BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 6(1)

Structural inferences for Cholera toxin mutations in *Vibrio cholerae*

Gunasagaran Shamini[1, 5], Manickam Ravichandran[1], John T Sinnott[2, 3], Charurut Somboonwit[2, 3], Harcharan S Sidhu[1], Paul Shapshak[2, 4], Pandjassarame Kangueane[1, 5]*

[1]Department of Biotechnology, AIMST University, Semeling 08100, Kedah, Malaysia; [2]Division of Infectious Disease, Department of Internal Medicine, Tampa General Hospital, University of South Florida, College of Medicine, Tampa, Florida 33606, USA; [3]Clinical Research Unit, Hillsborough Health Department, Tampa, Florida 33602, USA; [4]Department of Psychiatry and Behavioral Medicine, University of South Florida, College of Medicine, Tampa, Florida 33613, USA; [5]Biomedical Informatics, Pondicherry, India 607402; Pandjassarame Kangueane -Email: kangueane@bioinformation.net; Phone: +6012 436 2581; *Corresponding author

Received January 20, 2011; Accepted February 01, 2011; Published March 02, 2011

Abstract:
Cholera is a global disease that has persisted for millennia. The cholera toxin (CT) from Vibrio cholerae is responsible for the clinical symptoms of cholera. This toxin is a hetero-hexamer (AB_5) complex consisting of a subunit A (CTA) with a pentamer (B_5) of subunit B (CTB). The importance of the AB_5 complex for pathogenesis is established for the wild type O1 serogroup using known structural and functional data. However, its role is not yet documented in other known serogroups harboring sequence level residue mutations. The sequences for the toxin from different serogroups are available in GenBank (release 177). Sequence analysis reveals mutations at several sequence positions in the toxin across serogroups. Therefore, it is of interest to locate the position of these mutations in the AB_5 structure to infer complex assembly for its functional role in different serogroups. We show that mutations in the CTA are at the solvent exposed regions of the AB_5 complex, whereas those in the CTB are at the CTB/CTB interface of the homo-pentamer complex. Thus, the role of mutations at the CTB/CTB interface for B_5 complex assembly is implied. It is observed that these mutations are often non-synonymous (e.g. polar to non-polar or vice versa). The formation of the AB_5 complex involves inter-subunit residue-residue interactions at the protein-protein interfaces. Hence, these mutations, at the structurally relevant positions, are of importance for the understanding of pathogenesis by several serogroups. This is also of significance in the improvement of recombinant CT protein complex analogs for vaccine design and their use against multiple serogroups.

Keywords: Cholera toxin (CT), *Vibrio cholerae*, O1/O139, non O1/O139, mutation, protein-protein interfaces

Background:
Vibrio cholerae is the cause of a waterborne disease affecting thousands of life every year **[1]**. The outbreak in October, 2010, in Haiti demonstrates the global issue of cholera and resulted in approximately 1,000 deaths within a month **[2]**. Cholera is an acute diarrheal disease caused by the gram-negative bacterium, *Vibrio cholerae*. There are more than 200 serogroups of *Vibrio cholerae* present in the natural environment **[3]**. However, two serogroups, O1 (widespread with El Tor and classical biotypes) and O139 (colonizes few regions of Asia) have been associated with the epidemics and pandemics of cholera during the last 25 years **[4-6]**. The O1 (El Tor biotype – Ogawa serotype) serogroup is responsible for the recent outbreak in Haiti **[7]**. The symptoms of cholera are caused by the secretion of an entero-toxin called cholera toxin (CT) **[8-9]** which is encoded by virulence factor genes; ctxA and ctxB **[10-11]**. These genes are acquired from a lysogenic filamentous bacteriophage (CTXΦ) through CTXΦ DNA integration into the host *Vibrio cholerae* genome **[12-14]**. It should be noted that the incidence of cholera outbreaks with serogroups other than O1/O139 (collectively referred as non O1/non O139) has also been recorded **[5, 10, 15-17]**. These strains are responsible for the sporadic outbreaks **[18-22]**. It is known that the virulent factors for non-(O1/O139) are different from the O1/O139 strains **[5, 23, 24]**. However, non-(O1/O139) strains with ctxA and ctxB genes also have been observed **[25-28]**.

CT, also known as choleragen, is a hetero-hexameric AB_5 complex in structure **(Figure 1) [29-31]** and is composed of an enzymatic A subunit (CTA) and a cell targeting B subunit (CTB) **[32-34]**. The enzymatically activated A subunit catalyzes adenylate cyclase to cause massive excretion of electrolytes from bowel **[35, 36]**. However, the homo-pentamer B subunit is mandatory for pathogenesis because of its vital role in binding to receptors of target cells **[37-39]**. The B complex binds to the intestinal epithelium and the A molecule then detaches and enters the cell via endocytosis. The A molecule then goes onto ribosylating the Gs alpha-subunit of G proteins that results in constitutive production of cAMP. The result is excretion of bicarbonate, chloride, potassium, and sodium ions as well as water from these cells **[40]**. Thus, the AB_5 complex assembly is critical for pathogenesis. The virulence factors in both O1/O139 and non-(O1/O139) strains have been identified **[8, 9, 16, 17, 24, 28, 41]**. It should be noted that information on the diarrheagenic potential of non-(O1/O139) is limited. The effect of mutations in the toxin from all

known serogroups is not available. Therefore, it is of importance to describe the virulence factors in both O1/O139 and non-(O1/O139). This is possible with the help of known structural complexes available in Protein databank (PDB). A comprehensive analysis of the AB_5 CT structures from PDB describing the nature of A and B_5 interface has been documented elsewhere **[42]**. Here, we describe the significance of mutations in CT among serogroups based on their residue positional occurrence (either at solvent exposed or interface regions of the AB_5 complex).

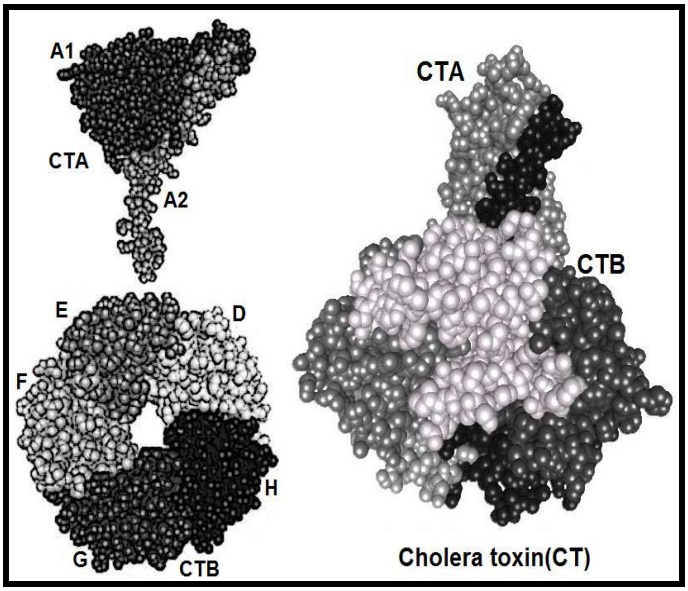

Figure 1: Structural model of a cholera toxin (CT). CT is a hetero-hexameric complex (AB_5) consisting of CTA (194 residues A1 and 46 residues A2) and CTB (103 residues) pentamer with D, E, F, G and H chains.

