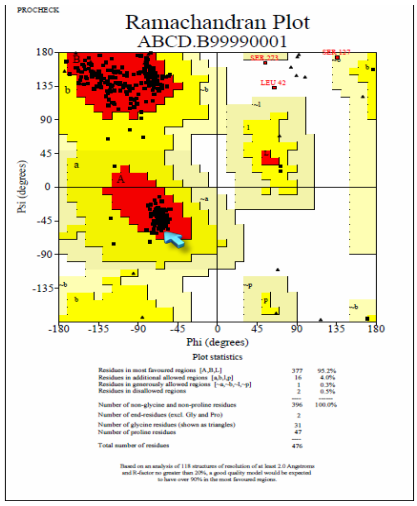


Figure 2(c): Ramachandran plot for Modeller

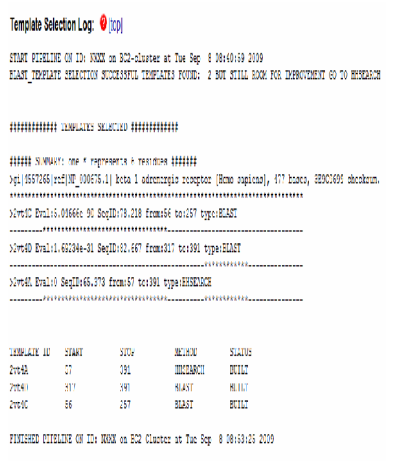


Figure 3(a): Result page of Swiss model server

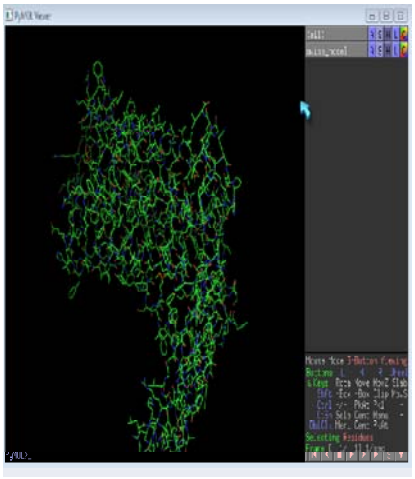


Figure 3(b): 3-dimensional structure of model

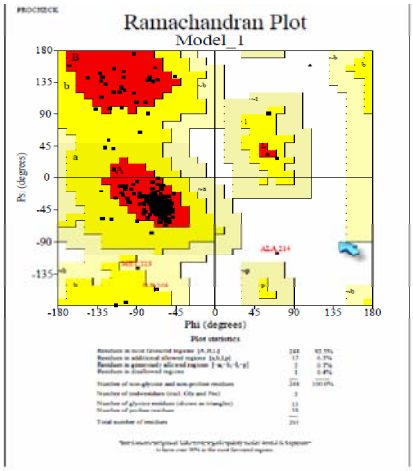


Figure 3(c): Ramachandran plot of Swiss-Model server

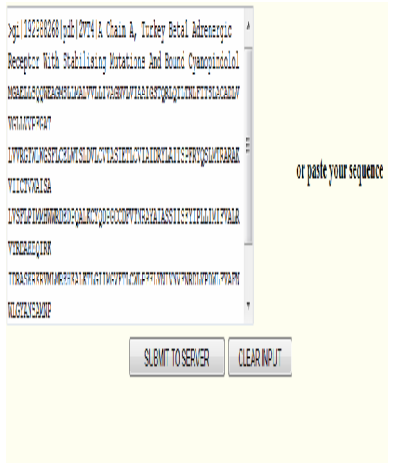


Figure 4(a): Submission form for Esypred 3D Server

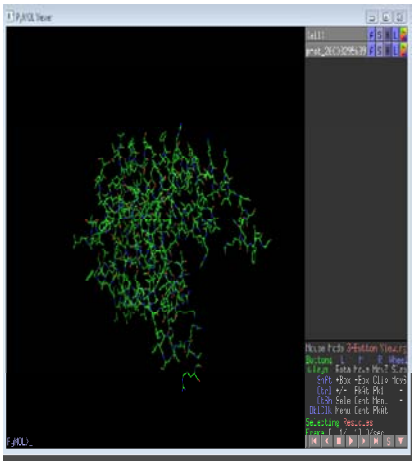


Figure 4(b): Three-dimensional structure of model

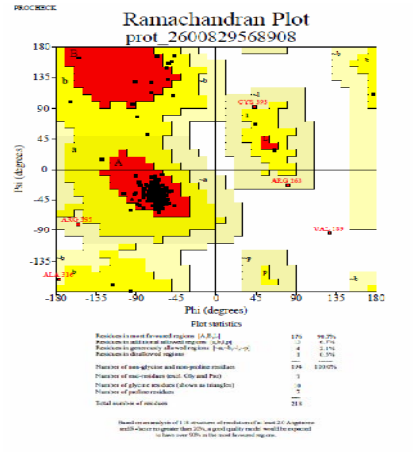


Figure 4(c): Ramachandran plot for Esypred 3D

Table 3: Summary of successfully produced models through MODELLER

File name	Mol pdf
ABCD.B9999001.pdb	1886.91333
ABCD.B9999002.pdb	1940.59387
ABCD.B9999003.pdb	1891.17175
ABCD.B9999004.pdb	2049.98804
ABCD.B9999005.pdb	1903.56274
ABCD.B9999006.pdb	2003.89868
ABCD.B9999007.pdb	2101.79688
ABCD.B9999008.pdb	1929.20789
ABCD.B9999009.pbb	1959.43909
ABCD.B9999010.pdb	1970.66272

The sequence of target molecule in FASTA format was retrieved from NCBI database. Then the sequence of template molecule in FASTA format was generated using BLAST. The 3D (Three Dimensional) structure of template molecule was obtained from PDB (Protein Data Bank) and viewed by using Pymol as shown in Figure 2(a). The alignments of target and template sequences are performed by the CLUSTALW. After the execution of modeller the summary of modeller result is represented as shown in Table 3. The model with lowest molecular probability density function was obtained and its 3D structure was viewed using Pymol as shown in Figure 2(b). The energy minimized values of the model are analyzed by Rama chandran plot as