

In Silico Structural and Functional Prediction of *Phaseolus vulgaris* Hypothetical Protein PHA VU_004G136400g

Zohaib Bashir^{1*}, Muhammad Rizwan^{1,2}, Kanwal Mushtaq¹, Anum Munir^{1,3} and Ishtiaq Ali¹

¹Department of Bioinformatics, Government Post Graduate College Mandian, Abbottabad 22010, Pakistan

²Department of Bioinformatics, Hazara University Dodial, Mansehra 21120, Pakistan

³Department of bioinformatics, Muhammad Ali Jinnah University, Islamabad 44000, Pakistan

Abstract

Hypothetical proteins (HPs) are the proteins whose occurrence has been predicted, yet *in vivo* function has not been manufactured up. Illustrating the structural and functional confidential visions of these HPs might likewise prompt a superior understanding of the protein-protein relations or networks in diverse varieties of life. Common bean (*Phaseolus vulgaris*) is the most important food legume for direct human consumption in the world, most important grain legume for human consumption and has a role in sustainable agriculture helping to its ability to fix atmospheric nitrogen. However, in spite of the importance of this crop species, its genetics have been poorly characterized. In the present study, the hypothetical protein of *Phaseolus vulgaris* (Common bean) was chosen for analysis, and modeling by distinctive Bioinformatics apparatuses and databases. As indicated by primary and secondary structure analysis, XP_007152511.1 is a stable hydrophobic protein containing a noteworthy extent of α -helices; Homology modeling was utilized SWISS-MODEL server where the templates identity with XP_007152511.1 protein was less which demonstrated novelty of our protein. The Ab initial strategy was conducted to produce its 3D structure. A few evaluations of quality assessment and validation parameters determined the generated protein model as stable with genuinely great quality. Functional analysis was completed by ProtFun 2.2, and KEGG (KAAS), recommended that the hypothetical protein is a translation factor with nuclear domain. The protein was observed to be energetic for translation process, involved in trans-membrane barriers, signaling and cellular processes, and protein binding. It is suggested that further test approval would help to anticipate the structures and functions of other uncharacterized proteins of different plants and living being.

Keywords: *Phaseolus vulgaris*; Homology modeling; Functional annotations; Ab initio

Introduction

In latest times, genome sequences of different organisms are available in numerous databases by applying next generation sequencing technology to collect information to a larger quantity. As a result, increasing quantities of records approximately hypothetical proteins deposited in sequence databases instead of experimentally isolated facts in Protein data bank (PDB). Hypothetical proteins are generally fore casted to be expressed from an open reading frame (ORF). Those proteins don't have any experimental proof concerning their functions. At gift, it's far assumed that 50% protein of a genome is hypothetical proteins. This encourages *in silico* take a look at of a hypothetical protein utilizing experimental information [1]. Valid structural and functional annotations of HPs of unique genome may also spark off the finding of latest structures and also new functions and help to offer extra protein pathways and cascades, in this way finishing our rough records on the variety of proteins [2]. The fully sequenced genomes of numerous organisms offer huge quantities of facts approximately cellular biology (see the genomes indexed at the internet website of The Institute for Genome studies: www.tigr.org). It's far an imperative venture of bioinformatics to apply this fact in coming across the function of proteins. Functional assignments of genes come primarily from biochemical experimentation, which may be prolonged through matching recently sequenced proteins to people who have already been characterized [3].

The common bean is the most significant food bank for direct human consumption in the world. The crop is a main creation in worldwide commerce and is created and consumed by huge numbers of the rural and town unfortunate in Latin America and Africa [4]. Common bean (*Phaseolus vulgaris*) is the major green bean for human

digestion and has a role in maintaining agriculture outstanding to his capacity to fix atmosphere nitrogen (N₂). We accumulated 473 MB of the 587-Mb genome and hereditarily anchored 98% of this sequence in 11 chromosome-scale pseudo molecules [5]. However, in unkindness of the significance of this yield species, its genetics have remained poorly characterized. For illustration, only a limited morphological and seed and flower color markers have remained used to grow a rudimentary linkage map. More lately, a few isozyme and protein markers take remained additional to the map. Although cytogenetic analysis takes to remain troubled by the minor size of the chromosomes, five primary trisomics take remained characterized [6].

