JOURNAL OF PURE AND APPLIED MICROBIOLOGY, June 2015. Vol. **9**(2), p. 919-926

Computational Approaches Predict the Reliable Three Dimensional (3D) Structure and Possible Involvement of a Hypothetical Protein of *Mycobacterium leprae*, ML-1369 in Its Cell Division

Sudip Paul[1*], Moumoni Saha[1], Sattya Narayan Talukdar[2], Susmita Rajbongshi[3], Rubait Nazneen Akhand[4] and Md. Aminul Islam[1]

[1]Department of Biochemistry and Molecular Biology, Jahangirnagar University, Savar, Dhaka-1342, Bangladesh.
[2]Department of Biochemistry, Primeasia University, Banani, Dhaka-1213, Bangladesh.
[3]Department of Pharmacy, LM College of Pharmacy, Gujrat Technological University, Gujrat, India.
[4]Department of Chemistry and Biochemistry, Sylhet Agricultural University, Sylhet-3100, Bangladesh.

(Received: 12 January 2015; accepted: 28 March 2015)

***In silico* structural and functional analyses of hypothetical proteins may unravel the hidden information seeded inside the protein sequence. This information could be utilized in finding new drug targets and drug designing. In the current study, we used different bioinformatics algorithms to conduct structural and functional annotations of a hypothetical protein in *Mycobacterium leprae*, ML-1369. In brief, at first the sequence analysis and secondary structure prediction were conducted and found that the studied hypothetical protein is a stable hydrophilic protein with a significant proportion of alpha helix and random coil. The protein was predicted to be a cytoplasmic protein. After that, we built the 3D model of the hypothetical protein using SWISS-MODEL from the complete amino acid sequence by homology modeling method. Then, we employed several quality and structural assessment programs such as ERRAT, Q-MEAN, ProSA, Procheck and found that the generated 3D model was structurally good and reliable. We also found that the protein contained the domain of unknown function (DUF) 387 and might belong to chromosomal condensation and segregation protein ScpB. STRING analysis also suggested the interaction of this protein with several others in cell division. The hypothetical protein is essential for the organism, although showed no homology to any human protein. Furthermore, KEGG analysis recommended the involvement of the protein in genetic information processing, environmental information processing, cellular processes and organisomal systems of *M. leprae*. The findings of the study will help better understanding of the functional protein network in the *M. leprae* that might lead to valuable drug development.**

Key words: Hypothetical protein, *Mycobacterium leprae*, ML-1369, Cell division.

Mycobacterium leprae is a Gram positive, rod-shaped acid fast bacterium that causes leprosy, also known as Hansen's disease. Still suffering thousands of people worldwide, leprosy is an infectious disease mainly affecting skin, peripheral nerves and mucosa[1, 2]. In 2012, the total number of chronic cases of leprosy recorded worldwide was 189,000 and the number of fresh cases was 230,000[3]. Due to the inherent characteristics of *M. leprae*, i.e., slow growth and weak pathogenicity, leprosy has taught to be still prevalent. It takes 12 to 14 days for *M. leprae* to replicate, so it is predicted that 2 to 5 years are necessary for the clinical manifestations to appear after an infection[4, 5].

A significant reduction in the number of registered leprosy patients had been observed by deployment of the intensive chemotherapy program. However, a frightening number of new leprosy cases have been reported even after the introduction of multidrug therapy (MDT) by the

* To whom all correspondence should be addressed.
Tel: +880-1674389745;
E-mail: sudippaul.bcmb@gmail.com

World Health Organization (WHO)[6]. Although MDT can kill the bacilli within a short time without relieving symptoms, the dead cells could lead to several leprosy symptoms. Nowadays, the MDT based control measures for treating leprosy are mainly intended to prevent the drug-resistant *M. leprae* from spreading [7].

The first genome sequence of a strain of *M. leprae*, was completed in 1998[8]. The genome sequence of a strain originally isolated in Tamil Nadu, India and designated as TN, was completed in 2013. Cole *et al*[9] have found that, the *M. leprae* genome of 3,268,203 bp encodes only 1,604 proteins and contains 1,116 pseudogenes. They have annotated and classified all these genes into various functional categories. The complete genome contains about 360 conserved hypothetical and 141 unknown proteins[10].

