Protein Science (1994), *3*:166-175. Cambridge University Press. Printed in the USA.

Crystal structure of cholera toxin B-pentamer bound to receptor G_{M1} pentasaccharide

ETHAN A. MERRITT,[1] STEVE SARFATY,[1] FOCCO VAN DEN AKKER,[1] CÉCILE L'HOIR,[2] JOSEPH A. MARTIAL,[2] AND WIM G.J. HOL[1]

[1] Department of Biological Structure, University of Washington, Seattle, Washington 98195
[2] Laboratoire de Biologie Moléculaire et de Génie Génétique, B6, Université de Liège, B-4000 Sart-Tilman, Belgium

(RECEIVED November 4, 1993; ACCEPTED December 7, 1993)

Abstract

Cholera toxin (CT) is an AB_5 hexameric protein responsible for the symptoms produced by *Vibrio cholerae* infection. In the first step of cell intoxication, the B-pentamer of the toxin binds specifically to the branched pentasaccharide moiety of ganglioside G_{M1} on the surface of target human intestinal epithelial cells. We present here the crystal structure of the cholera toxin B-pentamer complexed with the G_{M1} pentasaccharide. Each receptor binding site on the toxin is found to lie primarily within a single B-subunit, with a single solvent-mediated hydrogen bond from residue Gly 33 of an adjacent subunit. The large majority of interactions between the receptor and the toxin involve the 2 terminal sugars of G_{M1}, galactose and sialic acid, with a smaller contribution from the *N*-acetyl galactosamine residue. The binding of G_{M1} to cholera toxin thus resembles a 2-fingered grip: the Gal(β1-3)GalNAc moiety representing the "forefinger" and the sialic acid representing the "thumb." The residues forming the binding site are conserved between cholera toxin and the homologous heat-labile enterotoxin from *Escherichia coli*, with the sole exception of His 13. Some reported differences in the binding affinity of the 2 toxins for gangliosides other than G_{M1} may be rationalized by sequence differences at this residue. The CTB_5:G_{M1} pentasaccharide complex described here provides a detailed view of a protein:ganglioside specific binding interaction, and as such is of interest not only for understanding cholera pathogenesis and for the design of drugs and development of vaccines but also for modeling other protein:ganglioside interactions such as those involved in G_{M1}-mediated signal transduction.

Keywords: cholera toxin; crystal structure; ganglioside G_{M1}; sugar binding specificity

Cholera is a severe disease that can lead to death within a few hours. The major clinical symptoms are caused by a toxin released after adhesion of the noninvasive *Vibrio cholerae* bacteria to the proximal small intestine of the host. Cholera toxin (CT) acts intracellularly to catalyze ADP-ribosylation of residue Arg 187 in the α subunit of the trimeric protein G_s. The modified $G_{s\alpha}$ loses its GTPase activity and remains constitutively in its GTP-bound state (Cassel & Pfeuffer, 1978), which in turn causes a continuous stimulation of adenylate cyclase. The resulting elevated levels of cyclic AMP lead to massive loss of fluids, the characteristic pathology of enterotoxigenic disease.

Cholera toxin is an AB_5 hexamer consisting of 5 identical B subunits and a single A subunit. It is structurally and functionally related to a larger group of bacterial enterotoxins that includes the closely related *Escherichia coli* heat-labile enterotoxin (LT) as well as pertussis toxin, diphtheria toxin, shigella toxin, and *Pseudomonas aeruginosa* exotoxin A. In this class of toxins, the biological functions of target cell recognition and enzymatic activity are separated into distinct domains. In the case of CT and LT, cell recognition and binding are carried out by the B-pentamer. After the holotoxin binds to the cell membrane, the enzymatically active A subunit is directed to the interior of the target cell by a mechanism that as yet is not known in any detail. The cell membrane receptor for CT and LT is the pentasaccharide attached to the headgroup of ganglioside G_{M1} (Eidels et al., 1983).[3] The binding interaction is quite specific: cells lacking G_{M1} do not bind CT; conversely, the addition of G_{M1} can induce sensitivity in cells normally lacking the receptor (Eidels et al., 1983; Van Heyningen, 1983). The saccharide moiety of G_{M1} is bound by the complete AB_5 hexamer and also by the B-pentamer but not by monomeric B subunits (de Wolf et al., 1981). There are 5 binding sites on the toxin, and the bind-

Reprint requests to: Wim G.J. Hol, Department of Biological Structure SM-20, University of Washington, Seattle, Washington 98195; e-mail: hol@xray.bchem.washington.edu.

[3] Gangliosides referred to in this paper: G_{M1}, Gal(β1-3)GalNAc(β1-4){NeuAc(α2-3)} Gal(β1-4)Glc(β1-1)ceramide; G_{M2}, GalNac(β1-4){NeuAc(α2-3)} Gal(β1-4)Glc(β1-1)ceramide; G_{D1b}, Gal(β1-3)GalNAc(β1-4){NeuAc(α2-8)NeuAc(α2-3)} Gal(β1-4)Glc(β1-1)ceramide; G_{b3}, Gal(α1-4)Gal(β1-4)Glc(β1-1)ceramide.

ing of G_{M1} to the 5 sites is known to be cooperative (Schön & Freire, 1989). Ganglioside G_{M1} has been found to be selectively expressed in the surface cell membrane of a variety of cell types, particularly in nervous tissue. However, the physiological function of G_{M1} in the intestinal epithelial cell membrane, the site of attack by CT during cholera infection, is not known. Structural study of the cholera toxin may permit design of molecules that bind strongly or block access to either the receptor binding site on the B-pentamer or the catalytic site on the A subunit, and are thus potential anticholera drugs.

We report here the crystallographically determined structure of the cholera toxin B-pentamer bound to the G_{M1} pentasaccharide. This structure contains the receptor binding domain of the toxin and the specificity-determining portion of the corresponding cell-surface receptor; it therefore constitutes the first crystallographic analysis of the specific binding interactions between a protein and a receptor ganglioside.

Results

Structure of the cholera toxin B-pentamer

Crystallographic refinement of the CTB_5:G_{M1} pentasaccharide complex using data to 2.2 Å resolution yielded a standard residual $R = 0.171$ (Table 1). Good electron density is observed for the entire cholera toxin B-pentamer, with the exception of the C-terminal residue Arg 103 and the extra Met residue at the N-terminus. Because the 5-fold symmetry of the pentamer is not reflected in the crystallographic lattice, each monomer is in a different crystal packing environment. The 5 monomers are thus crystallographically distinct and will be referred to as subunits B#1 through B#5 in further discussion.

Each subunit consists of a short helix $\alpha 1$ at the N-terminus, a long helix $\alpha 2$, and two 3-stranded antiparallel β-sheets (note: residue assignments to secondary structural elements are $\alpha 1$ [5–9], $\beta 1$ [15–23], $\beta 2$ [26–30], $\beta 3$ [38–41], $\beta 4$ [47–50], $\alpha 2$ [61–78], $\beta 5$ [81–88], and $\beta 6$ [94–102]). In the pentamer, the β-sheets from adjacent monomers combine to form a continuous 6-stranded antiparallel sheet across each monomer–monomer interface. The 5 $\alpha 2$ helices of the pentamer surround a central pore whose diameter is roughly 12 Å. This is the same secondary structure observed for *E. coli* LT (Sixma et al., 1991, 1993a), which shares 80% sequence identity with the CT B subunit (Yamamoto & Yokota, 1983). In the LT holotoxin, the central pore is filled, in a remarkable mode of association, with the 18-residue carboxy-terminal tail of the A2 chain of the toxic A subunit (Sixma et al., 1991, 1993a), and a similar mode of association is seen in the AB_5 cholera toxin assembly (E. Westbrook, pers. comm.). Despite the absence of the A subunit, the CT B-pentamer in the present structure is very similar to the pentamer of the LT AB_5 holotoxin. The RMS difference for the backbone atoms in the CT and LT pentamers is only 0.55 Å. The largest differences ($\leq$1.6 Å) occur in the loop formed by residues 42–46 and may be due to sequence differences at positions Asn 44 (Ser 44 in LT) and Ala 46 (Glu 46 in LT). The position and conformation of the central $\alpha 2$ helices are very similar indeed; after least-squares superposition of the entire pentamer backbone, the C^{α} coordinates of the $\alpha 2$ helices differ by only 0.45 Å RMS and the coordinates for side chain atoms through $C\gamma$ differ by 0.51 Å RMS. Clearly the interactions with the A2 chain after holotoxin assembly have only marginal structural consequences for the B-pentamer conformation.

