An *in silico* Approach for Structural and Functional Annotation of *Salmonella enterica* serovar *typhimurium* Hypothetical Protein R_27

Arif Khan[1], Hossain Ahmed[2], Nasreen Jahan[3], Syed Raju Ali[4], Al Amin[4], Mohammad Neaz Morshed[1*]

[1]*Department of Science and Humanities*
Military Institute of Science and Technology (MIST)
Mirpur Cantonment, Dhaka 1216, Bangladesh
Emails: arifkhanbge35@gmail.com, mmneaz@hotmail.com

[2]*Department of Biotechnology and Genetic Engineering*
University of Development Alternative
Dhanmondi, Dhaka 1209, Bangladesh
Email: ahmedshohag.bd@gmail.com

[3]*Department of Biotechnology and Genetic Engineering*
Mawlana Bhashani Science and Technology University
Santosh, Tangail 1902, Bangladesh
Email: naza.bge@gmail.com

[4]*Department of Biotechnology*
Bangladesh Agricultural University
Mymensingh 2202, Bangladesh
Emails: m15121307@bau.edu.bd, josnasnan@gmail.com

[*]*Corresponding author*

Received: October 11, 2015 **Accepted: February 15, 2016**

Published: March 31, 2016

Abstract: *Typhoid fever is a major cause of illness in most developing countries, including Bangladesh. In quest of new potential drug against Typhoid fever, the current study was designed to elucidate structural and functional details of S. typhi hypothetical protein (HP) R_27. HP R_27 has the primary amino acid sequences available only. The structural annotation was determined by ProtParam, SOPMA, and CELLO. The three-dimensional (3D) structure of HP R_27 predicted through homology modeling by using Phyre2. The 3D structure then refined and verified by ModRefiner, PROCHECK, ERRAT, QMEAN. The functional annotation was also performed by InterProScan, SMART, Pfam, NCBI-CDD and found Phospholipase D-like and DNA repair activity. Multiple sequence alignment also supported the existence of PLD-like domain and DNA repair protein domain in the selected hypothetical protein sequences. Finally, the cavity of drug binding was also identified to assist further molecular docking study and potent inhibitor identification. This in silico approach can be further utilized in molecular drug design for other clinically significant pathogens.*

Keywords: *Typhoid fever, Hypothetical protein, Homology modeling, PLD-like activity.*

Introduction

Salmonella enterica serovar *typhimurium* (*S. typhi*) is an enteric gram-negative bacteria agent of typhoid fever and other related clinical diseases that are distributed worldwide mainly in regions with limited sanitary conditions [28]. Typhoid fever is a serious invasive bacterial

disease of humans with an annual global burden of approximately 16 million cases, leading to 600,000 fatalities [31].

Typhoid fever caused an anticipated 21.7 million cases and 217,000 deaths in 2000. Adolescent's children and infants in south-central and Southeast Asia experience the most burden of illness. About 400 incident cases are accounted and the fever is projected to occur in about 6,000 people per year in the United States [7, 27, 38]. Invasive strains of non-typhoidal salmonella, for instance *S. typhi* ST313 have recently been labeled as rising diseases in Africa. Associations among malaria, HIV and malnutrition, have thrown in to spread of this fever, explicit the necessity of using costly antimicrobial drugs in the poorest health services in the world [10].

Typhoid fever is take epidemic proportions in Bangladesh as it is a round the year problem. The specific reason for such kind of occurrence are having tainted water, unhygienic food managing practice and flawed sewerage system. A high frequency of multidrug resistant typhoid fever has been reported. So, new and cost effective drug development is a basic need for developing country like Bangladesh [26, 29, 33].

In more recent time, genome sequences of different organisms are available on various databases by utilizing next generation sequencing technology to gather information to a greater extent. As a result, increasing amounts of information about hypothetical proteins deposited in sequence databases rather than experimentally isolated data in Protein Data Bank (PDB) [3, 4]. Hypothetical proteins are generally forecasted to be expressed from an open reading frame (ORF). These proteins have no experimental evidence regarding their functions [20]. At present, it is assumed that 50% proteins of a genome are hypothetical proteins [35]. This encourages *in silico* study of a hypothetical protein utilizing experimental data [30].

In the current study, the *S. typhi* hypothetical protein (HP) R_27 was selected as the primary amino acid sequences is available but there have no structural and functional details. *S. typhi* HP R_27 was subjected to extensive *in silico* study to determine its molecular and structural properties. Further, we tried to predicted good quality model of HP R_27 using protein homology modeling techniques and successive computer aided active site prediction for the development of an effective drug against *S. typhi*.