shown in Figure 2(c).The SWISS MODEL Work space was opened in an automated modeling model. FASTA format of the template sequence (2vt4) was submitted and the result page of SWISS MODEL server with the pdb file of the model along with alignment file, 3D log, modeling log, template selection log as shown in Figure 3(a). The 3D structure of model was viewed using Pymol viewer in Figure 3(b). The energy minimized values of the model are analyzed by Rama chandran plot as shown in Figure 3(c). The Esypred3D web server was opened; data and parameters as input are submitted. The sequence was pasted and submitted to the server as shown in Figure 4(a). 3Dimensional structure of model was viewed using Pymol viewer as shown in Figure 4(b). The energy minimized values of the model are analyzed by Ramachandran plot as shown in Figure 4(c).

CONCLUSION

The modeling accuracies of a molecular modeling package, MODELLER is evaluated and compared against SWISS – MODEL server and Esypred 3D server. The study revealed that the MODELLER outperforms the similar programs in almost all the parameters tested. However, the study also revealed that although lagging behind in accuracy, results from SWISS – MODEL server and Esypred 3D server are biologically meaningful. This taken together with the fact that SWISS – MODEL server and Esypred 3D server are offered in an intuitive, easy to use graphical user interface, suggests that it can be employed as an effective teaching tool to demonstrate molecular modeling to beginners in this area. More research and development of the various modeling and scoring modules of SWISS – MODEL server and Esypred 3D server should definitely make this software a more popular one in educational circles.

