ORIGINAL ARTICLE

In Silico Structural and Functional Annotation of Hypothetical Proteins of *Vibrio cholerae* O139

Md. Saiful Islam, Shah Md. Shahik, Md. Sohel, Noman I. A. Patwary, Md. Anayet Hasan*

Department of Genetic Engineering & Biotechnology, Faculty of Biological Sciences, University of Chittagong, Chittagong 4331, Bangladesh

In developing countries threat of cholera is a significant health concern whenever water purification and sewage disposal systems are inadequate. *Vibrio cholerae* is one of the responsible bacteria involved in cholera disease. The complete genome sequence of *V. cholerae* deciphers the presence of various genes and hypothetical proteins whose function are not yet understood. Hence analyzing and annotating the structure and function of hypothetical proteins is important for understanding the *V. cholerae*. *V. cholerae* O139 is the most common and pathogenic bacterial strain among various *V. cholerae* strains. In this study sequence of six hypothetical proteins of *V. cholerae* O139 has been annotated from NCBI. Various computational tools and databases have been used to determine domain family, protein-protein interaction, solubility of protein, ligand binding sites etc. The three dimensional structure of two proteins were modeled and their ligand binding sites were identified. We have found domains and families of only one protein. The analysis revealed that these proteins might have antibiotic resistance activity, DNA breaking-rejoining activity, integrase enzyme activity, restriction endonuclease, 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 cholera.

Keywords: cholera, computational tools, docking, drug discovery, *Vibrio cholerae* O139

Introduction

Vibrio cholerae is a gram-negative, highly motile, curved or comma-shaped rod with a single polar flagellum [1]. *V. cholerae* is transmitted by the fecal-oral route, mainly found in unhygienic environment. *V. cholerae* secretes enterotoxin that induces a life-threatening secretory diarrhea called cholera. Cholera is a major epidemic disease. The cholera toxin binds to the plasma membrane of intestinal epithelial cells and releases an enzymatically active subunit which causes a escalation in cyclic adenosine 5-monophosphate (cAMP) production. The resulted high cAMP level inside the cell causes massive secretion of electrolytes and water into the intestinal lumen. Other *Vibrios* may also be clinically significant for human and some are well-known to cause diseases in domestic animals as well. Nonpathogenic *Vibrios* are widely dispersed in the environment, mostly in estuarine waters and seafood's [2]. *V. cholerae* comprises nearly 200

serogroups based on the O antigenic structures [3]. Among them two serogroups of *V. cholerae* O1 and O139 cause widespread cholera epidemics [4]. The emergence in 1992 of a *V. cholerae* non-O1 serovar, labeled *V. cholerae* synonym O139 Bengal, in Bangladesh and India and its subsequent appearance in Southeast Asia, displacing *V. cholerae* O1 El Tor, was well known causative agent in the history of cholera [5]. In the autumn of 1993, *V. cholerae* serogroup O139 (Bengal), was implicated in outbreaks of cholera in Bangladesh and India. *V. cholerae* serogroup O139 (Bengal), causes characteristic severe cholera symptoms and has been implicated in a case of a traveler returning from India to the United States [6]. *V. cholerae* O139 serogroup strains showed susceptibility to 22 anti-bacterials in various regions of the world and an increase in resistant markers with resistance to fluoroquinolones [7].

During recent years, hundreds of bacterial genomes are available, while their annotation is of interest [8]. However, many of these protein functions are still unknown. For this

reason, there is an increasing demand for the annotation of the functions of uncharacterized proteins, called "hypothetical proteins" [9], but the structures of which are known though. Structural genomics initiatives deliver plenty of structures of hypothetical proteins at a constantly growing rate. However, without function annotation, this huge structural storage is of little use to biologists who are interested in particular molecular systems. Additionally some of the proteins, which are known to be sound annotated, may have further functions beyond their listed archives. About half of the proteins in genomes are candidates for hypothetical proteins (HPs) [10]. Many of the "hypothetical proteins" occur in fact in more than one bacterial species, which increases the probability that they are indeed protein coding genes and not the consequence of erroneous gene predictions. Proteins that occur in diverse species can be combined into orthologous groups, which are known to be suitable for functional analyses and annotations of newly sequenced genomes [11].

