Chapter 9

In-silico structural and functional analysis of Tdh and Trh protein of *Vibrio parahaemolyticus*

In-silico structural and functional analysis of Tdh and Trh protein of *Vibrio parahaemolyticus*

9.1 Introduction

The global spread of pandemic strains of *Vibrio parahaemolyticus* is a serious concern (Nair et al., 2007). Most of the pathogenic isolates carrying the *tdh* gene showed Kanagawa Phenomenon (KP). The important marker for the identification of pathogenic strains is the *tdh* gene. Before 1987, Tdh was considered as a major virulent factor in *V. parahaemolyticus* but in the year 1987, KP-negative strain of *V. parahaemolyticus* was isolated responsible for an outbreak of gastroenteritis in the Maldives and found to produce a Tdh-related hemolysin (Trh) (Honda et al., 1987; Honda et al., 1988). The *trh* gene was also reported from different *Vibrio* species and non-*Vibrio* sp., namely *Vibrio mimicus, Vibrio hollisae, V. cholerae* non-*O*1, *Aeromonas hydrophila* and *Aeromonas veronii*. This was reported from samples isolated from the aquatic environment and from the diarrheal stool. Thermo Stable Direct Haemolysin (Tdh) and Tdh-related hemolysin (Trh) are the major virulence factor responsible for hemolysin of erythrocytes, cardiotoxicity, cytotoxicity and enterotoxicity (Ceccarelli et al., 2013; Raghunath 2015). Trh is heat liable and lost its activity upon heating at 90 °C. Trh has similar physicochemical and immunological properties like Tdh (Honda et al., 1988). TRH formed tetramer in solution (Ohnishi et al., 2011). The nucleotide sequence analysis of *tdh* and *trh* showed that *trh* gene shares 68 % homology with *tdh* gene (Nishibuchi et al., 1989). There was significantly broader variation in nucleotide sequence of *trh* genes from different strain of *V. parahaemolyticus* leading to two subgroups (*trh*1 and *trh*2) which shared 84 % identity (Kishishita et al., 1992). The *trh* positive strain of *V. parahaemolyticus* has enterotoxic effects which lead to fluid accumulation in the intestine

(Xu et al., 1994). The *trh* increases the Ca^{+2} concentration which acts as a secondary messenger and activates the Cl^{-} channel, resulting in elevation of Cl^{-2} secretion which may be responsible for watery diarrhea (Takahashi et al., 2000).

Most of the pathogenic strains of *V. parahaemolyticus* carry either *tdh* or *trh* or both the hemolysin gene (Zhang et al., 2005). Different types of antibiotics are commonly used to control the disease caused by *V. parahaemolyticus*. The emergence of multi-drug resistant strains of *V. parahaemolyticus* was identified from environmental and clinical samples around the world (Odeyemi et al., 2016).

The uses of antibiotics for the treatment of bacterial infections in humans are very common. The major concern of infections with *V. parahaemolyticus* is antibiotic resistance and this could facilitate an increase in the rate of casualty (Daniels et al., 2000). It has become essential to explore alternative drugs that could replace the existing antibiotics due to the emergence of multi-drug resistant strains globally. Gao et al. (2007) showed that the use of high-affinity antibody leads to inhibition of Tdh and thermolabile hemolysin (Tlh). Thus, reduces Tdh-induced cytotoxicity and neutralizes the effects of Tlh in cells under culture. In this context, a thorough study of the Trh protein is necessary for the development of biomolecules that can inhibit the virulent properties of the Trh protein. A lot of information is available on the structure-function relationships of the Tdh protein and its pathological significance in the field of drug discovery. But little is known regarding the structural and functional properties of the Trh protein. The present study analyses the primary sequence and multiple sequences to make out the significant biological properties responsible for the virulence of the Trh protein. These analyses help to understand the intra and inter-species evolutionary relationships of Trh protein

collected from various species of *Vibrio* sp. and non-*Vibrio* sp. Homology modeling approach was used to predict the 3D structure of the Trh protein. Molecular dynamics (MD) simulation over a time frame of 60 ns were analysed to understand the structural stability and residual fluctuations. Trh protein is similar to that of Tdh, possessing all the essential properties required for a drug target for the treatment of *V. parahaemolyticus* infections. Structural analyses like active site identification and analysis of the druggability of protein pockets will also help in understanding the structure-function relationships of the Tdh and Trh protein. This will lead to the generation of information that would be useful to the development of therapeutics against *V. parahaemolyticus* infection.

9.2 Materials and Methods

9.2.1 Bacterial culture, DNA isolation, and amplification of the *tdh* and *trh* gene

The genomic DNA isolation was isolated from pure culture of *tdh* (S24P132) and *trh* (AP429) positive strain using the methodologies described in Chapter 4. The full-length *tdh* and *trh* gene were amplified by PCR using gene-specific primers. The *tdh* gene was amplified using the primer tdhcdsF 5′-AAGTACCGATATTTTGC-3′ and tdhcdsR 5′-TTGTTGATGTTTACATTCAA-3′ as forward and reverse primers, respectively (Bechlars et al., 2013) The *trh* gene was amplified by using trhseqF 5′-TGCTTTCCTTTATCTCGAGC-3′ and trhseqR 5′-CATATGAAAACAAATGCTTTTTTAG-3′ as forward and reverse primers, respectively (Nilsson and Turner, 2016). The PCR reaction mixture was prepared by mixing the reagents as described in Chapter 4. The PCR program for amplification are denaturation for 2 min at 95 °C followed by denaturation at 94 °C for 30 s, annealing at 54 °C for *tdh*

and 52 °C for *trh* for 45 s and extension at 72 °C for 45 s for 35 cycles with final extension for 3 min. at 72 °C. PCR product was visualized on 1.8 % agarose gel stained with ethidium bromide.

9.2.2 Sequencing of the amplified *tdh* and *trh* gene

The amplified gene product was sequenced using an ABI 3730xl capillary sequencer (Applied Biosystems, Foster City, CA) in forward and reverse directions (Behera et al., 2014). The contig was prepared by aligning the forward and reverse sequences using DNA baser 7.0.0. software. The reverse sequence was used to proofread the forward sequence during contig preparation. The assembled full-length *tdh* and *trh* gene sequences were then submitted in GenBank Database.

9.2.3 Primary information of Tdh and Trh protein of *Vibrio* sp.

The primary information that is the amino acid sequence information about Tdh and Trh proteins of different *Vibrio* sp as well as non-*Vibrio* sp such as *Vibrio parahaemolyticus, Vibrio alginolyticus, Vibrio. diabolicus* and non-*Vibrio* sp *Aeromonas veronii* has been retrieved from UniProt database.

9.2.4 Sequence based analysis

After retrieving the primary information from the database, some sequence-based analysis of Trh protein has been carried out using different popular web-base tools such as InterProScan, Psypred server, CD-search tools.

The phosphorylation sites of Tdh and Trh protein was predicted using Netphos Server. CLC Genomics Workbench-8.5 was used to predict glycosylation sites and surface regions of Tdh and Trh protein which is very important in contrast to their structural behavior. A local hydropathy plot was generated on the bases of the Kyte-

Doolittle scale and Eisenberg scale to predict the transmembrane region of Tdh and Trh protein using CLC Genomics Workbench-8.5 (https://www.qiagenbioinformatics.com/). Physico-chemical properties such as Isoelectric pH, aliphatic index, molecular weight, hydrophobicity and hydrophilicity of Tdh and Trh proteins of *Vibrio parahaemolyticus* has been analyzed using CLC Genomics.

9.2.5 Multiple sequence alignment and phylogenetic analysis

Applying MUSCLE algorithm a multiple sequence alignment analysis of Tdh and Trh protein has been carried out to identified the evolutionarily conserved sequence pattern in different *V. parahaemolyticus* species. A phylogenetic tree was constructed applying the Neighbor-Joining method, with a bootstrap value 1000 using CLC Genomics work branch 8.5.

