Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 05, May-2020

Functional Annotation and Molecular Modeling of Hypothetical Proteins (HPS) from P. Aeruginosa Plasmid Pum505: An in-Silico Approach

Srikant Awasthi[1-2], Pragya Saxena[1,] Hillol Chakdar[1], Alok K Srivastava[1] and Salman Akhtar[2*]
[1]Department of Bio-Engineering, Integral University, Lucknow, INDIA, 226026
[2]Microbial Genomics Laboratory, National Bureau of Agriculturally Important Microorganisms (NBAIM), Mau, Uttar Pradesh, INDIA, 275101

Abstract:- **Structural and function annotation of *P. aeruginosa* plasmid pUM505 is essentially required to facilitate the understanding of mechanisms of pathogenesis and biochemical pathways important for selecting novel therapeutic target. In present study, randomly selected twelve hypothetical protein sequence of *P. aeruginosa* plasmid pUM505 has been annotated using various *In-Silico* tools and databases to determine domain family, solubility of protein, ligand binding sites etc. Six out of 12 proteins have been putatively annotated, in which four have been annotated with high confidence. Physio-chemical characterization revealed two proteins are stable. The three-dimensional structure of two important annotated proteins were modeled and their ligand binding sites were identified. Domains and families for six proteins have been found. The analysis revealed that these proteins have antitoxin activity, integrase enzyme activity, conjugal DNA transfer activity, etc. Structural prediction of these proteins and detection of binding sites from this study would indicate a potential target aiding docking studies for therapeutic designing against various diseases.**

Keywords - Hypothetical Protein, Functional Annotation, Molecular Modeling, Docking

1. INTRODUCTION

Pseudomonas aeruginosa, a gram-negative bacterium is well known for its environmental versatility. Diverse growing habitat includes soil, coastal marine, plant and animal tissues [1]. *P. aeruginosa* is also well known for its multidrug resistant and is a global threat towards many infections disease. *P. aeruginosa* is a major opportunistic pathogen in humans, causing serious complications caused by infections in patients particularly susceptible like people with immune system deficiencies, victims of skin burns, catheterized patients who suffer urinary tract infections and patients with respirators, causing nosocomial pneumonia [2]. It is the major cause of mortality in patients with cystic fibrosis colonizing the lungs [3]. Role of plasmids in antibiotic resistance are well establish. Plasmids are circular deoxyribonucleic acid molecules that exist in bacteria, usually independent of the chromosome. The study of plasmids is important to medical microbiology because plasmids can encode genes for antibiotic resistance or virulence factors [4]. The pUM505 plasmid contains a genomic island with sequence similar to islands found in chromosomes of virulent *P. aeruginosa* clinical isolates. Plasmid pUM505 contains several genes that encode virulence factors, suggesting that the plasmid may contribute directly to bacterial virulence [5]. The bacterium's virulence depends on a large number of cell-associated and extracellular factors. The virulence factors play an important pathological role in the colonization, the survival of the bacteria and the invasion of tissues [6].

Due to cost-efficiency throughput of genome sequencing has increased enormously resulting thousands of bacterial genomes now available and this number is increasing enormously day by day. Functional annotation of proteomes is a demanding problem [7]. A large fraction of proteins is still labeled as ‘‘hypothetical protein’’,‘‘unknown function’’ or with similar terms that imply that there is no functional indication for the ORF. Function annotation of putative uncharacterized HPs for their possible biological activity has emerged as an important focus for computational biology [8,9,10]. The pUM505 sequence contained 138 complete coding regions, the majority of them encoded on the complementary DNA strand (75%), with respect to the predicted origin of replication (oriV). Most of the identified genes (46%) encode hypothetical proteins (HSPs). Proper structural and functional determination of this huge fraction (46%) is very important to reveal complete understanding of virulence mechanism in *P. aeruginosa.* Therefor an improved functional annotation of its proteome is of particular urgency. In present study, 12 randomly selected HPs from *P.aeruginosa* plasmid pUM505 have been annotated with the help of various bioinformatics resources. Moreover, two important annotated HPS have been structurally modeled and characterized.

