Int J Pharm Bio Sci 2012 July; 3(3): (B) 280 - 288

Research Article **Bioinformatics**

International Journal of Pharma and Bio Sciences **ISSN 0975-6299**

IN SILICO IDENTIFICATION AND COMPARISON OF SURFACE BINDING POCKETS OF *VIBRIO PARAHAEMOLYTICUS* THERMOSTABLE DIRECT HEMOLYSIN PROTEINS

MALATHI SHEKAR, IDDYA KARUNASAGAR AND INDRANI KARUNASAGAR*

UNESCO- MIRCEN for Marine Biotechnology, Department of Fishery Microbiology, Karnataka Veterinary, Animal and Fisheries Sciences University, College of Fisheries, Mangalore – 575 002, India

ABSTRACT

V.parahaemolyticus pathogenicity has been associated with its ability to produce thermostable direct hemolysin (TDH) encoded by the *tdh* gene. Although all *V.parahaemolyticus* strains possess this gene some strains are hemolytic, intermediate or non-hemolytic. To assess the differences seen, we screened the surface of the TDH proteins showing variable hemolytic activity for identification of possible pockets/cavities. We report here the probable large binding pockets associated with TDH, the key residues involved in their formation and their physicochemical parameters. Several 'hotspot' residues associated with binding sites identified are also reported. This report is the first to study the surface of the TDH protein, the results of which could be used in planning experiments for further understanding the way the protein expresses and interacts.

KEY WORDS; ***Vibrio parahaemolyticus*****, Thermostable direct hemolysin, Surface pockets, binding sites, hotspots**

INDRANI KARUNASAGAR
UNESCO- MIRCEN for Marine Biotechnology, Department of Fishery Microbiology, Karnataka Veterinary, Animal and Fisheries Sciences University, College of Fisheries, Mangalore – 575 002, India

INTRODUCTION

Vibrio parahaemolyticus is a Gram negative halophilic bacterium that causes gastrointestinal illness in humans[1]. Among the several putative virulence factors associated with pathogenesis of this organism, the exotoxin, thermostable direct hemolysin (TDH) has been considered as the key virulence factor and an important virulence marker associated with illness[2,3]. TDH-producing *V.parahaemolyticus* strains are distinguished from their non-TDH-producing counterparts by testing for 'Kanagawa phenomenon' (KP), a β-hemolysis detected on a special blood agar, induced by the production of extracellular TDH[4]. TDH is encoded by the *tdh* gene and studies show that *V.parahaemolyticus* acquired these genes through horizontal gene transfer[5,6]. A typical *V. parahaemolyticus* KP^+ phenotype carries two copies of this gene (*tdh*1 and *tdh*2) on its chromosome, while a KP^- phenotype in addition to having a single copy (*tdh*3) on its chromosome also carries a copy of the gene on its plasmid, designated *tdh*4[5]. A *tdh* in KP^- strain designated *tdh*5 that produced hemolysin at a very low level and having 98% similarity to *tdh*2 is also reported[7]. The recent availability of complete genome sequencing data reveals that *V. parahaemolyticus* carries two copies for this gene designated *tdh*A and *tdh*S on its smaller chromosome[8]. TDH peptide is composed of 189 amino acids[4]. Although the protein sequences corresponding to the TDH phenotypes are highly conserved with a >97% sequence identity[5], it has been seen that only *tdh*1 and *tdh*2 are responsible for the hemolytic phenotype[9] while *tdh*3 and *tdh*4 are the non-hemolytic phenotypes. Further, hemolysis by TDH has been thought to be caused by a serial process of binding and post binding events[10,11] wherein the toxin binds to the erythrocyte surface, forming pores[12] culminating in swelling and osmotic lysis of cells[13]. However, the precise mechanism for its pore-forming ability is still unknown. Studies have shown that the manifestation of KP phenotype to depend on a single base difference at a particular position in the promoter of the TDH gene[14]. The N–terminal and C-terminal regions of the TDH are implicated to be involved in binding and post binding process respectively[11]. Except for these studies, there is no published evidence for any association between the TDH hemolytic activity to the residue differences seen within the protein sequence itself.