9.2.6 Structure prediction through homology modeling:

To predict the three-dimensional (3D) structure of Trh, protein a homology modeling approach was used. The first and most important part of homology modeling is the identification of suitable template, BLASTP (Altschul et al., 1990) search against PDB database (http://www.rcsb.org/) was used for template identification and PDB ID: 3a57 was chosen as a template. The protein model for the target sequence was generated using MODELLER9.11 (Sali et al., 1993). The protein model was evaluated by calculating Discrete Optimized Protein Energy (DOPE score) and by plotting the energy using MODELLER. After the generation of model, some additional analysis like RMSD value, target-template alignment TM score and sequence logo based on target-template alignment was carried out using CLC genomics.

9.2.7 Structural quality assessment

The overall quality of the Trh model was evaluated by using different model validation tools. The consistency and reliability of the Trh model was analyzed by PROCEHCK (Laskowski et al., 1993). The PROCEHCK was used to understand stereochemical quality. Ramachandran plot was used to quantify the residues present in favored regions, allowed regions and generously allowed regions. Furthermore, the overall model quality was evaluated by different parameters like model quality confidence score, global quality Z score, overall model quality score and RMSD deviation by using different web application such as ModFOLD (McGuffin et al., 2013) Verify3D (Eisenberg et al., 1997), ProSA (Markus et al., 2007) and PSVS server (Bhattacharya et al., 2007) respectively.

9.2.8 Molecular dynamics (MD) simulations

To understand the structural dynamics, stability as well as residual fluctuation of TRH protein, a molecular dynamics simulation was conducted using the four most popular force-fields (Amber, Gromos, OPLS and Charmm) with all-atom explicit water model. The simulation of Trh protein follows CABS simulation procedure (Jamroz et al., 2013) running on a high-performance cluster computer server (http://biocomp.chem.uw.edu.pl/CABSflex). The simulation time was set to 60 ns to obtain the best possible convergence and the sampling is realized by the Monte Carlo method (Kolinski et al., 2004). CABS pipeline incorporates multiscale reconstruction and optimization procedures (Kmiecik et al., 2011; Gront et al., 2012) where the results have represented an ensemble of protein models (in all-atom resolution) reflecting the flexibility of the input structure. The generated trajectory was analyzed using VMD and

VAGA ZZ. The residual fluctuation of the entire atom provides information on protein behavior as well as its functional property. The following equation was used to calculate mean-square fluctuation:

$$\langle (\Delta R)_i^2 \rangle = \frac{1}{N} \sum_{j}^{N} (xi(j) - \langle xi \rangle)^2$$

Where < > denotes the average over a whole trajectory, and x is the position of particle ***i*** in the frame ***j***. Clustering of protein dynamics trajectory was done by using the K-means clustering method in such a way that more similar model (in the sense of RMSD measure) retained in the same group. Protein dynamics trajectory cluster was then analyzed to screen the best conformation. The superimposition of each cluster with the representative model was performed by the Theseus application.

9.2.9 Pocket identification

The presence of different protein pockets and their volume were identified in trh protein model using CASTp server (Dundas et al., 2006). The amino acids involved in the pocket are also identified using this server.

9.3 Results

9.3.1 Amplification and sequencing of the full-length *tdh* and *trh* gene

The *tdh* and *trh* gene was amplified by specific primers, which generated 450 bp and 765 bp fragments and was sequenced in both directions. The full-length sequences of the *tdh* and *trh* gene were submitted to the NCBI, and the assigned the accession number was MG772656 and MH332788, respectively. The CDS of the *tdh* and *trh* gene

comprises of a polypeptide chain of 189 amino acids, was subjected to further *in silico* analyses.

9.3.2 Sequence based analysis

The conserved sequence motif, domain as well as a superfamily of Trh protein, has been detected by popular web application InterPro scan and CD-search tool, respectively. Conserved domain analysis revealed that our targeted Trh protein share significance conserveness with Tdh protein domain, Pfam 03347 superfamilies. Two motifs region, PD006802 and PF03347 were identified in Trh protein sequence after InterProScan search against several databases (Figure. 9.1). PSIPRED web server was used to predict the secondary structure composition of Trh protein of *V. parahaemolyticus.* There were ten β-sheets and four α-helices present the seccondary structure Trh protein of *V. parahaemolyticus* (Figure 9.2). All the α-helices and β-sheets were connected with short loop, except three loops. The connection from the first α-helix at the N-terminal, the second loop from the β-strand (residues 32-41) at the N-terminal and the third loop from the α-helix (residues 117-121) towards the C-terminal were longer than the other loops (Figure 9.3).

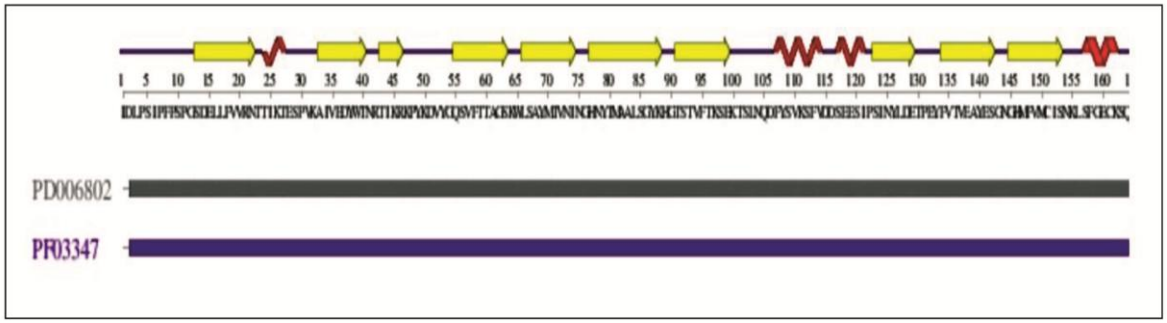

Figure 9.1 Position specific structural motifs identified by InterPro scan tool. Gray colour represent hit with ProDom and violet colour represent hit with Pfam database. Yellow aerrow indecates beta plate and red indecates alpha helix.

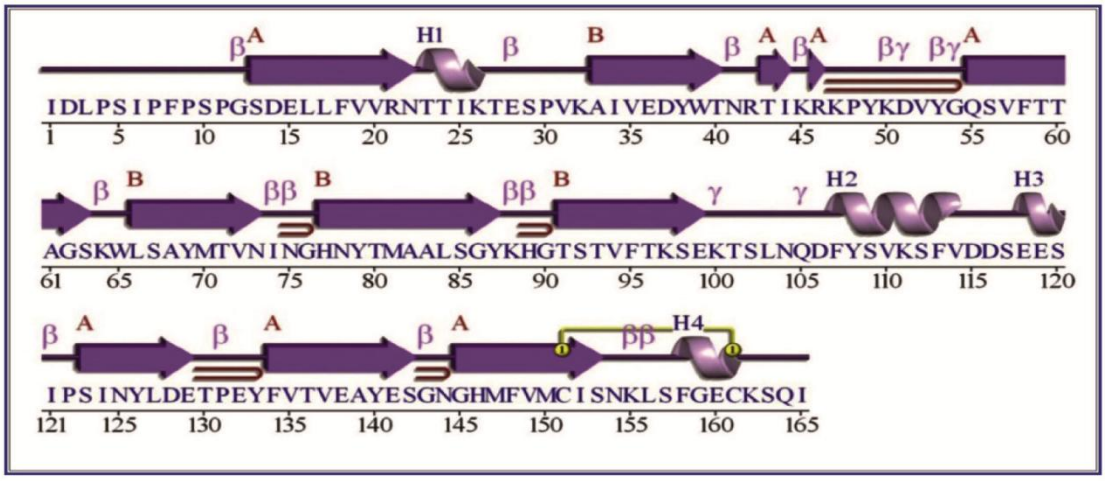

Figure 9.2 Representation of the secondary structure composition of the Trh protein. The α-helices are labeled as H and the β-strands are represented by purple arrows. The different structural motifs represent as β-turns (β), γ-turns (γ), β-hairpins (⊃), and disulfide bridge (Yellow line).