Materials and Methodology:

CT sequence dataset:

We created a dataset of 27 CTA (O1: 14; O139: 5; non-O1/O139: 8) and 165 CTB (O1: 121; O139: 37; non-O1/O139: 7) sequences as available from GenBank (release 177; year 2010 **[43]** using the procedure outlined in **Figure 2**. The number of sequences in the datasets is stated in **Table 1 (see Supplementary material)**. There are more CTB sequences than CTA sequences suggesting a higher frequency of mutations in CTB. Some partial sequences have been included in the dataset due to the non-availability of their full-length sequences in GenBank. In addition, these partial sequences also harbored mutations compared to wild type sequences.

Multiple Sequence Alignment (MSA) of CTA and CTB:

MSA is performed using ClustalX 2.0.12 **[44]** with the substitution matrix PAM 80. A gap-opening penalty of 10 and extension-penalty of 0.2 were used for the alignment. Sequences of CTA and CTB with known structures (PDB ID: 1XTC **[45]**) belonging to the O1 classical 569B strain were used as reference sequences in this alignment. The alignment was used to identify mutations in CTA **(Figure 3)** and CTB **(Figure 4)** among the different serogroups. Mutations were identified at six residue positions (7, 28, 112, 134, 163 and 222) in CTA **(Figure 3)** and at 13 residue positions (3, 7, 13, 15, 18, 22, 25, 34, 46, 47, 52, 60 and 94) in CTB **(Figure 4)** among O1/O139 and non-(O1/O139) strains.

CT structures:

The formation of the AB_5 complex is critical for pathogenesis. This is achieved through the formation of B_5 and AB_5 complexes. The B_5 complex is formed through the assembly of 5 monomeric B subunits arranged in a circle with a central groove in the first stage. This results in an assembly with each B subunit juxtaposed on either side with two other B subunits with a stable interface as shown in **Figure 5**. Mutations in the B subunit and the potential occurrence at the CTB/CTB interface influence the formation of the B_5 complex. The formation of AB_5 complex occurs through the interaction of CTA and B_5 complex. Thus, mutations in either CTA or CTB among the different serogroups have effect at the interface of CTA/ CTB complex.

Interface residues:

Interface residue positions were identified using the change in solvent accessible surface area (ASA) upon complex formation from a monomer state to a dimer state both within B_5 complex and between CTA/CTB. ASA is calculated using an algorithm developed by Lee and Richard (1971) implemented in the software SURFACE RACER with a probe radius of 1.4 Å **[46]**. We identified the interface residues between CTA/CTB complex and within the B_5 complex in respective serogroups using the procedure described elsewhere **[42]**. In this procedure, interface residue positions were identified using ASA analysis of subunits in the AB_5 structural complexes.

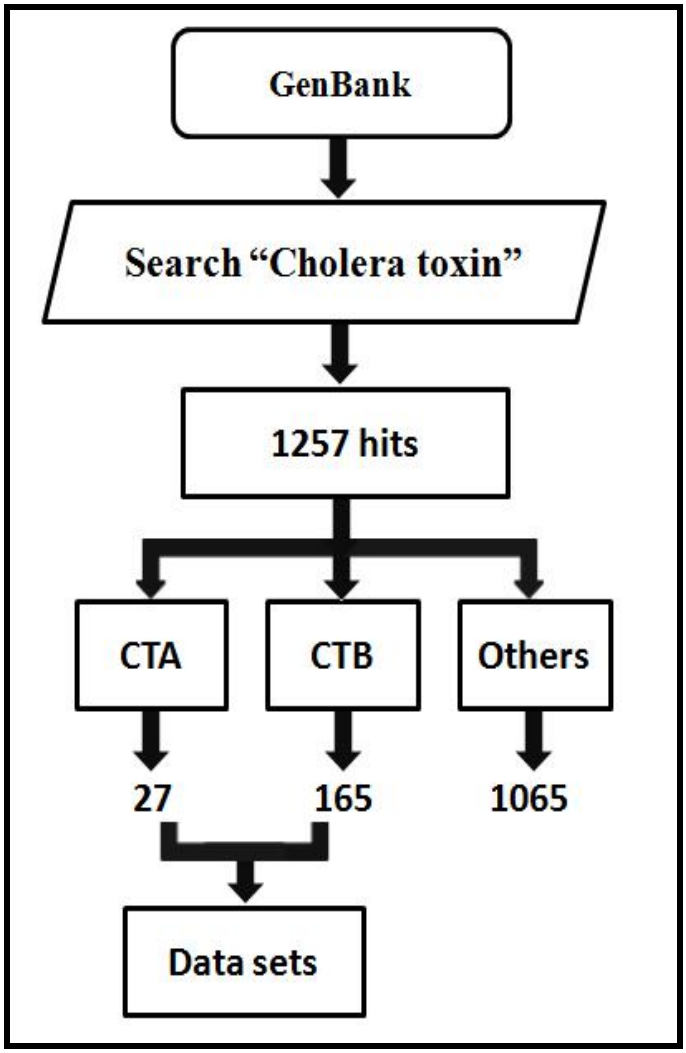

Figure 2: Creation of sequence dataset for CTA and CTB. A sequence dataset of CTA and CTB was derived from GenBank (release 177) using KEYWORD search as illustrated in the flowchart. The KEYWORD search "cholera toxin" resulted in 1257 hits. This set consists of 27 CTA sequences, 165 CTB sequences according to GenBank description and available annotations. The remaining 1065 sequences with descriptions such as secretion protein, cholera toxin transcriptional activator, ADP-ribosylation factor, GNAS complex, dopamine receptor, Pertusis toxin, Shiga-like toxin and the like are eliminated from the dataset. Thus, a CT sequence dataset of 192 sequences (**Table 1**) consisting of 27 CTA and 165 CTB was created. The CTA and CTB sequences are included in the dataset as available in the GenBank. The biased availability on the amount of CTA and CTB sequences in GenBank is attributed to the likely observation of frequent mutations in CTB.