2.MATERIALS AND METHODS

2.1 Sequence retrieval and physiochemical characterization

Randomly selected twelve HPs which contain standard number of amino acids sequences of *P.aeruginosa* plasmid pUM505 were retrieved from NCBI

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 05, May-2020

(http://www.ncbi.nlm.nih.gov/) [11].For Physio-chemical characterization EXPAXY ProtParam [12] tool were used.

2.2 Functional characterization

Functional annotation was performed using conserved domain database (CDD)[13] Pfam [14] and EGGNOG [15]. CELLO [16] and PSORT B [17] were used for subcellular localization of proteins. Signal Peptide [18] and SecretomeP were used for nonclassical secretory pathway in proteins.

2.3 3D structure modelling

Homology models of HPs were determined using SWISSMODEL [19] and protein homology recognition engine Phyre2 [20]. The Phyre server uses a library of known protein structures taken from the structural classification of proteins (SCOP) database and the PDB [20]. The top ten scoring alignments were used to construct the three-dimensional structure of each HP.

2.4 Validation of predicted model

The stereo-chemical quality of the modeled structure for HPs was validated with verification server PDBSUM [21] using PROCHECK [22]. PROCHECK validates the stereochemical quality of a protein structure by analyzing the overall structure and residue-by-residue geometry of proteins. ERRAT server [23] was used for structure validation.

2.5 Active site prediction

MetaPocket 2.0 (http://projects.biotec.tu-dresden.de/metapocket/help.php) was used to find out the ligand binding sites. Proteins are primarily scanned for ligands and it uses the interaction energy between the protein and a simple van der Waals probe to locate vigorously favorable binding sites [24].

3. RESULTS AND DISCUSSION

3.1 Sequence analysis and functional annotation

Physicochemical studies of selected proteins revealed that two out of 12 proteins are stable. Most of the proteins are of acidic in nature (Supplementary Table S1).

In pursuit of assigning putative function to hypothetical proteins, available sequences and functional information from various resources has been interrogated. Six out of 12 selected hypothetical proteins have been putatively annotated using sequence/domain information from PFAM, functional information from EGGNOG, pathway (Table1). The presence of different domains in varying combinations in different proteins gives rise to the diverse repertoire of proteins found in nature. Identifying the domains present in a protein can provide insights into the function of that protein. Protein YP_004928038.1 has been annotated as PIN3 family protein, YP_004928003.1 as TraU family, YP_004927980.1 as NA-37 family, and YP_004927989.1 as HNH endonuclease (Table 1).

S.No.	GenBank ID	Functional Annotation	
		EGGNOG	**Pfam**
1.	YP_004927991.1	Function Unknown	Not found
2.	YP_004928098.1	No Ortholog	Not found
3.	YP_004927980.1	Nucleoid-associated protein	NA-37 family
4.	YP_004928038.1	PIN3 domain protein	PIN3 family
5.	YP_004928012.1	Secreted protein	DUF2895
6.	YP_004927986.1	No Ortholog	Not found
7.	YP_004927975.1	Function Unknown	Not found
8.	YP_004927989.1	phage protein	HNH endonuclease
9.	YP_004928003.1	TraU	TraU family
10.	YP_004928109.1	Function Unknown	DUF 1845 (family of unknown function)
11.	YP_004927994.1	No Ortholog	Not found
12.	YP_004928060.1	Function Unknown	DUF 1302 (family of unknown function)

Table1. Predicted Functions of HPs in *P.aeruginosa* plasmid pUM505

Four proteins have been annotated with high confidant. PIN3 family protein plays a very important role and function as nuclease enzymes that cleave single stranded RNA in a sequence dependent manner. PIN domain contain three nearly invariant acidic residues. PIN-domain proteins found in prokaryotes are the toxic components of toxin-antitoxin operons. These loci provide a control mechanism that helps free-living prokaryotes cope with nutritional stress [25]. HNH proteins are involved in DNA homing, restriction, repair, or chromosome degradation. The HNH proteins are not only involved in homing but also carry other biological functions, such as DNA degradation, repair, and restriction [26]. TraU is an essential protein to conjugal DNA transfer [27]. All these annotated HPs play crucial role. Sub-cellular localization of protein is an important parameter for protein function in various cellular processes. Sub-cellular localization analysis revealed that 9 proteins are cytoplasmic, one outer membrane protein and two unknown (Table 2).