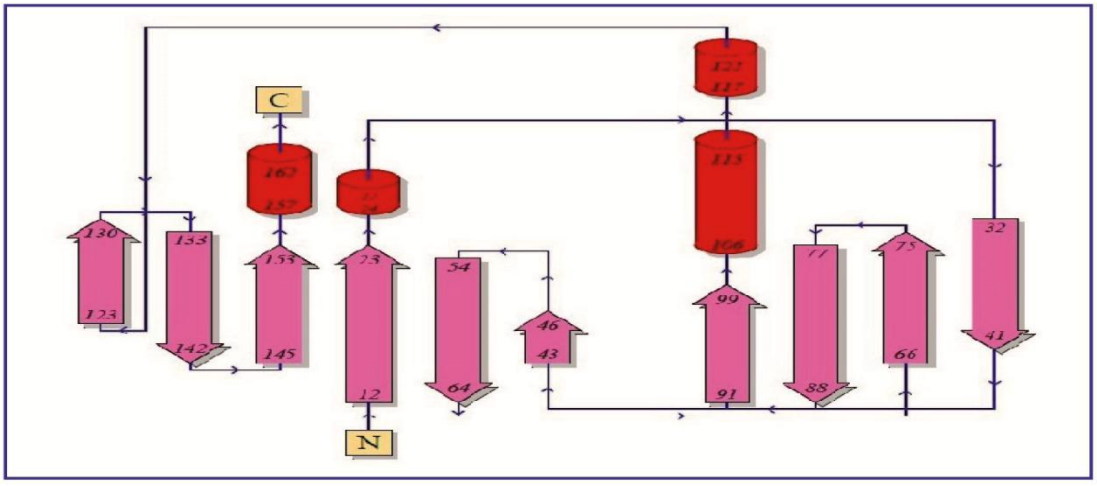

Figure 9.3 Topology diagram of the Trh protein of *V. parahaemolyticus* was generated using the PDBSUM server. The ten β–strands are represented by pink arrows, and α-helices are represented by red cylinders.

The important phosphorylation sites of Tdh and Trh protein has been predicting by Netphose web application, from the figure it is clear that there are 21 phosphorylation sites in Tdh and 24 phosphorylation sites Trh, which cross the threshold value represented in pink line. There are seven serine phosphorylation sites, eight threonine phosphorylation sites and six tyrosine phosphorylation site are detected in Tdh protein. However, nine serine phosphorylation sites, ten threonine phosphorylation sites and five tyrosine phosphorylation site are detected in Trh protein (Figure 9.4). The glycosylation sites of both the protein were predicted and represented in the Figure 9.5. Three N-glycosylation sites were detected which represent as a yellow arrow. To understand the structural importance transmembrane region of Trh protein was predicted using Kyte-Doolittle hydropathy plots which identified six transmembranes with a peak value of more than 1.8. Some important physiochemical properties such as isoelectric point, molecular weight, hydrophobicity value of Tdh and Trh protein were calculated.

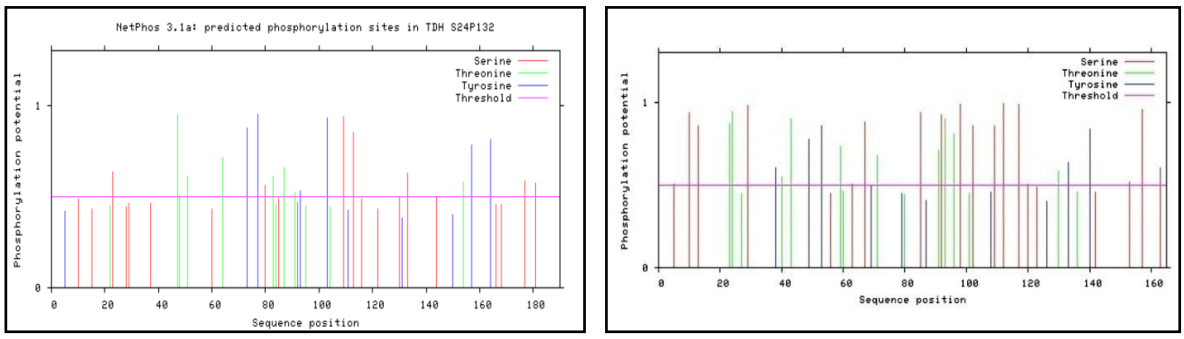

Figure 9.4 Phosphorylation sites predicted in the Tdh and Trh protein. The amino acids sequences of the protein are represented X-axis and the Y-axis represents the phosphorylation potential. The threshold level represented by the horizontal pink line parallel. Residues showing values above threshold level are potential phosphorylation sites.

Three important N-glycosilation sites were identified in the Trh protein using CLC genomics. However no N-glycosilation sites were identified in the Tdh protein of *V. parahaemolyticus* though Tdh and Trh share 85 % sequence similarity.

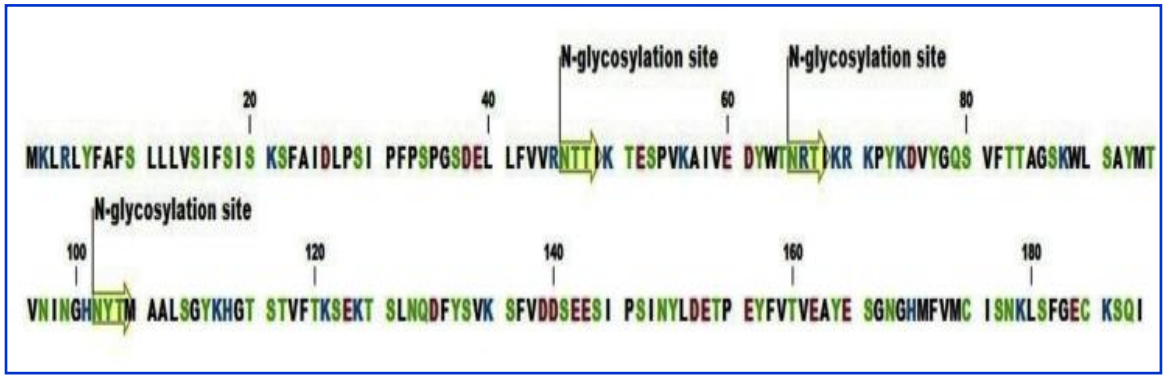

Figure 9.5 Representation of N-glycosylation sites predicted in the Tdh and Trh protein. The glycosylation sites within the protein are represented with yellow arrows.

9.3.3 Multiple sequence analysis and phylogenetic analysis

To detect evolutionarily conserved sequence pattern in Tdh and Trh protein among different vibrio and non-Vibrio species, a multiple sequence analysis has been carried out. From MSA, it was found that most of the sequences are highly conserved throughout the evolution, the residual conserveness is a highlighted in the Figure. 11. Sequence of Tdh protein belonging to five different *Vibrio* sp. from two different clusters. The cluster 2 represent the largest group which includes eight sequences. The Tdh protein of *V. diabolicus* (L7XA04; L7X8V6) and *V. parahaemolyticus* (P19250) showed high evolutionary divergence. The 21 sequences of the Trh protein, belonging to five different species, form three different clusters. 11 sequence of Trh protein of *V. parahaemolyticus* together with our sequence (MH332788) form largest cluster 2. The Trh protein sequences of different *Vibrio* sp including our sequence (highlighted in red) present in one cluster clearly indicated that these sequences are evolutionary conserved

and very close. Furthermore, four different Trh protein sequence from two *Vibrio* species and one sequence from a non-*Vibrio* species from cluster 1, while cluster 3 is formed with three Trh protein sequence from two different *Vibrio* species. Trh protein sequence with accession number Q27VZ9 evolutionary highly diverge.