The hypothetical proteins are so named because it is unclear whether their genes encode actual proteins. These are also referred to as "uncharacterized" or "unknown" proteins[11]. As of April 25, 2006, the NCBI protein database contained 19,85,480 protein sequences from ~373 completely sequenced genomes; one out of three proteins had no assigned function and one out of ten proteins was annotated as "conserved hypothetical"[12]. Therefore, it is clear that a large portion of the different genomes remains unrevealed at time. However, the proper structural and functional annotation of hypothetical proteins may lead to better understanding of the complex protein cascades, protein-protein interaction or networks in several biological systems[13]. This understanding will be helpful in both diagnostic and therapeutic purposes[14].

Protein characterization is important for determining the current state of any protein, which has significant implications to several biological processes. The major shortcomings of experimental methods that have been used to characterize the proteins of various organisms are time consuming, costly and not amenable to high throughput techniques [15]. Computational or *in silico* approaches provide a viable solution to these problems. The amino acid sequence provides most of the information required for determining and characterizing molecule's function along with physical and chemical properties. Computational based characterization of proteins found or predicted in completely sequenced proteomes is an important task in search for the knowledge of protein function[16].

In this study, for the first time we analyzed a hypothetical protein ML-1369 of *M. leprae* by *in silico* methods. Different computational tools were used to generate the three-dimensional (3D) model of the protein. Moreover, the functional and comparative genomic analyses were also conducted to propose the possible functions of the protein and research in drug design planning.

METHODS

Protein sequence retrieval

The sequence of the screened protein was recovered from the Uniprot databse (http://www.uniprot.org/).

Physiochemical analysis and subcellualr localization prediction

The primary sequence analysis was conducted using Protparam server (http://web.expasy.org/protparam/). The parameters computed by ProtParam included the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index, and grand average of hydropathicity (GRAVY)[17]. Subcellular localization of ML-1169 was envisaged using CELLO[18] and confirmed by Psortb[19].

Secondary structure prediction

The secondary structure of the query protein was analyzed using SOPMA[20] and PSIPRED server[21, 22].

Homology modeling, energy minimization and validation

The 3D model for the protein was generated using SWISS-MODEL server by homology modeling method[23, 24]. Once the 3D model was generated, energy minimization was performed by GROMOS96 force field in a Swiss-PdbViewer [25]. The model generated was checked for errors and reliability by ERRAT in SAVES server[26]. The quality of the model was also assessed by QMEAN[27] and ProSA web[28]. The stereochemistry of the structure was analyzed by PROCHECK [29]. The modeled structure was

visualized by UCFS chimera 1.8.1[30]. The generated structure was superimposed on the template structure using UCFS chimera for RMSD values that indicate similarities between two structures.

Functional annotations

The domain analysis of the protein was performed by Interproscan[31], Pfam[32] and NCBI-CDD[33].

Protein-protein interaction analysis

Protein residues interact with each other for their accurate functions. In this study, we employed the STRING v9.1 database (http://string-db.org/) of known and predicted proteins interactions. Currently, this database covers 5,214,234 proteins from 1133 organisms[34].

Comparative genomics analysis

BLASTP search was performed against human proteome to find whether or not the protein is homologous to any human protein. Database of essential genes (DEG) is employed to check whether the queried protein is essential for the bacterium or not[35]. To check the involvement of ML-1369 into metabolic pathways of *M. leprae*, KEGG automatic annotation server (KAAS) was used[36].

Submission of the model in protein model database (PMDB)

The model generated for *M. leprae* hypothetical protein ML-1369 was successfully submitted in protein model database (PMDB) (http://bioinformatics.cineca.it/PMDB/).

RESULTS AND DISCUSSION

The complete protein sequence of the hypothetical protein ML-1369 in FASTA format was successfully retrieved from the Uniprot database.

After that we calculated the theoretical physiochemical characteristics of the hypothetical protein ML-1369 by Expasy's Protparam server. The protein was predicted to be consisting of 231 amino acids, with a molecular weight of 25329.1 Daltons and an isoelectric point (P^I) of 4.77 indicating a negatively charged protein. The instability index of the protein was computed to be 27.70, indicating this protein as stable[37]. The negative grand average of hydropathicity (GRAVY) index of -0.009 is an indicator of a hydrophilic and soluble protein[38].