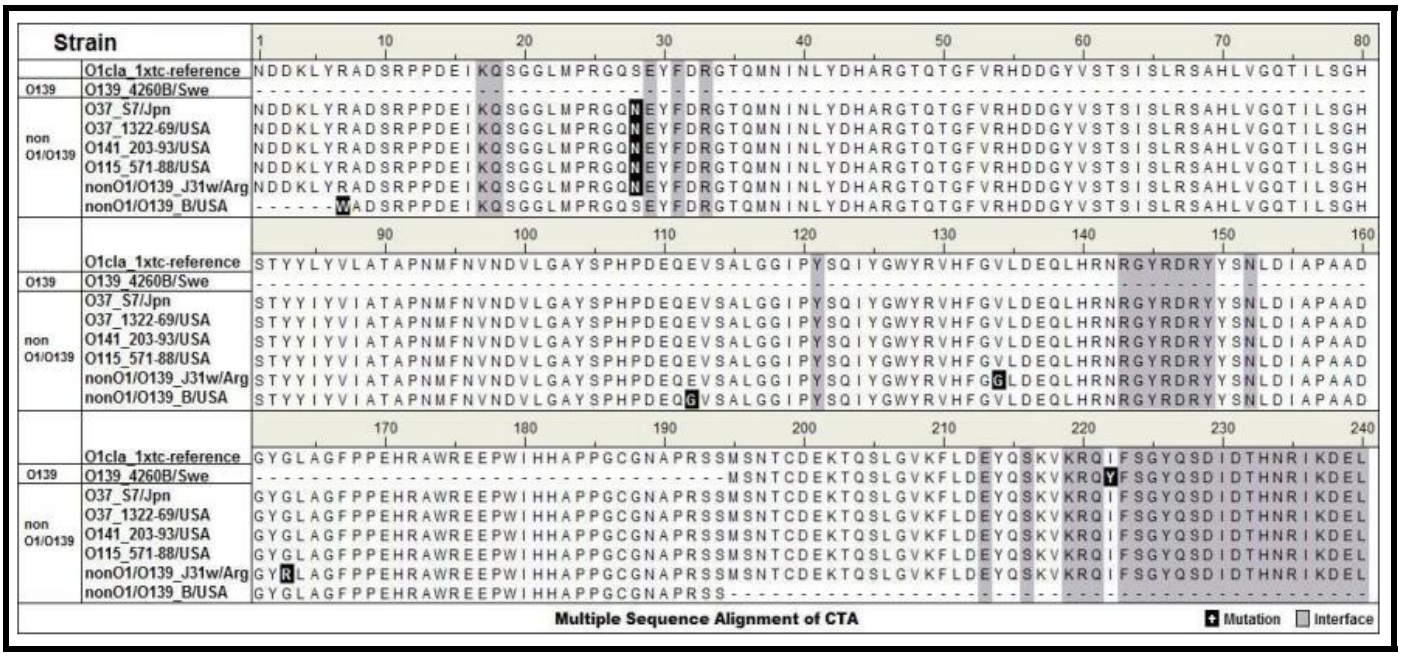

Figure 3: MSA for the CTA subunit of different serogroups. The MSA was performed using the wild type O1 classical strain sequence with known structure (PDB ID: 1XTC) as reference. The position specific mutations among the available CTA sequences (27) with reference to the classical sequence are indicated using dark shades. CT is an AB_5 hetero-hexamer and hence, the CTA/CTB interface residues in CTA are indicated using light shades.

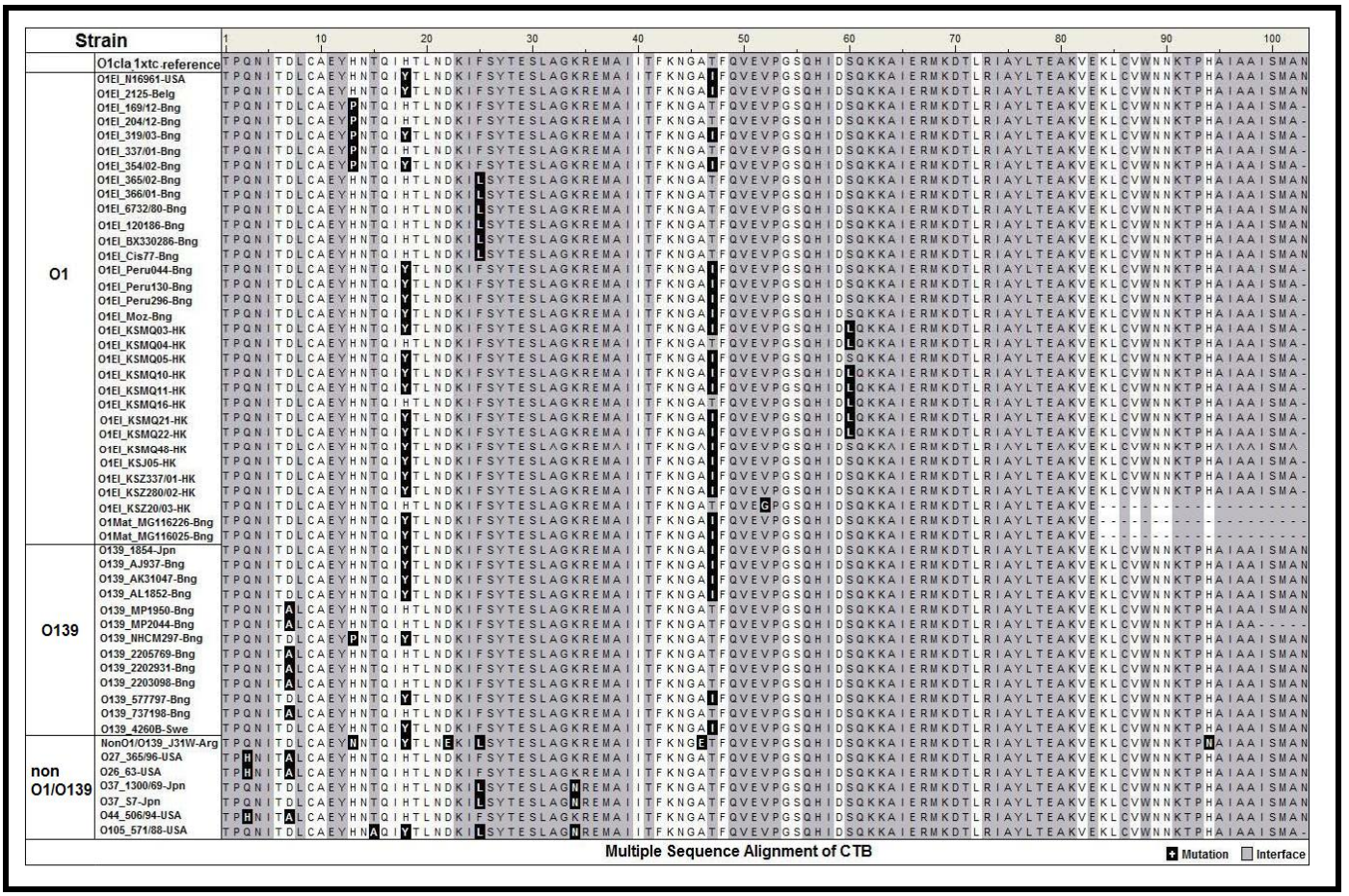

Figure 4: MSA for the CTB subunit of different serogroups. The MSA was performed using the wild type O1 Classical strain with known structure (PDB ID: 1XTC) as reference. The position specific mutations among the available CTB sequences (165) with reference to the Classical sequence are indicated using dark shades. B_5 is a homo-pentamer and hence, the CTB/CTB interface residues in B_5 are indicated using light shades. It should be noted that the mutated residues at the CTB/CTB interfaces in B_5 are highlighted using both dark and light shades at their corresponding position specific residues.

Mapping mutated residue positions to structures:
The structures for AB_5 and B_5 complexes of the wild type O1 strain are available at the PDB. It is of interest to infer structural effect caused by the mutations in other known serogroups. It is well known that homologous sequences have similar structures and they differ only in side chain details. Therefore, mapping of mutated residue positions from MSA **(Figure 3** and **Figure 4)** to known structural regions (exposed, buried, interface) provide the opportunity to identify mutations at the interface of CTA/CTB and within B_5 **(Figure 6)**. This approach identified mutations (at six residue positions such as 7, 28, 112, 134, 163 and 222) in CTA that are located at structurally solvent exposed regions of the complex **(Figure 6a)**. It also helped to locate several mutations (seven residue positions such as 3, 15, 25, 34, 47, 52 and 60) in CTB that are at the B_5 homo-pentamer subunit interfaces **(Figure 6b** and **6c)**.

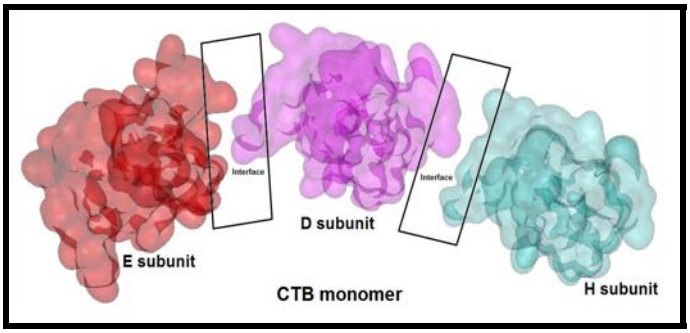

Figure 5: Structural model of CTB/CTB interfaces in B_5. B_5 is a homo-pentamer and each CTB subunit (D) is juxtaposed by two other CTB units on either side (E and H). Thus, the D subunit creates two different types of interfaces (D-E and D-H) on either side. This subsequently results in two different "position specific interacting" patterns in sequence for subunit D.