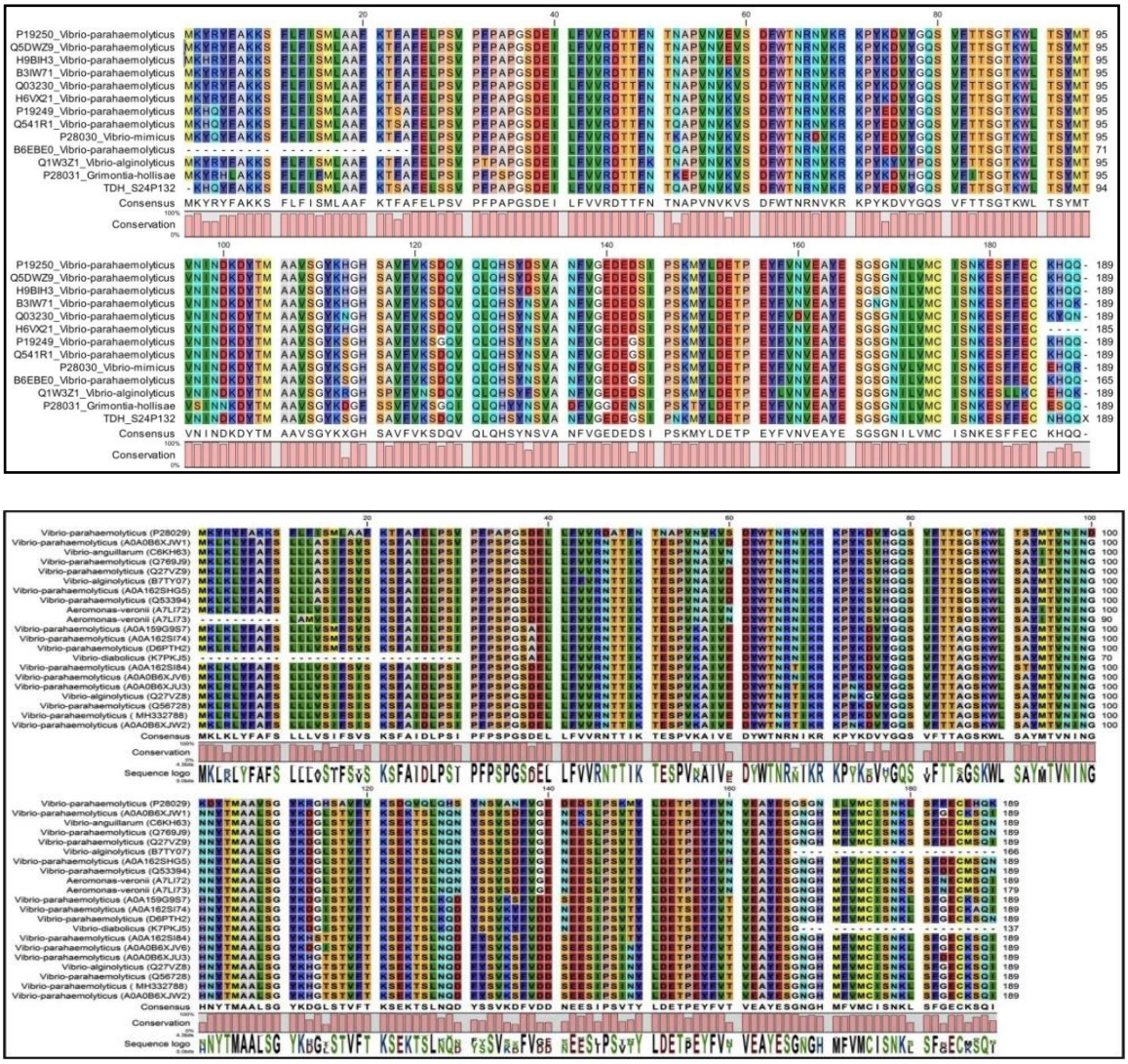

Figure 9.6 Representation of multiple sequence alignment of Tdh and Trh protein sequences of different *Vibrio* sp. and non-*Vibrio* sp. The global consensus sequence is giving at the base. Red bar at the base represents the conservedness of the sequences.

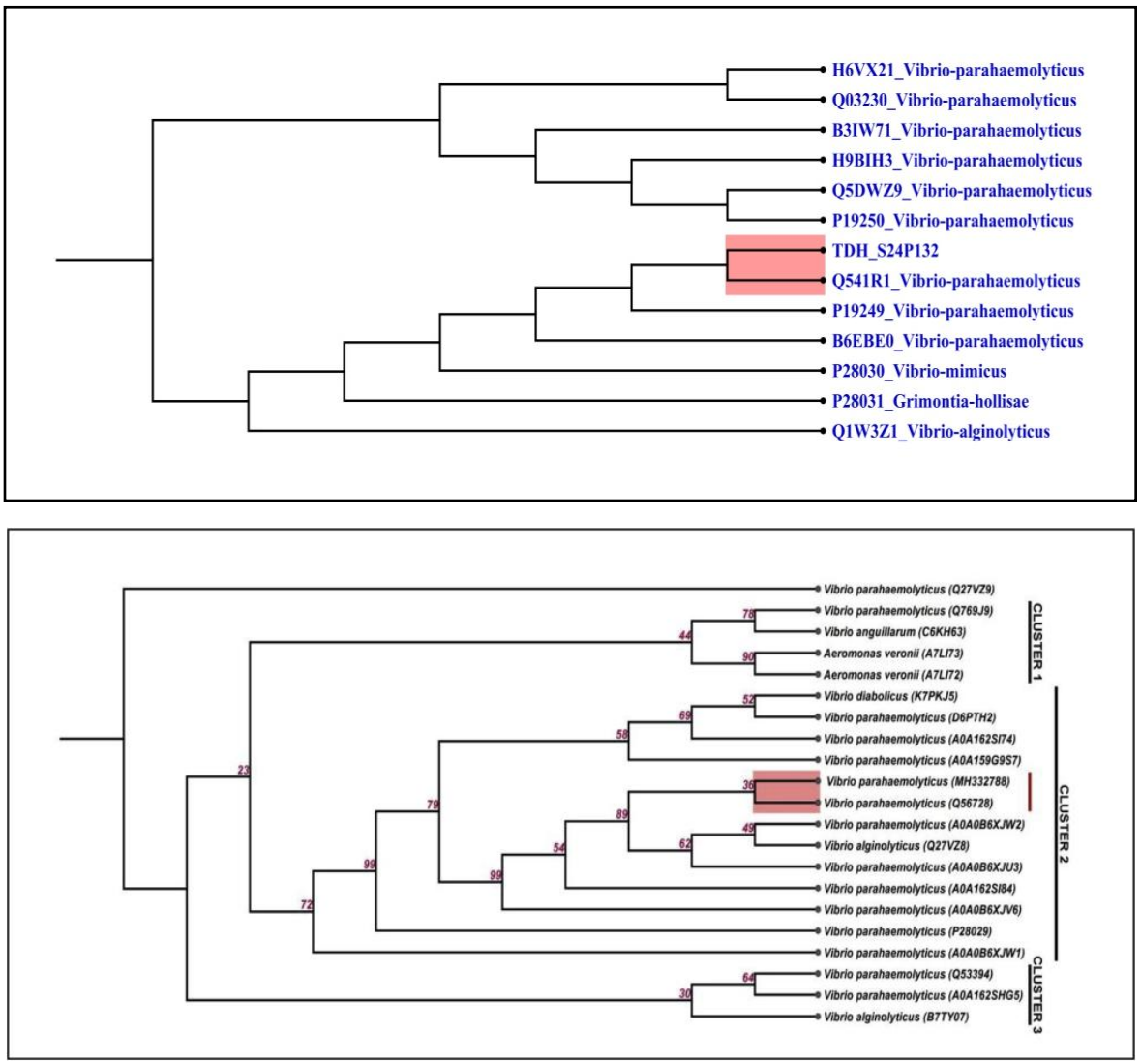

Figure 9.7 Representation phylogenetic cladogram of the Tdh and Trh protein sequences from *Vibrio* sp. and non-*Vibrio* sp. The bootstrap percentage obtained from 1000 samples.