Table 1. *Crystallographic summary of CTB_5:G_{M1} pentasaccharide*

Space group	C2
Cell dimensions	
a	101.9 Å
b	67.58 Å
c	80.47 Å
β	105.69°
Residual	
For 21,985 reflections, 10–2.2 Å, $I/\sigma(I) > 1.0$ with no overall B correction	$R = 0.177$
With overall anisotropic B correction	$R = 0.171$
Model	
Atoms contributing to F_c	4,582
Water molecules in solvent model	284
Thermal parameters: Average B (Å^2)	
All protein atoms	17.7
284 Solvent molecules	33.2
Saccharide atoms	
B#1	38.3
B#2	29.5
B#3	43.6
B#4	23.6
B#5	21.9
Stereochemistry	
RMS deviation from ideal bond length	0.017 Å
RMS deviation from ideal bond angle	3.33°
RMS ΔB across bonds	2.9 Å^2
RMS ΔB across angles	4.4 Å^2

Interaction of the pentasaccharide with the toxin

Electron density for the G_{M1} pentasaccharide is visible in all 5 copies of the binding site, but the quality of the density varies. It was possible to unambiguously build models for the first 2 sugar residues, Gal(β1-3)GalNAc, in all 5 copies and for the sialic acid residue in 4 of the 5 copies. In subunits B#1 through B#4, electron density for the remainder of the saccharide, although present, is not of sufficient quality to build a complete model with confidence. In subunit B#5, however, the sugar conformation is stabilized by lattice interactions to a symmetry-related protein molecule, resulting in excellent electron density for the entire pentasaccharide (Fig. 1).

We have left the sugar model for 4 of the 5 copies of the binding site incomplete rather than impose a single conformation onto all copies of the pentasaccharide. Residual electron density for the binding site in subunits B#2, B#3, and B#4 is entirely consistent with the conformation fit to B#5, but is sufficiently weak to preclude independent fitting of the remaining sugar residues. In subunit B#1 alone, it appears possible that lattice interactions have distorted the sugar somewhat from the conformation seen in the model fit to B#5. Note that the hydrogen bonding shown in Table 2 for the partial sugar models should be considered less reliable than that given for the complete saccharide bound to B#5, as the incomplete models are not as stereochemically constrained during refinement. The remaining electron density at all sites is substantial enough to hold out the hope that higher resolution data, perhaps in combination with

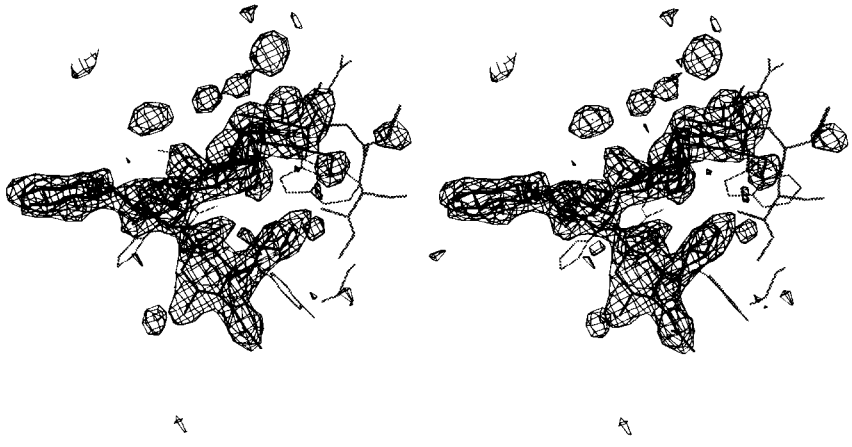

Fig. 1. Stereo pair showing electron density for G_{M1} pentasaccharide in subunit B#5. Contour level is 3σ in an $F_o - F_c$ map with sugar atoms and nearby solvent atoms omitted from F_c. Final saccharide model is shown in heavy lines; nearby protein residues are in shaded lines. The glucose residue at the far left of the figure would be the attachment point of the saccharide to the ceramide headgroup of the intact G_{M1} ganglioside.

data collection at low temperature, will eventually allow independent confirmation of the entire pentasaccharide conformation in all 5 copies of the binding site.

Both terminal sugars of the branched G_{M1} pentasaccharide, galactose and sialic acid, exhibit substantial specific binding interactions with the toxin (Kinemage 1). A smaller contribution to the binding interactions comes from the *N*-acetyl galactose residue. The remaining 2 sugars apparently interact only indirectly with the toxin, explaining the relatively poor electron density seen for these sugars in 4 of the 5 copies of the binding site. Approximately 350 Å^2 of the solvent-accessible surface of the pentasaccharide is buried upon binding. Of this area, roughly 43% is attributable to interactions of the sialic acid, 39% to interactions of the terminal galactose, and 17% to the *N*-acetyl galactosamine.

The G_{M1} terminal galactose residue interacts very extensively with the toxin. Sugar hydroxyl groups O2, O3, and O4 of this galactose accept hydrogen bonds directly from nitrogen donors

Table 2. *Potential hydrogen bonds involving the G_{M1} pentasaccharide*[a]

Donor/acceptor pair				B#1	B#2	B#3	B#4	B#5
Direct hydrogen bonds between CTB_5 and G_{M1} pentasaccharide								
Glu 11	O	Sia 108	O4		2.90		3.00	(3.27)
Glu 11	O	Sia 108	N5		2.99	2.91	2.95	3.07
His 13	N	Sia 108	O1B		2.87	2.90	2.85	2.81
His 13	ND1	Sia 108	O1B				3.01	(3.45)
Glu 51	OE1	Gal 104	O4	2.96	2.76	3.15	2.78	2.76
Gln 56	O	Gal 104	O6	3.11			3.10	(3.53)
Gln 61	OE1	Gal 104	O6	3.09	3.18	3.17		(3.39)
Asn 90	OD1	Gal 104	O3	2.70	2.66	2.79	2.83	2.78
Asn 90	ND2	Gal 104	O2	3.04	2.73	2.70	2.78	2.80
Asn 90	ND2	Gal 104	O3	2.74		3.00		(3.35)
Lys 91	NZ	Gal 104	O3	2.96	2.66	3.20	2.98	2.93
Lys 91	NZ	Gal 104	O4	2.73	3.13			2.99
Solvent-mediated hydrogen bonds between CTB_5 and G_{M1} pentasaccharide								
Solvent	248	Sia 108	O1A		2.79	2.66	2.72	2.88
Solvent	17	Gal 104	O2			2.83	2.66	
Solvent	207	Gal 104	O6		2.64	2.77	2.57	2.49
Solvent	207	Sia 108	O9		2.65	2.89	2.51	2.75
Solvent	206	Gal 104	O6		2.81	2.77	2.95	
Solvent	206	Sia 108	O8		2.80	3.03	2.69	2.80
Intrasaccharide hydrogen bonds								
Sia 108	O7	Sia 108	O10		2.77	2.98	2.81	2.77
GalNAc 105	N2	Sia 108	O1A		2.93	2.82	2.98	2.93
GalNAc 105	O6	Sia 108	O9			2.86	3.03	(3.41)
Sia 108	O6	Sia 108	O8			3.10	3.14	(3.26)

