ISSN: 2394-2371
Research Paper

In-silico 3D Structure Modeling of IpLA7 Protein of Borrelia burgdorferi using Homology Modeling

Sumit Govil[1], Anop Singh Ranawat[1], Mohan Kumar Yadav[1], Mahima Tiwari[1], Divya Srivastava[1], Shailesh Kumar[2]*

[1]Jaipur National University, Jaipur, Rajasthan, India
[2]Amity Institute of Biotechnology, Amity University Rajasthan, Jaipur, India

ABSTRACT

The lyme disease is the most common infectious disease in the world, which usually happens in temperate region which is carried by Nymphal and adult deer ticks. The target protein responsible for lyme disease is IpLA7 and also called as outer surface 22 Kda lipoprotein which was retrieved from the National Center for Biotechnology Information database. The appropriate template for homology modeling was determined by using BLASTP. The structure of the protein is important because the antibody against is only secreted when the disease is at last stage, detection at early stage for this disease is not possible [8] thus if structure is modeled we can detect the infection at early stage. 3D structure of unknown protein sequence was determined using Modelle9.11 software. The model generated was tested by using ProsA for the prediction of z-score and VADAR server for the analysis of Ramachandran Plot where maximum amino acids were plotted under core region. The antigen IpLA7 modeled structure had 9 different protein binding regions and 1 polynucleotide binding region, which can be further utilized to form drug against the antigen as the PeptideCutter server predicted presence of proteolysis sites on the amino acid sequence using different enzymes.

Key words: IpLA7 Protein, Borrelia burgdorferi, BLASTP, VADAR.

INTRODUCTION

Borrelia burgdorferi is a spirochaete which is responsible for causing Lyme disease which is transmitted from byte of Nymphal and adult deer ticks. The early age symptoms are fever, headache, fatigue and skin rashes, whereas the late age symptoms affect permanent impairment or disable activities of various organs like brain, nerves, joints and heart. Few Antibiotics like doxycycline and amoxycilline are useful in eradicating the population of Borrelia burgdorferi with few weeks of antibiotic dosages. Insect repellant, applying pesticides and reducing tick habitat can be few steps to prevent lyme disease [9].

The detection of lyme disease occurs after the antibody is secreted against the antigen IpLA7 and the techniques which are used are ELISA, Western blot, Polymerase Chain Reaction. But early age detection is not possible, so there is a need to have a tertiary structure model of the antigen to understand its function, cleavage site and easy detection.

A comparative genomic analysis of these spirochetes represents a very unique feature of

***Corresponding Author:**
Shailesh Kumar
Assistant Professor
Amity Institute of Biotechnology
Amity University Rajasthan
Jaipur-303002, Rajasthan, India
E.Mail: shailesh_iiita@hotmail.com
Article Received on: 25-09-2014
Revised on: 29-09-2014
Published on: 30-09-2014

Borrelia genomes where is was found that they carry a large number of linear and circular plasmids, and this stud was carried on strains N40, JD1, 297 and B31 showing related but non-identical sets of 16, 20, 19 and 21 plasmids, respectively, that plasmids are composing abotu 33–40% of their genomes [13]. Litrature on immunological studies suggested that Borrelia burgdorferi regulates synthesis of the IpLA7 lipoprotein during mammalian infection i.e. this protein is over expressed in the infection which shows the importance of this protein in the infection cycle [12]. In the present studies we have targeted this protein as an important part to study disease propagation and its infection. It is found that proteins function can easily be determined if their tertiary structure are known, but it may happen that similar structured protein have different function. Thus in order to understand the drug and disease interaction there is need to have an idea of how the protein's structure are formed, there are four levels of structure primary, secondary, tertiary and quaternary. Primary structure is the linear sequence of amino acids. Secondary structure is the helix, sheet and loops/coil. Different analyzing factor are used in determining secondary structure of a protein. The angle between two adjacent amino acids is called torsion angle, which determine the turn of the sequence. Tertiary structure is generally formed by the combination of secondary structure elements at stable configuration. Quaternary structure is the combination of more than one subunit or chains. Nowadays protein sequencing techniques has increased rapidly but till date only 30% of the total submitted sequences 3D structures are available in PDB. Determining the 3D structure of a protein is very difficult and complex task. Generally X-ray crystallography or NMR (Nuclear Magnetic Resonance) technique are used, which are very expensive, time consuming. Therefore computational methods are used which are based on Homology Modeling, Threading or Fold Recognition and Ab initio approaches. Among all of these methods, Homology modeling predicts structure accurately. It is based on alignment of known protein structure called templates with the unknown protein sequence which has matching similarity more than 30%. PROSA web server is used to validate the modeled protein structure with available protein structure derived from PDB on the basic of z-score. VADAR server used for the validation of 3d structure modeled by plotting Ramachandran plot, Solvent Accessible area etc.