9.3.4 Structure prediction through homology modeling:

To predict the three-dimensional structure of Trh protein, a biologically most significant homology modeling approach was used. In this approach, model was generated based on the sequence identity between target and template protein. PDB-blast analysis helps to identify the suitable template (PDB ID: 3A57) with a significant sequence identity of the target protein. Selected template (PDB ID: 3A57) is a crystal structure of Thermostable Direct Hemolysin at $1.5A^0$ resolution was used to construct the model structure of Trh protein with a reliable DOPE score value (−16845.50195). Pairwise sequence alignment between target and template revealed the consensus sequence and sequence conservation. Structural alignment analysis of Trh of *Vibrio parahaemolyticus* with template structure (PDB ID: 3A57) revealed the RMSD value was 0.173, TM-score was 0.932 and sequence identity was 68.2 % over align region

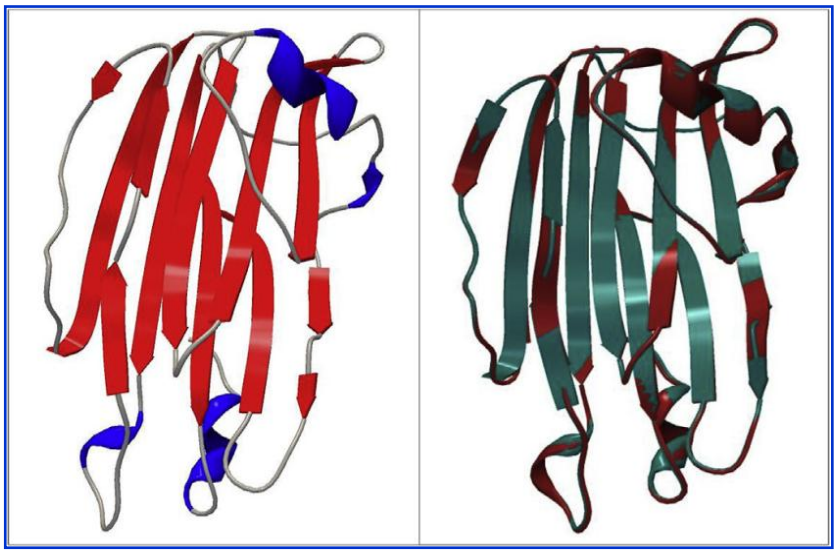

Figure 9.8 Representation of homology model of the Trh protein (A). The α-helices and β-sheets are represented in blue and red color, respectively. The coils are represented in grey color. The structures of Trh protein are superimposed with template (PDB ID:3A57) (B). The template structure showing in maroon colour.

9.3.5 Structural quality assessment

After the generation of successfully completed 3D model of Trh protein, the model structure was analyzed by different protein structure validation tools to find out the reliability of the model protein. Ramachandran plot analysis revealed that 91.1 % of all residues were in allowed regions with 7.5 % and 1.4 % residues in generously allowed and disallowed regions, respectively. ProSA program was used to estimate the overall model quality by measuring the Z score value (Figure 9.9). The ProSA analysis showed the Z score of -6.32 for the model trh which reflecting the good quality of the model.

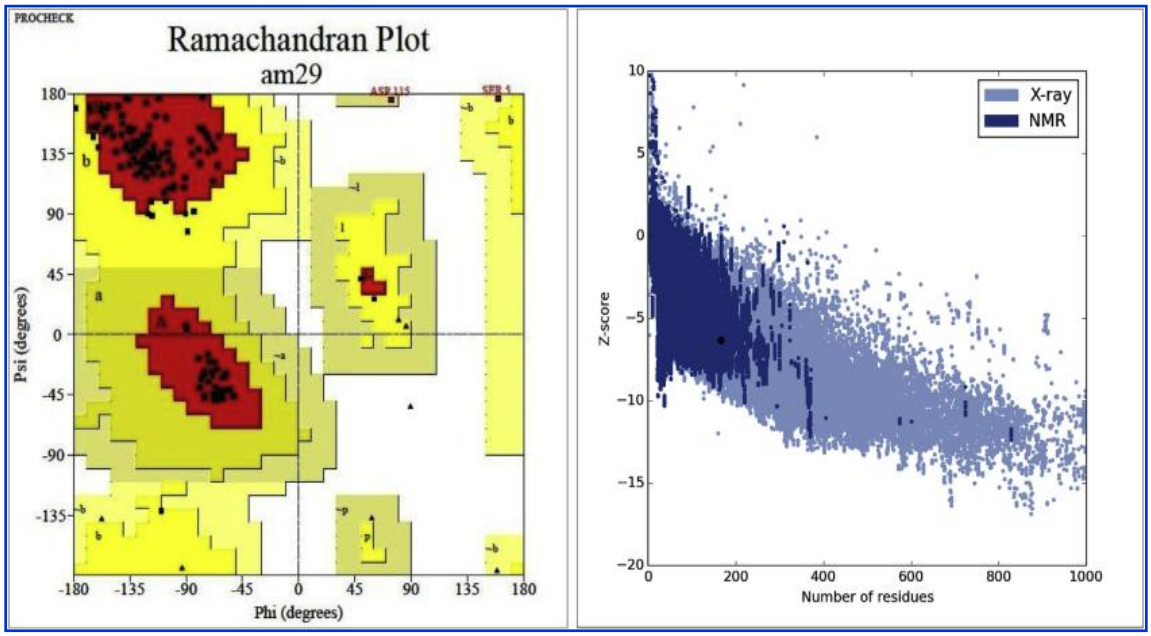

Figure 9.9 Ramachandran plot of the modeled Trh protein (left). The red regions within the plot symbolize the favored regions; the yellow regions symbolize the allowed regions, while the regions in pale yellow color symbolize the generously allowed regions. Validation of the modeled structure of the Trh protein by the ProSAZ-score (right).The Z-score of all the proteins determined by X-ray crystallography in the Protein Data Bank (PDB) is represented by light blue dots while that of the structures determined by NMR are represented as dark blue dots. The Z-score of the Trh protein is represented as a large black dot.

9.3.6 Molecular dynamics (MD) simulations

Molecular dynamics simulations over a time of 60 ns were carried out to understand the structural stability and the dynamics of the model of Trh protein generated in the present study. To understand the structural stability of Trh protein, the structural RMSD of simulation trajectory was monitored over time and found after 10 ns of the simulation, the protein becomes stable. The radius of gyration of Trh protein was calculated over the time to analyze the compactness of the protein. The root means square fluctuation (RMSF) and the atomic flexibility of Trh protein over the simulation time of 60 ns was measured. The most flexible region (flopping region) of the Trh protein was determined and indicated in Figure 9.10 in a different color. a k-means clustering algorithm was applied to identify the best conformation of the Trh protein from the trajectory.