Materials and methods

Sequence retrieval

The UniProt Knowledge Base (UniProtKB) database for *S. typhi* was primarily explored to find out hypothetical proteins with possible interest in research and application. The amino acid sequence of the HP R_27 was retrieved from the Uniprot database (http://www.uniprot.org/). UniProtKB is used to collect functional information on any kind of proteins, with accurate and rich annotation [24]. The primary accession ID of the selected protein is Q9L5J1 and it contains 259 amino acids. The amino acid sequence was then collected as a FASTA format for further proteomic analysis.

Structural annotation

A tool of ExPASy server, ProtParam (http://web.expasy.org/protparam/), was used for the analysis of the physicochemical properties from the retrieved protein sequence. This tool can predict physicochemical properties for instance the molecular weight, theoretical pI, aliphatic index, amino acid composition, grand average of hydropathicity (GRAVY), instability index, and extinction coefficients.

Self-optimized prediction method with alignment (SOPMA) was used for secondary structure prediction [12]. Protein's secondary structural properties are including α helix, pI helix, 3_{10} helix, Beta bridge, Extended strand, Bend region, Beta turns, Ambiguous states, Random coil and Other states. In addition, the PSIPRED ((http://bioinf.cs.ucl.ac.uk/psipred/) [5] server was also exploited to authenticate the results found from SOPMA.

Subcellular localization of R_27 was predicted by CELLO v.2.5 [39]. Results were also cross-checked with subcellular localization predictions obtained from PSORTb version 3.0.2 [40], TBpred [34], Signal Peptide (Signal P 4.1) [9], Secretory Protein (Secretome P) [1], HMMTOP [21], TMHMM [36], SOSUI [15].

Homology modeling

Phyre2 (Protein Homology/Analogy Recognition Engine) is a protein fold identification server [17]. It was used to predict the three-dimensional (3D) homology model of HP R_27. The input data of this protein was in FASTA format. After homology modeling, it is necessary to refine the structure for better accuracy.

ModRefiner (http://zhanglab.ccmb.med.umich.edu/ModRefiner/) is an algorithm for high-resolution protein structure refinement, which used to refine the predicted protein model. Amino acid sequences were given in the FASTA format. Model refinement was done for couple of times to get the best structure [36]. The predicted model was evaluated to verify the stereo chemical quality with PROCHECK by Ramachandran plot which was done through "Protein structure and model assessment tools" [21]. The best model was selected from PROCHECK depending on overall G-factor, number of amino acids in different regions. ERRAT and QMEAN were used for further analysis of the selected protein structure [2, 6].

Functional annotation

Hypothetical protein R_27 was analyzed due to find the presence of conserved domains based on sequence similarity search with close orthologous family members. Four bioinformatics tools and databases including InterProScan [40], Simple modular architecture research tool (SMART) [22], Proteins Families Database (Pfam) [11], and NCBI Conserved Domains Database (NCBI-CDD) [24] were used for functional annotation.

Active site identification

Active site is defined as a definite region of protein which is responsible for its functional activity. This consists of several amino acid residues. The active site of the selected hypothetical protein predicted by Computed Atlas of Surface Topography of proteins (CASTp) [8].

Results and discussion

UniProtKB delivers an authoritative resource for protein sequences and functional information. Sequences of HP R_27 of *S. typhi* were obtained from UniPortKB.

The blastp result against non-redundant database showed homology with DNA repair protein and Phospholipase D (PLD)-like protein (Table 1). Pfam server predicted the PLD-like domain at 14-135 amino acid residues with an e-value of 2.6e-11. The PLD-like domain domain was also found in InterProScan server and NCBI-CDD at 14-116 amino acid residues.

Table 1. Analogous proteins obtained from nonredundant database

Entry name	Organism	Protein name	Score	E-value
gi|742850038|	*Escherichia coli*	DNA repair protein	343	1e-116
gi|633894482|	*Klebsiella pneumoniae*	DNA repair ATPase	326	3e-108
gi|490126160|	*Escherichia coli*	PLD-like domain protein	325	4e-108
gi|695806886|	*Klebsiella oxytoca*	DNA repair protein	325	1e-107
gi|490302223|	*Klebsiella pneumoniae*	PLD-like domain protein	324	1e-107
gi|608787356|	*Raoultella planticola*	DNA repair ATPase	317	7e-107
gi|779902985|	*Vibrio campbellii*	DNA repair protein	224	9e-69
gi|659054291|	*Vibrio parahaemolyticus*	DNA repair protein	223	3e-68
gi|701174448|	*Vibrio harveyi*	DNA repair protein	222	4e-68

The ExPASy's ProtParam server was used to analyze the theoretical physiochemical characteristics from the plain amino acid sequence of the hypothetical protein R_27. The protein was predicted to be consisting of 259 amino acids, with a molecular weight of 29617.6 Daltons and pI of 5.92 indicating a negatively charged protein. The instability index of the protein was computed to be 23.17, classified this protein as stable. The negative GRAVY index of -0.531 is indicative of a hydrophilic and soluble protein. The most abundant amino acid residue was found to be Leucine (34), followed by Lysine, Asparagine and Glutamate (24 each). The sequence had 33 positively charged residues and 36 negatively charged residues. The molecular formula of the protein was found as C1315H2108N352O413S5 (Table 2).