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 05, May-2020

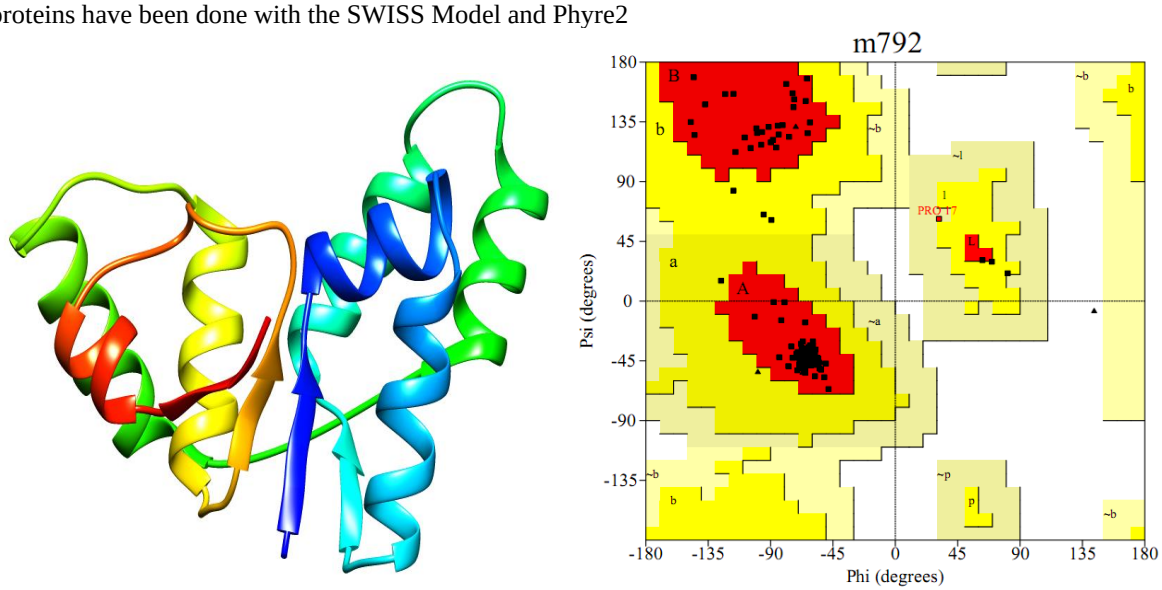
S.No	GeneBank ID	Sub-cellular localization			SignalPeptide	Secretory Protein
		PSORT B	PSLpred	CELLO		
1	YP_004927991.1	Cytoplasmic	Cytoplasmic protein	Cytoplasmic	No	No
2	YP_004928098.1	Cytoplasmic	Cytoplasmic protein	Cytoplasmic	No	No
3	YP_004928060.1	Outer membrane	Inner membrane protein	Outer membrane	Yes	No
4	YP_004928038.1	Cytoplasmic	Cytoplasmic protein	Cytoplasmic	No	No
5	YP_004927994.1	Cytoplasmic	Periplasmic protein	Cytoplasmic	No	No
6	YP_004927986.1	Unknown	Cytoplasmic	Cytoplasmic	No	No
7	YP_004927980.1	Cytoplasmic	Inner membrane protein	Cytoplasmic	No	No
8	YP_004927975.1	Cytoplasmic	Inner membrane protein	Cytoplasmic	No	Yes
9	YP_004927989.1	Cytoplasmic	Cytoplasmic protein	Cytoplasmic	No	No
10	YP_004928003.1	Unknown	Cytoplasmic protein	Extracellular	Yes	Yes
11	YP_004928012.1	Cytoplasmic	Inner-membrane protein	Cytoplasmic	No	No
12	YP_004928109.1	Cytoplasmic	Inner membrane protein	Cytoplasmic	No	No

Table2.Subcellular localization of HPs predicted by different bioinformatics tool

Motifs are signatures of protein families and can be preferably used to define the protein function, particularly in enzyme where motifs are associated with the catalytic function [28].