The most abundant amino acid residue was found to be leucine (13.9%), followed by alanine (12.4%). The sequence had 37 negatively charged residues (Aspartic acid+Glutamic acid) and 27 positively charged residues (Arginine+Lysine). The aliphatic index reflecting the relative volume occupied by the side chains of amino acids (alanine, leucine, valine and isoleucine) of the query protein was -106.45[39]. The extinction coefficient indicating the light

Table 1. Parameters used for structural assessment of model and template proteins

Protein	Structural assessment methods	Ramachandran plot (%)	Z-score	Q value	RMSD (A°)
ML-1369	PROCHECK analysis	90.9%[A] 7.7[B] 1.4[C] 0.0[D]			
	ProSA		-5.33		
	QMEAN			0.695	
	RMSD				0.000
2Z99	PROCHECK analysis	92.4[A] 6.2[B] 0.7[C] 0.7[D]			
	ProSA		-5.27		
	QMEAN			0.744	
	RMSD				0.067[E]

[A]Residues in favorable regions; [B]residues in allowed regions; [C]residues in generally allowed regions; [D]residues in disallowed regions; [E]root mean square deviation between CαML-1369 structure and 2z99 template

Fig. 1. Predicted secondary structure of *Mycobacterium leprae* hypothetical protein ML-1369 by PSIPRED server.

absorption capacity was found to be 11460 M^{-1} cm^{-1}, at 280 nm measured in water for the queried protein[40].

We found that the predicted subcellular localization of the query protein is cytoplasm. It was analyzed by CELLO and validated by Psortb v 2.0 sever. The secondary structure of the protein was predicted by SOPMA and PSIPRED server (Figure 1). The alpha helix was found to be the most predominant (48.05%), followed by random coil (31.17%) and extended strand (11.69%). Also, beta turn was found as 9.09%.

In the study, we used SWISS-MODEL web server to identify the best possible template

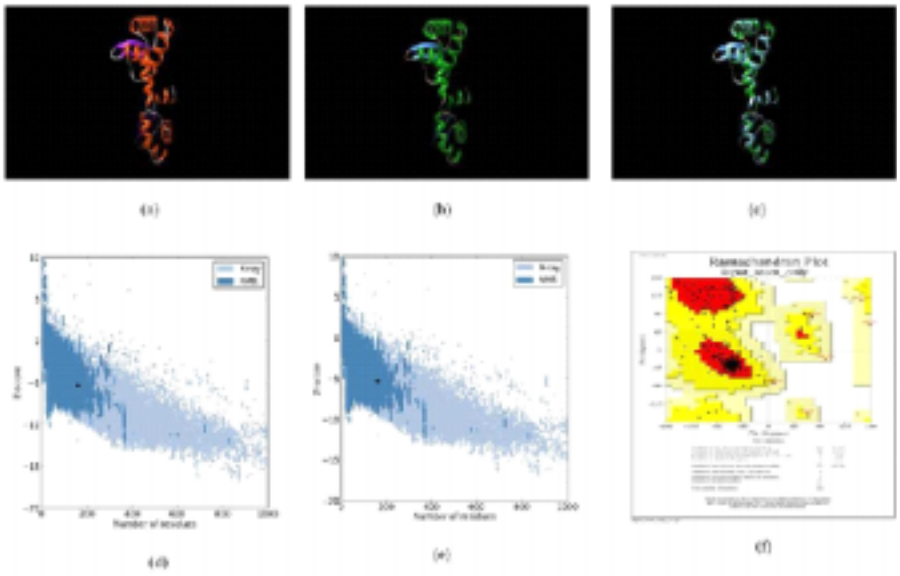

Fig. 2. Structural analysis of *Mycobacterium leprae* hypothetical protein ML-1369. (a) the predicted 3D model of ML-1369; (b) the 3D model of the template protein (2z99) used for homology modeling; (c) Superimposed structure of hypothetical and template proteins; (d) and (e) the Z-scores for the hypothetical and template proteins respectively; (f) the representative Ramachandran plot for ML-1369.