REFERENCES

1. Bajorath J, Stenkamp R, Aruffo A. Knowledge-based model building of proteins: concepts and examples. Protein Sci. 2, 1994, 1798-1810.
2. Blundell TL, Sibanda BL, Sternberg MJE, Thornton JM. Knowledge based prediction of protein structures and the design of novel molecules. Nature. 326, 1987, 347-352.
3. Johnson MS, Srinivasan N, Sowdhamini R, Blundell TL. Knowledge based protein modeling. CRC crit. Rev. Biochem. Mol. Biol. 29, 1994, 1-68.
4. Sali A. Modelling mutations and homologous proteins. Curr. Opin. Bio tech. 6, 1995, 437–451.
5. S´anchez R, Sali A. Advances in comparative protein-structure modeling. Curr. Opin. Struct. Biol. 7, 1997, 206–214.
6. Chothia C, Lesk AM. The relation between the divergence of sequence and structure in proteins. *EMBO J.* 5, 1986, 823–826.
7. Koehl P, Levitt M. A brighter future for protein structure prediction. *Nat. Struct. Biol.* 6, 1999, 108–111.
8. Lesk AM, Chothia C. How different amino acid sequences determine similar protein structures: The structure and evolutionary dynamics of the globins. *J. Mol. Biol.* 136, 1980, 225–270.
9. Fischer D, Eisenberg D. Assigning folds to the proteins encoded by the genome of *Mycoplasma genitalium*. *Proc. Natl. Acad. Sci. USA* 94, 1997, 11929–11934.
10. Huynen M, Doerks T, Eisenhaber F, Orengo C, Sunyaev S, Yuan Y, Bork P. Homology-based fold predictions for *Mycoplasma genitalium* proteins. *J. Mol. Biol.* 280, 1998, 323–326.
11. Jones DT. An efficient and reliable protein fold recognition method for genomic sequences. *J. Mol. Biol.* 287, 1999, 797–815.
12. Rychlewski L, Zhang B, Godzik A. Fold and function predictions for *Mycoplasma genitalium* proteins. *Folding Design* 3, 1998, 229–238.
13. S´anchez R, Sali A. Large-scale protein structure modeling of the *Saccharomycescerevisiae* genome. *Proc. Natl. Acad. Sci. USA* 95, 1998, 13597–13602.
14. Bairoch A, Apweiler R. The SWISSPROT protein sequence data bank and its supplement TrEMBL in 1999. *Nucleic Acids Res.* 27, 1999, 49–54.
15. Abola EE, Bernstein FC, Bryant SH, Koetzle TF, Weng J. Protein data bank. In *Crystallographic Databases—Information, Content, Software Systems, Scientific*

Applications, ed. FH Allen, GBergerhoff, R Sievers, 1987, 107–132. Bonn/Cambridge/Chester. Data Commission Int. Union of Crystallography.

16. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucleic Acids Res.* 28, 2000, 235–242.
17. Zhang ZT. Relations of the numbers of protein sequences, families and folds. *Protein Eng.* 10, 1997, 757–761.
18. Holm L, Sander C. Mapping the protein universe. *Science.* 273, 1996, 595–602.
19. Jeff JM, Donahue BS, Brown-Gentry K, Roden DM, Crawford DC, Stein CM, Kurnik D. Genetic variation in the β1-adrenergic receptor is associated with the risk of atrial fibrillation after cardiac surgery. Am Heart J. 167, 2014, 101-108. e1. doi: 10.1016/j.ahj.2013.09.016.
20. Venkateswarulu TC, Kodali VP, Ranganadha Reddy A, John DB, Rahul RN, Indira M, Narayana AV. Comparative Studies of H5N1 Gene Segments with other Subtypes of Influenza-A Virus. Research Journal of Pharmaceutical, Biological and Chemical Sciences. 5, 2014, 1417–1429.
21. Bairoch, A.; Apweiler, R. "The SWISS-PROT protein sequence data bank and its new supplement TREMBL". Nucleic Acids Research. 24, 1996, 21–25. doi: 10.1093/nar/24.1.21. PMC 145613. PMID 8594581.
22. Bairoch, A. "Serendipity in bioinformatics, the tribulations of a Swiss bioinformatician through exciting times!". Bioinformatics 16, 2000, 48–64. doi: 10.1093/bioinformatics/16.1.48. PMID 10812477.
23. Séverine Altairac, "Naissance d'une banque de données: Interview du prof. Amos Bairoch". Protéines à la Une, August 2006. ISSN 1660-9824.
24. Berman, H. Acta Crystallogr A: Foundations of Crystallography. 64, 2008, 8–95.
25. Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. J Mol Biol. 112, 1977, 535–542.
26. Berman, H. M.; Henrick, K.; Nakamura, H. Nat Struct Biol. 10, 2003, 980.
27. Ranganadha Reddy A, Sreedhara. R. Voleti, Lakshmi Padma Ch. Towards the prediction of a three-dimensional structure of human glp-1receptor. Research Journal of Pharmaceutical, Biological and Chemical Sciences, 4, 2013, 151-168.
28. Alexander Pertsemlidis and John W Fondon Having a BLAST with bioinformatics (and avoiding BLASTphemy) *Genome Biology* 2001, 2: reviews2002-reviews2002. 10 doi: 10.1186/gb-2001-2-10-reviews2002.
29. Ranganadha Reddy A, Venkateswarulu TC, JohnBabu D, Shyamala Devi N. Homology Modeling, Simulation and Docking Studies of Tau-Protein Kinase. Research Journal of Pharmacy and Technology. 7, 2014, 376-388.

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