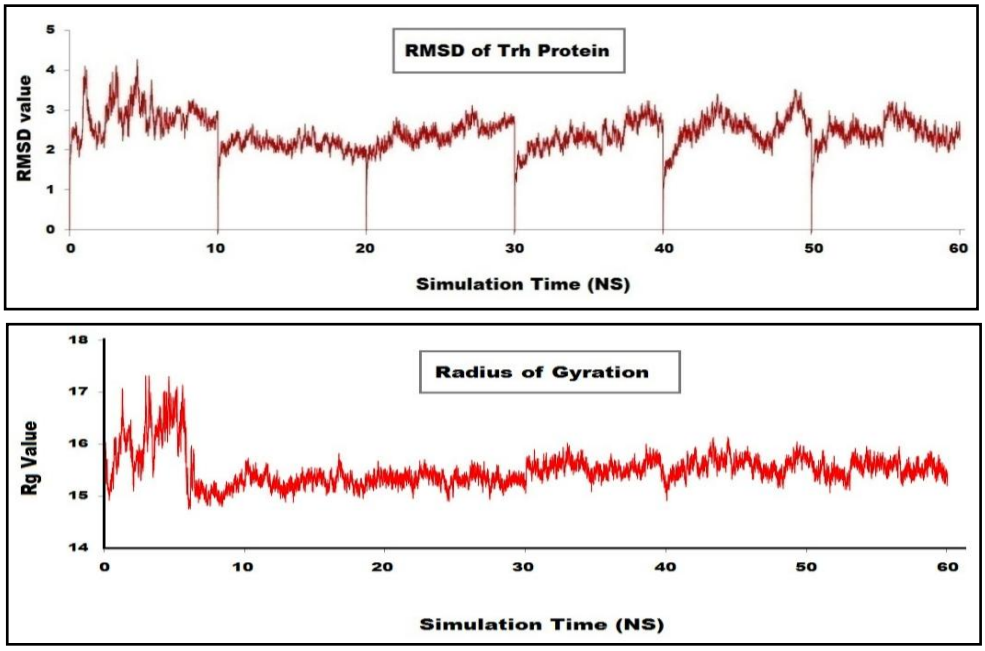

Figure 9.10 Root mean square deviation (RMSD) of the Trh protein over the 60 ns simulation time. The radius of gyration (Rg) of the modelled Trh protein over the 60 ns trajectory.

9.3.7 Pocket identification

Pocket identification of Tdh and Trh protein was carried out by the CASTp server. The analysis identified nineteen different pockets within Tdh and Trh (Figure 9.11). The largest pocket has a volume of 45.96 and 238.006 incase of Tdh and Trh, respectively. The 2nd largest pocket has a volume of 15.306 and 36.493 in the case of Tdh and Trh, respectively. The smallest pocket within the Tdh has a volume of 0.002 and only two amino acids are involved in the formation this pocket where as in Trh the volume of smallest pocket is 0.026 and only four amino acids are involved in the pocket formation. The details of pockets present in the Tdh and Trh protein given in Table 9.1 and 9.2

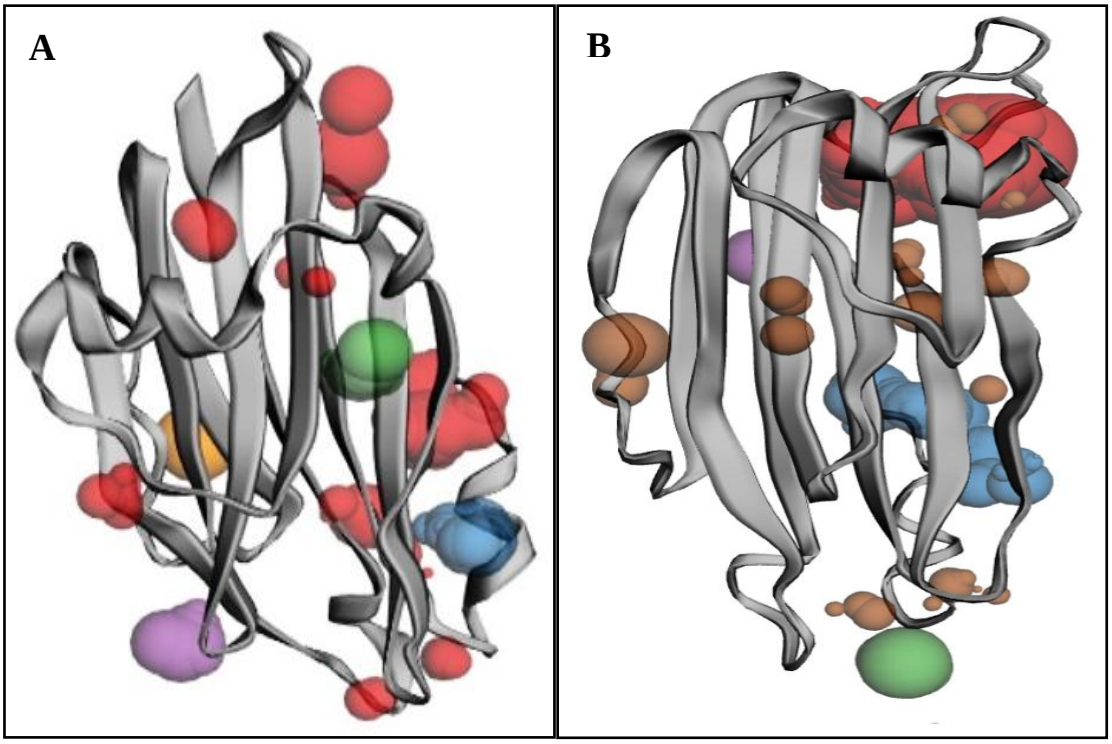

Figure 9.11 Nineteen pockets have been predicted in the Tdh (A) and Trh (B) protein. Four pockets were identified in the structure of the Tdh and Trh protein, which had a large volume, and have been represented using four colors, namely, red, blue, green, and violet. The other small pockets have been shown in brown color.

Table 9.1 The pockets present in the structure of the Tdh protein, along with the volume, residual composition, and interaction sites

Pocket ID	Area ($Å^2$)	Volume ($Å^3$)	SeqID	Amino acids	Atom compositions
1	112.979	45.96	21	ARG	CD, NE,NH2
			124	LYS	CD,CE,NZ
			136	ASN	OD1
			138	GLU	CB,CGCD,OE1
			140	TYR	CE2,CZ,OH
			149	VAL	O,CB,CG1,CG2
			151	CYS	CB,SG
			160	GLU	O
			161	CYS	CA,O
			162	LYS	CA,C,O
			163	HIS	O,CD2
			164	GLN	CA,CB,CG,NE2,N,O,CG
2	58.207	15.306	126	TYR	CB,CD1,CE1
			128	ASP	O
			129	GLU	CB,CD,OE1
			134	PHE	C,O,CB,CD1
			135	VAL	N,CA,C,O
			136	ASN	CB,ND2,
			151	CYS	O
			153	SER	OG
			158	PHE	O,CD1,CE1
			161	CYS	CB
			162	LYS	CD,CE,NZ
3	39.538	11.379	96	VAL	O,CB,CG2
			97	LYS	CA
			98	SER	N,CB,OG
			120	SER	O,CB
			123	SER	O,OG
			125	MET	CB,CE
			137	VAL	O,CG1
			139	ALA	N,CB
			148	LEU	CD2
4	21.637	7.694	32	ASN	C,O,CB
			34	GLU	CB,CG,CD,OE
			73	ASN	O,OD1
			75	ASN	N,CA
			76	ASP	N,OD1
5	11.465	5.062	46	ARG	NE,CZ,NH2
			47	LYS	O
			48	PRO	CA
			49	TYR	N,CB
			56	SER	CA
			57	VAL	N,O,CB,CG2
6	8.833	2.086	92	SER	O,OG
			93	ALA	CA
			94	VAL	CG2
			141	GLU	OE1
7	11.613	4.110	21	ARG	CB,CD,NH1
			22	ASP	N,C,O
			23	THR	CA,CG2