[a] All distances are given in Å. Only donor–acceptor distances less than 3.2 Å are considered potential hydrogen bonds, except that longer distances are listed in parentheses for monomer copy B#5 if the equivalent interatomic distance is less than 3.2 Å in another monomer copy. The potential intrasaccharide hydrogen bonds observed only in subunits B#1 through B#4 may be artifacts of the incomplete sugar model for those sites.

in the side chains of Asn 90 and Lys 91, and are additionally involved with direct and solvent-mediated hydrogen bonds to residues Asn 14 and Glu 51 (Fig. 2; Table 2). The participation of galactose hydroxyl O6 is somewhat ambiguous. The consensus position of the galactose residue seen in the 5 copies of the binding site suggests additional direct hydrogen bonding to the side chain of Gln 61 and to the carbonyl O of Gln 56, as seen in the LT:lactose complex (Sixma et al., 1992). However, in the single copy (B#5) of the site for which it was possible to model a complete sugar, these galactose:protein distances are too long for hydrogen bonding (Table 2). It is therefore possible that the consensus interactions for the galactose O6 are an artifact of the incomplete sugar model; clarification of this issue must await higher resolution structural study. The galactose sugar ring additionally makes extensive hydrophobic contact with Trp 88. These features of the galactose:toxin binding interaction are also

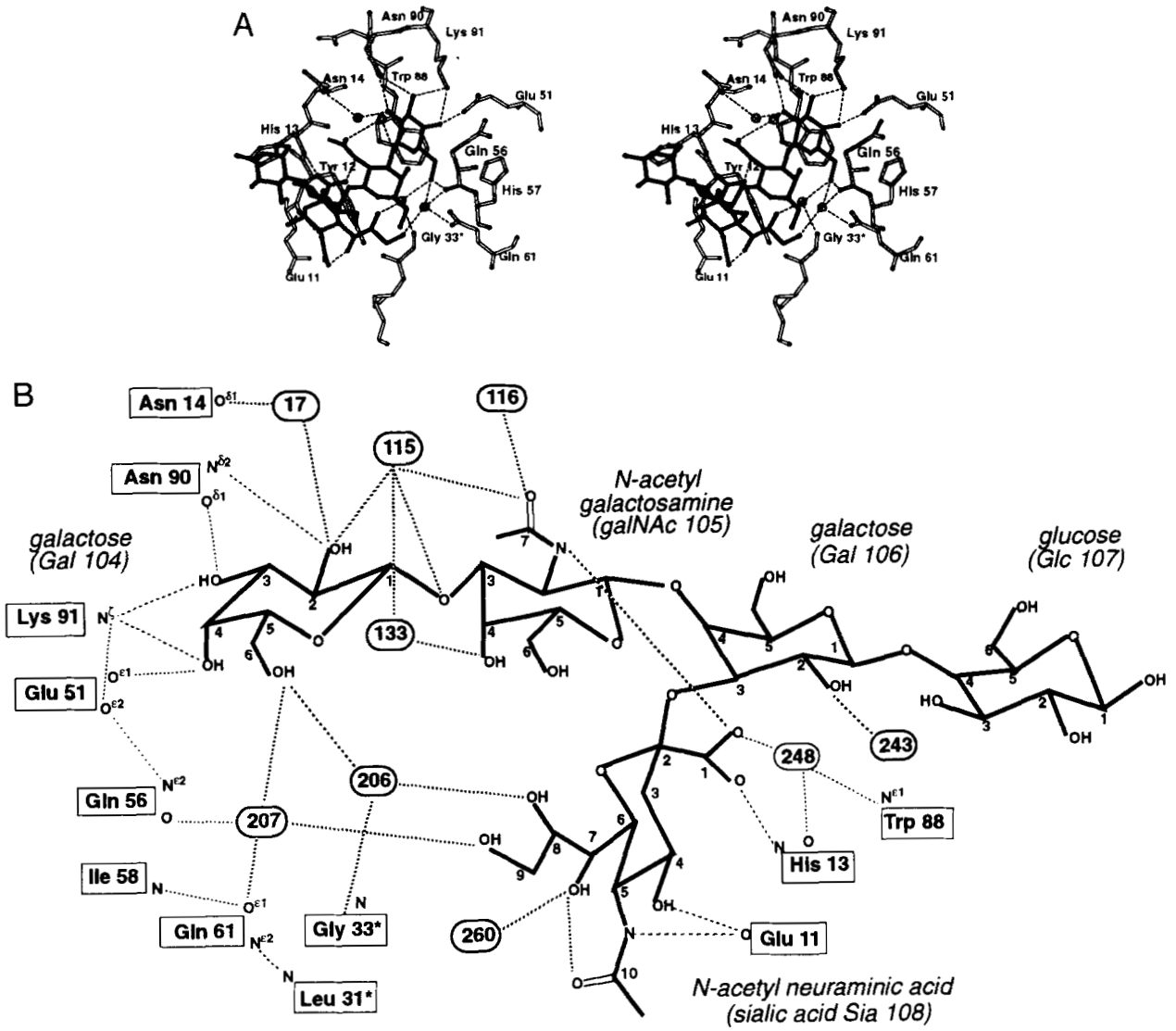

Fig. 2. G_{M1} pentasaccharide binding site. **A:** Stereo representation of the G_{M1} pentasaccharide binding site, showing both direct and solvent-mediated hydrogen bonding interactions between sugar and protein residues. The viewpoint is from "underneath" the bottom, membrane binding, surface of the B-pentamer. Solvent molecules that do not directly mediate sugar/protein interactions have been omitted for clarity. Starred residues are from an adjacent monomer. The terminal galactose residue of the pentasaccharide, at the upper right, is most deeply inserted into the binding site and is involved in the greatest number of identifiable binding interactions. The sialic acid residue, near the bottom of the figure, is also involved in a substantial number of hydrogen bonding interactions. Hydrophobic interactions include the approach of the sialic acid acetyl group to the edge of the Tyr 12 phenyl ring, and the positioning of the terminal galactose sugar ring in parallel to the indole ring of Trp 88. Figure generated using the program MOLSCRIPT (Kraulis, 1991). **B:** Schematic representation of hydrogen bonding interactions involving the G_{M1} pentasaccharide in subunit B#5 of the pentamer. The peptide residues shown belong to a single monomer, except for the involvement of Gly 33 from an adjacent monomer. Solvent molecules are depicted as ovals. Electron density at the other 4 copies of the pentamer is insufficient to model the full pentasaccharide. In subunit B#5, however, the conformation of the glucose end of the pentasaccharide is stabilized by hydrogen bonds (not shown) to a separate molecule related by the crystallographic symmetry operation $(x + 1/2, y + 1/2, z)$.

observed in the LT:lactose complex (Sixma et al., 1992) and are typical for protein:saccharide interactions (Quiocho, 1988; Vyas, 1991).

The G_{M1} sialic acid associates less intimately with the toxin than has been reported for other protein:sialic acid complexes (e.g., Weis et al., 1988; Wright, 1990; Varghese et al., 1992), making hydrophobic interactions with Tyr 12 and direct hydrogen bonds only with the backbone of residues Glu 11 and His 13. The sialic acid also makes at least 1 direct hydrogen bond with the *N*-acetyl galactosamine and 2 water-mediated interactions with the terminal galactose, as well as an internal hydrogen bond as discussed below. This intrasaccharide hydrogen bonding presumably constrains the conformational flexibility of the ganglioside even in the absence of toxin binding.