Table 2. Different physio-chemical properties of HP R_27

Parameter	Value
Molecular weight	29617.6
Extinction coefficients Abs 0.1% (= 1 g/l) 1.088, assuming all pairs of Cys residues form cystines	29005
Ext. coefficient Abs 0.1% (= 1 g/l) 1.078, assuming all Cys residues are reduced	28880
Theoretical pI	5.92
Total number of negatively charged residues (Asp + Glu)	36.00
Total number of positively charged residues (Arg + Lys)	33.00
Instability index	23.17
Grand average of hydropathicity (GRAVY)	-0.53
Aliphatic index	95.68

The secondary structure of HP R_27 was predicted by SOPMA with standard parameters which are presented in a tabulated form in Table 3 respectively. It contains 35.14% α helix, 21.62% extended strand, 9.65% β turn and 33.59% random coil. The graphical secondary structure presentation of HP R_27 obtained from PSIPRED server is shown in Fig. 1.

Predicting subcellular localization of hypothetical proteins can give information about their cellular functions. This information could be utilized in understanding disease mechanism and developing drugs [41]. It was analyzed by CELLO and authenticated by PSORTb v3.2.0 and

Predict Protein severs. The subcellular localization of the query protein was predicted to be a cytoplasmic protein.

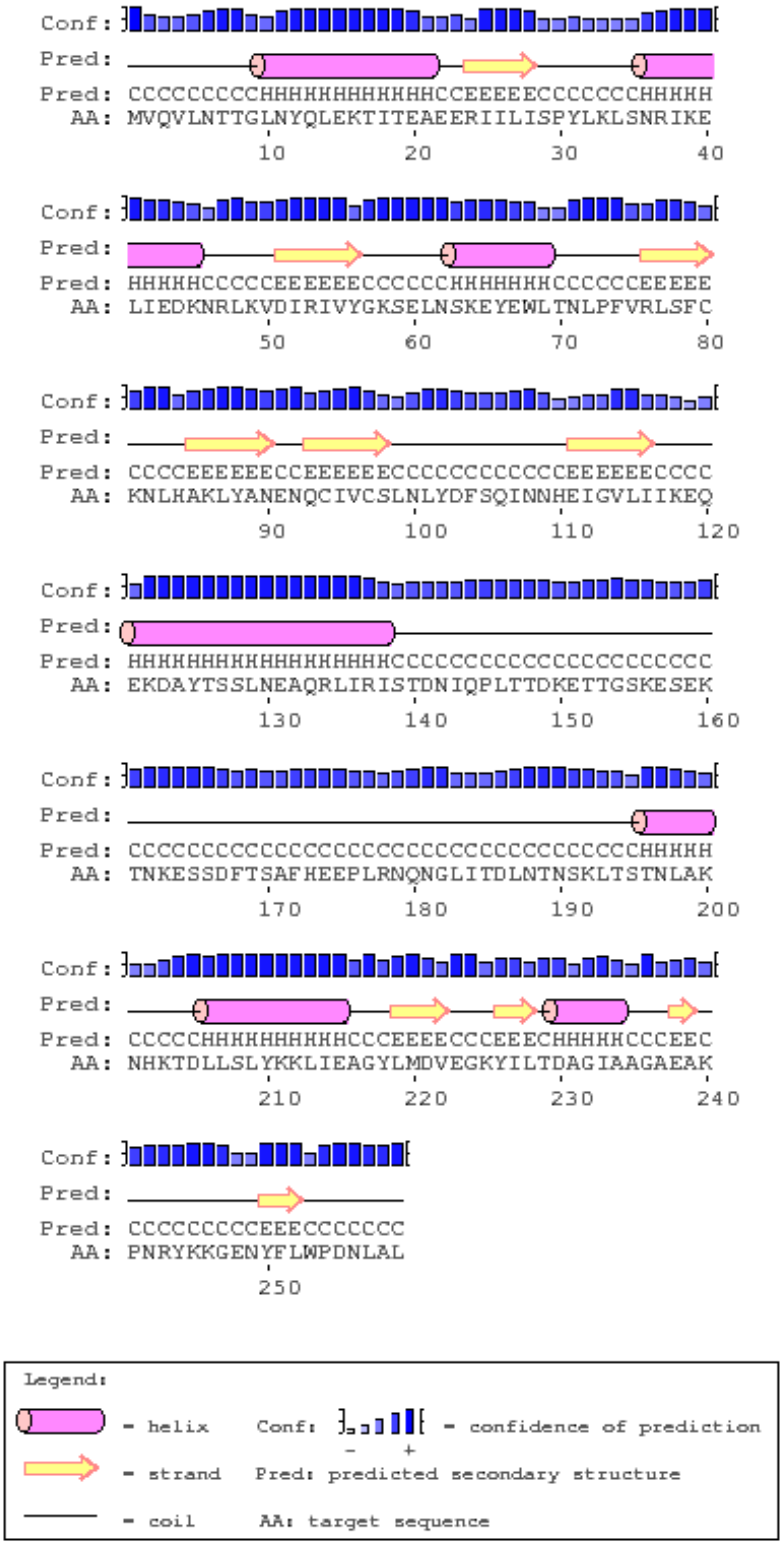

Fig. 1 Predicted secondary structure of HP R_27 generated by PSIPRED server

Table 3. Calculated secondary structure elements of by SOPMA

Secondary structure	No. of residue	Percentage, (%)
α helix (Hh)	91	35.14
3_{10} helix (Gg)	0	0.00
pI helix (Ii)	0	0.00
Beta bridge (Bb)	0	0.00
Extended strand (Ee)	56	21.62
Beta turn (Bt)	25	9.65
Bend region (Ss)	0	0.00