Improving the functional annotation is of great importance for many follow up studies and we here apply computational tools for function prediction for one of the most devastating human pathogens *V. cholerae* O139, the causative agent of cholera especially in Southeast Asia. Therefore an improved functional annotation of its proteome is of particular urgency. The annotation of these HPs may be helpful as markers and pharmacological targets. With the overall faith that the majority of hypothetical proteins are the product of pseudogenes, it is necessary to have a tool with the capability of analyzing the minority of hypothetical proteins with a high probability of being expressed [10]. So far, there is no classification of HPs and functioning terms are swapping definitions of hypothetical proteins. Here, we combined physiochemical properties with protein-protein interaction (PPI) based function predictions. Our present study is mainly aimed to predict the structure, function and binding sites of these HPs which are important for docking studies for drug designing.

Methods

Sequence retrieval

Six randomly selected HPs which contain standard number of amino acids sequences of *V. cholerae* O139 were randomly retrieved from NCBI (http://www.ncbi.nlm.nih.gov/) for annotation. Moreover these were supposed to find out interactions between these proteins as they are both from chromosomal and plasmid DNA. The sequence IDs of those 6 HPs were gi|84095108, gi|163644906, gi|163644912, gi|163644916, gi|84468567, and gi|84468557. Various computational tools and databases were used to analyze the

different properties i.e., physicochemical, functional, and structural characteristics of HPs.

Physicochemical and functional categorization

By using the Expasy's Protparam server (http://us.expasy.org/tools/protparam) physicochemical characterization, molecular weight, theoretical isoelectric point (pI), total number of positive and negative residues, extinction coefficient [12], instability index [13], aliphatic index (AI) [14] and grand average hydropathy (GRAVY) [15] of HPs were analyzed.

Pfam

Pfam (http://pfam.sanger.ac.uk/) is designed as a comprehensive and accurate collection of protein domains and families [16, 17]. Pfam families are typed as Pfam-A and Pfam-B. Each Pfam-A family possess a curated seed alignment containing a small set of envoyed members of the family and an automatically created full alignment which contains all noticeable protein sequences belonging to the family, as defined by profile Hidden Markov Models searches of primary sequence databases. On the other hand, Pfam-B entries are automatically created from the ProDom database and are shown by a single alignment [18].

CDD-BLAST

CD-Search (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi/) was done to find out the conserved domain of these protein sequences. This was performed with the use of RPSBLAST, a modified version of PSI-BLAST, to quickly scan a set of predetermined position-specific scoring matriceswith a protein query [19].

PPI prediction

STRING (http://string.embl.de/) is a database of known and predicted protein interactions by using four sources: Genomic Context, (Conserved) Co-expression, High-throughput Experiments, and Previous Knowledge. STRING currently contains the databases of 5,214,234 proteins from 1,133 organisms [20].

Proteins location prediction

PSORTB (http://www.psort.org/psortb/) server was used to predict the cellular locations of HPs and then SOSUI server was used to find out whether the protein is soluble or trans-membrane in nature (http://bp.nuap.nagoya-u.ac.jp/sosui/sosui_submit.html).

Detection of disulfide bridges

DISULFIND (http://disulfind.dsi.unifi.it/) server was used to predict the presence of any disulfide bond state

between cysteine residues in the amino acid sequences of HPs. Moreover, disulfide bridges play a key role in the stabilization of the folding process for many proteins. We analyzed the data using this software. The disulfide bridges are very important finding in the study of structural and functional properties of specific proteins [21].

Protein structure prediction

(PS)2 (pronounced PS square) was used for the prediction of the tertiary structures of HPs (http://www.ps2.life. nctu.edu.tw/). This method combined PSI-BLAST [22, 23], IMPALA [24], and T-Coffee [25] by using an effective accord strategy in both target-template selection and target-template alignment. Three dimensional structures were constructed further using the modeling package MODELLER [26-28]. The predicted structures obtained from the PS square were saved in the Protein Data Bank (PDB) formats.

Active site prediction

Q-SiteFinder (http://www.modelling.leeds.ac.uk/qsitefin-der/) was used to find out the ligand binding sites. It works by finding clusters of probes and binding hydrophobic (CH3) probes to the protein with most favorable binding energy. Q-SiteFinder requires uploading a PDB file or selecting one from the Protein Database. 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 [29]. We used this tool for evaluating these features including the active site in the desired sequence.