3.2 3D Structure prediction and Validation

Three dimensional structures of two important annotated proteins have been done with the SWISS Model and Phyre2 server. Based on the results, the stereo chemical evaluation of backbone Ψ and Φ dihedral angles of the HSPs revealed that for model HP38; 94.6, and 5.4 % residues were within the most favored regions, additionally allowed regions respectively (Fig 1b). No residue was found in generously allowed or disallowed regions.



(a) **(b)**

Fig1. (a-b):- (a)Three dimensional structure of HP38 model (b) their stereo-chemical property by Ramchandran plot. All the residues are in most favored region

For model HP80, 84.6,12.8, 1.7 and 0.9% residues fall within the most favored regions, additionally allowed regions, generously allowed regions and disallowed regions respectively (Fig 2 a-b). Overall 100 % implies best stereo-chemical quality for both the model. Based on the stereo chemical validation model HP38 showed more robust model than HP80.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 05, May-2020

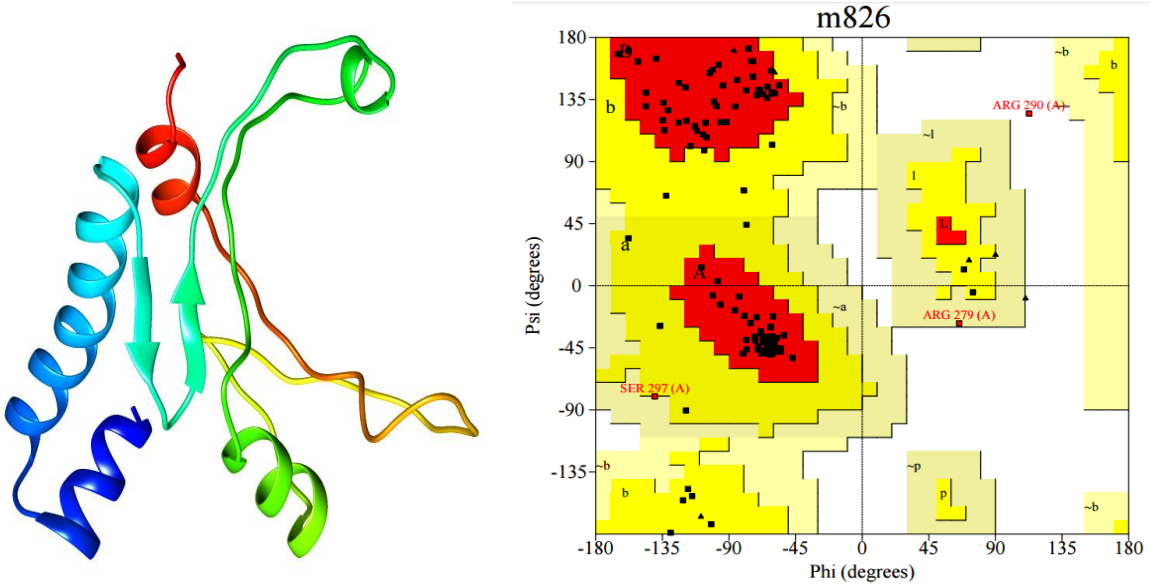
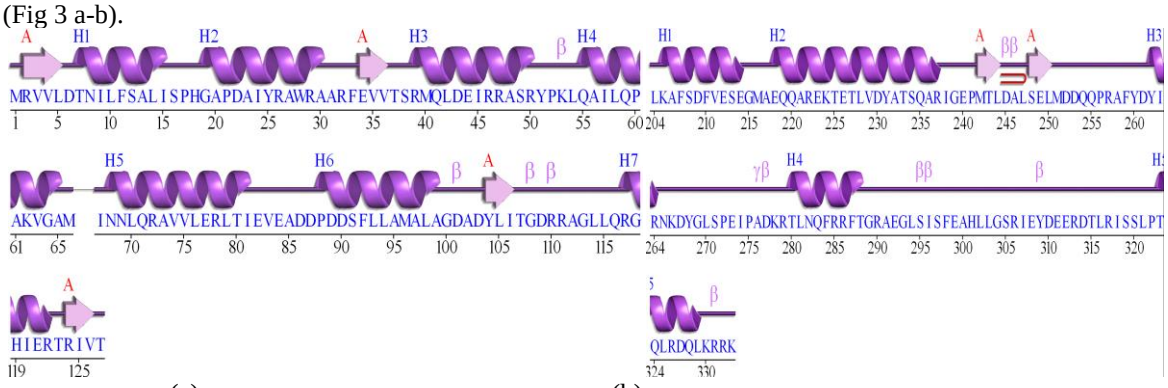