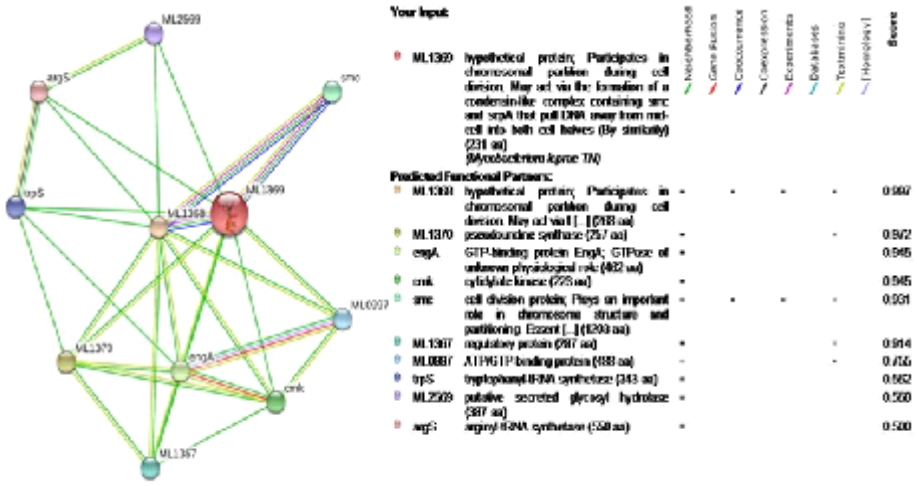

Fig. 3. STRING network representing the predicted functional partners of the hypothetical protein ML-1369.

for homology modeling. Hence, the 2z99.1.A template with 80.56% target-template sequence identity was used (GMQE: 0.67; QMEAN4: -1.61 Sequence similarity: 0.52). The template was X-ray diffraction model of the crystal structure of putative uncharacterized protein ScpB from *Mycobacterium tuberculosis* with a resolution of 2.30 A°.

To obtain an accurate homology model, it is very important that appropriate steps are built into the process to assess the quality of the model. Therefore, the accuracy of the predicted models was checked through a series of tests such as the PROCHECK, ERRAT, ProSA, and QMEAN methods. UCSF Chimera and Swiss-PdbViewer were used to view the models and image preparation.

As indicated by the Ramachandran plot (Figure 2(f)), 90.9% residues in ML-1369 model were within the most favored regions, 7.7% residues were in the additional allowed region, 1.4% residues were in the generously allowed regions and 0.0% residues were in the disallowed regions; 92.4% residues in 2z99.1 (coordinate file) template were within the most favored regions, 6.2% residues werein the additional allowed regions, 0.7% residues were in the generously allowed regions and 0.7% residues were in the disallowed regions (Table 1). This indicated that the backbone dihedral angles, phi and psi, in the ML-1369 3D model, were reasonably accurate.

As indicated by the ERRAT program, the result showed that the overall quality factor is 95.395 for ProSA and 90.132 for 2z99 meaning that both of the two structures have good high resolution. ProSa analysis showed Z-scores of 5.27 and 5.33 for ML-1369 and 2z99 (Table 1), respectively. Q values for ML-1369 and 2z99 structures were found to be 0.695 and 0.774, respectively. Structural similarity was further compared by superimposition of modeled structure with template. The modeled structure of the hypothetical protein closely resembled the template structure (2z99.1) and it had good similarity with the template upon superimposition as reflected by root mean square deviations (RMSD) value of 0.067 A°(Figure 2(c) and Table 1). Based on these validations, it is shown that the homology model was adopted for this study.

In the present study, three web servers were employed to find out the conserved domains and potential function of ML-1369. Based on consensus predictions made by Pfam, NCBI-CDD and InterProScan, ML-1369 was suggested to contain domain of unknown function DUF 387 Putative transcriptional regulators (Ypuh-like) by Pfam server. This family of conserved bacterial proteins is thought to possibly be helix-turn-helix type transcriptional regulators. InterProscan server showed that the queried hypothetical protein might belong to the chromosome segregation/condensation protein ScpB. This family represents ScpB, which along with ScpA, interacts with SMC *in vivo* forming a complex essential for chromosome condensation and segregation. From this, it could be suggested that the hypothetical protein may participate in chromosomal partition during cell division. NCBI-CDD also showed the similar results found by aforementioned servers.

To visualize the protein-protein interaction network of the protein ML-1369, STRING was employed, and the obtained network is shown in Figure 3. STRING interaction network revealed that our targeted protein (ML-1369) interacted with 10 different proteins for its functioning. The target protein showed interactions with hypothetical proteins, proteins required for cell division, regulatory proteins and many others.

Comparative genomics analysis of the hypothetical protein was then conducted to further characterize ML-1369. The hypothetical protein was found to be a unique protein of *M. leprae* and showed no homology to any of the human proteins by performing a BLASTP search against human proteome. The protein was also found to be essential for *M. leprae* by performing microbial BLASTP search in DEG database. Essential proteins of a pathogen regulate key factors, such as metabolism, nutrient uptake, virulence and pathogenicity. Therefore, pathogenic proteins that fulfill the characteristics of being both unique and essential at the same time represent more striking drug targets as these will not show any side effects as well as offer immense potentials in disrupting pathogen functions and existence[41].