The *N*-acetyl galactosamine, a third specificity determinant, does not exhibit any direct or water-mediated hydrogen bonds with the protein although its methyl group is in contact with Cβ of His 13. Neither the central galactose nor the glucose residues of the sugar make any direct interaction with the protein.

Conformation of bound pentasaccharide

The φ/ψ torsion angles describing the sugar linkages in the saccharide are given in Table 3. All are in good agreement with the NOE distance constraints found in solution NMR studies (Acquotti et al., 1990), with the exception of the Gal(β1-4)Glc linkage, which has been reported to exhibit great flexibility in solution (Poppe et al., 1990). Such flexibility for the glucose is consistent with the lack of good electron density for 4 copies of this sugar in the present structure and probably facilitates the orientation of the remainder of the saccharide chain against the protein surface during binding. That is, the terminal 4 sugars are relatively free to move with respect to the glucose, which is attached to the membrane-bound ceramide group. The positions of the terminal galactose and sialic acid residues, however, are remarkably similar in the 5 monomers. The conformation of the sialic acid residue is stabilized by an internal hydrogen bond from O7 to O10, and the pentasaccharide as a whole is constrained by several intrasaccharide hydrogen bonds (Fig. 2; Table 2). The bound pentasaccharide is positioned so that the longest extent of the sugar chain lies diagonally away from the bottom surface of the toxin (Fig. 3). The O1 hydroxyl group of the glucose residue is thus furthest from the convoluted surface of the B-pentamer, which is entirely consistent with its being the point from which the ceramide lipid tail of the intact G_{M1} ganglioside is anchored in the cell membrane.

Receptor binding site

Each of the 5 identical G_{M1} binding sites lies primarily within a single monomer of the B-pentamer, in the cleft formed on one side by the loop (residues 51-58) connecting strand β_4 to the central helix and on the other by the 2 loops connecting α_1 to β_1 (residues 10-14) and β_5 to β_6 (residues 89-93). At one end of this cleft the β_2-β_3 loop (residues 31-36) from an adjacent monomer forms the remainder of the binding site, although these residues do not interact directly with the bound oligosaccharide.

The location of the galactose moiety of the bound G_{M1} saccharide coincides very closely with the galactose binding site observed in the LT:lactose and LT:galactose complexes (Sixma et al., 1992; Merritt et al., submitted for publ.). The set of CT B-pentamer C^{α} carbons within 10 Å of the galactose binding site superimpose onto the corresponding C^{α} atoms in the LT:galactose complex to 0.41 Å RMS and onto the LT:lactose atoms to 0.36 Å RMS. Given this superposition of the pentamers, the atoms of the terminal galactose in G_{M1} superimpose to within 0.52 Å RMS onto the galactose atoms in the LT:galactose complex, and to within 0.39 Å RMS onto the galactose moiety of the LT:lactose complex. Clearly the terminal galactose has a very well-defined binding mode in both CT and LT.

Discussion

Involvement of adjacent monomer in sugar binding

The only contribution to the binding interactions from the neighboring subunit visible in the X-ray structure is the solvent-mediated hydrogen bond from the backbone N of Gly 33 to O6 of the terminal galactose. This same interaction was previously observed in the structures of LT complexed with galactose (Mer-

Table 3. *Sugar conformation in G_{M1} pentasaccharide*[a]

Sugar linkage	Angle	B#1	B#2	B#3	B#4	B#5	NMR model
Gal(β1-3)GalNAc	φ	45.2°	58.4°	58.9°	59.9°	57.7°	25 to 30°
	ψ	5.2°	−4.1°	14.1°	−6.2°	3.0°	30 to −40°
GalNAc(β1-4)Gal	φ					48.2°	30°
	ψ					6.8°	25°
Gal(β1-4)Glc	φ					−48.7°	Multiple
	ψ					−6.7°	Multiple
Sia(α2-3)Gal	φ					−168.5°	−165°
	ψ					−30.8°	−18°
Sia	θ_1		−48.7°	−52.5°	−49.7°	−54.2°	−60°
	θ_2		−178.9°	−160.2°	−173.7°	−178.3°	−160°

[a] Torsion angles are defined as for NMR studies, e.g., φ for the (β1-3) sugar linkage is the dihedral angle defined by H1-C1-O3-C3, and ψ by C1-O3-C3-H3; torsion angles of the glycerol moiety of sialic acid are defined as θ_1 = H6-C6-C7-H7 and θ_2 = H7-C7-C8-H8. Except for the Gal(β1-4)Glc linkage, which has been reported to exhibit multiple conformations in solution (Poppe et al., 1990), the X-ray structure of the bound saccharide agrees well with NMR models for the structure of the free saccharide (Acquotti et al., 1990).

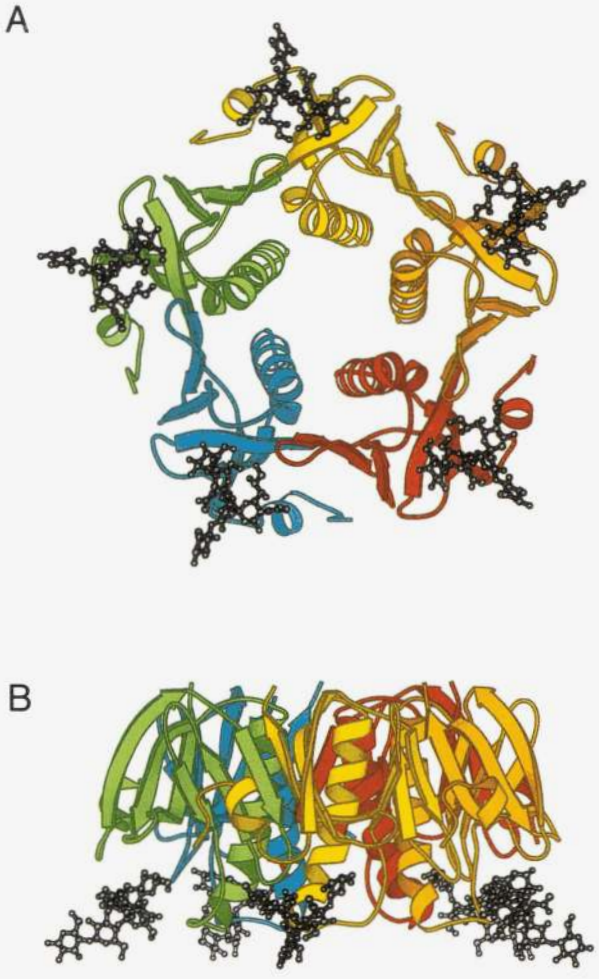

Fig. 3. Cholera toxin B subunit pentamer (CTB_5) with G_{M1} pentasaccharide bound. To generate this figure, the complete pentasaccharide model from subunit B#5 has been duplicated 5 times and superimposed onto the partial saccharide model built in each of the other 4 copies. The sugar binding site is primarily formed by the cradle of loops connecting secondary structural elements within a single monomer (seen most easily in the green monomer), with a secondary interaction contributed by residues 34–36 at the very tip of a loop on an adjoining monomer (e.g., the blue copy). The 5 α2 helices line a pore through the center of the B-pentamer into which the C-terminal segment of the A subunit anchors. **A:** View from the cell membrane toward the binding surface. The distance from the 5-fold axis to the glucose O1 atoms, the attachment site to the remainder of the G_{M1} molecule, is 31 Å. **B:** View perpendicular to the 5-fold axis, with the binding surface downward.