The biological activity of proteins is largely governed by the inter-molecular interactions that take place on the protein molecule. Protein surface regions are carved by numerous concavities and projections, which offer the necessary environment for binding interactions, thereby fulfilling their cellular roles and other fundamental biological processes through interaction with other macromolecules[15]. The structural units on the surface region of the protein, that involve only few key residues determine the spatial patterns and orientation of these units providing favourable locations for functional interactions. The protein cavities are known to either accommodate prosthetic groups or functional water molecules, or serve the function of allowing conformational flexibility[16]. Therefore, computation of the geometric properties such as surface area and molecular volume related to protein surfaces has been the focus of research for some time. In this study, we screened the surface topography of the modeled TDH protein for accessible surface pockets that act as binding sites. The large binding sites identified for each of the modeled protein was compared to find the spatial differences between them and for the identification of potential residues that may act as 'hotspots' which could possibly hold clues for further understanding the molecular interactions and recognition of this toxin protein.

MATERIALS AND METHODS

Twenty deduced amino acid sequences (AB112355, AY044107, AY044109, AY044110, AY044111, AY044113, AY044114, AY249144, D90100, D90101, D90238, GU971653, NC4605A, NC4605S, S67841, S76724, X54340, X54341, X54342, X53343) corresponding to the *tdh*

genes and one protein sequence (3A57) were downloaded from GenBank and protein data bank respectively. The programs Multalign[17] and Signal IP ver.3.0[18] was used in multiple alignment and identifying the signal peptide region respectively. Swiss model server[19] was used to generate 3-D structure models of the proteins. The surface pockets were computed using Castp[20], with default probe radius of 1.4°A optimized for water molecule. The default values of the Metapocket 2.0 (http:// projects.biotec.tu-dresden.de/metapocket/index.php) program were employed for identification of binding sites. This program