MATERIALS AND METHODS

Model Building

The linear amino acid sequence of antigen IpLA7 (outer surface 22 kda lipoprotein) of Bacterium Borrelia burgdorferi ZS7 was retrieved from protein sequence database of NCBI (http://www.ncbi.nlm.nih.gov/). The Accession No: ACK74785.1, Version: ACK74785.1, GI: 218164724, of 196 amino acids were taken as target protein sequence. As the template target sequence alignment is more than 30% hence we choose Advanced homology based modeling methodology of the MODELLER 9.11[2] to produce the tertiary structure of the protein in the sequential order as: selecting template from PDB [1] (Protein Data Bank) by using BLASTp algorithm [3] for protein-protein sequence alignment, after excluding the sequences obtained from uncultured environmental samples. The hits selected (Figure:1) as templates used as frame for modeling the structure of protein IPLA7 are (PDB ID: 1DDV, 1DDW, 1I2H, 4LZI, 2QE7, 2P8V, 1PBO, 3MQ1, 1I7A). The software MODELLER9.11 version's compary.py script was used to align the templates sequence first. The file generated was used to align the multiple templates and target sequence using the script align2d.py. The alignment file was further used as an input to generate 10 models for the target protein (IpLA7) by using the script build_model.py. The model structures were built on the basis of target sequence alignment with multiple 3D-structures of template taken as reference.

Evaluation

In order to choose the best modeled structure developed by homology modelling the lowest

energy confirmation is taken for further analysis. The MODELLER provides statistical parameters which are dope score and molpdf score for every modeled structure, the lowest molpdf score is thus selected from the available structures for further evaluation, in this case Modeled structure 4 (Figure 2) with molpdf score 3034.95898 was selected as the predicted structure for the target sequence. Loop refinement was not performed as the structure predicted had no gaps and rather the energy plotted per residue [5] (Figure 3) was in between -0.01 to -0.07. the structure selected was further processed for energy minimization process using SPBV [9] with default up to two times and the graph was plotted to compare the structures scores on different parameters (Figure 4).

Model evaluation and analysis

In order to provide structure reliability and confidence to the modeled structure we need to evaluate it, various tools are available for protein structure evaluation but we have used VADAR [4] (web based tool) to predict the reliability of the protein. In VADAR server Ramachandran plot (in Figure 5) based conformational analysis is visualize by the positions of amino acids present in on plot under favored, allowed, and generously allowed regions marked as red, yellow, green color respectively. PorsA [5] server used to predict the z score of the modeled structure without energy minimization (Figure 6) is 0.3, and after energy minimization [11] (Figure 7) is observed as 0.2 which indicated the overall modeled structure quality using different sources (X-Rays and NMR) differentiated on the basis of different colors as light blue and dark blue. The second plot (Figure 8) shows local model quality by plotting energies as a function of amino acid sequence position i. In general, positive values correspond to problematic or erroneous parts of the input structure. A plot of single residue energies usually contains large fluctuations and is of limited value for model evaluation. Hence the plot is smoothed by calculating the average energy over each 40- residue fragment s (i, i+39), which is then assigned to the 'central' residue of the fragment at position i+19. A second line with a smaller window size of 10 residues is shown in the background of the plot. [5,6]

Predicting properties and active site prediction

The Protoparm web server (http://web.expasy.org/protparam/) [7] available on ExPASy predicted protein's Molecular weight: 21865.7 and Theoretical pI: 5.53. The online software Predict Protein (http://ppopen.informatik.tu-muenchen.de/) showed that there are 9 different protein binding regions and 1 polynucleotide binding region at position 66 (Figure 7), the server even classified the submitted sequence as: 'all-alpha': %H > 45% AND %E < 5%, 'all-beta': %H < 5% AND %E > 45%,'alpha-beta': %H > 30% AND %E > 20%, 'mixed': All others.