ritt et al., submitted for publ.) and lactose (Sixma et al., 1992). Gly 33 has previously been implicated as being important to the binding affinity of CT for G_{M1}, but the reason for this importance bears some examination. Jobling and Holmes (1991) characterized the binding affinity of 11 mutants of CT each with a single substitution at Gly 33. Negatively charged and some hydrophobic substitutions (Glu, Asp, Ile, Val, Leu) markedly reduced the binding affinity, whereas small or positively charged substitutions (Ala, Lys, Arg) left the binding affinity unchanged. Various substitutions at the adjacent residues Lys 34 and Arg 35 did not significantly affect binding affinity. The observed structure is somewhat surprising, therefore, in that one would predict that such a solvent-mediated interaction from a backbone amide would be relatively insensitive to amino acid substitutions at that site. The φ/ψ angles observed for Gly 33 ($-67° < \varphi < -53°$; $125° < \psi < 138°$) in all 5 copies are well within the allowed region for all amino acids, so considered by itself the substitution of a C^{β}-containing amino acid for Gly 33 should be allowed sterically (Fig. 4). Residues 33–35, however, comprise a sharp bend at the tip of a loop between 2 strands of β-sheet. Perhaps because of this, Lys 34 is notable for lying in a normally unfavorable conformation at ($\varphi = 60°$, $\psi = 0°$), and Arg 35 is also observed to adopt a mildly unusual conformation at ($\varphi = -140°$, $\psi = 50°$). The glycine at position 33 may serve to minimize the energy penalty associated with the atypical conformation of the following 2 residues. It may be that the sensitivity of binding affinity to substitutions at Gly 33 reported by Jobling and Holmes (1991) is actually due to disruption of the secondary structure of this loop rather than to local perturbation of the hydrogen bonding at the sugar binding site. We are pursuing structure determination of point mutants at residue 33, which may clarify the role played by the 33–35 bend in forming the receptor binding site.

Structural basis of binding specificity

In addition to the strong binding affinity for ganglioside G_{M1}, both CT and the human-derived isolate of LT (hLT-I) bind weakly to ganglioside G_{D1b} (Fukuta et al., 1988). However, LT, unlike CT, also binds weakly to G_{M2} and asialo-G_{M1} (Fukuta et al., 1988). G_{D1b} differs from G_{M1} by the addition of a second sialic acid residue to O8 of the G_{M1} sialic acid; G_{M2} differs from G_{M1} by the absence of the terminal galactose residue. That is to say, both CT and LT can tolerate the addition of a second sialic acid residue; LT but not CT can tolerate the removal of either the sialic acid or the terminal galactose. Consideration of these observations together with analysis of the binding interactions seen in the present structure may afford some insight into the structural determinants of receptor binding specificity in these toxins.

The toxin residues involved in direct or solvent-mediated hydrogen bonding to the receptor are almost entirely conserved between CT and LT. The 1 exception is residue 13, which is a histidine in CT but may be either histidine or arginine in LT, depending on the strain of *E. coli* from which the toxin is isolated. Sequence variation at other sites more remote from the receptor binding site may, of course, induce structural differences that propagate to induce minor conformational changes at the binding site. As a first approach, however, we may attempt to rationalize the greater tolerance of LT for alternate receptors in terms of sequence variation at His/Arg 13.

The sequence at residue 13 in LT varies depending on the strain of *E. coli*. Yamamoto and Yokota (1983) found the codon for this residue to be CGC (Arg) in *E. coli* H10407, an isolate from a strain in Bangladesh infectious for humans. This matches the sequence reported for the porcine isolate P307 from Great Britain (Dallas, 1983) and corresponds to the sequence of our crystallographically determined pLT-I structures. Human infectious strain 240-3 also has Arg 13 (Tsuji et al., 1987). However, Leong et al. (1985) found this codon to be CAC (His) in *E. coli* strain H74-114, which is infectious for humans and was isolated in the United States. This matches the residue His 13

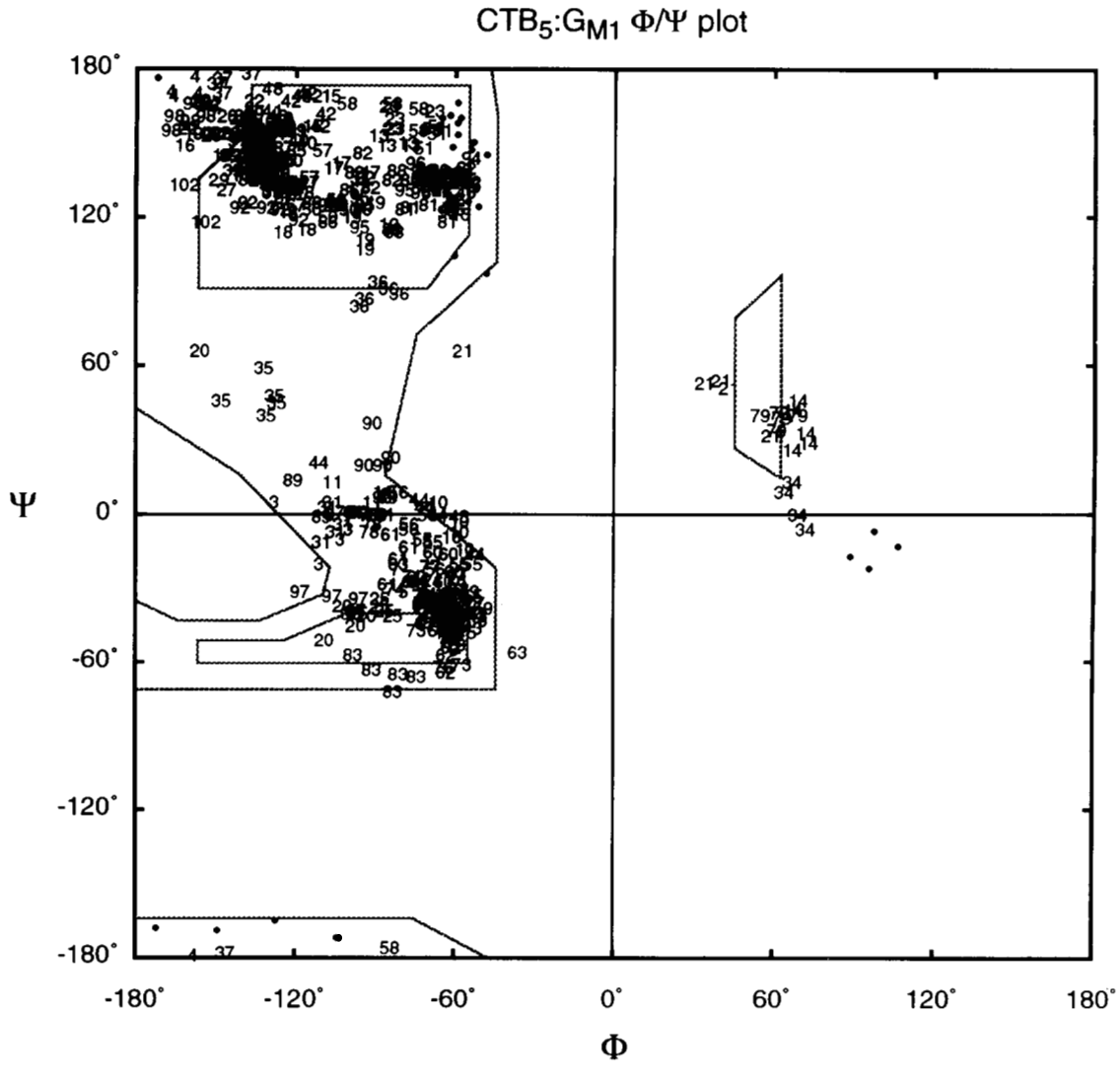

Fig. 4. φ/ψ angles for B-pentamer in CTB_5:G_{M1} pentasaccharide complex. Backbone torsion angles are shown for the 5 chains forming the cholera toxin B-pentamer. Glycine and proline residues are indicated by •; other residues are indicated by sequence number. Note particularly that, as discussed in the text, the loop comprised of residues 33–35 exhibits somewhat unusual backbone torsion angles. The backbone carbonyl O of residue 33 is involved in receptor binding, and point mutants for sites in this loop exhibit altered receptor binding activity.

found at this position in CT and hence in the present structure. Leong et al. (1985) proposed that the LT variants are evolutionarily related, with a common ancestral sequence most similar to that of strain H10407. This strain-specific sequence variability at residue 13 may be crucial to analysis of the reported differences in saccharide binding specificity between CT and LT. It is unfortunate, therefore, that it is not always possible to determine from reports in the literature which specific LT sequence (or even which strain of *E. coli*) is represented by the toxin used for binding assays. It seems imperative that in the future, reports of binding constants for these toxins should state unambiguously the bacterial strain from which the toxin was derived and give, if possible, a reference to the corresponding gene or amino acid sequence.