1

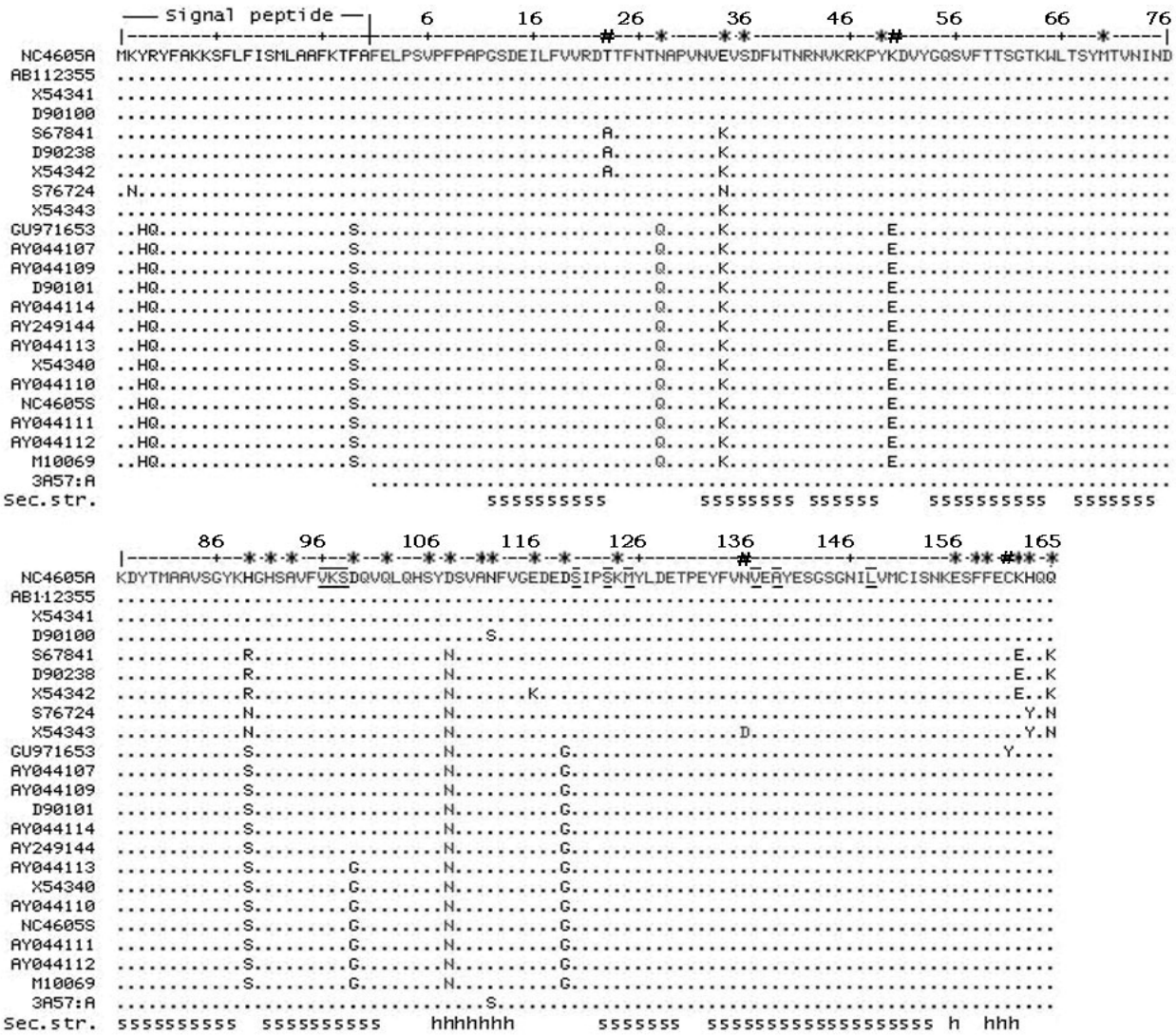

Figure 1

Multiple sequence alignment for thermostable direct hemolysin sequences.

**** highly reliable mutable 'hotspot' residues; # mutable but not reliable 'hotspot' residues***

Table 1
Classification of TDH sequences analysed in this study

TDH group	Representative TDH sequence	Variable amino acid residues in the TDH sequence at positions in the														
		Signal sequence			Mature peptide											
		3	4	23	23#	28#	34*	50#	89*	108*	99*	116*	119*	162*	163*	165*
TDH-S	NC004605S (=*tdh*1)	H	Q	S	T	Q	K	E	S	N	G	E	G	K	H	Q
	AY249144 (=*tdh*1)	H	Q	S	T	Q	K	E	S	N	D	E	G	K	H	Q
TDH-A	NC004605A (=*tdh*2)	Y	R	F	T	N	E	K	H	D	D	E	D	K	H	Q
	X54342 (=*tdh*3)	Y	R	F	A	N	K	K	R	N	D	E	D	E	H	K
	X54343 (=*tdh*4)	Y	R	F	T	N	K	K	N	N	D	E	D	K	Y	N

TDH-S group sequences: AY044107, AY044109 - AY044114, AY249144, D90101, GU971653, X54340, NC004605S. TDH-A group sequences: NC004605A, AB112355, X54341, D90100, S76724, S67841, X54342, X54343, D90238, 3A57. * highly reliable 'hotspsot' residues # high but unreliable 'hotspot' residues

is a consensus based program and does not require any prior knowledge of ligands. The binding residues identified for each pocket using the two programmes were combined and separated into a new pdb file for further analysis. The geometric calculations for the pockets identified were done using the program Molecular geometry (http://www.molmovdb.org/geometry) and characterized using VADAR program[21]. Pocket structures were visualized using Pymol (www.pymol.org) and POCASA 1.0[22]. The Hotspot wizard[23] was employed to identify functional mutational sites in the modeled TDH proteins.