Structural 3D visualization of mutated residue positions in serogroups:
We used Discovery Studio Visualizer (v2.5.5.9350) to illustrate the mutated residue positions in CTA **(Figure 7a)** and CTB **(Figure 7b)** among the serogroups. The mutated residue positions at the interface of CTA/CTB **(Figure 8a)** and with B_5 **(Figure 8b)** is also shown.

Results:
Table 1 describes the dataset of CTA and CTB sequences retrieved from GenBank (release 177; year). The dataset consists of CTA and CTB sequences from O1 (El Tor, Classical, Matlab), O139 and non-(O1/O139) serogroups. We compared the CT sequences for O139 and non-(O1/O139) with the wild type Classical O1 serogroup. **Figure 3** and **Figure 4** show the results of MSA for CTA (27 sequences) and CTB (165 sequences), respectively. The wild type O1 Classical sequence with known structure (PDB ID: 1XTC) from strain 569B was used as reference in the alignment. The alignment is showed only for sequences with mutations (7 CTA and 52 CTB mutants) to the wild type reference sequence **(Figure 3** and **Figure 4)**. The mutations observed from the MSA of known CTA and CTB sequences are summarized in **Table 2** and **Table 3 (see Supplementary material)**, respectively. The mutations in CTA are found at 6 residue positions (7, 28, 112, 134, 163 and 222) among serogroups in the dataset. The mutations in CTB sequences are at 13 residue positions (3, 7, 13, 15, 18, 22, 25, 34, 46, 47, 52, 60 and 94) in the dataset.

Table 2 shows that the I222Y mutant is in the O139 strain (4260B) and the other six are in non-(O1/O139) strains. Data also shows that all strains except for strains B (2 positional mutations) and J31W (3 positional mutations) have only one positional mutation (**Table 2** and **Figure 3**). Similarly, mutations were seen at one position in 18 strains (O1: 12; O139: 6), at two positions in 26 strains (O1: 14; O139: 7; non-(O1/O139): 5), at 3 positions in 6 O1 strains, 4 positions and 6 positions in one non-(O1/O139) strain for CTB (**Figure 4** and **Table 3**). The non-(O1/O139) serogrouped J31W strain carried the maximum number of mutations in CT (CTA and CTB). Thus, the position specific mutations for CTA and CTB sequences were observed.

The availability of CT structure (PDB ID: 1XTC) provides an opportunity to map position-specific mutations in different serogroups to its structural preference (solvent exposed, buried, interfaces). Therefore, the significance of these mutations in the formation of the AB_5 assembly could be subsequently inferred. The mutations (dark shades as background) in the CTA and the CTB are shown in **Figure 3** and **Figure 4** along with corresponding interface residues (light shades). This helps to relate the consequence of mutations in structure. CT is an AB_5 complex **(Figure 1)** consisting of several layers of subunit protein interfaces formed by non-covalent interactions. Therefore, it is of interest to map the mutations in serogroups to their structural positions (interior, interface, surface).

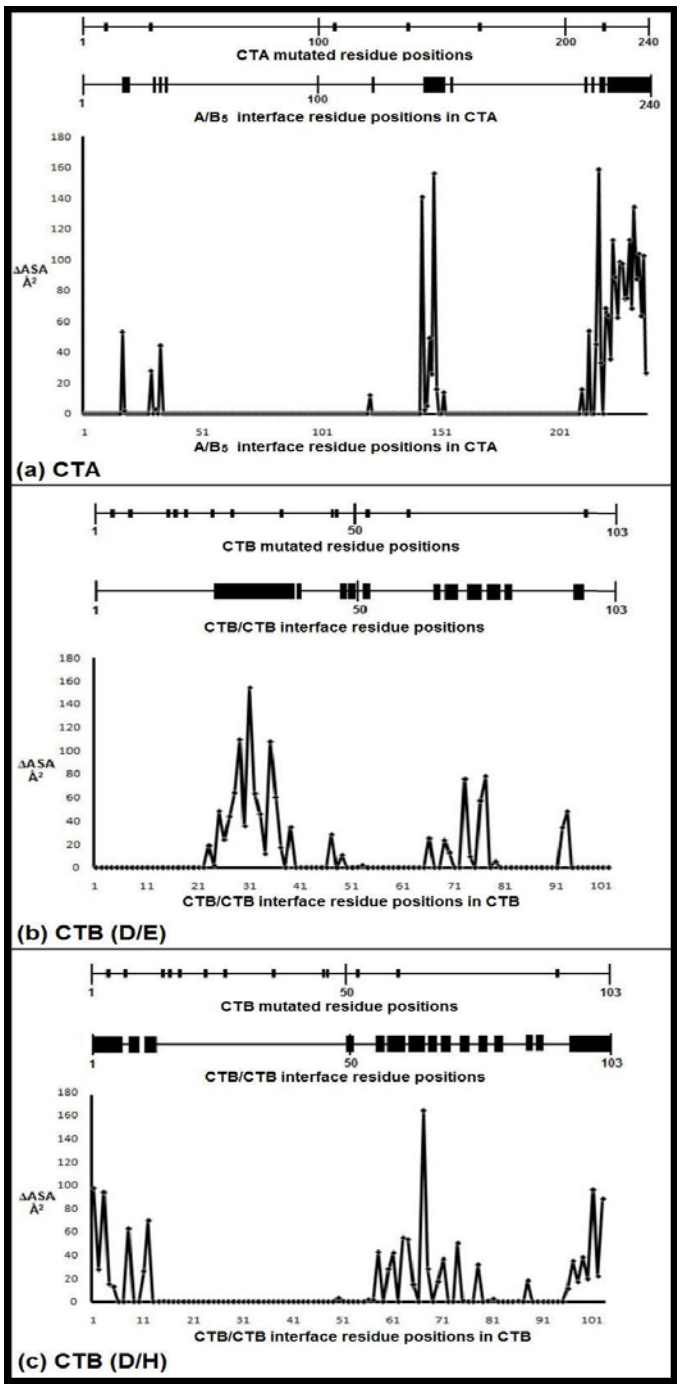

Figure 6: Representation of mutated residue positions in serogroups to interface residues in CT complex as a function of their residue position identified using ΔASA measure. (a) Mapping of CTA mutations to CTA/CTB interface residues in CTA (Please refer to **Figure 1** for the visual illustration of CTA/CTB interface). (b) Mapping of CTB mutations to CTB (D subunit)/CTB (E subunit) interface residues (Please refer to **Figure 5** for the visual illustration of D-E interface). (c) Mapping of CTB mutations to CTB (D subunit)/CTB (H subunit) interface residues (Please refer to **Figure 5** for the visual illustration of D-H interface). It should be noted that mutated residue positions are mapped on to corresponding interface residue positions in all the three cases (a), (b) and (c).