We first consider the involvement of the residue 13 backbone atoms, whose conformation should be largely sequence-independent. The carbonyl O of Arg 13 in the LT:lactose complex (Sixma et al., 1992) was observed to interact through a solvent-mediated hydrogen bond with O2 of the terminal galactose. The same solvent molecule also mediated an interaction from the side chain of Asn 14. In the present complex, the Asn 14···solvent···O2(Gal 104) hydrogen bonding chain remains, but the His 13 carbonyl O now interacts via a bridging solvent molecule to $N^{\epsilon 1}$ of Trp 88 (Fig. 2B). This pair of solvent-mediated hydrogen bonds seen in the present complex is also observed in the LT:galactose complex (Merritt et al., submitted for publ.), which may indicate that this aspect of the geometry observed in the LT:lactose complex was an artifact induced by the "incorrect" second sugar present in lactose. The present structure is our first opportunity to observe interactions with the remaining sugars in the G_{M1} pentasaccharide, and we now observe that His 13 also donates a direct hydrogen bond

from the backbone amide nitrogen to O1 of the G_{M1} sialic acid residue. Further, we see that C^{β} of His 13 is in hydrophobic contact with the G_{M1} GalNAc residue. None of the above interactions would be expected to distinguish between His or Arg at this position.

It is particularly interesting to note, however, that although the remainder of the His 13 side chain does not contribute to the hydrogen bonding network at the sugar binding site in CT:G_{M1} as seen in the present structure, the corresponding Arg 13 side chain in the LT structures is in an ideal position to hydrogen bond to the proximal 3 sugar residues, GalNAc(β1–4)Gal(β1–4)Glc, of ganglioside G_{M1} (Kinemage 2). This suggests 1 possible explanation for reports that LT will bind to G_{M2} and asialo-G_{M1} whereas CT will not. In the CT:G_{M1} structure, essentially all of the binding energy is due to interactions with Gal 104 and Sia 108. It is thus quite plausible that drastically lower binding affinity would result from truncation of either branch of the G_{M1} saccharide, i.e., removal of either the galactose, to yield G_{M2}, or the sialic acid, to yield asialo-G_{M1}. We can hypothesize that LT containing Arg 13, on the other hand, potentially forms additional hydrogen bonds from the Arg 13 side chain to O7 of GalNAc 105, to O2 of Gal 106, and/or to O6 of Glc 107 (Kinemage 2). The binding energy from these additional hydrogen bonds to proximal sugar residues on the ganglioside would offset the loss of hydrogen bonding due to truncation of the terminal sugar on either branch of the sugar chain. Thus, the known sequence variability of LT at residue 13 may, in fact, explain the differing cross-reactivity of CT and LT for gangliosides other than G_{M1}.

Cell-surface binding of AB_5 toxins

The association of G_{M1} to cholera toxin may be thought of as a 2-fingered grip: the Gal(β1–3)GalNAc "forefinger" is inserted in a deep pocket of the binding site, while the sialic acid "thumb" occupies a shallower depression on the surface of the toxin. This conformation places the ceramide lipid tail (attached at the glucose O1 in the intact G_{M1} ganglioside) far from the central pore, on the "convoluted" surface of the B-pentamer (Fig. 2). Combining this result with the fact that the A subunit is located on the opposite, "flat," surface of the pentamer (Sixma et al., 1991, 1993a), it is now definitely established that LT and CT bind to cells with the A subunit pointing away from the membrane (Fig. 5). Note that the G_{M1} binding mode seen here leaves the carboxy-terminal KDEL sequence of the CT A2-chain (an endoplasmic reticulum retention signal) close to the membrane surface, which suggests a possible role for these residues in the membrane translocation process. This still leaves as a marvelous puzzle precisely how the A1 fragment eventually crosses the membrane and exerts its toxic function inside the cell.

Recently the 3-dimensional structure of the B-pentamer of verotoxin, also called shiga-like toxin, from *E. coli* was determined. It was found to be strikingly similar in structure to the LT B-pentamer, in spite of virtually no sequence homology between the two (Stein et al., 1992; Sixma et al., 1993b). The comparison of the known lactose binding site of LT with the verotoxin structure appeared to show that the galactose binding site of LT is entirely absent in verotoxin. Knowledge of the sialic acid binding site in CT and LT (Fig. 2) now allows us to

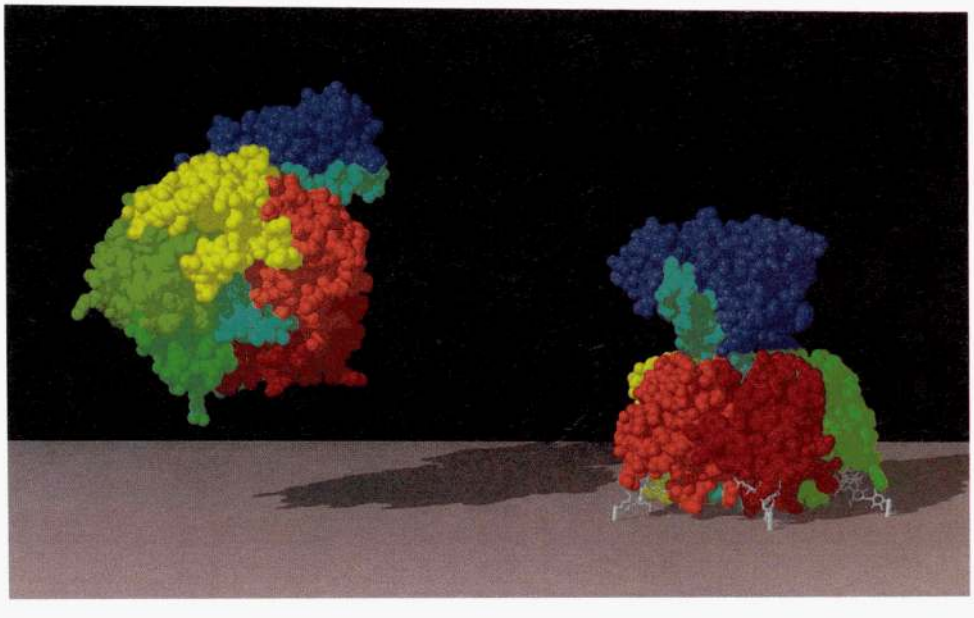

Fig. 5. Model of cholera toxin binding to 5 G_{M1} gangliosides on cell surface. The enzymatically active fragment A1 is shown in purple, the linker fragment A2 in blue, the G_{M1} pentasaccharides in white, the B subunits in other colors. The model shown for the AB_5 holotoxin is obtained by combining the cholera toxin B-pentamer:G_{M1} complex depicted in Figure 2 with the enzymatically active A subunit from the homologous (80% identical) LT holotoxin (Sixma et al., 1991, 1993a). In the toxin approaching the membrane surface (left), the blue A2 linker domain can be seen to extend all the way through the central pore to the convoluted side of the pentamer. After binding (right), the carboxy-terminal residues (KDEL) of A2 are in close proximity to the membrane surface, where they may be functionally important in membrane translocation of the entire A subunit. The A1 fragment points away from the membrane.

note that residues 10–14, most deeply involved in sialic acid binding, have no counterpart in verotoxin. The receptor for verotoxin is ganglioside G_{b3}, whose trisaccharide (Gal(α1–4) Gal(β1–4)Glc) is equivalent to the 3 sugars *not* involved in substantial binding interactions in the present CT:G_{M1} structure. Nevertheless, the lack of structurally equivalent residues at the receptor binding sites of the 2 toxins is intriguing in view of the remarkable similarity of the remainder of the respective B-pentamers.