Results and Discussion

We analyzed the physiochemical properties of these HPs of cholera for the first time. In Table 1 the physicochemical properties of HPs are tabulated. Isoelectric point (pI) of the HP ranges from 4.62 to 9.78. pI is the pH at which the amino acid of protein tolerates no net charge and hence does not

move in a direct current electrical field. The determined pI will be handy as solubility is minimum and in an electro focusing system mobility is zero at pI. Moreover proteins become stable and compact at isoelectric pH, for this reason computed pI will be helpful for developing a buffer system for purification by isoelectric focusing method.

At 280 nm, the extinction co-efficient of HPs ranges from 23295 to 62005 M cm computed by Expasy's Protparam instead of 276, 278, 279, and 282 nm. The presence of high concentration of Cys, Trp, and Tyr indicates a higher extinction coefficient of HPs. The quantitative study of protein-protein and protein-ligand interactions in solution can be done by using this computed extinction coefficients. The instability index value of the HP was found to be ranging from 30.44 to 50.35. It is predicted that a protein will be stable whose instability index is smaller than 40, a value above 40 predicts that the protein will be unstable [13].

Another parameter of structure identification of protein is instability index. Proteins, gi|163644906, gi|163644912, and gi|163644916 were stable and others were unstable. The instability index indicates an approximate stability of proteins in a test tube.

The AI is the relative volume of a protein occupied by aliphatic side chains (A, V, I, and L) and is considered as a positive factor for the raise of thermal stability of globular proteins. The range of the AI for the HPs is from 64.14 to 82.92. The proteins with very high AI may show stability in a wide temperature range where lower AI proteins are not thermal stable and show more flexibility.

The GRAVY of HPs is ranging from −0.304 to −0.633. The better interaction of protein and water is occurring in low GRAVY. The GRAVY value for a protein is calculated by adding the values of hydropathy of all the amino acids and dividing it by the number of residues in the sequence [14].

To study the functional analysis conserved domains were observed because conserved domains are functional units within a protein that act as building blocks in molecular evolution and recombine in various arrangements to make

Table 1. Physicochemical properties of hypothetical proteins of Protparam tool

Sequence ID	No. of AAs	MW	pI	(+)R	(−)R	EC	II	AI	GRAVY
gi|84095108	161	18,290.0	8.30	19	16	38,555	49.29	66.65	−0.429
gi|163644906	284	31,693.3	8.78	45	40	23,295	30.44	82.92	−0.385
gi|163644912	209	23,395.4	4.62	18	35	39,670	35.51	81.20	−0.304
gi|163644916	210	23,085.9	7.65	26	25	32,680	31.51	64.14	−0.633
gi|84468567	183	21,248.2	9.20	26	21	30,035	50.35	80.98	−0.591
gi|84468557	208	24,368.6	9.78	29	17	62,005	46.71	67.50	−0.629

AA, amino acid; MW, molecular weight; pI, isoelectric point; (+)R, total number of positively charged residues (Arg + Lys); (−)R, total number of negatively charged residues (Asp + Glu); EC, extinction coefficient; II, instability index; AI, aliphatic index; GRAVY, grand average hydropathy.

proteins with different functions. The data are then used for putative functional annotation of protein query sequences based on matches to specific super-families history, identification of proteins with similar domain. The proteins have been classified into particular families based on the presence of specific domains in the sequence [19]. In our study we used 6 HPs but found only 1 protein gi|84468567 possessing specific domains which were DNA_BRE_C super-family, Topoisomer_IB_N, DUF3946 domains and they were classified as super-families accordingly. The presence of these domains in the HPs indicates that the protein might do the same function. The domains of the HP gi|84468567 and their super-family is given by function in Tables 2 and 3.

Domains and families present in HPs were identified by the Pfam database research (Tables 4 and 5). They are PhnA Zn ribbon, prokaryotic membrane lipoprotein lipid attachment site, Phage integrase family and integrase core domain.

To explain the protein functions involved in various cellular processes it is important to know the sub-cellular localization of that protein. During the drug discovery process knowledge of the sub-cellular localization of a protein play a very significant role in target identification.

In our study, we have found two proteins gi|163644906, gi|184468567 are cytoplasmic as their best performing sites. The remaining other protein localization was not found. The server SOSUI differentiates whether the HPs are membranous or soluble. No trans-membrane protein was found and all were soluble.

Moreover, DISULPHIND server revealed no disulphide bonds were present in any of those proteins which indicate that they were thermally unstable. Moreover, disulfide bridges play a major role in stabilizing the folding process of many proteins. Disulfide bridges are very important finding in the study of structural and functional properties of specific proteins [21].