Random coil (Cc)	87	33.59
Ambigous states (?)	0	0.00
Other states	0	0.00

Homology modeling of the selected protein was done by using Phyre2 in order to obtain 3D structure. 3D structure of proteins gives important insights about the molecular basis of function and thereby permits an effective design of experiments [13]. That is why; the high resolution 3D structure of a protein is the main key in the understanding and manipulation of biochemical and cellular functions of proteins, [18]. Modrefiner refine again the Phyre2 generated model. Refined model is depicted in Fig. 2.

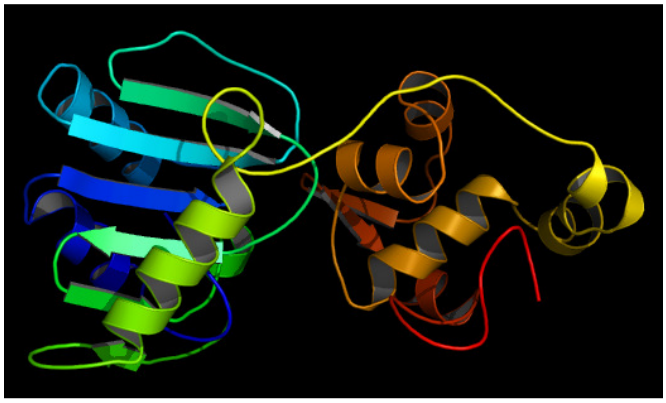

Fig. 2 Refined model of HP R_27 generated by Modrefiner

ModRefiner derived refined model of HP R_27 was analyzed. The percent of residues in different regions remain same in the final model. The predicted structure then validated through Ramachandran plot analysis. The results of this analysis are depicted in Fig. 3 and Table 4. PROCHECK, another tool was used to measure the accuracy of protein models. Ramachandran plot statistics of HP R_27 revealed that 90.04% of amino acid residues were found in the most favored regions. Thus, the selected protein model was very good as all of the amino acid residues were within the limits of Ramachandran plot. Verification was also done by ERRAT and QMEAN server. The respective values *Z*-scores of Cbeta interaction energy, torsion angle energy, solvation energy, secondary structure, and solvent accessibility in case of HP R_27 are -2.06, -2.86, -2.2, -2.26 and -3.62. The overall QMEAN score for HP R_27 is -4.71. QMEAN generated results confers HP R_27 as a qualified model for drug target scopes.