The structure determination of the CTB_5:G_{M1} pentasaccharide complex described in this paper provides important evidence regarding the position of the A subunit with respect to the target cell membrane after receptor binding in the AB_5 bacterial toxins. It seems likely that the same mode of binding, with the catalytic A subunit initially on the opposite side of the receptor-bound B-pentamer from the cell membrane, is shared by even those AB_5 toxins with limited homology to cholera toxin.

Implications of observed toxin structure

The B-pentamers of both CT and LT present inviting targets for protein engineering. They are particularly attractive as components for the construction of orally or nasally administered vaccines because of their potent activity as immune adjuvants, their ability to induce an exceptional mucosal immune response, and their abrogation of tolerance induction (Elson & Ealding, 1984a, 1984b; Clements et al., 1988; De Aizpurua & Russell-Jones, 1988). Several groups have reported that high adjuvant efficacy is retained by the nontoxic B-pentamer alone (e.g., Chen & Strober, 1990; Wu & Russell, 1993), although Lycke et al. (1992) claimed a requirement for the presence of the catalytically active A subunit. Linking the antigen to the toxin molecule can noticeably potentiate its adjuvant effect (e.g., Liang et al., 1988). This set of properties suggests a design scheme for oral vaccines against a wide variety of diseases: that of a chimeric protein combining the B-pentamer with a modified A subunit carrying epitopes characteristic of the target disease. Alternatively, selected epitopes may be engineered onto the surface of the B-pentamer itself (Czerkinsky et al., 1989; Dertzbaugh et al., 1990; Schödel et al., 1990). The present structure will serve as a framework for further refinement of the design of such engineered constructs. In particular, knowledge of the complete G_{M1} binding site is of the greatest importance in designing CT- and LT-based vaccines that should leave the G_{M1} binding affinity intact.

The structure we report here also provides the first detailed view of the molecular interactions underlying specific recognition of a ganglioside by a protein. As such, it may help to identify binding regions in other proteins that act through specific recognition of ganglioside receptors. Although the normal function of the gangliosides is so far poorly understood, they have recently been implicated in various signal transduction pathways. Exogenous G_{M1} affects Ca^{2+} signaling (e.g., Hilbush & Levine, 1992), modulation of CD4 expression (Morrison et al., 1990, 1991), and modulation of the tyrosine protein kinase activity of epidermal growth factor (Weis & Davis, 1990). The cholera toxin B subunit itself is known to interfere with or potentiate the effect of several growth factors (see, for example, Spiegel, 1989). The pentamer:pentasaccharide structure revealed by this study is thus important not only for our understanding of the action of cholera toxin as a disease agent, and as a starting point for vaccine design, but also for probing the signal transduction effects of CTB_5:G_{M1} interactions.

Materials and methods

Recombinant CTB corresponding in amino acid sequence to that of *V. cholerae* Ogawa 41 (classical biotype) was expressed in *E. coli* as described by L'hoir et al. (1990). The recombinant protein contains an additional Met residue at the N-terminus as compared to the natural mature CT B subunit. Purified protein was initially concentrated to 4 mg/mL in HEPES buffer containing 1 mM pentasaccharide at pH 7.5. Crystals of the CT B-pentamer complexed with the G_{M1} pentasaccharide were grown in sitting drop trays from a starting drop containing 2.5 μL protein solution and 1.25 μL well solution (100 mM cacodylate, 200 mM $MgCl_2$, 15% PEG 1000, pH 5.0). Fresh crystals of the complex of CTB_5 with the G_{M1} pentasaccharide diffract to better than 1.8 Å resolution; crystallographic data to 2.2 Å resolution were measured for the work presented here (Table 1). Data were collected from a single crystal using a Siemens X100 detector mounted on a Rigaku RU200 anode. Observations (72,029) were integrated using the program XENGEN and scaled using the local program MACRO to yield 26,010 unique reflections from ∞ to 2.2 Å with an overall merging $R_I = 0.064$. The initial structural model was obtained by molecular replacement using as a probe the B-pentamer of the homologous *E. coli* heat-labile enterotoxin (LT). Probe coordinates were taken from the crystallographically determined structure the LT:lactose complex after truncation of all residues known to differ between the LT and CT sequences. Molecular replacement searches as well as all subsequent refinement were performed using X-PLOR v.2.1 (Brünger, 1990). Positional and thermal refinement of the transformed probe coordinates brought R from 0.43 to 0.28. All protein atoms missing from this initial model were hand-built into a $(2F_o - F_c)$ electron density map at this point, and the resulting coordinates were subjected to simulated annealing using a slow-cooling protocol (Brünger, 1990). Sugar residues were built into $(F_o - F_c)$ density maps and included in subsequent refinement of positional and thermal parameters. A model for the discretely ordered solvent atoms was constructed by iterative selection and refinement of peaks greater than 3σ from $(F_o - F_c)$ density maps that exhibited plausible hydrogen bonding geometry. Potential solvent peaks near the saccharide binding sites were not accepted until all sugar residues that could be built unambiguously into electron density had been placed and refined.

Acknowledgments

We acknowledge support from special funds from the University of Washington School of Medicine, from the NIH (award GM50386), from the Belgian Region Wallonne, and from the Medical Research Fund of the Faculty of Medicine of the University of Liège. It is a pleasure to thank Stewart Turley for assistance in data collection and to thank Ingeborg Feil and Christophe Verlinde for stimulating discussions.

References

Acquotti D, Poppe L, Dabrowski J, von der Lieth CW, Sonnino S, Tettamanti G. 1990. Three-dimensional structure of the oligosaccharide chain of GM1 ganglioside revealed by a distance-mapping procedure: A rotating and laboratory frame nuclear Overhauser enhancement investigation of native glycolipid in dimethyl sulfoxide and in water-dodecylphosphocholine solutions. *J Am Chem Soc 112*:7772–7778.

Brünger A. 1990. *X-PLOR, version 2.1. Manual.* New Haven, Connecticut: Yale University.

Cassel D, Pfeuffer T. 1978. Mechanism of cholera toxin action: Covalent modification of the guanyl nucleotide-binding protein of the adenylate cyclase system. *Proc Natl Acad Sci USA 75*:2669-2673.

Chen KS, Strober W. 1990. Cholera holotoxin and its B subunit enhance Peyer's patch B cell responses induced by orally administered influenza virus: Disproportionate cholera toxin enhancement of the IgA B cell response. *Eur J Immunol 20*:433-436.

Clements JD, Hartzog NM, Lyon FL. 1988. Adjuvant activity of *Escherichia coli* heat-labile enterotoxin and effect on the induction of oral tolerance in mice to unrelated protein antigens. *Vaccine 6*:269-277.