Finally, we utilized KEGG to identify the involvement of ML-1369 in *M. leprae* metabolic pathways. Based on search performed via KAAS, ML-1369 was found to be involved in genetic

information processing, environmental information processing, cellular processes and organismal systems of the organism. Furthermore, the generated 3D model of the hypothetical protein was successfully submitted to the PMDB and assigned the PMDB id PM0079888.

CONCLUSION

The study was designed to predict the 3D structure and biological function of the hypothetical protein ML-1369 of *M. leprae*. Our study proposed a reliable and good quality 3D structure of the hypothetical protein. Moreover, a number of analyses suggested the possible involvement of the target protein in chromosomal partitioning during cell division. Furthermore, the essentiality of the protein for the organism and uniqueness among different proteomes make it a good candidate for drug target. It needs further verification through laboratory experiments to validate all the findings for the protein ML-1369. Similar extensive studies on hypothetical proteins may create breakthrough in the fields of future biomedical research.

REFERENCES

1. Ryan, K.J. and Ray, C.G. (ed.): Sherris Medical Microbiology, 4th edn. New York: McGraw Hill, 2004; pp 451-453.
2. McMurray, D.N. : Mycobacteria and Nocardia. In: Medical Microbiology, 4th edn. (Baron S, ed). Galveston (TX): University of Texas Medical Branch at Galveston, 1996; Chapter 33. Available from: http://www.ncbi.nlm.nih.gov/books/NBK7812/
3. Leprosy Fact sheet N°101. WHO, 2014.
4. Cho, S.N. and Brennan, P.J. New biological tools for leprosy surveillance. *Int. J. Lepr. Other Mycobact. Dis.* , 1999; **67**: 59-62.
5. Salazar, L., Fsihi, H., De Rossi, E., Riccardi, G., Rios, C., Cole, S.T. and Takiff, H.E. Organization of the origins of replication of the chromosomes of Mycobacterium smegmatis, *Mycobacterium leprae* and *Mycobacterium tuberculosis* and isolation of a functional origin from M. smegmatis. *Mol. Microbiol.*, 1996; **20**(2): 283-293.
6. Leprosy—global situation. *Wkly Epidemiol. Rec.*, 2000; **75**(28): 226-231.
7. Maeda, S., Matsuoka, M., Nakata, N., Kai, M., Maeda, Y., Hashimoto, K. and Kashiwabara, Y. Multidrug resistant Mycobacterium leprae from patients with leprosy. *Antimicrob. Agents Chemother.*, 2001; **45**(12): 3635-3639.
8. Cole, S.T., Brosch, R. and Parkhill, J. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature*, 1998; **393**(6685): 537-544.
9. Cole, S.T., Eiglmeier, K., Parkhill, J., James, K.D., Thomson, N.R., Wheeler, P.R., Honore, N., Garnier, T., Churcher, C. and Harris, D. Massive gene decay in the leprosy bacillus. *Nature*, 2001; **409**(6823): 1007-1011.
10. Vissa, V.D. and Brennan, P.J. The genome of Mycobacterium leprae: a minimal mycobacterial gene set. *Genome Biol.*, 2001; **2**(8): 1023.1-1023.8.
11. Sivashankari, S. and Shanmughavel, P. Functional annotation of hypothetical proteins–A review. *Bioinformation*, 2006; **1**(8): 335-338.
12. Kolker, E., Makarova, K.S., Shabalina, S., Picone, A.F., Purvine, S., Holzman, T., Cherny, T., Armbruster, D., Munson, R.S. Jr., Kolesov, G., Frishman, D. and Galperin, M.Y. Identification and functional analysis of 'hypothetical'genes expressed in Haemophilus influenzae. *Nucleic Acids Res.*, 2004; **32**(8): 2353-2361.
13. Loewenstein, Y., Raimondo, D., Redfern, O.C., Watson, J., Frishman, D., Linial, M., Orengo, C., Thornton, J. and Tramontano, A. Protein function annotation by homology-based inference. *Genome Biol.*, 2009; **10**(2): 207.
14. Lubec, G., Afjehi-Sadat, L., Yang, J.W. and John, J.P. Searching for hypothetical proteins: theory and practice based upon original data and literature. *Prog. Neurobiol.*, 2005; **77**(1): 90-127.
15. Alterovitz, G., Afkhami, E. and Ramoni, M. : Robotics, automation, and statistical learning for proteomics. In: New Developments in Robotics Research (Liu JX, ed). New York: Nova Science Publishers, Inc., 2005; pp 217-252.
16. Sahay, A. and Shakya, M. In silico analysis and homology modelling of antioxidant proteins of spinach. *J. Proteomics Bioinform.*, 2010; **3**: 148-154.
17. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D. and Bairoch, A. : Protein identification and analysis tools on the ExPASy server. In: The Proteomics Protocols Handbook (Walker JM, ed). New York: Springer, 2005; pp 571-607.
18. Yu, C.S., Chen, Y.C., Lu, C.H. and Hwang, J.K. Prediction of protein subcellular localization. Proteins: Structure, Function, and *Bioinformatics*, 2006; **64**(3): 643-651.