In current centuries, numerous hypothetical proteins have been created in the genome of several life systems. In some circumstance, because of a few limitations, for illustration, the expenditure and time required for examining methodologies, whole genome annotations have not accomplished until now. In adding, the wide amount of theoretical proteins in a genome creates their analysis a difficult task. Bioinformatics approaches using distinguishing algorithms and databases to evaluation protein capacity would be a decent altered option for laboratory exploration ability based procedures. As these algorithms

***Corresponding author:** Zohaib Bashir, Department of Bioinformatics, Government Post Graduate College Mandian, Abbottabad 22010, Pakistan, Tel: +92-3315460614; E-mail: zohaibbashir001@gmail.com, mrizwanhu@gmail.com

Received June 08, 2017; **Accepted** August 31, 2017; **Published** September 07, 2017

Citation: Bashir Z, Rizwan M, Mushtaq K, Munir A, Ali I (2017) *In Silico* Structural and Functional Prediction of *Phaseolus vulgaris* Hypothetical Protein PHA VU_004G136400g. J Proteomics Bioinform 10: 207-211. doi: 10.4172/jpb.1000443

Copyright: © 2017 Bashir Z, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

and databases are in graceful of experimental results analysis, they can be a convincing aims to complete functionality and structural annotation of hypothetical proteins [7].

In the present study, the Common bean (*Phaseolus vulgaris*) hypothetical protein>XP_007152511.1, belongs to AAA group was selected as the primary amino acid sequence of the protein is available however structural details are not available. This study meant to examine the physiochemical and secondary structure mechanisms, to produce the first three-dimensional (3D) model completed by Ab initio method, and finally to complete functional annotations. The result of this work resolve suitable for better comprehension of the tail of this protein and determining other novel proteins and their functions by the same way that we have accomplished for Common bean (*Phaseolus vulgaris*) protein.

Materials and Methods

Sequence retrieval

The amino acid sequence of the common bean (*Phaseolus vulgaris*) hypothetical protein>XP_007152511.1 was retrieved from the Uniprot database (www.uniprot.org).

Physiochemical analysis of the protein

Examination of the physiochemical properties of the studied protein such as molecular weight, theoretical pI, amino acid composition, atomic composition, instability index, and grand average of hydrophobicity (GRAVY) remained to complete using ProtParam tool (<http://web.expasy.org/protparam/>) [8].

Secondary structure analysis

The server SOPMA was applied for secondary structure calculations (helix, sheets, and coils) of the hypothetical protein [9]. In adding to that, the PSIPRED [10] and Predict Protein [11] servers were also exploited to confirm the results achieved from SOPMA.

Subcellular localization prediction

Subcellular localization of common bean (*Phaseolus vulgaris*) was predicted by PSORT [12]. Results remained also cross-checked through subcellular localization predictions acquired from SOSUI server and Predict-Protein servers [13].

Homology modeling of the hypothetical protein

The possible 3D structure of the protein>XP_007152511.1 was created over alignment approach in protein structure homology modeling server SWISS-MODEL [2,14] with the full amino acid sequence of the protein in FASTA format.

Quality assessment of the 3D model and visualization

The early structural model achieved, was checked for acknowledgement of mistakes in 3D structure [11] by ERRAT and Verify3D programs involved in structural examination and confirmation server SAVES (<http://nihserver.mbi.ucla.edu/SAVES/>) [15,16]. The Ramachandran plots for the model remained built using the RAMPAGE server [17], viewing the percentage of protein residues in the favored, allowed and outlier sections. The visualization of creating model was accomplished by Discovery studio 4.1 [14].

Functional annotation of the protein

Common bean (*Phaseolus vulgaris*) hypothetical

protein>XP_007152511.1 was examined for the function. Three different bioinformatics tools and databases including ProtFun 2.2 [15], ProFunc [18], and NCBI Conserved Domains Database (NCBI-CDD) [19] remained utilized for this goal. Moreover, KEGG automatic annotation server (KAAS) was used to examine the association of (*Phaseolus vulgaris*) hypothetical protein in the metabolic pathways [14].

Submission of the model in protein model database

The model created for Common bean (*Phaseolus vulgaris*) hypothetical protein>XP_007152511.1 was successfully submitted in Protein Model Database (PMDb) (<http://bioinformatics.cineca.it/PMDB/>).

