IDENTIFY THE 3D STRUCTURE OF ENVELOPE SMALL MEMBRANE (sM PROTEIN) PROTEIN OF HUMAN SARS CORONAVIRUS (SARS-CoV; SEVERE ACUTE RESPIRATORY SYNDROME CORONAVIRUS) USING HOMOLOGY MODELING

Dr. Harish Kumar Dhingra[1], Pratima Bajpai[2]

[1]Assistant Professor, Department of Biosciences, College of Arts, Science and Humanities, Mody University, Rajasthan.
[2]Assistant Professor, Department of Biosciences, College of Arts, Science and Humanities, Mody University, Rajasthan.

ABSTRACT

BACKGROUND

Homology modelling can be used to determine the 3D structure of proteins. It uses available high-resolution protein structures to produce a model of a protein of similar, but unknown structure. Herein, the article describes the essential steps in the process and discusses the circumstances in which homology modelling is likely to give a useful result. Homology modelling plays a valuable role in drug designing. The article describes the drug designing concept using an example of anti-SARS inhibitors.

KEYWORDS

Molecular Modeling, Model Building.

HOW TO CITE THIS ARTICLE: Dhingra HK, Bajpai P. Identify the 3D structure of envelope small membrane (sM protein) protein of human SARS coronavirus (SARS-CoV; severe acute respiratory syndrome coronavirus) using homology modeling. J. Technological Advances and Scientific Res. 2017;3(1):11-13, DOI: 10.14260/jtasr/2017/04

BACKGROUND

It is almost 50 years since the first protein crystal structure of myoglobin was solved.[1,2] With the advances in techniques of experimental structural biology as well as in whole-genome sequencing in the late decade of twentieth century, the excitement of solving a single-crystal structure has been replaced by determining protein structures on a large scale. This was a start of new era in structural biology of determining the protein structure.[3-6] The Protein Data Bank (PDB), which is an electronic repository for obtaining 3D structures of proteins and nucleic acids is very popular tool to know the structure of protein and now a days these initiatives contribute very much to the newly structured families of proteins.[7]

There are currently (5th January, 2017) 116511 protein structures in the PDB (Table 1). Although, this number is increasing rapidly, there remains a vast gap between the number of available gene sequences and experimentally solved protein structures by applying physical techniques.

Applied techniques carry on much more slowly than genome sequencing; and since many more genome sequences are in the pipeline, this gap must surely grow. Various projects on structural genomics want to determine the 3D model of the various proteins, this means that instead of trying to characterise the structure of every protein experimentally, it would be advantageous to start from already available protein structures and employ various computation techniques to determine the structure for the related proteins. This method is known as homology modelling.

Financial or Other, Competing Interest: None.
Submission 07-01-2017, Peer Review 16-01-2017,
Acceptance 30-01-2017, Published 21-02-2017.
Corresponding Author:
Dr. Harish Kumar Dhingra,
Assistant Professor,
Department of Biosciences,
College of Arts, Science and Humanities,
Mody University, Rajasthan.
E-mail: harishdhingra2000@gmail.com
DOI: 10.14260/jtasr/2017/04

According to New York Structural Genomics Research Consortium, every new protein structure could be modelled to many fold level without any prior structural characterisation.[8]

Molecular Type				
Experimental Method	Proteins	Nucleic Acids	Protein/Nucleic Acid Complexes	Total
X-ray diffraction	105028	1796	5389	112217
NMR	10239	1187	237	11671
Electron microscopy	966	30	335	1331
Hybrid	97	3	2	103
Other	181	4	6	204
Total	116511	3020	3481	125526

Table 1. Number of Proteins/Nucleic Acid Complex Structures Obtained by Various Experimental Methods, Available in the PDB as on 5th January 2017 (Taken from www.rcsb.org/pdb)

Steps in Homology Modelling

Homology modelling employs to predict the 3D structure of a protein based on its sequence similarity to one or more proteins of known structure. The method relies on the observation that the amino acid sequence of a protein is more liable to be present in variable conformation than structural conformation of a protein. Homology modelling can be divided into four steps- 1. Template identification; 2. Alignment; 3. Model building and Refinement and 4. Validation (Figure 1). This can be done with various computational tools (Table 2) available for each step.