PeptideCutter[7] http://web.expasy.org/peptide_cutter/ server predicted the cleavage of IPLA7 protein on C terminal site and the enzymes predicted to cleave the protein are Arg-C proteinase, Asp-N endopeptidase, CNBr, Formic acid, Glutamyl endopeptidase, Hydroxylamine,Proline-endopeptidase, Trypsin, Tehrmolysine.

RESULTS AND DISCUSSION

The protein sequence antigen IpLA7 of Bacterium Borrelia burgdorferi ZS7 (Accession No: ACK74785.1, Version: ACK74785.1, GI: 218164724) was modeled by software MODELLER9.11. The modeled structure has 6 helices, 1 beta sheet, and rest loops and turns, the Ramachandran Plot plotted by VADAR showed only 2 amino acids in the disallowed region where maximum of them were plotted in core region, when comparing with pdbsum http://www.ebi.ac.uk/thorntonsrv/databases/pdbsum/Generate.html/ [9] plot of Ramachandran plot showed only 2 amino acids in disallowed region PHE123 and GLU120 thus the predicted structure can nearly exits in natural form, rest was same as VADAR. Z-score predicted by PROSA web server was 0.3, this protein acts as an antigen hence its binding region were predicted which showed protein and polynucleotide binding sites which can help us to predict future drug

interaction analysis, as the protein enhances the lyme disease so cleavage sites were predicted for in-silico protease containing drug design which can cleave the protein and cure the disease. In future docking [11] of active compounds will be carried out on modeled structure to find binding pattern on drug to the active site.

REFERENCES

1. F.C.Bernstein, T.F.Koetzle, G.J.Williams, E.E.Meyer Jr., M.D.Brice, J.R.Rodgers, O.Kennard, T.Shimanouchi, M.Tasumi, "The Protein Data Bank: A Computer-based Archival File For Macromolecular Structures," J. of. Mol. Biol., 112 (1977): 535.
2. Methods Mol Biol. 2014; 1137:1-15. doi: 10.1007/978-1-4939-0366-5_1. Protein structure modeling with MODELLER.
3. Webb B1, Sali A.Altschul S.F., Gish W., Miller W., Myers E.W. and Lipman D.J. (1990) Basic local alignment search tool. J. Mol. Biol. 215: 403-410
4. Leigh Willard, Anuj Ranjan,Haiyan Zhang,Hassan Monzavi, Robert F. Boyko, Brian D. Sykes, and David S. Wishart "VADAR: a web server for quantitative evaluation of protein structure quality" Nucleic Acids Res. 2003 July 1; 31 (13): 3316.3319.
5. Wiederstein & Sippl (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Research 35, W407-W410.
6. Kumar S, Nath O, Govil S, Pathak AN. Computational 3D Structure Prediction, Evaluation and Analysis of Pyruvate Dehydrogenase an Effective Target for Filarial Infection by Brugia pahangi Using Homology Modeling Approach. International Journal of Pharmaceutical Sciences and Drug Research 2014; 6(2): 120-123
7. Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.; Protein Identification and Analysis Tools on the ExPASy Server; (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press (2005). pp. 571-607
8. Lam, T. T., Nguyen, T.-P. K., Fikrig, E. & Flavell, R. A. (1994). A chromosomal Borrelia burgdorferi gene encodes a 22-kilodalton lipoprotein, P22, that is serologically recognized in Lyme disease. J Clin Microbiol 32, 876–883.
9. Guex, N. and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis 18, 2714-2723
10. Roger Sayle and E. James Milner-White. "RasMol: Biomolecular graphics for all", Trends in Biochemical Sciences (TIBS), September 1995, Vol. 20, No. 9, p. 374. Herbert J. Bernstein, "Recent changes to RasMol, recombining the variants, Trends in Biochemical Sciences (TIBS), September 2000, Vol. 25, No. 9, pp. 453-455
11. Kumar, S., A. Tyagi, S. Govil and S. Mishra, 2012. Computational studies on Ribavirin binding to RNA-dependent RNA polymerase, inducing Crimean-Congo hemorrhagic fever. ISABB J. Biotechnol. Bioinformat., 2: 1-5
12. Von Lackum, Kate, Kristina M. Ollison, Tomasz Bykowski, Andrew J. Nowalk, Jessica L. Hughes, James A. Carroll, Wolfram R. Zückert, and Brian Stevenson. "Regulated synthesis of the Borrelia burgdorferi inner-membrane lipoprotein IpLA7 (P22, P22-A) during the Lyme disease spirochaete's mammal–tick infectious cycle." Microbiology 153, no. 5 (2007): 1361-1371.
13. Casjens SR, Mongodin EF, Qiu W-G, Luft BJ, Schutzer SE, et al. (2012) Genome Stability of Lyme Disease Spirochetes: Comparative Genomics of Borrelia burgdorferi Plasmids. PLoS ONE 7(3): e33280. doi:10.1371/journal.pone.003328.