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 6(1): 1- 9 (2011)

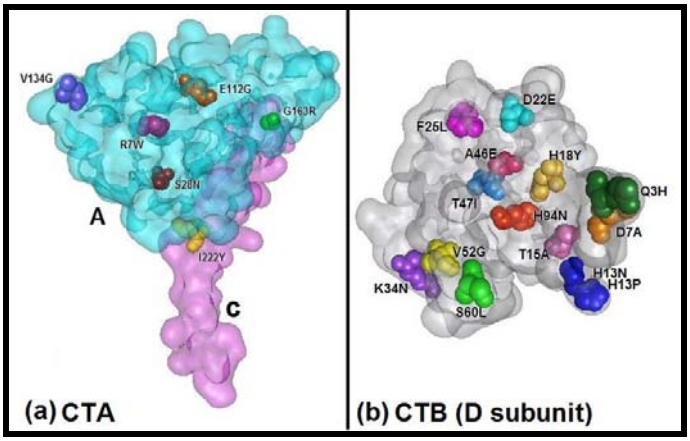

Figure 7: Structural models of CTA (a) and CTB (b) subunits with known mutations among archived serogroups. We used the structure with PDB entry (1XTC) for generating this visual using the freeware Discovery studio from Accelrys Inc. (a) A total of 6 unique mutations thus observed among the known CTA sequences **(Table 2)** from several serogroups are shown at their corresponding 6 residue positions using the Corey-Pauling-Kultun (CPK) residue model representation. (b) Fourteen unique mutations thus observed among the known CTB sequences **(Table 3)** from several serogroups are shown at their corresponding 13 residue positions using the CPK residue model representation.

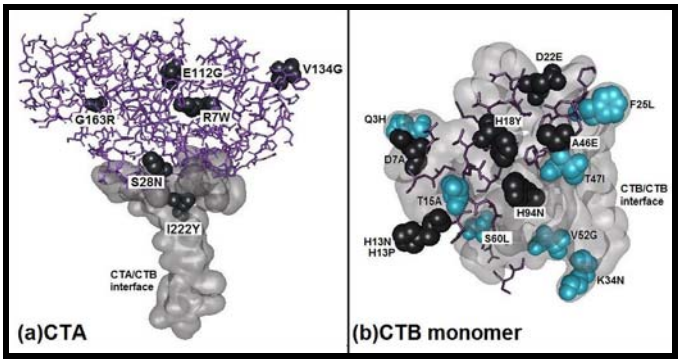

Figure 8: Structural models of CTA (a) and CTB (b) subunits with known mutations at the respective structural interfaces or solvent accessible regions in the complex among archived serogroups. We used the structure with PDB entry (1XTC) for generating this visual using the freeware Discovery studio from Accelrys Inc. (a) A total of 6 unique mutations thus observed among the known CTA sequences **(Table 4)** from several serogroups are shown at their corresponding 6 residue positions using the CPK residue model representation. All of these 6 mutated positions are present at the solvent exposed regions of CTA in both monomer and CTA/CTB complex state. (b) A total of 7 out of 14 unique mutations thus observed among the known CTB sequences **(Table 4)** from several serogroups are shown at their corresponding 7 (3, 15, 25, 34, 47, 52 and 60) out of the 13 residue positions using the CPK residue model representation are at the CTB/CTB interfaces in the B_5 complex.

Protein-Protein interfaces are formed between A and B_5 as well as within B_5. The mutations in A and B will potentially affect A/B_5 interface **(Figure 1)**. B_5 is a homo-pentamer and each B subunit is juxtaposed with similar CTB units on either side **(Figure 5)**. Similarly, mutations within B will possibly affect the formation of B_5 such that the D-E and D-H interfaces are affected **(Figure 5)**. Nevertheless, these interfaces should be translated into sequence positions using ΔASA in solved CT structures as described in **Figure 6**. Moreover, Fig. 6 maps the mutations in CTA to their occurrence at the CTA/CTB interface **(Figure 6a)** and in CTB to their possible occurrence at the D-E and D-H interfaces **(Figure 6b** and **6c)** in B_5. This comparison helps to identify the presence of mutated residues **(Figure 7)** in CTA **(Figure 7a)** and CTB **(Figure 7b)** at their respective CTA/CTB **(Figure 8a)** interface and CTB/CTB **(Figure 8b)** interfaces. The 6 mutations (R7W, S28N, E112G, V134G, G163R and I222Y) in CTA are positioned at the solvent exposed regions of the subunit **(Figure 6)** with no mutations at the CTA/CTB interface. However, it should be noted that the S28N and I222Y mutation were closely located to the CTA/CTB interface **(Table 4)**. The role of these mutations in CT complex assembly is of interest. A number of mutations in the CTB sequence are positioned at the CTB/CTB interfaces unlike the mutations in CTA sequence. The mutations at residue positions 3, 15, 25, 34, 47, 52 and 60 are within CTB/CTB interfaces **(Figure 8)**. The nature of amino acid mutations in CTA and CTB among O1/O139 and non-(O1/O139) serogroups are given in **Table 4 (see Supplementary material)**.

Discussion:

Choleragen (CT) and Choleragenoid (CTB) have been used as cholera vaccine candidates **[47]**. A number of subunit vaccine candidates using CTA ((S63K, R192G, R192N) **[48]**, (I16A or V72Y, I16A+Y68S, V72Y+Y68S) **[49]**, (V53D, V53E, V53Y, S63K, V97K, V97Y, Y104K, Y104D, Y104S, P106S) **[50]**) mutants and CTB recombinants have been developed in addition to heat killed attenuated *Vibrio cholerae* as vaccines. Sequence and structural studies of CT offer tremendous opportunity for the improvements in vaccine candidate design and development. The presence of CT epitypes **[51]** and heterogeneity in CTB subunit **[52]** also need to be considered from a vaccine perspective. A vaccine for cholera must target O1, O139 as well as non-O1 and non-O139 strains to have effective control over cholera outbreaks. Moreover, different serogroups of non-(O1/O139) strains (with ctxAB genes **[25, 41, 53, 54]**) and newly emerging *Vibrio cholerae* strains (O1 Matlab **[55-57]**, O1 El Tor with altered CTB **[58, 59]**) must be taken into consideration in cholera vaccine design. Hence, comparison studies on CTA and CTB sequences from various *Vibrio cholerae* serogroups provide insights in developing an effective toxin analog for vaccine design against multi serogroups.

A number of sequence comparison studies show CT sequence homology among various *Vibrio cholerae* serogroups. Recently, Kumar et al. (2009) documented a new CT variant of the *Vibrio cholerae* O1 El Tor biotype isolated from Orissa (India) **[60]**. The study highlighted a novel mutation (H20N) in CTB and the presence of altered CTB of the Classical biotype in the El Tor clinical isolates. Raychoudhuri and team (2009) conducted a study to attest the replacement of El Tor biotype ctxB allele by Classical biotype ctxB allele in O1 strains **[61]**. A study by Ansaruzzaman and colleagues (2004) reported H18Y and T47I substitutions in CTB of El Tor strain and these sequences are similar to CTB of Classical biotype **[62]**. Previously, a study demonstrated the emergence of new El Tor strains with a modified Classical biotype CT **[60]**. Thus, a dataset of sequences **(Table 1)** for CTA and CTB representing diverse serogroups isolated at various periods of time from a variety of sources and locations available in GenBank (release 177) is created for this study. The nature of mutations **(Table 4)** among the serogroups is presented for CTA **(Table 2)** and CTB **(Table 3)** sequences. Several studies have demonstrated the effects of site directed mutations in CTA as well as in CTB subunits for the wild type O1 strain **(Table 5 see Supplementary material)**. Manufactured site directed mutants leading to decrease or loss in toxicity has been reported for CTA (R7K, R11K, I16A, R25G, E29H, S68Y + V72Y, E112Q, F223D) and CTB (R35D, H57A, L77D, I74D, T78D). Thus, the role of site directed mutants in the loss of toxicity is known for the wild type O1 strain. Therefore, it is important to evaluate the effect of mutations caused by natural selection pressure among serogroups.