RESULTS AND DISCUSSION

The signal peptidase recognition site identified between residues A24 and F25 divided the 189 residues of TDH into, a signal peptide region (residues 1-24) and a mature peptide region (residues 25-189) (Figure 1). The signal sequence was found to be variable at residue positions 3, 4 and 23 (Figure1).

The genome of *V.parahaemolyticus* has two *tdh* gene copies annotated as *tdh*A and *tdh*S[8]. Based on the variable residues in the signal sequence, the TDH sequences analyzed in this study were seen to be identical to either *tdh*A or *tdh*S (Figure 1) and could be separated into two groups, TDH-S and TDH-A (Table 1).

Nine of the 20 sequences belonging to the TDH-A group had residues Y3, R4 and F23 in their signal peptide, while the remaining 11 sequences belonging to TDH-S had H3, Q4 and S23 (Figure 1 and Table 1). Our analysis also revealed that sequences previously annotated as *tdh*1 belonged to the TDH-S group, while sequences annotated as *tdh*2, *tdh*3, *tdh*4 grouped with TDH-A (Table 2). As seen from Figure1, sequences belonging to the TDH-S group were found to be highly conserved except for a single residue difference at the 99th position (aspartate or glycine). In contrast, *tdh*2, *tdh*3 and *tdh*4 gene sequences in the TDH-A group showed residue differences at positions 23(T/A), 34(E/K/N), 89(H/R/N), 108(D/N), 112(N/S), 116(E/K), 136(N/D), 162(K/E), 163(H/Y), 165(Q/K/N) (Figure 1 and Table 1). The computational method employed by the 'Hotspot wizard' identified 23 highly reliable and 4 high but unreliable mutable 'hotspots' in the mature peptide region of TDH-A and TDH-S sequences (Figure 1). Nine out of the 23 mutable 'hot spots' identified corresponded to the variable regions that showed residue substitutions, while the remaining although recognized as highly mutable were seen to be conserved residues (Figure 1 and Table 1).

The mutable residues identified were mostly in the 89-136 and 156-165 of the C-terminal region of the protein (Figure 1 and Table 1).

The computational analysis using Castp determined the TDH-S sequences to have 16-17 pockets while TDH-A sequences to have 14-16 pockets. The residues lining each pocket were from diverse regions in the primary sequence. Among this, six pockets (3 large and 3 small) were identified as binding pockets by the Metapocket program. Since larger pockets are known to be associated with active sites on proteins involved in binding of ligands, we report here only the three largest pockets (P1, P2, and P3) identified. Among the three pockets, pocket P1 was the largest comprising of 22 residues followed by pockets P2 (20 residues) and P3 (15 residues). Pocket P1 (area 2302.53 $Å^2$ and volume 2495.67$Å^3$) on the protein, formed a deep binding pocket involving a large number of residues belonging mainly to the C-terminal region. The pocket P2 was the second largest pocket (area 2103.85 $Å^2$ and volume 1708.47 $Å^3$). Pocket P3 (area 1508.32 $Å^2$ and volume 1380.51 $Å^3$) was in the close vicinity of pocket P1 having a few common residues lining their cavity walls. The amino acid composition and the physico-chemical properties such as H-bonds, electrostatic and hydrophobic interactions, disulphide bonds etc. for each binding pocket are listed out in Table 2 and the three-dimensional models of the pockets are shown in Figure 2.