			25	PHE	CD2,CE2
			54	GLY	CA,O
8	12.146	2.549	65	TRP	CZ3
			84	VAL	CG1,CG2
			95	PHE	CD2
			111	ALA	CA,O,CB
			115	GLY	CA,O
			116	GLU	CG,OE2
9	8.401	2.086	88	LYS	O,CB
			91	HIS	O,CB
			92	SER	O
			93	ALA	CB
			118	GLU	CD,OE2
10	9.522	1.28	36	SER	OG
			37	ASP	CB,OD2
			71	THR	CB,OG1,CG2
			102	GLN	CA,CB,CG,NE2
			103	LEU	N
11	7.274	0.931	130	THR	O
			131	PRO	C,O
			132	GLU	CA,C
			133	TYR	N,CA
			134	PHE	CE2,CZ
			154	ASN	CA,CB
			155	LYS	N,CG,CE,NZ
12	5.451	0.929	29	ALA	O,CB
			31	VAL	CG1
			132	GLU	CA,O,CB,OE2
13	2.076	0.286	141	GLU	CB,OE1
			143	GLY	O
			144	SER	CA, C
			145	GLY	O
14	2.271	0.285	23	THR	CA,O
			25	PHE	CE2
			26	ASN	OD1,ND2
			160	GLU	OE1
15	2.469	0.265	79	TYR	TYR
			99	ASP	CA,OD1
			100	GLN	N,CB
			127	LEU	O,CB,CD2
16	1.331	0.102	15	GLU	OE2
			17	LEU	CB
			145	GLY	CA
			146	ASN	N,O
			147	ILE	CG1
17	1.336	0.072	95	PHE	CB
			96	VAL	O
			114	VAL	O
			117	ASP	N,CB
			120	SER	OG
18	0.485	0.055	38	PHE	CE2
			48	PRO	O
			50	LYS	CA,CB
19	0.201	0.002	26	ASN	OD1,ND2
			160	GLU	OE1,OE2

Table 9.2 The pockets present in the structure of the Trh protein, along with the volume, residual composition, and interaction sites.

Pocket ID	Area ($Å^2$)	Volume ($Å^3$)	SeqID	Amino acids	Atom compositions
1	234.003	238.006	6	ILE	CG2, CD1
			8	PHE	CG, CD1, CD2
			9	PRO	CD,O
			10	SER	CA, OG
			11	PRO	N, -, CG, O
			14	ASP	CB, CG, OD1
			64	LYS	CE, NZ
			91	THR	CG2
			92	SER	SER, CB, OG
			93	THR	N, CA, OG1,CG2
			94	VAL	N, CB**,** CG2
			118	GLU	CG, CD, OE2
			121	ILE	CG1, CD1
			141	GLU	CD ,OE1, OE2
			143	GLY	C, O
			144	ASN	CA, OD1, O
			146	HIS	NE2, CE1
2	140.663	36.493	21	ARG	NE, CZ, NH2
			124	ILE	CG1, CD1
			126	TYR	CB, CD1, CE1
			128	ASP	N, O
			129	GLU	CG, OE1
			134	PHE	CB, C
			135	VAL	N, CA
			136	THR	OG1, CG2, O
			137	VAL	CA
			138	GLU	N, CB, OE1
			140	TYR	CE2,OH
			149	VAL	CB, CG1, CG2, O
			151	CYS	N, CB, SG
			153	SER	OG
			158	PHE	CE1,CZ
			161	CYS	CB,SG,O
			162	LYS	CA, CD, CE, NZ, O
			163	SER	O
			164	GLN	CA, CB, CG
			165	ILE	N, CG2, O
3	15.186	5.275	28	GLU	CG,OE2
			29	SER	CB,OG
			154	ASN	OD1,ND2
			155	LYS	CB,CD
			156	LEU	CG,CD2
4	12.240	3.352	15	GLU	CB, CG, CD, OE2, O
			17	LEU	N,CB
			142	SER	OG
			145	GLY	CA
			146	HIS	N,O
			147	MET	CG,CE
5	9.209	3.291	35	VAL	CB, CG1, O
			38	TYR	CD1,CE1

			48	PRO	CG,O
			50	LYS	CA, CB, CG, CD
6	11.926	1.575	81	MET	CB
			96	THR	CG2
			98	SER	OG
			125	ASN	ND2
			137	VAL	CB
			138	GLU	C,O
			139	ALA	N,CB
			148	PHE	CB,CD2
7	6.172	1.198	46	ARG	NE,NE2
			47	LYS	O
			48	PRO	CA
			49	TYR	N,CB
			56	SER	CA
			57	VAL	N
8	5.011	0.643	96	THR	CB, CG2, C, O
			97	LYS	CA
			120	SER	CB,O
			123	SER	OG
			125	ASN	ND2
			139	ALA	CB
9	3.871	0.388	69	TYR	CE2
			71	THR	CB,CG2
			102	SER	CA,CB
			103	LEU	N,O
10	3.538	0.303	29	SER	CB,O
			30	PRO	O
			31	VAL	CG1
			132	GLU	CA,OE1,O
			154	ASN	ND2
11	4.121	0.299	130	THR	O
			131	PRO	O
			132	GLU	C,O
			133	TYR	N,CA
			134	PHE	CE2,CZ
			154	ASN	CA,ND2
			155	LYS	N,CG
12	4.257	0.246	65	TRP	CZ3
			84	LEU	CB,CG
			95	PHE	CD2,CE2
			111	LYS	CG,CD
			115	ASP	O
			116	ASP	CG,OD2
13	3.101	0.231	121	ILE	CB,CG2
			139	ALA	CB
			140	TYR	N,C,O
			146	HIS	CB,CD2
			148	PHE	CE2
14	2.165	0.222	36	GLU	CG
			71	THR	CG2
			78	ASN	CB,CG,OD1
			102	SER	CB
15	1.292	0.106	79	TYR	CD1,CE1
			100	LYS	N,CB
			127	LEU	CB,O

16	0.145	0.044	130	THR	O
			131	PRO	CA,O
			134	PHE	CZ
			155	LYS	CG,CE
17	0.466	0.014	80	THR	CB
			101	THR	CB
			103	LEU	CG,CD2
			113	PHE	CE2,CZ
18	0.536	0.014	95	PHE	CB
			96	THR	O
			114	VAL	O
			117	SER	N,CB
			120	SER	OG
19	0.026	0.000	141	GLU	OE1
			143	GLY	O
			144	ASN	O
			145	GLY	O

9.4 Discussion

The Tdh and Trh are the major virulent protein secreted by *V. parahaemolyticus* cause gastrointestinal infection (Honda et al., 1988; Yoh et al., 1992 and Raghunath et al., 2015). The complete cDNA of *tdh* and *trh* gene of *Vibrio parahaemolyticus* was 491bp and 567 bp long, respectively which encoded 189 amino acids protein of which first 24 amino acids residues from n-terminal site belonged to signal sequence. There are two different types of *trh* reported (t*rh1* and *trh2*) which showed 84 % homology (Kishishita et al., 1992). Similarly, two different copies of *tdh* genes (*tdh*1 and *tdh*2) was reported in *V. parahaemolyticus* which share 97.2 % sequence similarity (Nishibuchi et al., 1991 and Bhowmik et al., 2014). The *tdh* gene of different *Vibrio* sp shared >96.7 % identity with the same biological activity (Okuda and Nishibuchi, 1998). The *trh* sequence of *V. parahaemolyticus* had high similarity with *trh* sequence of other *Vibrio* sp and non-*Vibrio* sp and shared 98.8 %, 66.1 % and 85.2 % identities with *trh* sequence of *Vibrio alginolyticus*, *Vibrio diabolicus* and *Aeromonas veronii*, respectively. The conserved domain analysis of Trh showed that it is under Tdh superfamilies. The pairwise alignment