19. Yu, N.Y., Wagner, J.R., Laird, M.R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S.C., Ester, M., Foster, L.J. and Brinkman, F.S. PSORTb 3.0: improved 713 protein subcellular localization prediction with refined localization subcategories and 714 predictive capabilities for all prokaryotes. *Bioinformatics*, 2010; **26**: 1608-1615.
20. Geourjon, C. and Deleage, G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics*, 1995; **11**(6): 681-684.
21. Jones, D.T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.*, 1999; **292**(2): 195-202.
22. Buchan, D.W.A., Minneci, F., Nugent, T.C.O., Bryson, K. and Jones, D.T. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res.*, 2013; **41**(W1): W349-W357.
23. Arnold, K., Bordoli, L., Kopp, L. and Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, 2006; **22**(2): 195-201.
24. Kiefer, F., Arnold, K., Kunzli, M., Bordoli, L. and Schwede, T. The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res.*, 2009; **37**(suppl 1): D387-D392.
25. Guex, N. and Peitsch, M.C. SWISS MODEL and the Swiss Pdb Viewer: an environment for comparative protein modeling. *Electrophoresis*, 1997; **18**(15): 2714-2723.
26. Colovos, C. and Yeates, T.O. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.*, 1993; **2**(9): 1511-1519.
27. Benkert, P., Tosatto, S.C. and Schomburg, D. QMEAN: A comprehensive scoring function for model quality assessment. *Proteins: Struct., Funct., Bioinf.*, 2008; **71**(1): 261-277.
28. Wiederstein, M. and Sippl, M.J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.*, 2007; **35**(suppl 2): W407-W410.
29. Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, 1993; **26**(2): 283-291.
30. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.*, 2004; **25**(13): 1605-1612.
31. Zdobnov, E.M. and Apweiler, R. InterProScan–an integration platform for the signature-recognition methods in InterPro. *Bioinformatics*, 2001; **17**(9): 847-848.
32. Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunasekaran,P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E.L., Eddy, S.R. and Bateman, A. The Pfam protein families database. *Nucleic Acids Res.*, 2010; **38**(suppl 1): D211-D222.
33. Marchler-Bauer, A., Lu, S., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Lu, F., Marchler, G.H., Mullokandov, M., Omelchenko, M.V., Robertson, C.L., Song, J.S., Thanki, N., Yamashita, R.A., Zhang, D., Zhang, N.N., Zheng, C. and Bryant, S. H. CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.*, 2011; **39**(suppl 1): D225-D229.
34. Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering, C. and Jensen, L. J. STRING v9. 1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.*, 2013; **41**(D1): D808-D815.
35. Zhang, R., Ou, H.Y. and Zhang, C.T. DEG: a database of essential genes. *Nucleic Acids Res.*, 2004; **32**(suppl 1): D271-D272.
36. Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C. and Kanehisa, M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.*, 2007; **35**(suppl 2): W182-W185.
37. Guruprasad, K., Reddy, B.B. and Pandit, M.W. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Eng.*, 1990; **4**(2): 155-161.
38. Kyte, J. and Doolittle, R.F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.*, 1982; **157**(1): 105-132.
39. Atsushi, I. Thermostability and aliphatic index of globular proteins. *J. Biochem.*, 1980; **88**(6): 1895-1898.
40. Gill, S.C. and Von Hippel, P.H. Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.*, 1989; **182**(2): 319-326.
41. Butt, A.M., Batool, M. and Tong, Y. Homology modeling, comparative genomics and functional annotation of Mycoplasma genitalium hypothetical protein MG_237. *Bioinformation*, 2011; **7**(6): 299-303.