Fig2 (a-b):-(a)Three dimensional structure of HP80 model (b) their stereo-chemical propertyby Ramachandran plot.

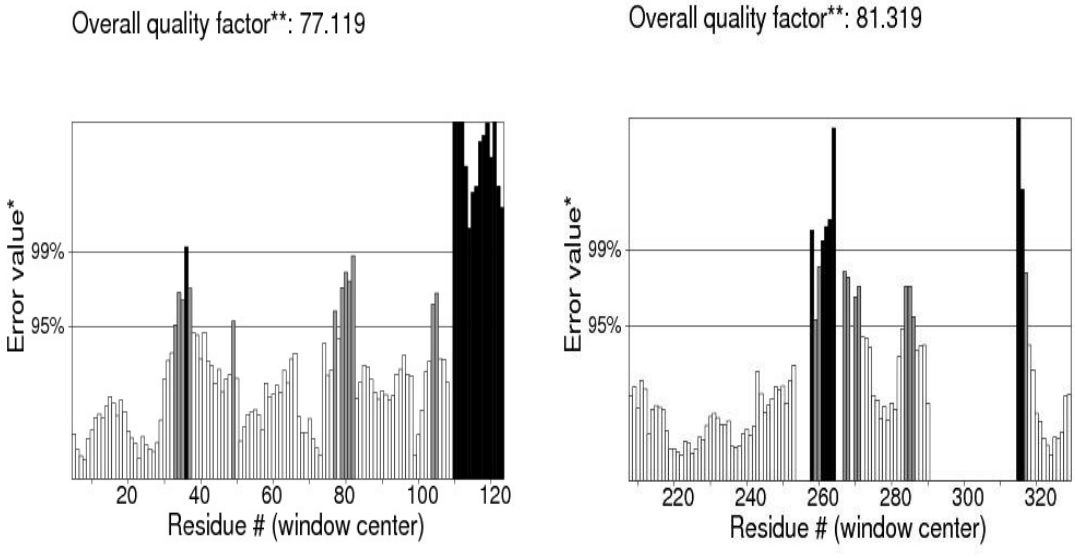
Secondary structure prediction showed 5 helices, 2 strands, 7 beta turns for model HP03 while 7 helices, 4 strands and 7 beta turns for HP38 (Fig 3 a-b).



(a) (b)

Fig. 3 (a-b) The secondary structure of (a) HP38 and (b) HP80 showing α helices, b-Sheets and b-hairpins.

ERRAT is a program for verifying protein structures determined by crystallography. ERRAT server was also used for model quality estimation which showed that overall quality factor was 77.11 and 81.32 % for model HP38 andHP80 respectively which confirms as a good model (Fig.4a- b).