ANNEXURE

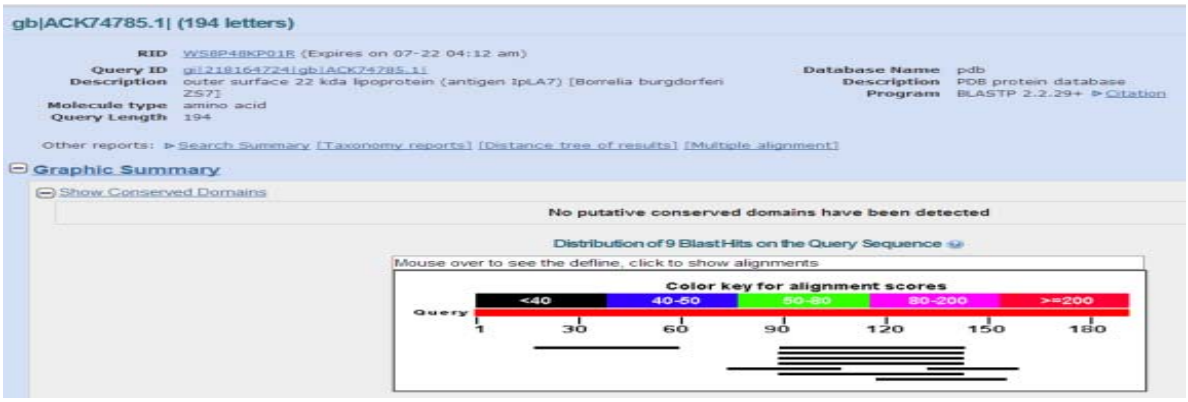

Figure 1: BLASTp graphical view of hits for target protein IpLA7.

Figure 2: Modeled target structure of protein IPLA7 using RASWIN. [10]

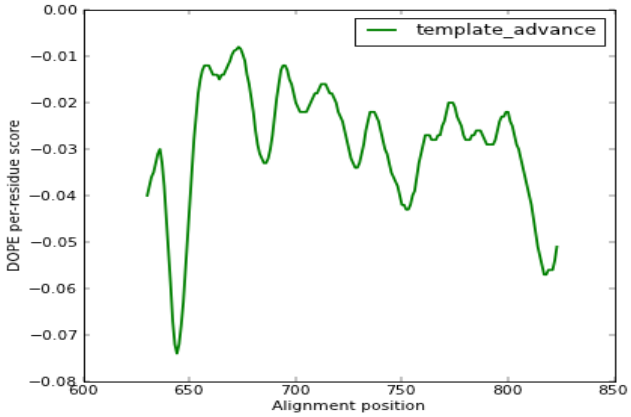

Figure 3: Energy Plot of lowest molpdf score structure.

STRUTCURE NO.	BONDS	ANGLES	TORSION	IMPROPER	NONBONDED	ELECTROSTATIC	CONSTRAINT	TOTAL
antigenIpLA.B99990004	1532.894	1295.214	715.959	438.097	5126.25	-1425.33	0	7683.081
ENERGY_MIZ_1	486.569	893.368	822.042	354.535	-2448.94	-725.16	0	-617.59
ENERGY_MIZ_2	227.765	779.903	883.69	381.872	-3282	-2135.77	0	-3144.55

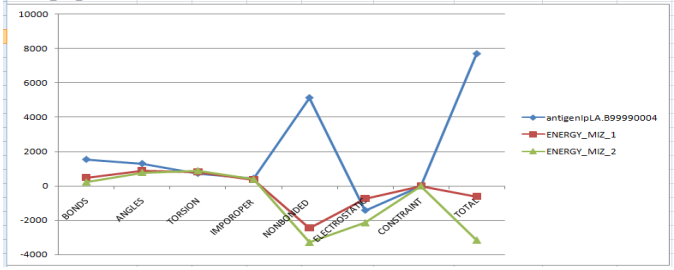

Figure 4: Comparisons of energies of energy minimized structure from SPDBV. [9]