For performing almost all the cellular functions the PPI are important. Proteins often interact with one another in a mutually dependent way to perform a common function. It is notable that translational factors interact among themselves to carry out the whole translation. The function of protein is predictable from this based on their interaction with other proteins. It is very rare that proteins bring out function with any interactions with other biomolecules. For this reason, in this post genomic era PPI databases have turned as a most important resource for searching biological networks and pathways in cells [29]. The proteins gi|163644906 and gi|163644912 were found to have interaction with 2 proteins signal peptide peptidase SppA domain-containing protein and DSBA like thioredoxin domain containing protein. gi|163644916 had interacted with 3 proteins such as IV

Table 2. Identification of domains by CDD-BLAST

Sequence ID	Domains
gi|84468567	DNA_BRE_C superfamily, Topoisomer_IB_N, DUF3946

Table 3. Functional description of superfamilies of hypothetical proteins

Superfamily	Description
DNA_BRE_C	DNA breaking-rejoining enzymes, C-terminal catalytic domain. The DNA breaking-rejoining enzyme superfamily includes type IB topoisomerases and tyrosine recombinases that share the same fold in their catalytic domain containing six conserved active site residues. The best-studied members of this diverse superfamily include human topoisomerase I, the bacteriophage lambda integrase, the bacteriophage P1 Cre recombinase, the yeast Flp recombinase and the bacterial XerD/C recombinases.
DUF3946	Protein of unknown function (DUF3946); a family of uncharacterized proteins found by clustering human gut metagenomic sequences. This family appears related to the N-terminal domain of phage integrases.
Topoisomer_IB_N	Topoisomer_IB_N: N-terminal DNA binding fragment found in eukaryotic DNA topoisomerase (topo) IB proteins similar to the monomeric yeast and human topo I and heterodimerictopo I from *Leishmania donvanni*.

Table 4. Families found in Pfam database

| Sequence ID	Pfam-A	Pfam-B	Domains
gi|84095108	PhnA Zn ribbon	Pfam-B_18384	-
gi|163644906	-	-	-
gi|163644912	-	-	-
gi|163644916	LPAM_1	Pfam-B_4989	-
gi|84468567	Phage integrase	Endonuc-PvuII	-
gi|84468557	Rev	Pfam-B_12598	Integrase core

conjugative transfersystem protein TraD & TraI, putative type IV conjugative transfer system coupling factor. gi|84468567 showed interaction with 6 proteins which were (1) ribosomal protein-alanine acetyl transferase, (2) re-

combination factor protein RarA, (3) ATP-dependent RNA helicase HrpA, (4) Zinc-binding domain-containing protein, (5) putative ATP-dependent helices, and (6) dihydroxy-acid dehydrates. gi|84468557 protein interacted only one protein ISVch4 transposes. Other HPs do not interact with any other proteins. Fig. 1 and Table 6 indicate the protein-protein interacting networks of HPs, which might have functions of their interacting proteins [30, 31].

PS square server (Fig. 2) was used to determine the three dimensional structure of the HPs. Out of 6 HPs, the PS square server could model only 2 proteins. Due to low sequence identity, the other four proteins could not be modeled. The server used templates to model those proteins