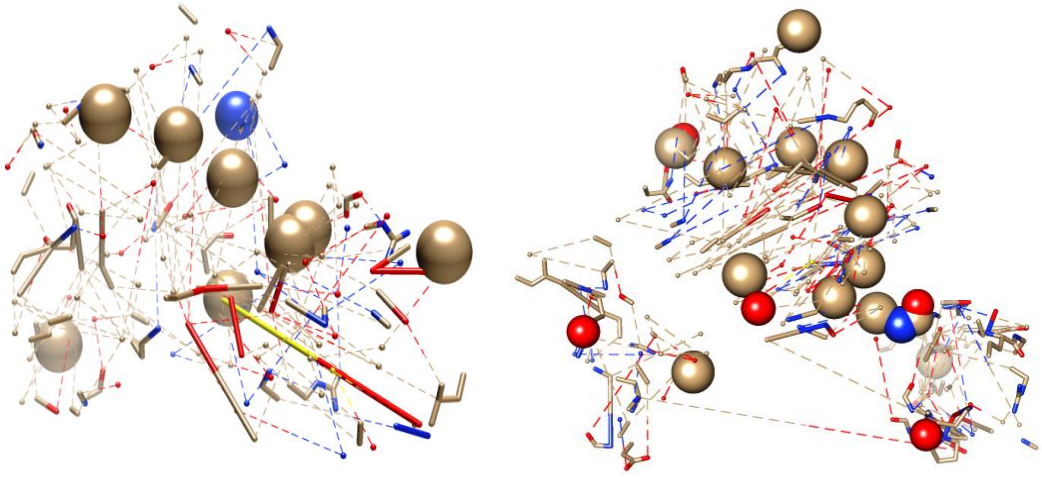
(a) (b)

Fig. 4 (a-b):-Model quality estimation plot obtained by ERRAT server for (a) model HP38 (b) model HP80

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 05, May-2020

3.3 Active sites in predicted models

Active sites on a protein is of fundamental importance for a range of applications including molecular docking, de novo drug design and structural identification and comparison of functional sites. During analysis, seven and four binding pockets were identified for model HP38 and HP80, respectively (Fig 4 c-d).



(C) (d)

Fig4(c-d). Active sites (shown in balls) identified in protein (c) HP38 (d) HP80

During analysis of NABPs in HP38 model, two largest positive patches were detected viz Patch 1: 'GHE', 'SFN', 'LCS', 'FPK', 'PAS', 'CON' Patch 2: 'FPK', 'PAS', 'LCS', 'CON', 'GHE'. For model HP38, two largest positive patches were detected viz Patch 1 'GHE', 'SFN', 'LCS', 'FPK', 'PAS', 'CON' Patch 2: 'PAS','FPK', 'LCS','SFN', 'GHE' Patch 3: 'PAS', 'LCS', 'GHE'. Results are shown in Table 3.

HP Model	Pocket No.	z-score	Pocket Sites
HP38 (YP_004928038.1**)**	1	11.99	'GHE-1', 'SFN-1', 'LCS-1', 'FPK-1', 'PAS-2', 'CON-1'
	2	3.52	'PAS-1', 'GHE-2'
	3	1.49	'LCS-2'
	4	0.93	'FPK-2', 'PAS-3', 'LCS-3', 'CON-2', 'GHE-3'
	5	0.74	'FPK-3'
	6	0.18	'SFN-2'
	7	0.04	'SFN-3'
HP80 (YP_004927980.1**)**	1	18.38	'GHE-1', 'SFN-1', 'LCS-1', 'FPK-1', 'PAS-1', 'CON-1'
	2	4.56	'PAS-2','FPK-2', 'LCS-2','SFN-2', 'GHE-2'
	3	1.49	'FPK-3', 'SFN-3'
	4	-0.02	'PAS-3', 'LCS-3', 'GHE-3'

Table 3. MetaPocket clusters and their functional residues

This data of active binding site residues will give insight into identifying binding interactions and docking with specific ligands.