Short Form	Full Name	Web address
BLAST	Basic Local alignment search tool	www.ncbi.nlm.nih.gov/BLAST
FASTA	Fast-Al	www.ebi.ac.uk/fasta
PDB	Protein Data Bank	www.rcsb.org/pdb
Pro-check	Protein structure check	www.biochem.ucl.ac.uk/procheck
What IF		http://swift.cmbi.knu.nl/whatif

Table 2. The Short Form, Full Names and Web Addresses of the Program and Web Servers in Alphabetical Order

Template Identification

The first step in molecular modelling is template identification, which is a critical step of the process. It lays the foundation by knowing the appropriate homologue(s) of a given protein structure, known as template(s), which are sufficiently similar to the target sequence that needs to be modelled. A simple search submits the target sequence to programs such as BLAST[9] or FASTA.[10]

These methods often suggest many templates. Out of these, the ideal template is that which has the highest percentage identity to the target and has the highest resolution and also has structures with (or without) appropriate ligands and/or cofactors. It maybe that there is no candidate template that is best according to all criteria; in that case, the choice is a matter of judgment and perhaps of trying different templates.

Alignment

The second step in homology modelling of protein structure involves creating an alignment of the target sequence with the identified template structure(s). This is a vital step and there are various ways to ensure high accuracy. The target and template sequence can be generated from all relevant sequences retrieved via BLAST.

Model Building and Refinement

Although, the theory behind building a protein homology model is complicated, however, using the available programs, it is relatively easy. Several modelling programs are available, using different methods to construct the 3D structures. In segment matching methods, the protein target is divided into short segments and alignment is allowed over these segments rather than over the entire protein.[11] Satisfying spatial restraints is the most common method. This method uses distances or optimisation techniques to satisfy various spatial restraints. The model building and refinement can be done using the popular program such as WHAT IF.[12] Web servers such as Swiss Model and the Rosetta server make it even easier to generate a model.

Validation

After building and refinement of model for 3D structure of protein, the model needs to be validated. One of the most important and thorough structure checking programs is Whatcheck.[13] Other programs such as Procheck, etc. are also available for validation of the model.[14] The best validation combines common sense, biological knowledge and results from analytical tools. Some models will require further refinement. There is a cycle between building-validating-refining. At the last, most refinement involves adjusting the alignment.

Advantages and Limitations of Homology Modelling

The first and foremost advantage of Homology modelling for obtaining 3D structure of proteins is that it is a relatively easy technique and do not require complex infrastructure facility. The method is user friendly and easy to understand than an experiment. As already discussed, the technique does not require any expensive experimental laboratory facility except a computer. Therefore, without any high-resolution experimental structures, homology modelling can be of high value to determine the 3D structure of proteins.

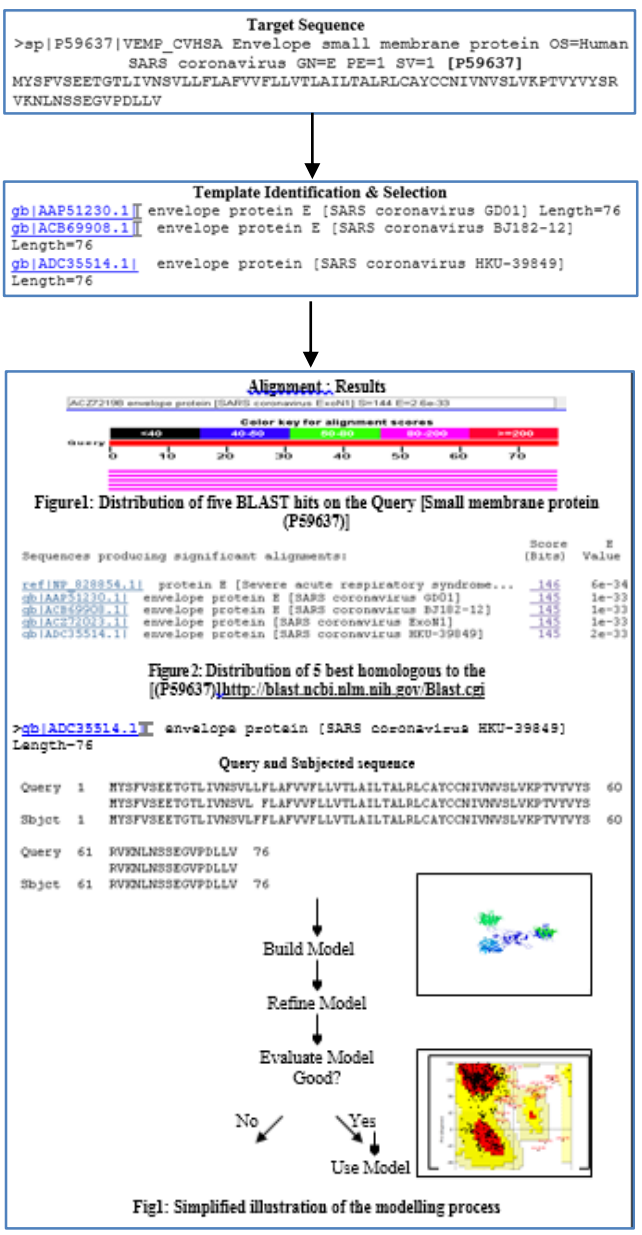

Fig1: Simplified illustration of the modelling process

As a part of limitation of the technique, it could be understood that the quality and accuracy of the homology model of the protein depend on various factors. This technique essentially requires a high-resolution experimental protein structure as a template, the accuracy of