Table 4. Ramachandran plot analysis of HP R_27

Ramachandran plot statistics	Residue	%
Residues in the most favored regions [A, B, L]	217	90.04
Residues in the additional allowed regions [a, b, l, p]	22	9.13
Residues in the generously allowed regions [a, b, l, p]	2	0.83
Residues in the disallowed regions [xx]	0	0.00
Number of non-glycine and non-proline residues	241	100.0
Number of end residues (excl. Gly and Pro)	2	-
Number of glycine residues	10	-
Number of proline residues	6	-
Total number of residues	259	-

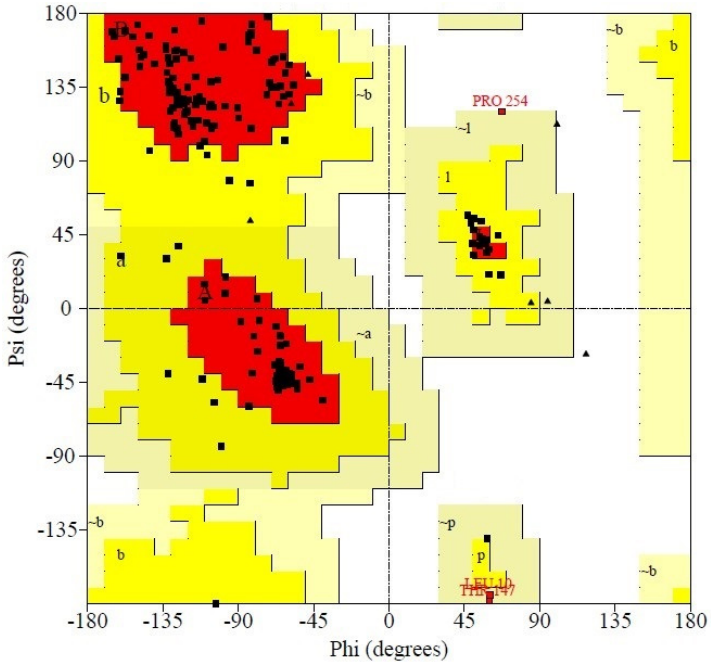

Fig. 3 Ramachandran plot analysis of HP R_27 through Swiss model workshop

The ligand binding site of the hypothetical protein R_27 was determined through CASTp server. It has been found that 57 amino acids are involved in potent active site. The best active site was found in areas with 1382.7 and a volume of 1754.3 amino acids. Some of them are also found in active site of other proteins [14, 16, 31]. The active site of hypothetical protein R_27 depicted in Fig. 4.

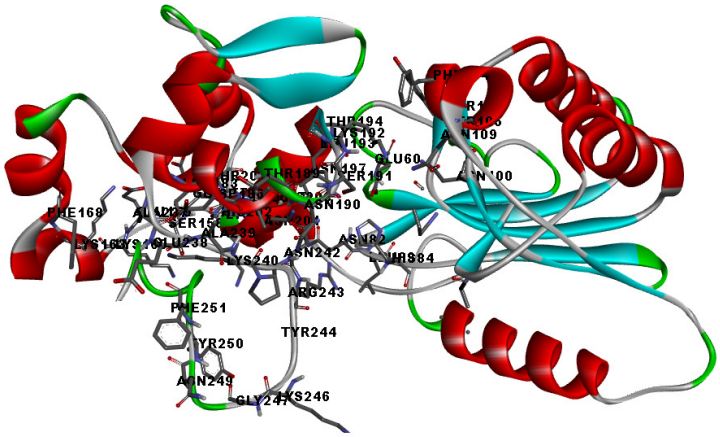

Fig. 4 Active site cavity identification of hypothetical protein R_27

Conclusion

We have analyzed HP R_27 from *Salmonella enterica* serovar *typhimurium* (*S. typhi*) through *in silico* approach. The physicochemical parameters prediction, sub-cellular localization and functional annotation are useful in distinguishing the hypothetical protein with DNA repair mechanism and PLD-like activity. The results might assist in developing drugs against *S. typhi*. This *in silico* approach can be further utilized in drug design to identify putative drug targets for other clinically significant pathogens.

Acknowledgements

We cordially thank Abdullah Al-Emran, Assistant Professor of the Department of Biotechnology and Genetic Engineering, Mawlana Bhashani Science and Technology University and Md. Anayet Hasan, Department of Genetic Engineering and Biotechnology, University of Chittagong, for their suggestions and inspiration during research proceedings.