4. CONCLUSION

Our primary sequence-based analysis led to the identification of two HPs as biologically significant, which might be involved as enzymes (antitoxins, conjugal DNA transferase, and oxido-reductase etc.). Furthermore, we successfully predicted the structure of two HPs to describe their functions at the molecular level. The outcome of the present study may facilitate better understanding of the mechanism of virulence, drug resistance, pathogenesis, adaptability to host, tolerance for host immune response and drug discovery for treatment of infections caused by *P. aeruginosa*.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 05, May-2020

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial assistance under project 'Application of Microorganisms in Agriculture and Allied Sectors (AMAAS)' from Indian Council of Agricultural Research (ICAR), India.

Conflict of interest The authors declare that they have no conflict of interest.

REFERENCES

[1] Khan, N. H., Ishii, Y., Kimata-Kino, N., Esaki, H., Nishino, T., Nishimura, M., Kogure, K., 2007. Isolation of Pseudomonas aeruginosa from open ocean and comparison with freshwater, clinical, and animal isolates. Microbial ecology, 53(2), 173-186.

[2] Lyczak. J. B., Cannon, C. L., Pier, G. B., 2000. Establishment of *Pseudomonas aeruginosa* infection: lessons from a versatile opportunist. Microbiology Infection, 2,1051–1060.

[3] Williams, B.J., Dehnbostel, J., Blackwell, T.S., 2010. *Pseudomonas aeruginosa*: host defence in lung diseases. Respirology 15, 1037– 1056.

[4] Mayer, L. W., 1988. Use of plasmid profiles in epidemiologic surveillance of disease outbreaks and in tracing the transmission of antibiotic resistance. Clinical microbiology reviews, 1(2), 228-243.

[5] Rodríguez-Andrade, E., Hernández-Ramírez, K. C., Díaz-Peréz, S. P., Díaz-Magaña, A., Chávez-Moctezuma, M. P., Meza-Carmen, V., Ramírez-Díaz, M. I., 2016. Genes from pUM505 plasmid contribute to *Pseudomonas aeruginosa* virulence. Antonievan Leeuwenhoek, 109(3), 389-396.

[6] Wang, H., Tu, F., Gui, Z., 2013. Virulence factors in *Pseudomonas aeruginosa*: mechanisms and modes of regulation. Indian Journal of Microbiology, 2, 163-167.

[7] Roberts, R. J., 2004. Identifying protein function -a call for community action. PloS 2, E42.

[8] Kumar K., Prakash A., Tasleem M., Islam A., Ahmad F., Hassan M.I., 2014. Functional annotation of putative hypothetical proteins from Candida *dubliniensis*. Gene 543,93–100.

[9] Loewenstein, Y., Raimondo, D., Redfern, O.C., Watson, J., Frishman D., Linial M., Orengo C., Thornton J., Tramontano A., 2009. Protein function annotation by homology-based inference. Genome Biology, 10,207.

[10] Shahbaaz M., Hassan, M. I., Ahmad, F., 2013. Functional annotation of conserved hypothetical proteins from *Haemophilus influenzae* Rd KW20. PLoS ONE 8, e84263.

[11] http://www.ncbi.nlm.nih.gov

[12] Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. E., Wilkins, M. R., Appel, R. D., Bairoch, A., 2005. Protein identification and analysis tools on the ExPASy server. Humana Press, (pp. 571-607).

[13] Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., Gwadz, M., 2001. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic acids research,39(suppl 1), D225-D229.

[14] Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., & Studholme, D. J., 2004. The Pfam protein families database. Nucleic acids research, 32(suppl 1), D138-D141.

[15] Powell, S., Szklarczyk, D., Trachana, K., Roth, A., Kuhn, M., Muller, J., Jensen, L. J., 2012. eggNOG v3. 0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. Nucleic acids research, 40(D1), D284-D289.

[16] Yu, C. S., Cheng, C. W., Su, W. C., Chang, K. C., Huang, S. W., Hwang, J. K., Lu, C. H., 2014. CELLO2GO: a web server for protein sub CELlular LOcalization prediction with functional gene ontology annotation. PloS one,9(6), e99368.