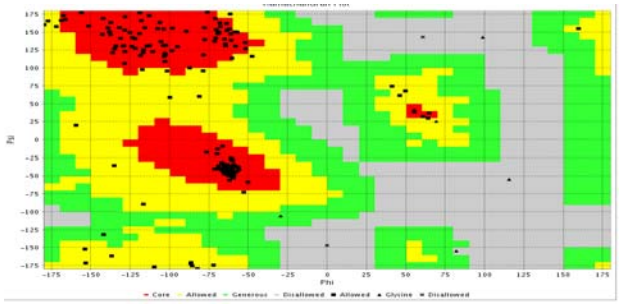

Figure 5: Ramachandran plot by VADAAR server.

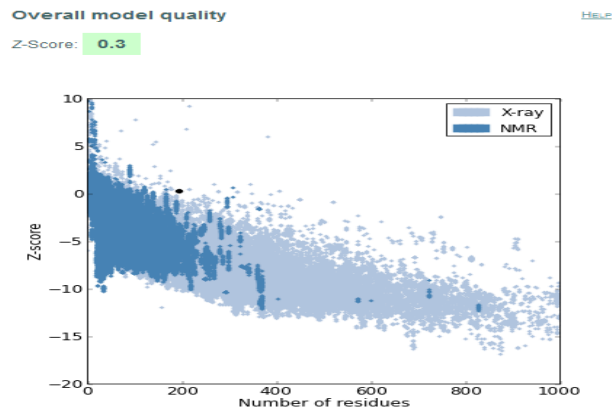

Figure 6: z score from PROSA web server before energy minimization of modeled structure.

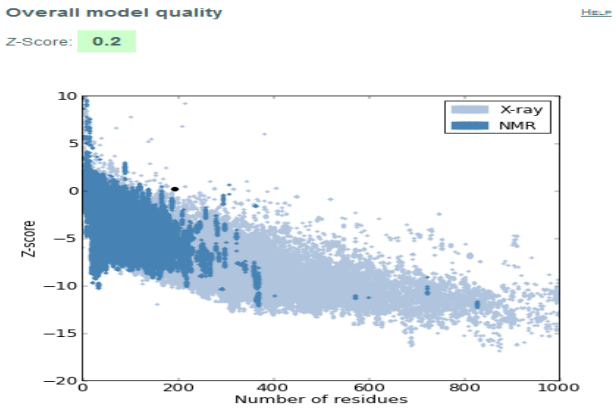

Figure 7: z score from PROSA web server after energy minimization of modeled structure.

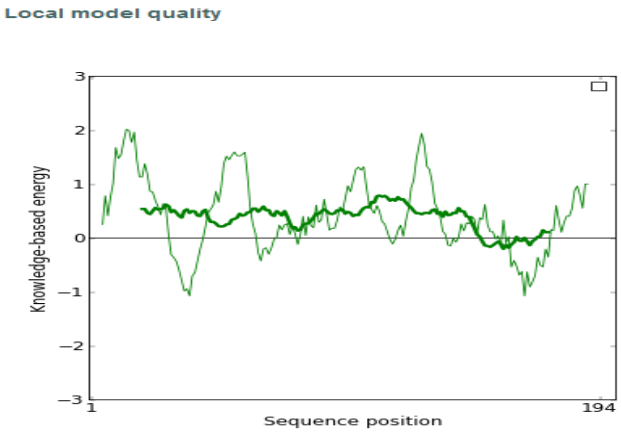

Figure 8: Energy of local model quality from PROSA.

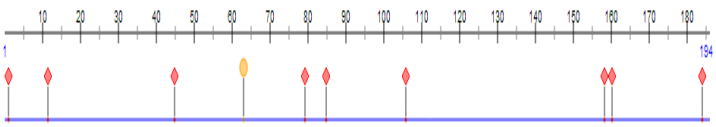

Figure 9: (protein and polynucleotide binding region prediction (http://ppopen.informatik.tu-muenchen.de/)