Table 5. Descriptions of Pfam families of hypothetical proteins

Sequence ID	Description
gi|84095108	PhnA Zn ribbon
gi|163644916	Prokaryotic membrane lipoprotein lipid attachment site
gi|84468567	Phage integrase family
gi|84468557	Integrase core domain
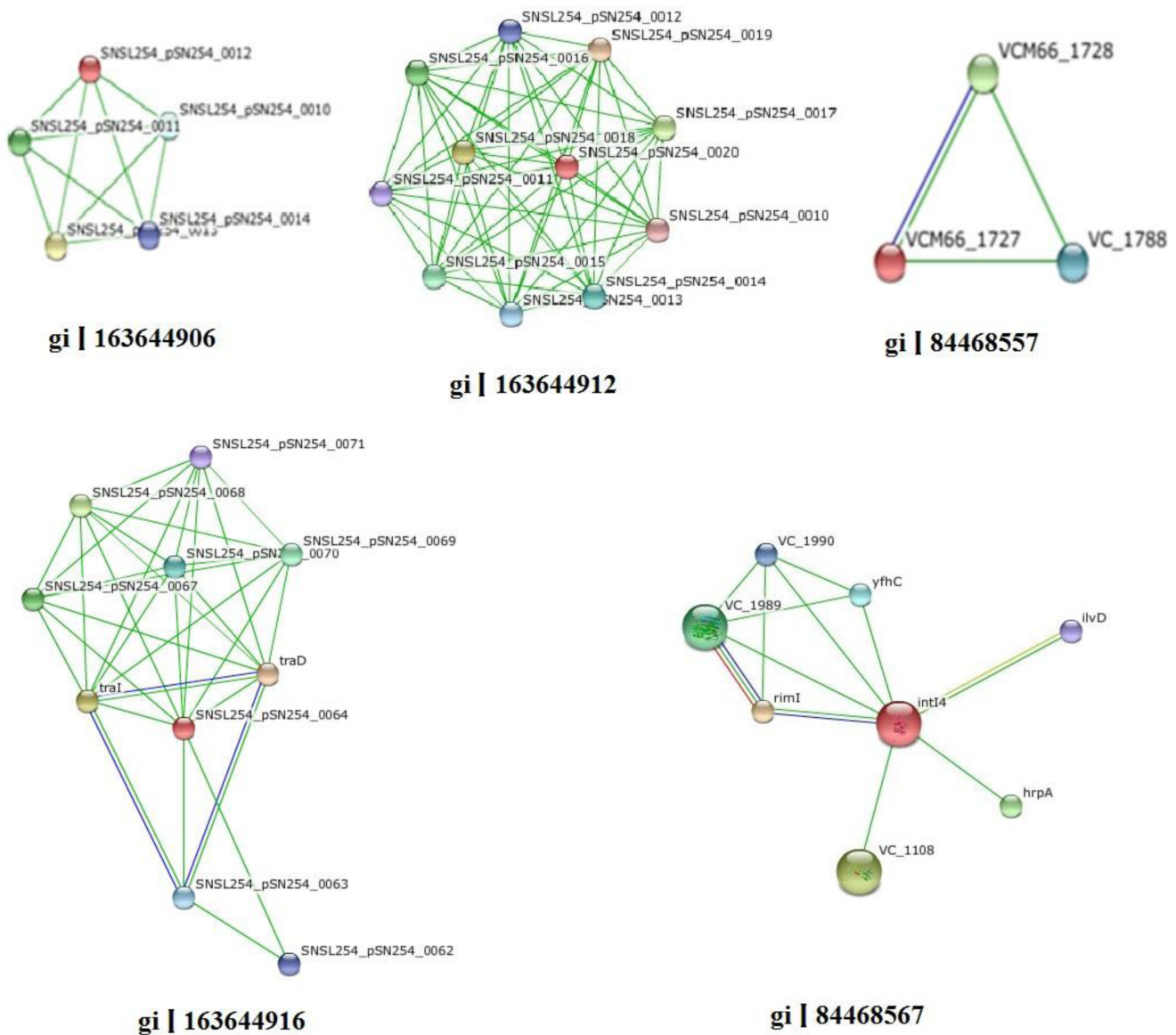
Fig. 1. Protein-protein interaction of hypothetical proteins.
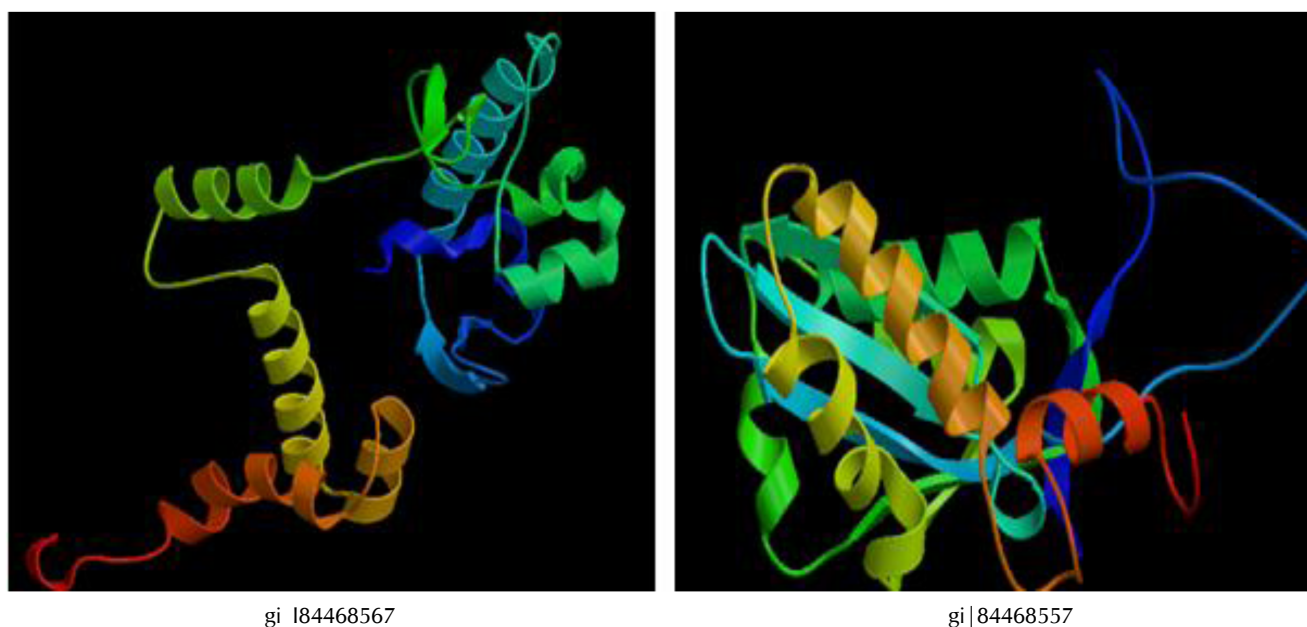
gi I84468567   gi|84468557