[17] Gardy, J. L., Spencer, C., Wang, K., Ester, M., Tusnady, G. E., Simon, I., Brinkman, F. S., 2003. PSORT-B: Improving protein subcellular localization prediction for Gram-negative bacteria. Nucleic acids research,31(13), 3613-3617. Yu

[18] Loewenstein, Yu., Raimondo, D., Redfern, O.C., Watson, J., Frishman D., Linial M., Orengo C., Thornton J., Tramontano A., 2009. Protein function annotation by homology-based inference. Genome Biology, 10,207.

[19] Kiefer, F., Arnold, K., Künzli, M., Bordoli, L., Schwede, T., 2009. The SWISS-MODEL Repository and associated resources. Nucleic acids research,37(suppl 1), D387-D392.

[20] Kelley, L. A., Sternberg, M. J., 2009. Protein structure prediction on the Web: a case study using the Phyre server. Nature protocols,4(3), 363-371.

[21] Laskowski, R. A., MacArthur, M. W., Moss, D. S., Thornton, J. M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. Journal of applied crystallography, 26(2), 283-291.

[22] Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R., Thornton, J. M., 1996. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. Journal of biomolecular NMR, 8(4), 477-486.

[23] Li, M., Wang, B., 2007. Homology modeling and examination of the effect of the D92E mutation on the H5N1 nonstructural protein NS1 effector domain. Journal of molecular modeling, 13(12), 1237-1244.

[24] Zhang, Z., Li, Y., Lin, B., Schroeder, M., Huang, B., 2011. Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. Bioinformatics, 27(15), 2083-2088.

[25] Arcus, V. L., McKenzie, J. L., Robson, J., Cook, G. M., 2011. The PIN-domain ribonucleases and the prokaryotic Vap BC toxin-antitoxin array. Protein Eng. Des. Sel. 24 (1-2), 33–40.

[26] Díaz, D. A., Barreto, G. E., Santos, J. G., 2014. Structural and Functional Prediction of the Hypothetical Protein Pa2481 in Pseudomonas Aeruginosa Pao1. Advances in Computational Biology, Springer International Publishing (pp. 47-55).

[27] Moore, D., Maneewannakul, K., Maneewannakul, S., Wu, J. H., Ippen-Ihler, K., Bradley, D. E., 1990. Characterization of the F-plasmid conjugative transfer gene traU, Journal of Bacteriology,172, 4263-4270.

[28] Bork, P., Koonin, E. V., 1996. Protein sequence motifs. Current opinion in structural biology, 6(3), 366-376.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 05, May-2020

Supplementary Tables

Table 1. Supplementary table S1. Physiochemical characterization of HPs protein

S.No.	Gene bank Accession Number	Sequence length	M. wt	pI	– R	+ R	EC	II	Protein Claas	AI	GRAVY
1.	YP_004927991.1	228	24996.02	4.80	32	22	50460	42.42	unstable	78.86	-0.375
2.	YP_004928098.1	110	12894.83	4.71	20	12	11920	59.63	unstable	101.18	-0.205
3.	YP_004928060.1	606	65668.79	4.58	66	41	106480	17.53	stable	77.81	-0.271
4.	YP_004928038.1	136	15199.66	6.73	17	17	9970	44.48	unstable	113.38	-0.267
5.	YP_004927994.1	245	27858.82	8.87	35	39	25565	57.76	unstable	94.86	-0.543
6.	YP_004927986.1	206	22816.97	6.12	21	17	40575	68.31	unstable	99.51	-0.127
7.	YP_004927980.1	340	38425.87	5.25	53	39	35870	48.97	unstable	76.12	-0.611
8.	YP_004927975.1	443	49901.31	7.18	56	56	52035	48.98	unstable	78.42	-0.691
9.	YP_004927989.1	242	27935.31	9.93	25	43	36690	45.12	unstable	81.03	-0.683
10	YP_004928003.1	312	33486.89	8.06	23	25	65360	24.16	Stable	76.38	-0.101
11	YP_004928012.1	219	25508.95	6.53	29	28	53650	47.42	unstable	76.58	-0.495
12	YP_004928109.1	289	32660.03	5.70	42	36	39545	51.49	unstable	84.15	-0.410