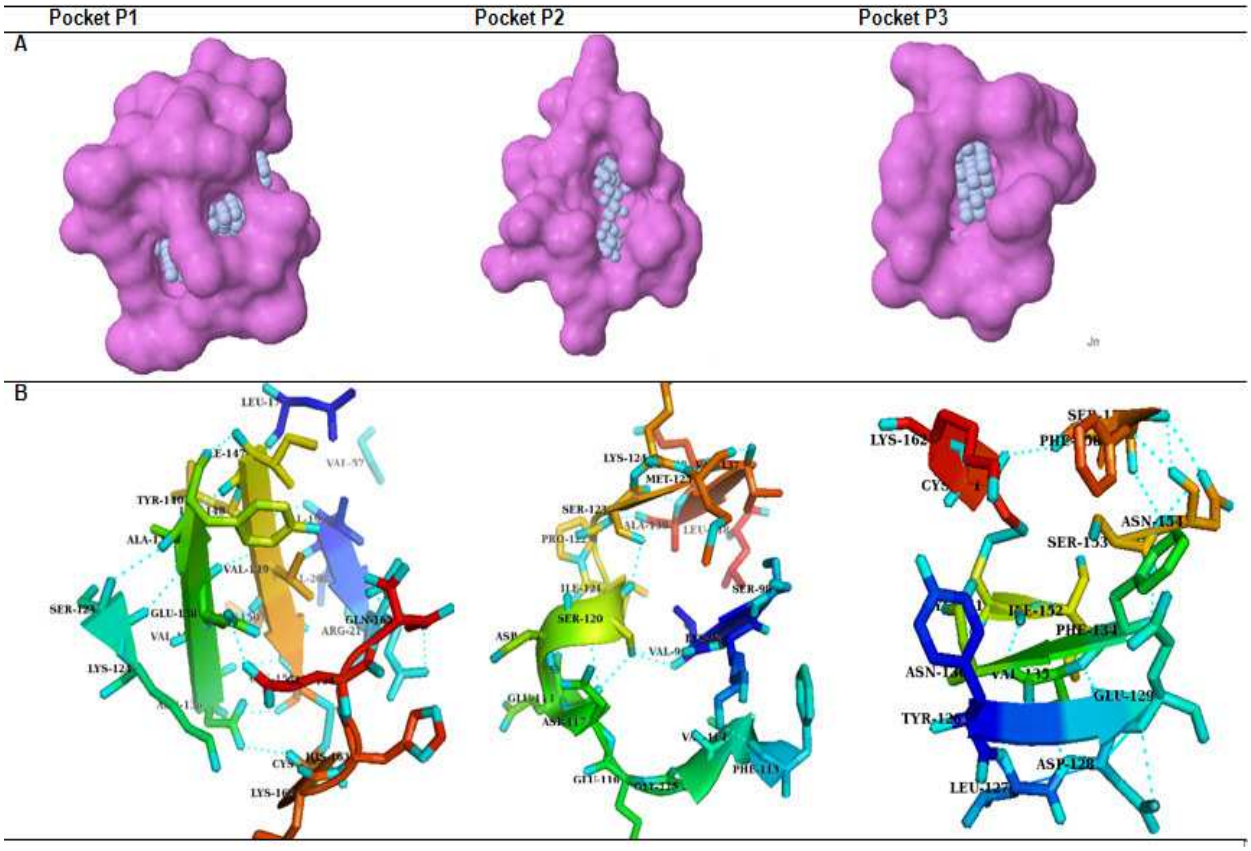

Figure 2

Binding pockets on the surface of the TDH protein

A.Three-dimensional molecular model of predicted Pockets with grids. B. Cartoon depiction showing the amino acid residues associated with binding pockets.

Table 2
Properties of binding pockets on the surface of TDH protein (PDB ID:3A57A)