The multiple sequence alignment (MSA) of these sequences showed mutations in CTA (at six residue positions such as 7, 28, 112, 134, 163 and 222) and CTB (at 13 residue positions such as 3, 7, 13, 15, 18, 22, 25, 34, 46, 47, 52, 60 and 94) among O1/O139 and non-(O1/O139) strains. The effects of these mutations in the formation of a clinically functional cholera toxin (AB_5 hetero-hexamer) are of significant importance. Reports describing the emergence of new serogroups with novel mutations in CTA and CTB are available. However, studies on the effects of mutations in CT relative to CTA/CTB-pentamer interface **(Figure 1)** and within CTB/CTB interfaces **(Figure 5)** are not yet available. Here, we present results of a comprehensive analysis of mutations in CTA and CTB sequences from several serogroups **(Table 4)**.

This mutational data is presented relative to A/B_5 and CTB/CTB interfaces for AB_5 assembly to understand its functions. Data suggest the presence of mutations in CTA **(Figure 8a)** and CTB **(Figure 8b)** at the solvent exposed, interior, subunit interface regions of the complex. The mutations (at 6 residue positions such as 7, 28, 112, 134, 163 and 222) in CTA are located at structurally solvent exposed regions of the subunit. However, several mutations (7 residue positions such as 3, 15, 25, 34, 47, 52 and 60) in CTB are at the B_5 homo-pentamer subunit interfaces. Thus, the role of these mutations in CTA and CTB towards the assembly of AB_5 CT among the O1/O139 and non-

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 6(1): 1- 9 (2011)

(O1/O139) strains is inferred from this study **(Figure 7** and **8)**. It should also be noted that some of these mutations (polar to non-polar or vice versa) are largely non-synonymous (causing physical and chemical property shift) in nature and have potential effect on protein-protein interactions of the CT subunits affecting AB_5 formation **(Table 5)**. Thus, data presented in **Table 4** is all-inclusive, updated, relevant and specific for several known serogroups. This is of significance towards the improvement of recombinant CT protein complex analogs for vaccine design against multi serogroups.

Conclusion:

The structural role of cholera toxin in pathogenesis is known for the wild type O1 strain. It was of interest to document its role in other known serogroups showing mutations with the wild type. We described the structural location of such mutations in the known serogroups to infer its functional role. We documented that mutations in CTA are at the solvent exposed regions of the AB_5 complex, while those in CTB are at the CTB/CTB interface of the homo-pentamer complex. It is observed that these mutations are also non-synonymous (i.e. polar to non-polar or vice versa) in property. Thus, the effect of these mutations in the AB_5 assembly is inferred. It is also of global importance to quantify precisely the structural effects caused by these mutations. The resulting data is relevant in designing a recombinant CT protein complex analog for vaccine design against multiple serogroups. Coupled to these analyses, it may be stated also that from a clinical perspective, the task of enhancing oral cholera vaccines entails reducing bacterial and Giardia infection and improving diet **[63]**.

Acknowledgements:

We wish to express our sincere appreciation to members of Biomedical Informatics (Pondicherry, India) for many discussions on the subject of this article. GS is thankful to the Department of Biotechnology (AIMST University, Malaysia) for support.

References:

[1] WHO. Cholera 2009. *Wkly Epidemiol Rec.* 2010 **85**: 293.
[2] Centers for Disease Control and Prevention (CDC). Update: Cholera Outbreak - Haiti, 2010, MMWR Morb Mortal Wkly Rep. 2010 **59**: 1473.
[3] Shimada T *et al. Curr Microbiol.* 1994 **28**: 175
[4] Bhattarcharya SK *et al. J Infect.* 1993 **27**: 11 [PMID: 8370939]
[5] Kaper JB *et al. Clin Microbiol Rev.* 1995 **8**: 48 [PMID: 7704895]
[6] Sack DA *et al. Lancet* 2004 **363**: 223 [PMID: 14738797]
[7] Centers for Disease Control and Prevention (CDC). *MMWR Morb Mortal Wkly Rep.* 2010 **59**: 1411 [PMID: 21048563]
[8] Cassel D & Pfeuffer T. *Proc Natl Acad Sci U S A.* 1978 **75**: 2669 [PMID: 208069]
[9] RA Finkelstein, Cholera, the cholera enterotoxins, and the cholera enterotoxin-related enterotoxin family, In Owen P, Foster TS (Eds): Immuno-chemical and Molecular Genetic Analysis of Bacterial Pathogens. Elsevier; 1988 85-102.
[10] Spangler BD. Microbiol Rev. 1992 **56**: 622 [PMID: 1480112]
[11] Davis BM & Waldor MK. *Curr Opin Microbiol.* 2003 **6**: 35 [PMID: 12615217]
[12] Pearson GD *et al. Proc Natl Acad Sci U S A.* 1993 **90**: 3750 [PMID: 8475125]
[13] Waldor MK & Mekalanos JJ. *Science* 1996 **272**: 1910 [PMID: 8658163]
[14] Karaolis DK *et al. Proc Natl Acad Sci U S A.* 1998 **95**: 3134 [PMID: 9501228]
[15] Kamble TK *et al. J Assoc Physicians India.* 2000 **48**: 505 [PMID: 11273144]
[16] Bag PK *et al. Appl Environ Microbiol.* 2008 **74**: 5635 [PMID: 18641168]
[17] Chatterjee S *et al. J Clin Microbiol.* 2009 **47**: 1087 [PMID: 19158257]
[18] Bagchi K *et al. J Clin Microbiol.* 1993 **31**: 1315 [PMID: 8501234]
[19] Ramamurthy T *et al. J Med Microbiol.* 1993 **39**: 310 [PMID: 8411093]
[20] Dalsgaard A *et al. J Clin Microbiol.* 1995 **33**: 2715 [PMID: 8567912]
[21] Sharma C *et al. J Clin Microbiol.* 1998 **36**: 756 [PMID: 9508308]
[22] Mitra RK *et al. J Med Microbiol.* 2001 **50**: 268 [PMID: 11232774]
[23] Nair GB *et al. Appl Environ Microbiol.* 1988 **54**: 3180 [PMID: 3223774]
[24] Chen Y *et al. Infect Immun.* 2007 **75**: 2645 [PMID: 17283087]
[25] Ghosh C *et al. Microb Pathog.* 1997 **22**: 199 [PMID: 9140915]
[26] Rivera IN *et al. Appl Environ Microbiol.* 2001 **67**: 2421 [PMID: 11375146]
[27] Sarkar A *et al. Infect Immun.* 2002 **70**: 4735 [PMID: 12117994]
[28] Jiang S *et al. Appl Environ Microbiol.* 2003 **69**: 7541 [PMID: 14660411]
[29] Gill DM, *Biochemistry* 1976 **15**: 1242 [PMID: 3214]
[30] Burnette WN, *Structure* 1994 **2**: 151 [PMID: 8069630]
[31] Zhang RG *et al. J Mol Biol.* 1995 **251**: 563 [PMID:7658473]
[32] Lai CY *et al. J Infect Dis.* 1976 **133**: S23 [PMID: 815448]
[33] Ohtomo N *et al. J Infect Dis.* 1976 **133**: S31[PMID: 815449]
[34] Sixma TK *et al. Nature* 1991 **351**: 371 [PMID: 2034287]
[35] Moss J & Vaughan M. *Mol Microbiol.* 1991 **5**: 2621 [PMID: 1779753]
[36] Moss J *et al. Methods Enzymol.* 1994 **235**: 640 [PMID: 8057933]
[37] Heyningen S Van. *Science* 1974 **183**: 656 [PMID: 4810267]
[38] Merritt EA *et al. Protein Sci.* 1994 **3**: 166 [PMID: 8003954]
[39] Rong W *et al. Colloids Surf B Biointerfaces.* 2004 **33**: 45.
[40] De Haan L & Hirst TR. *Mol Membr Biol*. 2004 **21**: 77 [PMID: 15204437]
[41] Chakraborty S *et al. Appl Environ Microbiol.* 2000 **66**: 4022 [PMID: 10966424]
[42] Shamini G *et al.* Protein-protein interfaces in cholera toxin complex, In Kangueane P (eds). Protein - Protein Interactions. Nova Science Publishers Inc. 2010; Chapter 12, pp251-272.
[43] http://www.ncbi.nlm.nih.gov/genbank.
[44] Larkin MA *et al. Bioinformatics* 2007 **23**: 2947 [PMID: 17846036]
[45] www.rcsb.org/pdb
[46] Lee B & Richard FM. *J Mol Biol.* 1971 **5**: 379 [PMID: 5551392]
[47] Vanden Broeck D *et al. Int J Biochem Cell Biol.* 2007 **39**: 1771 [PMID: 17716938]
[48] Pizza M, Fontana MR, Giannelli, Rappuoli VR, investors; United States patent US 7,632,513. 2009 Novartis Vaccines and Diagnostics SRL, assignee. Immunogenic detoxified mutants of cholera toxin.
[49] Green BA, Holmes RK, Jobling MG, Zhu D, investors; United States patent US 7,361,355. 2008 Wyeth Holdings Corporation and The Regents of the University of Colorado, A Body Corporate, assignee. Mutant forms of cholera holotoxin as an adjuvant.
[50] Domenighini M, Rappuoli R, Pizza M, Hol W, investors; United States patent US 6,149,919. 2000 Biocine S.p.A., assignee. Immunogenic detoxified mutants of cholera toxin and of the toxin LT, their preparation and their use for the preparation of vaccines.
[51] Finkelstein RA *et al. Rev Infect Dis.* 1987 **9**: 544 [PMID: 2440089]
[52] Olsvik O *et al. J Clin Microbiol.* 1993 **31**:22 [PMID: 7678018]
[53] Sharma A & Chaturvedi AN, *Int J Hyg Environ Health.* 2006 **209**: 521 [PMID: 16891156]
[54] Chun J *et al. Proc Natl Acad Sci U S A.* 2009 **106**: 15442 [PMID: 19720995]
[55] Nair GB *et al. J Clin Microbiol.* 2002 **40**: 3296 [PMID: 12202569]
[56] Safa A *et al. J Clin Microbiol.* 2005 **43**: 1401 [PMID: 15750117]
[57] Safa A *et al. J Med Microbiol.* 2006 **55**: 1563 [PMID: 17030917]
[58] Nair GB *et al. J Clin Microbiol.* 2006 **44**: 4211 [PMID: 16957040]
[59] Nguyen BM *et al. J Clin Microbiol.* 2009 **47**: 1568 [PMID: 19297603]
[60] Kumar P *et al. J Med Microbiol.* 2009 **58**: 234 [PMID: 19141742]
[61] Raychoudhuri A *et al. Emerg Infect Dis.* 2009 **15**: 131 [PMID: 19116078]
[62] Ansaruzzaman M *et al. Emerg Infect Dis.* 2004 **10**: 2057 [PMID: 16010751]
[63] Levine MM. *BMC Biology* 2010 **8**: 129 [PMID: 20920375]
[64] Burnette WN *et al. Infect Immun.* 1991 **59**: 4266 [PMID: 1937784]
[65] Hase CC *et al. Infect Immun.* 1994 **62**: 3051 [PMID: 8039872]
[66] Chan M *et al. Microb Pathog.* 2010 **49**: 211 [PMID: 20558271]
[67] Jobling MG & Holmes RK. *J Bacteriol.* 2001 **183**: 4024 [PMID: 11395467]
[68] Tinker JK *et al. Infect Immun.* 2003 **71**: 4093 [PMID: 12819100]
[69] Jobling MG & Holmes RK. *Infect Immun.* 2002 **70**: 1260 [PMID: 11854209]
[70] Silva A *et al. Curr Microbiol.* 1998 **37**: 231 [PMID: 9732528]
[71] Merritt EA *et al. Structure* 1995 **3**: 561 [PMID: 8590017]
[72] Aman AT *et al. Proc Natl Acad Sci U S A.* 2001 **98**: 8536 [PMID: 11447291]

Edited by VS Mathura
Citation: Shamini *et al.* Bioinformation 6(1): 1- 9 (2011)

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 6(1): 1- 9 (2011)

BIOINFORMATION

open access

Supplementary material:

Table 1: Dataset summary for CTA and CTB sequences from various serogroups

Serogroups		Number of sequences	
	Biotypes	CTA	CTB
O1	O1 El Tor	8	113
	O1 Classical	6	3
	O1 Matlab	0	5
O139	-	5	37
Non O1/O139	-	8	7
Total		27	165

Table 2: Mutations in CTA sequences among different serogroups as summarized from MSA.

Strain	Sero-group	Sero-type	Length	No. of mutated residues	Mutation	Accession number
4260B	O139	O139	46	1	I222Y	CAA53975
B		-	188	2	R7W, E112G	AAR29797
J31W		-	258	3	S28N, V134G, G163R	ACU00910
203-93	non	O141	258	1	S28N	AAL69945
571-88	O1/O139	O115	258	1	S28N	AAL69944
1322-69		O37	258	1	S28N	AAL60525
S7		O37	258	1	S28N	BAA06288

Table 3: Mutations in CTB sequences among different serogroups as summarized from MSA.

Strain	Serogroup	Biotype/Serotype	Length	No. of mutated residues	Mutation	Accession
N 16961	O1	El Tor	124	2	H18Y,T47I	AAF94613
2125	O1	El Tor	124	2	H18Y,T47I	CAA41593
169/12	O1	El Tor	123	1	H13P	ACH70469
204/12	O1	El Tor	123	1	H13P	ACH70470
319/03	O1	El Tor	123	3	H13P, H18Y, T47I	ACH70468
337/01	O1	El Tor	123	1	H13P	ACH70472
354/02	O1	El Tor	123	3	H13P, H18Y,T47I	ACH70467
365(2)	O1	El Tor	124	1	F25L	ACF35009
366(1)	O1	El Tor	124	1	F25L	ACF35007
6732/80	O1	El Tor	124	1	F25L	ACF35008
120186	O1	El Tor	124	1	F25L	ACF35006
BX 330286	O1	El Tor	124	1	F25L	ACF35005
Cis 77	O1	El Tor	124	1	F25L	ACF35010
Peru-044	O1	El Tor	115	2	H18Y,T47I	ACH70463
Peru-130	O1	El Tor	123	2	H18Y,T47I	ACH70464
Peru-296	O1	El Tor	123	2	H18Y,T47I	ACH70465
KSQM03	O1	El Tor	123	3	H18Y,T47I, S60L	ABV74277
KSQM04	O1	El Tor	123	1	S60L	ABV74281
KSQM05	O1	El Tor	123	2	H18Y,T47I	ABV74273
KSQM10	O1	El Tor	123	2	H18Y,T47I, S60L	ABV74278
KSQM11	O1	El Tor	123	3	H18Y,T47I, S60L	ABV74282
KSQM16	O1	El Tor	123	1	S60L	ABV74283
KSQM21	O1	El Tor	123	3	H18Y,T47I, S60L	ABV74274
KSQM22	O1	El Tor	123	3	H18Y,T47I, S60L	ABV74284
KSQM48	O1	El Tor	123	2	H18Y,T47I	ABV74285
KSJ05	O1	El Tor	123	2	H18Y,T47I	ABV74280
KSZ337/01	O1	El Tor	123	2	H18Y,T47I	ABV74276
KSZ280/02	O1	El Tor	123	2	H18Y,T47I	ABV74275
KSZ20/03	O1	El Tor	123	1	V52G	ABV74279
B65	O1	Mozambique	104	2	H18Y,T47I	AAV54184
MG116226	O1	Matlab	104	2	H18Y,T47I	ABG56879
MG116025	O1	Matlab	104	2	H18Y,T47I	ABG56881
1854	O139	-	124	2	H18Y,T47I	BAA06291
AJ_937	O139	-	124	2	H18Y,T47I	ACV81827
AK_31047	O139	-	124	2	H18Y,T47I	ACV81828
AL_1852	O139	-	124	2	H18Y,T47I	ACV81826
MP_1950	O139	-	124	1	D7A	ACV81829
MP_2044	O139	-	119	1	D7A	ACV81853
NHCM297	O139	-	124	2	H13P, H18Y	ACV81847
2205769	O139	-	124	1	D7A	ACV81837