Results and Discussion

Physiochemical characteristics of XP_007152511.1

The ExPASy's ProtParam server was developed to study the theoretical physiochemical characteristics of the amino acid sequence of hypothetical protein XP_007152511.1. The enormous majority of the calculations in this server show protein steadiness and stability, in light of the detail that the stability is recognized with its suitable function ability [3]. The protein was predicted to be contained 220 amino acids, with a molecular weight of 24330.23 Daltons and an isoelectric point (Pⁱ) of 6.38 demonstrated a negatively charged protein. The instability index of the protein was computed to be 47.86, which demonstrated this protein as stable.

The GRAVY index of -1.025 is expressive of a hydrophobicity and solubility of protein. The most plenteous amino acid residue was observed to be Alanine (20), followed by Leucine (19) and the most minimal amino acid as Cytosine and tryptophan (1). The sequence had 40 negatively charged residues (Aspartic acid+Glutamic acid) and 39 positively charged residues (Arginine+Lysine). The molecular formula of the protein remained found as C¹⁰⁵⁰ H¹⁷⁰¹ N³¹¹ O³⁴³ S⁵.

Subcellular localization of XP_007152511.1

Protein subcellular localization predictions contain the computational probability of somewhere protein survives in a cell. Predicting subcellular localization of unidentified proteins can provide information around their cellular functions. This information might be applied in understanding disease mechanism and developing drugs [20]. The subcellular localization of the query protein was anticipated to be a nuclear protein, analyzed by SOSUI and confirmed by PSORTb v3.2.0 and Predict Protein servers.

Secondary structure of XP_007152511.1

First the secondary structure of the protein remained predicted by SOPMA server. The alpha helix was found to be the most predominant (36%), followed by random coil (47%) and extended strand (8%). Likewise, beta turn was found as 7.73%. Second, the similar results stayed found from Predict-Protein and PSIPRED servers. The representative secondary structure of XP_007152511.1 obtained from the PSIPRED server is shown in Figure 1.

Homology modeling of XP_007152511.1

We assume these uncharacterized proteins infinite unexplored field with numerous occasions, both as medicinal and industrial tools. *In Silico* analysis may help with determining the biological functions of such un-characterized proteins. This can be encouraged by anticipating the 3D structure of the target protein. At the point when the uncertainly experimented structure is isolated, similar or homology modeling can