Property of binding pockets	Pocket P1	Pocket P2	Pocket P3
No. of residues involved	22 (184 atoms)	20 (157 atoms)	15 (122 atoms)
Residues involved	Leu17, Val19, Val20, Arg21, Val57, Ser123, Lys124, Asn136, Val137, Glu138, Ala139, Tyr140, Ile147, Leu148, Val149, Met150, Cys151, Cys161, Lys162, His163, Gln164, Gln165	Val96, Lys97, Ser98, Phe113, Val114, Gly115, Glu116, Asp117, Glu118, Asp119, Ser120, Ile121, Pro122, Ser123, Lys124, Met125, Val137, Glu138, Ala139, Leu148	Tyr126, leu127, Asp128, Glu129, Phe134, Val135, Asn136, Cys151, Ile152, Ser153, Asn154, Ser157, Phe158, Cys161, Lys162
No.of surface atoms	138	124	115
No.of buried atoms	30	21	7
No. of atoms with ASP=O	16	12	0
Secondary structure in binding site	Beta , Coil	Beta , Coil	Coil
Solvent Excluded Area ($Å^2$)	1868.87	1587.22	1186.65
Accessible Surface Area ($Å^2$)	2302.5	2103.85	1580.32
Exposed nonpolar ASA	1431.2	1248.3	1013.7
Exposed polar ASA	579.7	321.6	401.0
Exposed charged ASA	304.6	507.8	234.3
Voronoi Volume (A^3)	2495.67	1708.47	1380.51
No.of hydrogen bonds	10	10	9
Mean HBond distance	1.9	2.1	2.4
Mean HBond energy	-2.6	-2.1	-1.4
Residues with HBonds	13	7	10
Disulphide bonds	1	0	1
Hydrophobic interactions	4	5	2
Solvation free energy of folding (kJ/mol)	-10.56	-11.03	-7.43
Electrostatic potential	1.583	-3.464	-1.464

It has been suggested that the degree of residue conservation defining a protein cavity can be taken as a direct measure of the cavity's conservation[24]. Of the three pockets reported in this study, the TDH pockets P2 and P3 involved residues that were highly conserved. Further, pocket P2 was also identified as the functional pocket by the program 'hotspot wizard'[23] and was observed to be structurally conserved among all the TDH sequences analyzed. This assumes significance as protein pockets with greater residue conservation are well packaged and are important in maintaining the stability and type of interaction in the protein. It has also been suggested that the ligand binding sites tend to bind to the largest pocket/cavity on the protein surface which is also considered as the active site for the protein[25]. Further, for *tdh* the region near the C-terminal region has been implicated in post binding process of the toxin[11]. The 'hotspot' residues K/E/H162, H/Y163, K/N/Q165 in the C-terminal region of the tdh protein sequence have been associated with P1, the largest pocket reported in this study (Table 2). These mutable residues were observed to influence the geometric measurement of this cavity (Table 3).

Pocket P1 in all modeled sequences belonging to the TDH-S (=*tdh1*) group as well as NC004605A (=*tdh*2) from TDH-A group showed same geometric size (area 2302.53 Å^2 and volume 2561.96 Å^3). However, a reduction in pocket size (area 2247.63 Å^2 and volume 2349.81 Å^3) was seen in X54342 (=*tdh*3) brought about mutations at two residue positions i.e Lys162 to Glu162 and Gln165 to Lys165. In contrast, substitution of residues His163 to Tyr163 and Gln165 to Asn165 increased the pocket area (area 2366.70 Å^2 and volume 2473.35 Å^3) and decreased the volume in X54343 (=*tdh*4). Further, we see that variability in these three residues (K162, H163 and Q165) is seen only among sequences in the TDH-A group, while they are highly conserved among TDH-S sequences (Table 1). This indicates that these three amino acids at the C-terminal region are probably very important and there could be a possible link between the residues and to the altered phenotypes recorded for this gene. Since KP+ phenotype carry two copies of the gene *tdh*1 and *tdh*2, we do not know which among this is more hemolytic and whether both the copies or a single copy is necessary for bringing about hemolytic activity.

Pocket P3, the third largest pocket on the surface of TDH protein was adjacent to P1 and had in common residues N136, C151, C161, and E/H/K162 lining their cavity wall. Studies through mutagenesis experiments indicate the residues C151 and C162 lying adjacent to the side pore channel in tdh to have an influence on hemolytic activity[26] probably involved in membrane disruption brought about by this pore-forming toxin[27]. Extensive mutagenesis studies show that binding pockets on proteins that provide spatial flexibility bring about small structural changes enabling them to fit to different molecules bringing about alterations in the recognition patterns of the protein[16,28]. Probably the largest cavity observed in this study may be the active site which together with pocket P3 may be involved in bringing