References

1. Bendtsen J. D., L. Kiemer, A. Fausboll, S. Brunak (2005). Non-classical Protein Secretion in Bacteria, BMC Microbiology, 5, 58.
2. Benkert P., S. C. Tosatto, D. Schomburg (2008). QMEAN: A Comprehensive Scoring Function for Model Quality Assessment, Proteins, 71, 261-277.
3. Benson D. A., I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, B. A. Rapp, D. L. Wheeler (2002). GeneBank, Nucleic Acids Research, 30, 17-20.
4. Boeckmann B., A. Bairoch, R. Apweiler, M. C. Blatter, A. Estreicher, E. Gasteiger, M. J. Martin, K. Michoud, C. O'Donovan, I. Phan, S. Pilbout, M. Schneider (2003). The SWISS-PROT Protein Knowledgebase and Its Supplement TrEMBL in 2003, Nucleic Acids Research, 31, 365-370.
5. Buchan D. W. A., F. Minneci, T. C. O. Nugent, K. Bryson, D. T. Jones (2013). Scalable Web Services for the PSIPRED Protein Analysis Workbench, Nucleic Acids Research, 41(W1), W340-W348.
6. Colovos C., T. O. Yeates (1993). Verification of Protein Structures: Patterns of Nonbonded Atomic Interactions, Protein Science, 2, 1511-1519.
7. Crump J. A., E. D. Mintz (2010). Global Trends in Typhoid and Paratyphoid Fever, Clinical Infectious Disease, 50, 241-246.
8. Dundas J., Z. Ouyang, J. Tseng, A. Binkowski, Y. Turpaz, J. Liang (2006). CASTp: Computed Atlas of Surface Topography of Proteins with Structural and Topographical Mapping of Functionally Annotated Residues, Nucleic Acids Research, 34, W116-W118.

9. Emanuelsson O., S. Brunak, G. von Heijne, H. Nielsen (2007). Locating Proteins in the Cell Using TargetP, SignalP and Related Tools, Nature Protocol, 2, 953-971.
10. Feasey N. A., G. Dougan , R. A. Kingsley, R. S. Heyderman, M. A. Gordon (2012). Invasive Non-typhoidal Salmonella Disease: An Emerging and Neglected Tropical Disease in Africa, Lancet, 379, 2489-2499.
11. Finn R. D., J. Mistry, J. Tate, P. Coggill, A. Heger, J. E. Pollington, O. L. Gavin, P. Gunasekaran, G. Ceric, K. Forslund, L. Holm, E. L. Sonnhammer, S. R. Eddy, A. Bateman (2010). The Pfam Protein Families' Database, Nucleic Acids Research, 38(Database issue), D211-D222.
12. Geourjon C., G. Deléage (1995). SOPMA: Significant Improvements in Protein Secondary Structure Prediction by Consensus Prediction from Multiple Alignments, Comput Appl Biosci, 11, 681-684.
13. Hasan A., H. H. Mazumder, A. Khan, M. U. Hossain, H. K. Chowdhury (2014). Molecular Characterization of Legionellosis Drug Target Candidate Enzyme Phosphoglucosamine Mutase from *Legionella pneumophila* (strain Paris): An *in silico* Approach, Genomics and Informatics, 12, 268-275.
14. Hasan M. A., M. H. Mazumder, A. S. Chowdhury, A. Datta, M. A. Khan (2015). Molecular-docking Study of Malaria Drug Target Enzyme Transketolase in *Plasmodium falciparum* 3D7 Portends the Novel Approach to Its Treatment, Source Code of Biology and Medicine, 22(10), 7.
15. Hirokawa T., S. Boon-Chieng, S. Mitaku (1998). SOSUI: Classification and Secondary Structure Prediction System for Membrane Proteins, Bioinformatics, 14, 378-379.
16. Idrees S., S. Nadeem, S. Kanwal, B. Ehsan, A. Yousaf, S. Nadeen, M. I. Rajoka (2012). *In silico* Sequence Analysis, Homology Modeling and Function Annotation of *Ocimum basilicum* Hypothetical Protein G1CT28_OCIBA, International Journal Bioautomation, 16(2), 111-118.
17. Kelley L. A., M. J. Sternberg (2009). Protein Structure Prediction on the Web: A Case Study Using the Phyre Server, Nature Protocol, 4, 363-371.
18. Khan M. A., M. U. Hossain, S. M. Rakib-Uz-Zaman, M. N. Morshed (2015). Epitope-based Peptide Vaccine Design and Target Site Depiction against Ebola Viruses: An Immunoinformatics Study, Scandinavian Journal of Immunology, 82, 25-34.
19. Koonin E. V., M. Y. Galperin (2003). Principles and Methods of Sequence Analysis, US, Springer.
20. Krogh A., B. Larsson, G. von Heijne, E. L. Sonnhammer (2001). Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genomes, Journal of Molecular Biology, 305, 567-580.
21. Laskowski R. A., M. W. MacArthur, D. S. Moss, J. M. Thornton (1993). PROCHECK: A Program to Check the Stereochemical Quality of Protein Structures, J Appl Crystallogr, 26, 283-291.
22. Letunic I., T. Doerks, P. Bork (2012). SMART 7: Recent Updates to the Protein Domain Annotation Resource, Nucleic Acids Research, 40, D302-D305.
23. Magrane M., U. Consortium (2011). UniProt Knowledgebase: A Hub of Integrated Protein Data, Database: The Journal of Biological Databases and Curation, bar009, doi: 10.1093/database/bar009.
24. Marchler-Bauer A., S. Lu, J. B. Anderson, F. Chitsaz, M. K. Derbyshire, C. DeWeese- Scott, J. H. Fong, L. Y. Geer, R. C. Geer, N. R. Gonzales, M. Gwadz, D. I. Hurwitz, J. D. Jackson, Z. Ke, C. J. Lanczycki, F. Lu, G. H. Marchler, M. Mullokandov, M. V. Omelchenko, C. L. Robertson, J. S. Song, N. Thanki, R. A. Yamashita, D. Zhang, N. Zhang, C. Zheng, S. H. Bryant (2011).