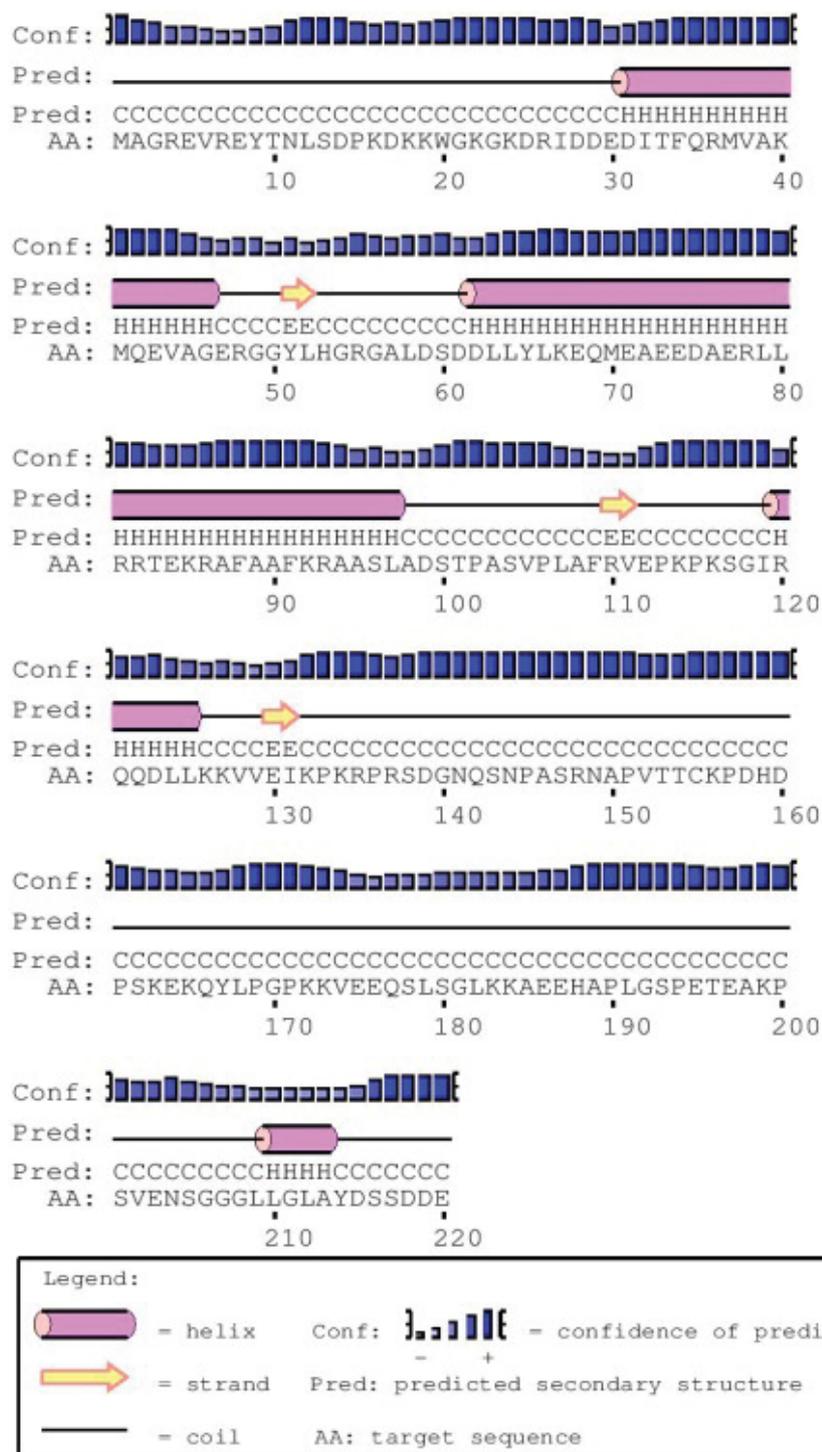


Figure 1: Predicted secondary structure of the Common bean (*Phaseolus vulgaris*) hypothetical protein >XP_007152511.1 by PSIPRED server.

now and then give a helpful 3D model to the protein of attention that is identified with no less than one known protein structure. Homology modeling predicts the 3D structure of an assumed protein sequence build principally with respect to its alignments to one or more proteins of known structure [21]. To perform the homology modeling, the query

sequence was given as input in SWISS-MODEL server. The server consequently performed BLASTP search for each protein sequence to identify templates for homology modeling. The highest template identity was 20% which showed that XP_007152511.1 hypothetical protein is novel and no similar template structure is present in

any databases. We predicted the 3D structure of XP_007152511.1 hypothetical protein by Ab initio method through Phyre2 server and 3D Jigsaw server which gave 99.3% confidence in model. The 3D model was viewed by Discovery studio 4.1 and shown in Figure 2.

Quality assessment and visualization

Reliability of the created model was initially checked by ERRAT that analyzed the statistics of non-bonded interactions between diverse atom types, based on characteristic atomic interactions. The overall quality factor was found as 26.88%, sufficient enough to use this model. As demonstrated by the Verify3D program, the results showed that 51.08% of residues had an average 3D (atomic model) – 1D (amino acid) score ≥ 0.2 meaning that this structures was compatible and genuinely good.

Ramachandran plots were resolved. Z-score of the query model was checked from PROSAweb. The model's Z-score was not shown due to novelty of XP_007152511.1 protein. The Z-score is used to estimation the quality of model using structured solved proteins as reference [22]. The stereo chemical quality of the model protein was examined using Ramachandran plots through the RAMPAGE server. Ramachandran plot analysis observed 81.2% of residues of the protein's model structure in the favored region, with 11.0% and 7.8% residues in allowed and outlier regions, individually, indicating that the model was reliable and of good quality shown in Figure 3. The final protein structure was deposited in PMDB and is available under ID PM0080295.