Table 3

Geometric area/volume in Pocket P1 in relation to residue substitutions

Pocket area and volume		Residue** substitutions in Pocket P1		
		KHQ	EHK	KYN
Analytical area (A^2) MSMS	Solvent excluded surface area	1868.87	1838.02	1911.58
	Solvent accessible surface area	2302.53	2247.63	2366.70
Voronoi Volume (Å3)*		2561.96	2349.81	2473.35

**** Substitutions involving residues 162, 163, 165 ** Voronoi volume calculated using Richard's radii***

about alteration in spatial conformations thereby influencing the hemolytic phenotype seen in *V. parahaemolyticus*.

CONCLUSION

The surface topography of sequences belonging to the TDH-S group (=*tdh*1) was observed to be stable with no geometrical alterations. In contrast, TDH-A sequences, especially X54342(=*tdh*3) exhibited the maximum alterations in surface topography (Table 2). The *tdh* genes *tdh*3 and *tdh*4 are non-hemolytic phenotypes associated with KP^- strains[5]. We therefore assume that the pockets identified in this study could hold important clues for understanding the way the TDH proteins facilitate interactions and perform their function. The conserved pockets identified in this study could be considered as potential allosteric sites, while the variable pocket (P1) identified could be evaluated for further understanding the protein structure or in designing molecules that can interfere with the function or modulate the activity of this protein. The 'hotspot' residues identified in this study could be used in planning mutagenesis experiments for further understanding the binding specificity of this protein toxin.

ACKNOWLEDGEMENTS

Financial support from the Department of Biotechnology, Government of India, under the Bioinformatics Centre programme is gratefully acknowledge.

REFERENCES

1. Yeung PS, Boor KJ, Epidemiology, pathogenesis, and prevention of food-borne *Vibrio parahaemolyticus* infections. Foodborne Pathog Dis, 1: 74–88, (2004).
2. Nishibuchi M, Kaper JB, Thermostable direct hemolysin gene of *Vibrio parahaemolyticus*: a virulence gene acquired by a marine bacterium. Infect Immun, 63: 2093–2099,(1995).
3. Shirai H, Ito H, Hirayama T, Nakamoto Y, Nakabayashi N et al., Molecular epidemiologic evidence for association of thermostable direct hemolysin (TDH) and TDH-related hemolysin of *Vibrio parahaemolyticus* with gastroenteritis. Infect Immun, 58: 3568–3573, **(**1990).
4. Honda T, Iida T, The pathogenicity of *Vibrio parahaemolyticus* and the role of the thermostable direct hemolysin and related hemolysins. Rev Med Microbiol, 4: 106 –113, (1993).
5. Nishibuchi M, Kaper JB, Duplication and variation of the thermostable direct haemolysin (*tdh*) gene in *Vibrio parahaemolyticus*. Mol Microbiol, 4: 87–99, (1990).
6. Terai A, Baba K, Shirai H, Yoshida O, Uda T et al., Evidence for insertion sequence-mediated spread of the thermostable direct hemolysin gene among *Vibrio* species. J Bacteriol, 173: 5036–5046, (1991).
7. Baba K, Shirai H, Terai A, Takeda Y, Nishibuchi M, Analysis of the *tdh* gene cloned from a *tdh* gene- and *trh* gene positive strain of *Vibrio parahaemolyticus*. Microbiol Immunol, 35: 253–258, (1991).
8. Makino K, Oshima K, Kurokawa K, Yokoyama K, et al., Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of *V. cholerae*. Lancet, 361: 743–749, (2003).
9. Nishibuchi M, Kumagai K, Kaper JB, Contribution of the *tdh*1 gene of Kanagawa phenomenon-positive *Vibrio parahaemolyticus* to production of extracellular thermostable direct hemolysin. Microbial Pathogenesis, 11: 453-460, (1991).
10. Tang GQ, Iida T, Yamamoto K, Honda T, A mutant toxin of *Vibrio parahaemolyticus* thermostable direct hemolysin which has lost hemolytic activity but retains ability to bind to erythrocytes. Infect Immun, 62: 3299-3304, (1994).
11. Tang G, Iida T, Yamamoto K, Honda T, Analysis of functional domains of *Vibrio parahaemolyticus* thermostable direct hemolysin using monoclonal antibodies.