CDD: A Conserved Domain Database for the Functional Annotation of Proteins, Nucleic Acids Research, 39(Database Issue), D225-D229.
25. Morshed M. G., W. A. Khan, H. Z. Khan, M. S. Akbar (1986). Multiple Drug Resistant *S. typhi* in Bangladesh, J Diarrhoeal Des Res, 4, 24.
26. Muyembe-Tamfum J. J., J. Veyi, M. Kaswa, O. Lunguya, J. Verhaegen, M. Boelaert (2009). An Outbreak of Peritonitis Caused by Multidrug-resistant *Salmonella typhi* in Kinshasa, Democratic Republic of Congo, Travel Med Infect Dis, 7, 3.
27. Neves-Ferreira A. G. C., C. M. de Andrade, A. M. Vannier-Santos, J. Perales, H. J. Nascimento, J. G. da Silva Jr. (2004). Complete Amino Acid Sequence and Location of Omp-28, an Important Immunogenic Protein from *Salmonella enterica* serovar *typhi*, Protein Journal, 23, 71-77.
28. Ochai R. L., C. J. Acosta, M. C. Danovaro-Holliday, D. Baiqing, S. K. Bhattacharya, M. D. Agtini, Z. A. Bhutta, G. Canh do, M. Ali, S. Shin, J. Wain, A. L. Page, M. J. Albert, J. Farrar, R. Abu-Elyazeed, T. Pang, C. M. Galindo, L. von Seidlein, J. D. Clemens, Domi Typhoid Study Group (2008). A study of Typhoid Fever in Five Asian Countries: Disease Burden and Implication for Controls, Bull World Health Organ, 86, 260-268.
29. Pandey G., V. Kumar, M. Steinbach (2006). Computational Approaches for Protein Function Prediction: A Survey. Twin Cities, MN: Department of Computer Science and Engineering, University of Minnesota.
30. Parkhill J., G. Dougan, K. D. James, N. R. Thomson, D. Pickard, C. Churcher, K. L. Mungall, S. D. Bentley, M. T. Holden, M. Sebaihia, S. Baker, D. Basham, K. Brooks, T. Chillingworth, P. Connerton, A. Cronin, P. Davis, R. M. Davies, L. Dowd, N. White, J. Farrar, T. Feltwell, N. Hamlin, A. Haque, T. T. Hien, S. Holroyd, K. Jagels, A. Krogh, T. S. Larsen, S. Leather, S. Moule, P. O'Gaora, C. Parry, M. Quail, K. Rutherford, M. Simmonds, J. Skelton, K. Stevens, S. Whitehead, B. G. Barrell (2001). Complete Genome Sequence of a Multiple Drug Resistant *Salmonella enterica* serovar *typhi* CT18, Nature, 413, 848-852.
31. Paul S., M. Saha, N. C. Bhoumik, S. N. Talukdar (2015). *In silico* Structural and Functional Annotation of *Mycoplasma genitalium* Hypothetical Protein MG_377, International Journal Bioautomation, 19(1), 15-24.
32. Rahman M., A. K. Siddique, F. C. Tam, S. Sharmin, H. Rashid, A. Iqbal, S. Ahmed, G. Nair, C. Balkrish, C.-L. Chaignat, P.-L. Lim (2007). Rapid Detection of Early Typhoid Fever in Endemic Community Children by the TUBEX 09-antibody Test, Diagn Microbiol Infect Dis, 58, 275-281.
33. Rashid M., S. Saha, G. P. S. Raghava (2007). Support Vector Machine-based Method for Predicting Subcellular Localization of Mycobacterial Proteins Using Evolutionary Information and Motifs, BMC Bioinformatics, 8, 337.
34. Roberts R. J. (2004). Identifying Protein Function a Call for Community Action, PLoS Biology, 2, E42.
35. Tusnady G. E., I. Simon (2001). The HMMTOP Transmembrane Topology Prediction Server, Bioinformatics, 17, 849-850.
36. Xu D., Y. Zhang (2011). Improving the Physical Realism and Structural Accuracy of Protein Models by a Two-step Atomic-level Energy Minimization, Biophys J, 101, 2525-2534.
37. Yap K. P., C. S. Teh, R. Baddam, L. C. Chai, N. Kumar, T. S. Avasthi, N. Ahmed, K. L. Thong (2012). Insights from the Genome Sequence of a *Salmonella enterica* serovar *typhi* Strain Associated with a Sporadic Case of Typhoid Fever in Malaysia, Journal of Bacteriology, 194, 5124-5125.