Functional annotation of XP_007152511.1

We utilized three web tools to search the potential functions of

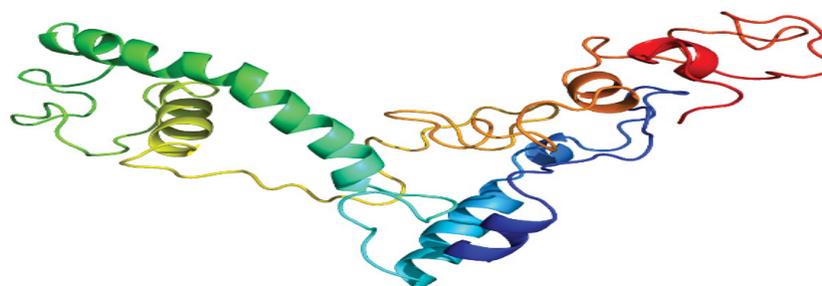


Figure 2: Structural analysis of Common bean (*Phaseolus vulgaris*) hypothetical protein >XP_007152511.1.

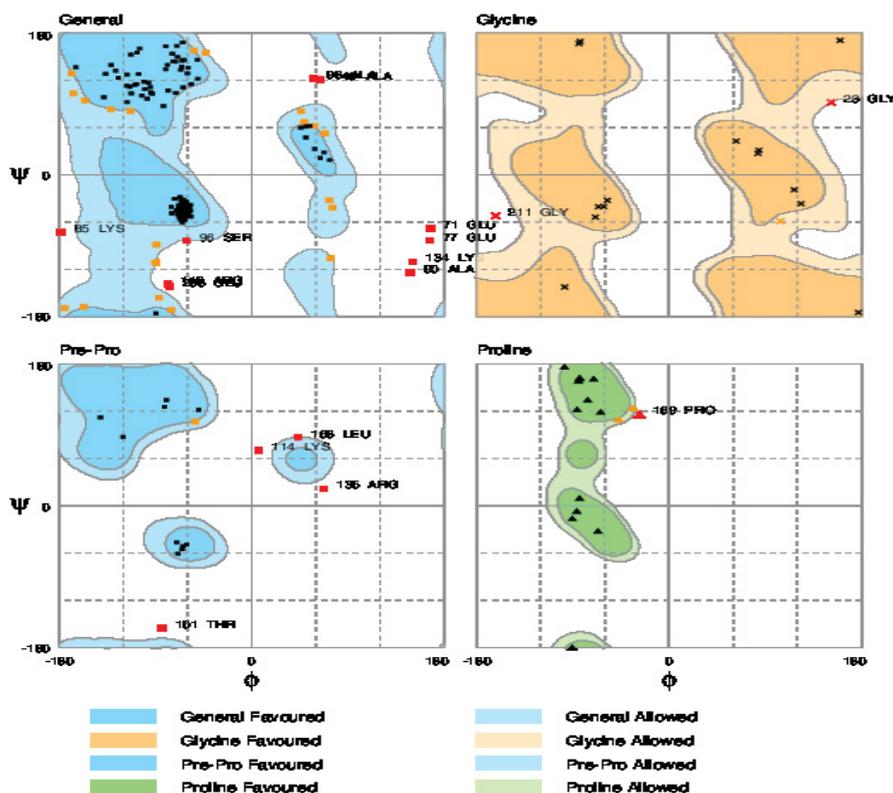


Figure 3: Ramachandran plot for the 3D model of the studied hypothetical protein XP_007152511.1 by RAMPAGE server.

XP_007152511.1. In light of predictions made by Prot Fun 2.2, and KEGG (KAAS), XP_007152511.1 was suggested as a translation and also acts as a non-enzyme. It helps in Trans membrane barriers, protein protein binding, metabolic process, cellular process. The protein was found to be central for the translation process, involved in signaling and cellular processes, transfrase activity, and catalytic activity.

Comparative genome analysis of XP_007152511.1

We utilized NCBI Blast search tool for comparative genome analysis of XP_007152511.1 hypothetical protein of *Phaseolus vulgaris* with other plant genomes. In the result the XP_007152511.1 showed highest similarities with other uncharacterized hypothetical proteins of several plants [23,24].

Conclusion

The present study was directed to create the first 3D structure and propose possible functions of the *Phaseolus vulgaris* hypothetical protein XP_007152511.1. The 3D model of the protein was constructed using Ab initio method as well as refined by few structural assessment methods and the final outcome was genuinely great. We observed that this novel protein is a stable nuclear protein and function as a translation factor process, non-enzyme act. The protein was observed to be central for the translation process, also involved in, trans membrane barriers, protein protein interaction, metabolic process, cellular process, transferase activity, catalytic activity processes and signaling and cellular process. From genomic similarities we conclude that this hypothetical protein may be checked for same function as of XP_007152511.1. Moreover, this sort of methodology could be helpful in the structure and functions prediction of other uncharacterized proteins.