which directly affects the quality of the model in question. Even more importantly, the quality of the model depends on the degree of sequence, identity between the template and protein to be modelled.[8,15,16-18] When the sequence identity is less than 30%, the possibility of alignment errors increases rapidly and if it is having about 30% and 50% sequence identity to the template, a medium accuracy homology occurs. They can facilitate structure-based prediction of target for 'drug ability', the design of mutagenesis experiments and the construction of in vitro test assays. Higher accuracy models are typically obtained when the level of sequence identity is more than 50%. This data can be used in the study of protein-legend interactions such as the prediction of the preferred sites for the metabolism of various small molecules, as well as structure-based drug designing.

The technique of Homology modelling for the membrane protein requires special care.[19] The available crystal structures are limited and modelling methods are mainly designed for water-soluble proteins. Another limitation of homology modelling is the presence of loops and inserts as they cannot be modelled without template data; however, it is possible to estimate length, location and distance from the active site if the target protein is an enzyme.

CONCLUSION

Homology modelling is very important tool of structural genomics, which can be efficiently used for predicting the 3D structure of protein without using very expensive resources. Homology modelling can also become a very important step in silico structural drug designing and drug discovery.

REFERENCES

1. Kendrew JC, Bodo G, Dintzis HM, et al. A three-dimensional model of the myglobin molecule obtained by x-ray analysis. Nature 1958;181:662-666.
2. Kendrew JC. The three-dimensional structure of a protein molecule. Sci Am 1961;205(6):96-110.
3. Cassman M, Norvell JC. Support for structural genomics and synchrotrons. Science 1999;286(5438):239-240.
4. Yokoyama S, Hirota H, Kigawa T, et al. Structural genomics projects in Japan. Nature Struct Biol Suppl 2000;943-945.
5. Bahar M, Ballard C, Coen SX, et al. SPINE workshop on automated X-ray analysis: a progress report. Acta Cryst D Biol Crystallogr 2006;62:1170-1183.
6. Williamson AR. Creating a structural genomics consortium. Nature Struct Biol Suppl 2000:953.
7. Marsden RL, Lewis TA, Orengo CA. Towards a comprehensive structural coverage of completed genomes: a structural genomics viewpoint. BMC Bioinformatics 2007;8:86.
8. Baker D, Sali A. Protein structure prediction and structural genomics. Science 2001;294(5540):93-96.
9. Altschul SF, Gish W, Miller W, et al. Basic local alignment search tool. J Mol Biol 1990;215(3):403-410.
10. Pearson WR. Rapid and sensitive sequence comparison with FASTP and FASTA. Methods Enzymol 1990;183:63-98.
11. Levitt M. Accurate modelling of protein conformation by automatic segment matching. J Mol Biol 1992;226(2):507-533.
12. Vriend G. What if: a molecular modeling and drug design program. J Mol Graph 1990;8(1):52-56.
13. Hooft RW, Vriend G, Sander C, et al. Errors in protein structures. Nature 1996;381(6580):272.
14. Laskowski RA, MacArthur MW, Moss DS, et al. Procheck: a program to check the stereochemical quality of protein structures. J Appl Cryst 1993;26:283-291.
15. Hillisch A, Pineda LF, Hilgenfeld R. Utility of homology models in the drug discovery process. Drug Discov Today 2004;9(15):659-669.
16. Marti-Renom MA, Stuart AC, Fiser A, et al. Comparative protein structure modeling of genes and genomes. Annu Rev Biophys Biomol Struct 2000;29:291-325.
17. Sanchez R, Sali A. Large-scale protein structure modeling of the saccharomyces cerevisiae genome. Proc Natl Acad Sci USA 1998;95(23):13597-13602.
18. Koehl P, Levitt M. A brighter future for protein structure prediction. Nat Struct Biol 1999;6(2):108-111.
19. Reddy ChS, Vijayasarathy K, Srivinas E, et al. Homology modeling of membrane proteins: a critical assessment. Comput Biol Chem 2006;30(2):120-126.