Fig. 2. Three-dimensional structure of hypothetical protein by PS square.

Table 6. Hypothetical proteins interacting with functionally important proteins

Sequence ID	Interacting protein
gi|163644906	Signal peptide peptidase SppA domain-containing protein DSBA-like thioredoxin domain-containing protein
gi|163644912	Signal peptide peptidase SppA domain-containing protein DSBA-like thioredoxin domain-containing protein
gi|163644916	Type IV conjugative transfer system protein TraD Type IV conjugative transfer system protein TraI Putative type IV conjugative transfer system coupling factor
gi|84468567	Ribosomal-protein-alanine acetyltransferase Recombination factor protein RarA ATP-dependent RNA helicase HrpA Zinc-binding domain-containing protein Putative ATP-dependent helicase Dihydroxy-acid dehydratase
gi|84468557	ISVch4 transposase

Table 7. Templates used by PS square server for modeling

Sequence ID	Template
gi|84468567	2a3vB
gi|84468557	1bcoA

Table 8. Residues involved in ligand binding sites predicted by QSITE finder

Sequence ID	Site volume	Residue
gi|84468567	499	MET 1, GLU 2, CYS 3, ARG 5, LEU 6, ARG 7 ,GLN 9,ASP 10, ARG 19, ILE 20, TRP 21, GLN 22, GLY 23, LYS 24, GLY 26, LYS 27, TRP 65, LEU 66, PRO 67, LEU 70, TRP 83, TYR 85
gi|84468557	493	GLY 45, ASP 46, VAL 47, ALA 60, VAL 61, VAL 62, SER 81, LEU 83, THR 84, GLY 85, ALA 87, LEU 88,SER 89, PHE 103, HIS 104, SER 105, GLN 107, THR 112, LYS 115, TYR 116, ILE 125, LYS 126, SER 128, LEU 129, ARG 132, TRP 136, ASP 137, ASN 138

which were tabulated in Table 7. The location of ligand binding site identification on protein is important for a wide range of applications including structural identification, comparison of functional sites, molecular docking and de novo drug design. Active site residues of the HPs are mentioned in Table 8. This data of active binding site residues will give insight into identifying binding interactions and docking with specific ligand.

We have retrieved 6 HPs from NCBI database and determined their physicochemical properties and identified domains and families using various Bioinformatics tools and

databases. The three dimensional structure of those HPs were modeled (only 2) and their ligand binding sites were identified. Among them we have found domains and families of only one HP, analysis showed that the domains and families are involved in DNA breaking-rejoining activities, integrase activity. All of these features from our findings may be used to design new potential drugs against this infectious bacterium.

References

1. Reidl J, Klose KE. *Vibrio cholerae* and cholera: out of the water and into the host. *FEMS Microbiol Rev* 2002;26:125-139.

2. Finkelstein RA. Cholera, *Vibrio cholerae* O1 and O139, and other pathogenic Vibrios. In: *Medical Microbiology* (Baron S, ed.). 4th ed. Galveston: University of Texas Medical Branch at Galveston, 1996. pp. 158-67.