References

1. Khan A, Ahmed H, Jahan N, Ali SR, Amin A, et al. (2016) An in silico Approach for structural and functional annotation of *Salmonella enterica serovar typhimurium* hypothetical protein R_27. Int J Bioautomation 20: 31-42.
2. Loewenstein Y, Raimondo D, Redfern OC, Watson J, Frishman D, et al. (2009) Protein function annotation by homology-based inference. Genome Biology 10: 207.
3. Paul S, Saha M, Bhoumik NC (2015) In silico structural and functional annotation of mycoplasma genitalium hypothetical protein MG_377. International Journal Bioautomation 19: 15-24.
4. Blair MW, Pedraza F, Buendia HF, Gaitán-Solis E, Beebe SE, et al. (2003) Development of a genome-wide anchored microsatellite map for common bean (*Phaseolus vulgaris* L). Theoretical and Applied Genetics, 107: 1362-1374.
5. Rost B, Liu J (2003) The predict protein server. Nucleic Acids Research 31: 3300-3304.
6. Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, et al. (2014) A reference genome for common bean and genome-wide analysis of dual domestications. Nature Genetics 46: 707-713.
7. Nimrod G, Schushan M, Steinberg DM, Ben-Tal N (2008) Detection of functionally important regions in "Hypothetical proteins" of known structure. Structure 16: 1755-1763.
8. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, et al. (2005) Protein identification and analysis tools on the ExpASY server. The Proteomics Protocols Handbook; pp: 571-607.
9. Geourjon C, Deleage G (1995) SOPMA: Significant improvements in protein secondary structure prediction by prediction from multiple alignments. Comput Applic Bioci 11: 681-684.
10. Buchan DWA, Minneci F, Nugent TCO, Bryson K, Jones DT (2013) Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Research 41: 349-357.
11. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. (2004) UCSF Chimera - A visualization system for exploratory research and analysis. Journal of Computational Chemistry 25: 1605-1612.
12. Lubec G, Afjehi-Sadat L, Yang JW, John JP (2005) Searching for hypothetical proteins: Theory and practice based upon original data and literature. Progress of Neurobiology 77: 90-127.
13. Vallejos CE, Sakiyama NS, Chase CD (1992) A molecular marker-based linkage map of *Phaseolus vulgaris* L. Genetics 131: 733-740.
14. Minion FC, Lefkowitz EJ, Madsen ML, Cleary BJ, Swartzell SM, et al. (2004) The genome sequence of *Mycoplasma hypopneumoniae* strain 232, the agent of swine mycoplasmosis. Journal of Bacteriology 186: 7123-7133.
15. Zdobnov EM, Apweiler R (2001) InterProScan: An integration platform for the signature-recognition methods in InterPro. Bioinformatics 17: 847-848.
16. Colovos C, Yeates TO (1993) Verification of protein structures: Patterns of non-bonded atomic interactions. Protein Science 2: 1511-1519.
17. Lovell SC, Davis IW, Adrendall WB, Bakker PIW, Word JM, et al. (2003) Structure validation by C alpha geometry: phi, psi and C beta deviation. Proteins-Structure Function and Genetics 50:437-450.
18. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, et al. (2014) Pfam: The protein families database. Nucleic Acids Research 42: 211-222.
19. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, et al. (2011) CDD: A conserved domain database for the functional annotation of proteins. Nucleic Acids Research 39: 225-229.
20. Yu CS, Chen YC, Lu CH, Hwang JK (2006) Prediction of protein subcellular localization. Proteins: Structure, Function and Bioinformatics.
21. Ar O, Sai A, Maa S, Hossain MU, Ferdoushi A (2014) Computational structure analysis and function prediction of an uncharacterized protein (I6U7D0) of *Pyrococcus furiosus* Com1. Austin J Comput Biol Bioinform 1: 5.
22. Benkert P, Biasini M, Schwede T (2011) Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics 27: 343-350.
23. Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates, et al. (1999). Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles. Biochemistry 96: 4285-4288.
24. Zhang R (2004) DEG: A database of essential genes. Nucleic Acids Research 32: 271-272.