Czerkinsky C, Russell MW, Lycke N, Lindblad M, Holmgren J. 1989. Oral administration of a streptococcal antigen coupled to cholera toxin B subunit evokes strong antibody responses in salivary glands and extramucosal tissues. *Infect Immun 57*:1072-1077.

Dallas WS. 1983. Conformity between heat-labile toxin genes from human and porcine enterotoxigenic *Escherichia coli*. *Infect Immun 40*:647-652.

De Aizpurua HJ, Russell-Jones GJ. 1988. Oral vaccination. Identification of classes of proteins that provoke an immune response upon oral feeding. *J Exp Med 167*:440-451.

Dertzbaugh MT, Peterson DL, Macrina FL. 1990. Cholera toxin B-subunit gene fusion: Structural and functional analysis of the chimeric protein. *Infect Immun 58*:70-79.

De Wolf MJS, Fridkin M, Kohn LD. 1981. Tryptophan residues of cholera toxin and its A and B protomers; intrinsic fluorescence and solute quenching upon interacting with the ganglioside G_{M1}, oligo-G_{M1} or dansylated oligo-G_{M1}. *J Biol Chem 256*:5489-5496.

Eidels L, Proia RL, Hart DA. 1983. Membrane receptors for bacterial toxins. *Microbiol Rev 47*:596-620.

Elson CO, Ealding W. 1984a. Generalized systemic and mucosal immunity in mice after mucosal stimulation with cholera toxin. *J Immunol 132*: 2736-2741.

Elson CO, Ealding W. 1984b. Cholera toxin feeding did not induce oral tolerance in mice and abrogated oral tolerance to an unrelated protein antigen. *J Immunol 133*:2892-2897.

Fukuta S, Magnani JL, Twiddy EM, Holmes RK, Ginsburg V. 1988. Comparison of the carbohydrate-binding specificities of cholera toxin and *Escherichia coli* heat-labile enterotoxins LTh-I, LT-IIa and LT-IIb. *Infect Immun 56*:1748-1753.

Hilbush BS, Levine JM. 1992. Modulation of a Ca^{2+} signaling pathway by G_{M1} ganglioside in PC12 cells. *J Biol Chem 267*:24789-24795.

Jobling MG, Holmes RK. 1991. Analysis of structure and function of the B subunit of cholera toxin by the use of site-directed mutagenesis. *Mol Microbiol 5*:1755-1767.

Kraulis P. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of proteins. *J Appl Crystallogr 24*:946-950.

Leong J, Vinal AC, Dallas WS. 1985. Nucleotide sequence comparison between heat-labile toxin B-subunit cistrons from *Escherichia coli* of human and porcine origin. *Infect Immun 48*:73-77.

L'hoir C, Renard A, Martial JA. 1990. Expression in *Escherichia coli* of two mutated genes encoding the cholera toxin B subunit. *Gene 89*:47-52.

Liang X, Lamm ME, Nedrud JG. 1988. Oral administration of cholera toxin-Sendai virus conjugate potentiates gut and respiratory immunity against Sendai virus. *J Immunol 141*:1495-1501.

Lycke N, Tsuji T, Holmgren J. 1992. The adjuvant effect of *Vibrio cholerae* and *Escherichia coli* heat-labile enterotoxin is linked to their ADP-ribosyltransferase activity. *Eur J Immunol 22*:2277-2281.

Morrison WJ, Offner J, Vandenbark AA. 1990. Transmembrane signalling associated with ganglioside-induced CD4 modulation. *Immunopharmacology 20*:135-141.

Morrison WJ, Offner J, Vandenbark AA. 1991. Ganglioside (GM_1)-treated T cells shed CD4. *Immunopharmacology 22*:77-84.

Poppe L, von der Lieth CW, Dabrowski J. 1990. Conformation of the glycolipid globoside head group in various solvents and in the micelle-bound state. *J Am Chem Soc 112*:7762-7771.

Quiocho FA. 1988. Molecular features and basic understanding of protein-carbohydrate interactions: The arabinose-binding protein-sugar complex. *Curr Topics Microbiol Immunol 139*:135-148.

Schödel F, Enders G, Jung MC, Will H. 1990. Recognition of a hepatitis B virus nucleocapsid T-cell epitope expressed as a fusion protein with the subunit B of *Escherichia coli* heat labile enterotoxin in attenuated salmonellae. *Vaccine 8*:569-572.

Schön A, Freire E. 1989. Thermodynamics of intersubunit interactions in cholera toxin upon binding to the oligosaccharide portion of its cell surface receptor, ganglioside G_{M1}. *Biochemistry 28*:5019-5024.

Sixma TK, Kalk KH, van Zanten BAM, Dauter Z, Kingma J, Witholt B, Hol WGJ. 1993a. Refined structure of *Escherichia coli* heat-labile enterotoxin, a close relative of cholera toxin. *J Mol Biol 230*:890-918.

Sixma TK, Pronk SE, Kalk KH, van Zanten BAM, Berghuis AM, Hol WGJ. 1992. Lactose binding to heat-labile enterotoxin revealed by X-ray crystallography. *Nature 355*:561-564.

Sixma TK, Pronk SE, Kalk KH, Wartna ES, van Zanten BAM, Witholt B, Hol WGJ. 1991. Crystal structure of a cholera toxin-related heat-labile enterotoxin from *E. coli*. *Nature 351*:371-378.

Sixma TK, Stein PE, Hol WGJ, Read RJ. 1993b. Comparison of the B-pentamers of heat-labile enterotoxin and verotoxin-1: Two structures with remarkable similarity and dissimilarity. *Biochemistry 32*:191-198.

Spiegel S. 1989. Inhibition of protein kinase-C-dependent cellular proliferation by interaction of endogenous ganglioside G_{M1} with the B subunit of cholera toxin. *J Biol Chem 264*:16512-16517.

Stein PE, Boodhoo A, Tyrrell GJ, Brunton JL, Read RJ. 1992. Crystal structure of the cell-binding B oligomer of verotoxin-1 from *E. coli*. *Nature 355*:748-750.

Tsuji T, Iida T, Honda T, Miwatani T, Nagahama M, Sakurai J, Matsubara H. 1987. A unique amino acid sequence of the B subunit of a heat-labile enterotoxin isolated from a human enterotoxigenic *Escherichia coli*. *Microb Pathol 2*:381-390.

Van Heyningen S. 1983. The interaction of cholera toxin with gangliosides and the cell membrane. *Curr Topics Membranes Transport 18*:445-471.

Varghese JN, McKimm-Breschkin JL, Caldwell JB, Kortt AA, Colman PM. 1992. The structure of the complex between influenza virus neuraminidase and sialic acid, the viral receptor. *Proteins 14*:327-332.

Vyas NK. 1991. Atomic features of protein-carbohydrate interactions. *Curr Opin Struct Biol 1*:732-740.

Weis FMB, Davis RJ. 1990. Regulation of epidermal growth factor receptor signal transduction. *J Biol Chem 265*:12059-12066.

Weis W, Brown JH, Cusack S, Paulson JC, Skehel JJ, Wiley DC. 1988. Structure of the influenza virus haemagglutinin complexed with its receptor, sialic acid. *Nature 333*:426-431.

Wright CS. 1990. 2.2 Å resolution structure analysis of two refined N-acetylneuraminyl-lactose-wheat germ agglutinin isolectin complexes. *J Mol Biol 215*:635-651.

Wu HY, Russell MW. 1993. Induction of mucosal immunity by intranasal application of a streptococcal surface protein antigen with the cholera toxin B subunit. *Infect Immun 61*:314-322.

Yamamoto T, Yokota T. 1983. Sequence of heat-labile enterotoxin of *Escherichia coli* pathogenic for humans. *J Bacteriol 155*:728-733.