of Trh with Tdh showed 68 % sequence homology as reported by Kishishita et al., 1992. The crystallographic structure of Tdh protein was present in the PDB database, therefore, sequence-based analysis and homology modeling were not performed. Post-translation modification like phosphorylation, glycosylation, methylation, acetylation and lipidation is an important process for protein activation, oligomerization, protein-protein interaction and sub-cellular localization (Kobir et al., 2011 and Whitmore et al., 2012). In bacteria, glycosylation of protein is associated with pathogenicity. The glycosylation of flagellin in *Pseudomonas aeruginosa* plays crucial role in virulence (Arora et al., 2005). Similarly, O-glycosylation in serine-rich adhesion proteins which is very important for the attachment of bacteria with the host cell (Zhou et al., 2009). In the present study, we have found three glycosylation sites within Trh protein. Phosphorylation is an important post-translation modification process in bacteria controlling many cellular functions (Whitmore et al., 2012). The phosphorylation of LUKS protein in *Staphylococcal leukocidin* is important for lysis of human and rabbit polymorphonuclear leukocytes (Nishiyama et al., 1998). In the present study, 24 phosphorylation sites were predicted within trh protein might be playing an important role in hemolysis and cytotoxicity.

The multiple sequence analysis showed the presence of five conserved amino acids residues at the position of K^{2}, S^{11}, K^{21}, F^{23} and A^{24} within the signal peptide. Such conserveness was not observed in the Tdh protein of pandemic strains and nonpandemic strains of *V. parahaemolyticus* as well as in other *Vibrio* species (Bhowmik et al., 2014). Like Tdh, the tetrameric form of Trh is important for the hemolytic activity of Trh protein. The sequence base analysis of Trh protein showed the presence of many conserved residues like R46, Y53, G62, W65, G90, C151, C161 and Q164 in Trh protein.

The similar amino acids residues at the same position were also reported in the Tdh protein which played an important role in the tetrameric assembly of Tdh protein and hemolytic activity (Kundu et al., 2017; Wang et al., 2011 and Baba et al., 1992). Arg46 formed π cation interaction with Tyr140 in the adjacent protomer which is very essential for the tetramer formation and the interaction between E138 and Q164 is also important for stabilization of tetrameric structure (Yanagihara et al., 2010). Arg46 and Tyr140 are conserved in all the Trh protein used in the present study. Similar conserveness of Arg46 and Tyr140 was observed in all the Tdh and Trh protein, reported by Yanagihara et al., 2010. The Cys151 and Cys161 were highly conserved in all the Tdh (Yanagihara et al., 2010) and play a critical role in the tetrameric assembly. The presence of Cys151 and Cys161 within Trh protein at the same position were detected and were conserved in all the Trh protein of *Vibrio* and non-*Vibrio* species. Kundu et al (2017) reported the presence of extended C-terminal region to contain nine amino acids in the position of 157-165 (157SFFECKHQQ165) is essential for tetrameric assembly. The intrapeptide disulfide bond between C151 and C161 restrict the movement of CTR region, important for the inter protomer interaction (Kundu et al., 2017). The multiple sequence analysis of all the Trh of the *Vibrio* and non-*Vibrio* sp. showed the presence of CTR region at the position of 157 to 165 with some variation of amino acids at the position 159, 162, 163 and 165 among all the Trh protein (157SFXECXXQX165). There are many amino acid residues which had been reported by several workers responsible for the hemolytic activity of Tdh protein. A study carried out by Yanagihara et al., 2010 showed that Trp65and Leu66 are important for the hemolytic activity of Tdh protein. A separate experiment carried out by Iida et al. (1995) and Baba et al. (1992) showed Arg46, Gly62,

Trp65, Thr67, Gly86, Gly 90 Gln116 and Glu138 play a vital role in hemolysis. Except for Thr67, all the amino acid residues were detected at the same position in Trh protein and are conserved in all the Trh protein used in the multiple sequence analysis. Site-directed mutagenesis showed that the substitution of W65L and T67I lost the hemolytic activity of Tdh protein (Yanagihara et al., 2010). However, in the present study, we found that T67 was not conserved in Trh protein of *Vibrio* and non-*Vibrio* sp. used in the present study.

Homology modeling (Krieger et al., 2005) is used to resolve the 3D structure of protein for structural and functional analysis of protein when X-Ray crystallographic or NMR structures of the protein are not present. In the present study, 3D structure of trh protein was generated using Tdh (PDB ID:) as a template and structural stability of the 3D model of Trh protein was determined by applying molecular dynamics simulation with 60 ns. The radius of gyration determines the protein structure compactness (Lobanov et al., 2008). The structural stability and compactness of a protein are very much essential for the functional analysis of protein (Henzler-Wildman et al., 2007). The 3D structure generated in the present study were further used to identify the druggable site within trh protein. In the present study, 19 pockets have been identified out of which pocket 2 is most important in Tdh and Trh protein. Twenty different amino acids are involved in the formation of this pocket in Trh protein out of which eleven amino acids (Glu^{138}, Tyr^{140}, Cys^{151}, Phe^{158}, Cys^{161}, Lys^{162}, Ser^{163}, Gln^{164}) are actively participating for the formation of tetrameric structure which is the active conformation for hemolysin activity. These eleven amino acids are also present in the pocket 2 of Tdh protein. So, identification of any chemical compound or designing of any peptide that can easily and

efficiently bind with this pocket will disrupt the formation of the tetrameric structure of trh protein. A study carried out by Gao et al., 2007 and Wang et al., 2012 showed that inhibition of Tdh and Tlh protein with monoclonal and high-affinity antibody inhibits the cytotoxic effects of Tdh and Tlh positive *V. parahaemolyticus* in culture cell. Arunkumar et al. (2017) identified two natural compound cyaniding and bergapten using *insilico* approaches which may be used against Tdh protein. Insilico molecular interaction studies of γ-hemolysin secreted by *Staphylococcus aureus* with flavonoids compounds showed that ponciretin would inhibit the cytotoxicity and hemolytic activity of the toxin which may be used as a drug (Mohan et al., 2013). Heptameric form of α-hemolysin is an important virulence factor for *S. aureus* infections. Rani et al (2014) designed a peptide that efficiently binds with chain A of α-hemolysin of *S. aureus* that disturbed the dimer.

Antibiotics are commonly used for the treatment of bacterial infection. Similarly, in case of infection with *V. parahaemolyticus*, antibiotics are commonly used. A multidrug-resistant pandemic strain of *V. parahaemolyticus* was isolated from different parts of the world. So, there is an urgent need for such for alternative drug molecules as a substitute of antibiotics commonly use in the treatment of *V. parahaemolyticus* infection. The findings of this study will enlight the path on the development of suitable drug molecules against Tdh and Trh to treat the infections in humans and other organisms caused by *V. parahaemolyticus.*

9.5 Conclusion

Trh is one of the major virulent proteins in *V. parahaemolyticus* and is evolutionarily conserved among *Vibrio* sp and non-*Vibrio* sp. The 3D model of the Trh protein was predicted using the comparative homology modeling approach. This

predicted Trh protein model was then validated using various protein validation methods such as Ramachandran plot analysis and ProSA analysis. Additionally, the study of the MD simulation, trajectory analysis by means of average RMSD/RMSF calculation revealed that the Trh protein showed residual fluctuation and stable throughout the simulation. Furthermore, druggability probability analyses revealed that the Trh protein possesses a total of 19 druggable sites. interestingly it was found that among the 19 drugable pockets, one pocket (pocket 2) which contains eleven residues that actively participate in the formation of the tetrameric structure and this tetrameric conformation is essential for the cytotoxicity of Trh protein. The present study provides an insight on the druggable pockets of Trh protein which open new vistas for drug designing and development of medical therapeutics against *V. parahaemolyticus* infections in near future.