38. Yu C. S., C. J. Lin, J. K. Hwang (2004). Predicting Subcellular Localization of Proteins for Gram-negative Bacteria by Support Vector Machines Based on N-peptide Compositions, Protein Science, 13, 1402-1406.
39. Yu N. Y., J. R. Wagner, M. R. Laird, G. Melli, S. Rey (2010). PSORTb 3.0: Improved Protein Subcellular Localization Prediction with Refined Localization Subcategories and Predictive Capabilities for All Prokaryotes, Bioinformatics, 26, 1608-1615.
40. Zdobnov E. M., R. Apweiler (2001). InterProScan: An Integration Platform for the Signature-recognition Methods in InterPro, Bioinformatics, 17, 847-848.
41. Zhang R., H. Y. Ou, V. T. Zhang (2004). DEG: A Database of Essential Genes, Nucleic Acids Research, 32(Database Issue), D271-D272.

Md. Arif Khan, B.Sc. (Hons.)
E-mail: arifkhanbge35@gmail.com

Md. Arif Khan is working as an M.Phil. (Chemistry) student at Chemoinformatics Laboratory of Department of Science and Humanities, Military Institute of Science and Technology. He completed his B.Sc. (Hons.) from the Department of Biotechnology and Genetic Engineering, Mawlana Bhashani Science and Technology University, Bangladesh. His areas of research are bioinformatics, chemoinformatics, medicinal chemistry, cancer biology, immunology, computational biology and computational drug design.

Hossain Ahmed
E-mail: ahmedshohag.bd@gmail.com

Hossain Ahmed is doing M.Sc. in Biotechnology and Genetic Engineering at the Department of Biotechnology and Genetic Engineering, University of Development Alternative, Bangladesh. Currently he is working in pharmacology, bioinformatics and medicinal plant.

Nasreen Jahan, M.Sc.
E-mail: naza.bge@gmail.com

Nasreen Jahan is working as a Lecturer in the Department of Biotechnology and Genetic Engineering, Mawlana Bhashani Science and Technology University, Bangladesh. He completed B.Sc. (Hons.) and M.Sc. in Biotechnology and Genetic Engineering from Mawlana Bhashani Science and Technology University, Bangladesh. Her areas of research are chemical biology, molecular genetics, bioinformatics, and immunology.

Syed Raju Ali, B.Sc.
E-mail: m15121307@bau.edu.bd

Syed Raju Ali is a M.Sc. student of Biotechnology Department, Bangladesh Agricultural University, Mymensingh, Bangladesh. He completed his B.Sc. in Biotechnology and Genetic Engineering from Mawlana Bhashani Science and Technology University. His scientific interests are in molecular biology, molecular biotechnology, human genetics, bioinformatics and biologics.

Md. Al Amin, M.Sc.
E-mail: josnasnan@gmail.com

Md. Al Amin completed his M.Sc. from Biotechnology Department, Bangladesh Agricultural University, Mymensingh, Bangladesh. He completed his B.Sc. in Biotechnology and Genetic Engineering from Mawlana Bhashani Science and Technology University. His scientific interests are in molecular genetics, cell and developmental biology, cellular biology of cancer and bioinformatics.

Assoc. Prof. Major Mohammad Neaz Morshed, Ph.D.
E-mail: mmneaz@hotmail.com

Mohammad Neaz Morshed is working as an Associate Professor in Department of Science and Humanities, Military Institute of Science and Technology. He has completed his Ph.D. in Medicinal and Pharmaceutical Chemistry from Korea Institute of Science and Technology, Seoul, South Korea. He completed his B.Sc. in Chemistry from Dhaka University and M.Sc. in Organic Chemistry from Cheju National University, South Korea. His research interests are chemoinformatics, computational chemistry, computer-aided drug design, virtual screenings, molecular modelling, QSAR/QSPR, pharmacophore modeling, chemical databases, molecular docking.