3. Shanan S, Abd H, Hedenström I, Saeed A, Sandström G. Detection of *Vibrio cholerae* and *Acanthamoeba* species from same natural water samples collected from different cholera endemic areas in Sudan. *BMC Res Notes* 2011;4:109.

4. Siddique AK, Baqui AH, Eusof A, Haider K, Hossain MA, Bashir I, *et al*. Survival of classic cholera in Bangladesh. *Lancet* 1991;337:1125-1127.

5. Alam AN, Alam NH, Ahmed T, Sack DA. Randomised double blind trial of single dose doxycycline for treating cholera in adults. *BMJ* 1990;300:1619-1621.

6. Large epidemic of cholera-like disease in Bangladesh caused by *Vibrio cholerae* O139 synonym Bengal. Cholera Working Group, International Centre for Diarrhoeal Diseases Research, Bangladesh. *Lancet* 1993;342:387-390.

7. Krishna BV, Patil AB, Chandrasekhar MR. Fluoroquinolone-resistant *Vibrio cholerae* isolated during a cholera outbreak in India. *Trans R Soc Trop Med Hyg* 2006;100:224-226.

8. Roberts RJ. Identifying protein function: a call for community action. *PLoS Biol* 2004;2:E42.

9. Galperin MY, Koonin EV. 'Conserved hypothetical' proteins: prioritization of targets for experimental study. *Nucleic Acids Res* 2004;32:5452-5463.

10. Friedberg I. Automated protein function prediction: the genomic challenge. *Brief Bioinform* 2006;7:225-242.

11. Eisen JA. Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. *Genome Res* 1998;8:163-167.

12. Gill SC, von Hippel PH. Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* 1989;182:319-326.

13. Guruprasad K, Reddy BV, Pandit MW. 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;4:155-161.

14. Ikai A. Thermostability and aliphatic index of globular proteins. *J Biochem* 1980;88:1895-1898.

15. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 1982;157:105-132.

16. Sonnhammer EL, Eddy SR, Durbin R. Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins* 1997;28:405-420.

17. Finn RD, Mistry J, Schuster-Böckler B, Griffiths-Jones S, Hollich V, Lassmann T, *et al*. Pfam: clans, web tools and services. *Nucleic Acids Res* 2006;34:D247-D251.

18. Bru C, Courcelle E, Carrère S, Beausse Y, Dalmar S, Kahn D. The ProDom database of protein domain families: more emphasis on 3D. *Nucleic Acids Res* 2005;33:D212-D215.

19. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, *et al*. CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res* 2011;39:D225-D229.

20. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, *et al*. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res* 2011;39:D561-D568.

21. Ceroni A, Passerini A, Vullo A, Frasconi P. DISULFIND: a disulfide bonding state and cysteine connectivity prediction server. *Nucleic Acids Res* 2006;34:W177-W181.

22. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, *et al*. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389-3402.

23. Schäffer AA, Aravind L, Madden TL, Shavirin S, Spouge JL, Wolf YI, *et al*. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res* 2001;29:2994-3005.

24. Notredame C, Higgins DG, Heringa J. T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 2000;302:205-217.

25. Marti-Renom MA, Stuart AC, Fiser A, Sánchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* 2000;29:291-325.

26. Fiser A, Do RK, Sali A. Modeling of loops in protein structures. *Protein Sci* 2000;9:1753-1773.

27. Hasan MA, Alauddin SM, Al Amin M, Nur SM, Mannan A. In silico molecular characterization of cysteine protease YopT from *Yersinia pestis* by homology modeling and binding site identification. *Drug Target Insights* 2014;8:1-9.

28. Chen CC, Hwang JK, Yang JM. (PS)2: protein structure prediction server. *Nucleic Acids Res* 2006;34:W152-W157.

29. Laurie AT, Jackson RM. Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics* 2005;21:1908-1916.

30. Peri S, Navarro JD, Amanchy R, Kristiansen TZ, Jonnalagadda CK, Surendranath V, *et al*. Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Res* 2003;13:2363-2371.

31. Hasan A, Mazumder HH, Khan A, Hossain MU, Chowdhury HK. Molecular characterization of legionellosis drug target candidate enzyme phosphoglucosamine mutase from *Legionella pneumophila* (strain Paris): an *in silico* approach. *Genomics Inform* 2014;12:268-275.