2202931	O139	-	124	1	D7A	ACV81825
2203098	O139	-	124	1	D7A	ACV81836
5777_97	O139	-	124	2	H18Y,T47I	ACV81833
7371/98	O139	-	124	1	D7A	ACV81822
4260B	O139	-	124	2	H18Y,T47I	CAA53973
J31W	non O1/O139	-	124	6	H13N,H18Y,D22E,F25L, A46E, H94N	ACU00911
365-96	non O1/O139	O27	104	2	Q3H, D7A	AAM22587
63	non O1/O139	O26	124	2	Q3H, D7A	AAL6052
1300-69	non O1/O139	O37	124	2	F25L, K34N	AAL60524
S7	non O1/O139	O37	124	2	F25L, K34N	BAA06289
506-94	non O1/O139	O44	124	2	Q3H, D7A	AAL69946
571-88	non O1/O139	O105	123	4	T15A, H18Y, F25L, K34N	AAL60523

Table 4: Mutations, their sequence positions, change in amino acid types, relative occurrence at surface, interface, interior regions of the complex in different serogroups is given.

		Mutation	Nature of Mutations	No. of Strains	CTA/CTB interface	Buried	Exposed	Strain(s)
CTA	**O1/O139**	I222Y	Non polar to Non polar	1			✓	4260B
	Non-(O1/O139)	R7W	Polar(+) to Non polar	1			✓	B
		S28N	Polar(0) to Polar(0)	5			✓	J31W, 203-93, 571-88, 1322-69, S7
		E112G	Polar(-) to Non polar	1			✓	B
		V134G	Non polar to Non polar	1			✓	J31W
		G163R	Non polar to Polar(+)	1			✓	J31W

					CTB/CTB interface	CTB/CTA interface	Buried	Exposed	Strains(s)
CTB		D7A	Polar(-) to Non polar	6	✓				MP1950, MP2044, 2205769, 2202931, 2203098, 737198
		H13P	Polar(+) to Polar(0)	6				✓	169/12, 204/12, 319/03, 337/01, 354/02, NHCM297
		H18Y	Polar(+) to Non polar	27				✓	N16961, 2125, 319/03, 354/02, Peru044, Peru130, Peru296, Moz, KSMQ03, KSMQ05, KSMQ10, KSMQ11, KSMQ21, KSMQ22, KSMQ48, KSJ05, KSZ337/01, KSZ280/02, MG116226, MG116025, 1854, AJ937, AK31047, AL1852, NHCM297, 577797, 4260B
	O1/O139	F25L	Non polar to Non polar	6	✓				365/02, 366/01, 6732/80, 120186, BX330286, Cis77
		T47I	Polar(0) to Non polar	23	✓				N16961, 2125, 319/03, 354/02, Peru044, Peru130, Peru296, Moz, KSMQ03, KSMQ05, KSMQ10, KSMQ11, KSMQ21, KSMQ22, KSMQ48, KSJ05, KSZ337/01, KSZ280/02, MG116226, MG116025, 1854, AJ937, AK31047, AL1852, 577797, 4260B
		V52G	Non polar to non polar	1	✓				KSZ20/03
		S60L	Polar(0) to Non polar	7	✓				KSMQ03, KSMQ04, KSMQ10, KSMQ11, KSMQ16, KSMQ21, KSMQ22
	Non-(O1/O139)	Q3H	Polar(0) to Polar(+)	3	✓				O26_63, O27_365/96, O44_506/94
		D7A	Polar(-) to Non polar	3	✓				O26_63, O27_365/96, O44_506/94

	H13N	Polar(+) to Polar(0)	1				✓	J31W,
	T15A	Polar(0) to Non polar	1	✓				O105_571/88
	H18Y	Polar(+) to Non polar	2				✓	J31W, O105_571/88
	D22E	Polar(-) to Polar(-)	1				✓	J31W
	F25L	Non polar to Non polar	4	✓				J31W, O37_1300/69, O37_S7,O105_571/88
	K34N	Polar(+) to Polar(0)	3	✓				O37_1300/69, O37_S7,O105_571/88
	A46E	Non polar to Polar(-)	1				✓	J31W
	H94N	Polar(+) to Polar(0)	1				✓	J31W

Table 5: Function inference to known site directed mutants with corresponding relative occurrence at surface, interface, interior regions of the complex is given.

	Mutation	Inference	Reference		CTA/CTB interface		Buried	Exposed	Partially buried
CTA	R7K	ADP-ribosyl-transferase activity of CTA disrupted	Burnette *et al*. (1991), Hase *et al*. (1994), Chan *et al*. (2010)	**[64,65,66]**					✓
	R11K	Reduced toxicity		**[67]**				✓	
	I16A	Reduced toxicity	Jobling & Holmes (2001)						✓
	R25G	and enzymatic						✓	
	E29H	activity				✓		✓	
	Y30W,A,H								✓
	S68Y + V72Y							✓	
	E110D							✓	
	E112D	Reduced toxicity and enzymatic activity	Hase *et al*. (1994), Jobling & Holmes (2001), Chan *et al*. (2010)	**[65,66, 67]**					✓
	F223D	Reduced stability and toxicity	Tinker *et al*. (2003)	**[68]**		✓		✓	
					CTB/CTB interface	**CTB/CTA interface**	**Buried**	**Exposed**	**Partially buried**
CTB	V10A	Affects immunoreactivity	Jobling & Holmes (2002)	**[69]**				✓	
	G33E	Over expressed CTB diminishes active CT production	Silva *et al.* (1998)	**[70]**				✓	
	G33D	Abolish receptor binding ability	Merritt *et al*. (1995)	**[71]**				✓	
	R35D	Reduced AB_5 assembly						✓	
	V46A	Affects immunoreactivity	Jobling & Holmes (2002)	**[69]**				✓	
	H57A	Loss of toxicity	Aman *et al*. (2001)	**[72]**				✓	
	L74D	No AB_5 formation	Tinker *et al*. (2003)	**[68]**	✓	✓			
	I77D	No B_5 assembly			✓	✓			
	T78D	No AB_5 assembly			✓	✓			