FEMS Microbiol Lett, 150: 289-296, (1997).
12. Honda T, Ni Y, Miwatani T, Adachi T, Kim J, The thermostable direct haemolysin of *Vibrio parahaemolyticus* is a pore forming toxin. Can J Microbiol, 38: 1175–1180, (1992).
13. Raimondi F, Kao JP, Fiorentini C, Fabbri A, Donelli G, et al., Enterotoxicity and cytotoxicity of *Vibrio parahaemolyticus* thermostable direct hemolysin in *in vitro* systems. Infect Immun, 68: 3180-3185, (2000).
14. Okuda J, Nishibuchi M, Manifestation of the Kanagawa phenomenon, the virulence-associated phenotype, of *Vibrio parahaemolyticus* depends on a particular single base change in the promoter of the thermostable direct haemolysin gene. Mol Microbiol, 30: 499–511, (1998).
15. Gerstein M, Lesk AM, Chothia C, Structural mechanisms for domain movements in proteins. Biochemistry, 33: 6739-6749, (1994).
16. Liang J, Edelsbrunner H, Fu P, Sudhakar PV, Subramaniam S, Analytical shape computation of macromolecules: II. In accessible cavities in proteins. Protein-Struct Funct Genet, 33: 18-29, (1998).
17. Corpet F, Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res, 16: 10881-10890, (1988).
18. Bendtsen JD, Nielsen H, von Heijne G, Brunak S, Improved prediction of signal peptides: SignalP 3.0. J Mol Biol, 340: 783-795, (2004).
19. Arnold K, Bordoli L, Kopp J, Schwede T The SWISS-MODEL Workspace: A web-based environment for protein structure homology modeling. Bioinformatics, 22: 195- 201, (2006).
20. Binkowski TA, Naghibzadah S, Liang J, CASTp: Computed atlas of surface topography of proteins. Nucleic Acids Res, 31: 3352-3355, (2003).
21. Willard L, Ranjan A, Zhang H, Monzavi H, Boyko RF, et al., VADAR: a web server for quantitative evaluation of protein structure quality. Nucleic Acids Res, 31: 3316-3319, (2003).
22. Yu J, Zhou Y, Tanaka I, Yao M, Roll: a new algorithm for the detection of protein pockets and cavities with a rolling probe sphere.Bioinformatics, 26: 46–52, (2010).
23. Pavelka A, Chovancova E, Damborsky J, HotSpot Wizard: a web server for identification of hot spots in protein engineering, Nucleic Acids Res, 37: 376-383, (2009).
24. Huang B, Schroeder M, LIGSITEcsc: predicting ligand binding sites using the Connolly surface and degree of conservation**.** BMC Struct Biol, 6: 19, (2006).
25. Laskowski RA, Luscombe NM, Swindells MB, Thornton JM, Protein clefts in molecular recognition and function. Prot Sci, 5: 2438-2452, (1996).
26. Baba K, Yamasaki S, Nishibuchi M, Takeda Y, Examination by site-directed mutagenesis of the amino acid residues of the thermostable direct hemolysis of *Vibrio parahaemolyticus* required for its hemolytic activity. Microb Pathog, 12: 279-287, (1992).
27. Yanagihara I, Nakahira K, Yamane T, Kaieda S, et al., Structure and functional characterization of *Vibrio para-haemolyticus* thermostable direct hemolysin. J Biol Chem, 285: 16267-16274, (2010).
28. Liang J, Edelsbrunner H, Woodward C, Anatomy of protein pocket and cavities: Measurement of binding site geometry and implications for ligands design. Protein Sci, 7: 1884-1897, (1998).