Simulation of Protein−Sugar Interactions: A Computational Model of the Complex between Ganglioside GM1 and the Heat-Labile Enterotoxin of *Escherichia coli*†

Anna Bernardi,* Laura Raimondi, and Fabio Zuccotto

Dipartimento di Chimica Organica e Industriale, Universita' di Milano, and Centro CNR per lo Studio delle Sostanze Organiche Naturali, via Venezian 21, 20133 Milano, Italy

Received September 30, 1996⊗

The cholera toxin from *Vibrio cholerae* (CT) and the 80% homologous heat-labile toxin of *Escherichia coli* (LT) are two well-known cases of sugar-binding proteins. The GM1:toxin complexes were chosen as test cases for the elaboration of a computational approach to the modeling of protein−saccharide interactions. The reliability of the method was evaluated on the LT:lactose complex. A model of this complex was built by performing a MC/EM conformational search of the sugar moiety within the binding pocket of LT, using the AMBER* force field and the GB/SA solvation model. The results are a reasonable reproduction of the reported X-ray structure of the complex. The same protocol was then applied to the LT:GM1 complex. The calculations were performed on a substructure that includes the protein shell within 5 Å from GM1, three water molecules solvating Glu-51 carboxylate, and two water molecules at crystallographic sites 2 and 3. A satisfactory agreement was found with the recently published X-ray structure of the CT:GM1 complex. All the relevant interactions between the sugar and the residues involved in binding are well reproduced by the calculations. These results suggest that the substructure here identified can be taken as a realistic representation of the toxin binding surface and that the method presented in this paper can be used as a predictive tool in designing artificial LT (CT) binders and thus potential anticholera drugs.

Introduction

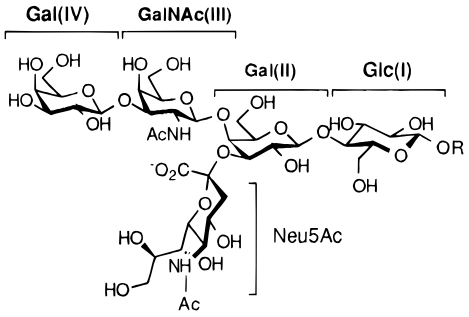
Cholera and related enterotoxigenic diseases are caused by secreted toxins. The cholera toxin from *Vibrio cholerae* (CT) and the closely related heat-labile toxin of *Escherichia coli* (LT) exhibit a 80% homologous structure, consisting of an AB_5 hexamer. Actual cell intoxication is carried out by a catalytic fragment of the A subunit, which reaches the interior of the target cell after the B_5 pentamer has docked to the cell membrane. The specific membrane receptor of both CT and LT is the pentasaccharide portion of ganglioside GM1 [Galβ1−3GalNAcβ1−4(NeuAcα2−3)Galβ1−4Glcβ1−1Cer], **1a**. The interaction of GM1 with CT and LT has been studied in great detail. The dissociation constant of the CT:GM1 complex is ca. 10^{-7}.[1] It has been proven that up to five GM1 molecules can bind to the B_5 pentamer. The binding data suggest that the presence of at least two B subunits is required for effective complexation of GM1.[2] This hypothesis was later supported by the X-ray structures of the toxins and various toxin−sugar complexes:[3] the sugar-binding site appears to be mostly formed by a single B unit, but the cavity is closed by a sharp loop centered on the Gly-33 residue of the adjacent B(+1) unit. Chemical modification and single-point mutation studies of the B subunits have identified the amino acids which are essential for GM1 binding.[4] More recently, a crystallographic study by Hol and co-workers on both the *V. cholerae* and *E. coli* toxins has led to the structures of unbound LT,[5] the LT:lactose[6] and LT:galactose[7] complexes, and the CT-B_5 pentamer bound to GM1.[8] This last structure shows that the large majority of interactions between the receptor and the toxin involves the two terminal sugars of GM1, galactose (Gal) and sialic acid (NeuAc), with a limited contribution from the *N*-acetylgalactosamine residue. This is in agreement with the biochemical data which indicate in Gal-IV and NeuAc the essential features for binding.[9] The structure of bound GM1 is very similar to the solution structure determined by NMR spectroscopy[10] and recently reproduced by molecular mechanics calculations.[11]

The binding of GM1 to the B_5 pentamers appears to be one of the best understood cases of those protein−saccharide interactions that have been shown to play a major role in the recognition processes at the cell surface.[12] We have recently initiated a program aimed to identify a viable computational approach to the study of such interactions and have chosen the GM1:toxin complexes as our test case. As a first step, a protocol was established that successfully reproduces the main features of the experimental (NMR) structure of GM1 pentasaccharide **1b** in water solution. Such protocol involved the use of an unconstrained Monte Carlo/energy minimization (MC/EM) conformational analysis of the sugar, using the continuum dielectric solvation model GB/SA implemented in MacroModel/Batchmin and the force field AMBER*, augmented by new parameters for the Neu-5-Ac residue. The search yielded a set of 10 low-energy conformations (within 1.5 kcal/mol from the global minimum), which nicely fitted the available NMR data.[11]

† Dedicated to Prof. Carlo Scolastico on the occasion of his 60th birthday.

⊗ Abstract published in *Advance ACS Abstracts,* March 1, 1997.

S0022-2623(96)00681-4 CCC: $14.00 © 1997 American Chemical Society



1a: R = 1-Ceramide
1b: R = Me

Figure 1. Ganglioside GM1.

In this paper we report how the same conformational search of the saccharide performed *within* the toxin binding pocket can be used to obtain a meaningful model of the GM1:LT and GM1:CT complexes. We adopted an approach that was first proposed by Guida and co-workers[13] to predict the binding conformations of enzyme inhibitors with high conformational flexibility and that is based on MC/EM techniques. The method starts from a structure of the ligand docked in the protein binding site and searches the conformational space available to the ligand in the confined space determined by the protein cavity. The starting point is generally provided by an X-ray crystal structure of the protein−ligand complex. In the present case the LT:lactose X-ray structure[6] was chosen as the starting one and docking of GM1 into the toxin was realized by superimposing Gal-IV of GM1 (see Figure 1) to the Gal residue of bound lactose. The efficiency of the conformational search was initially evaluated by performing it on the LT:lactose complex and comparing the results with the known X-ray structure. Various searches were carried out varying the number of crystallographic water molecules explicitly retained in the structure, with the aim of evaluating their relevance to the reproduction of the experimental data and the ability of the GB/SA implicit solvation model to substitute for them. Finally, GM1 was docked in the binding site, and the conformational search was performed on the LT:GM1 complex. The results were a fair reproduction of the CT:GM1 crystal structure published during the course of our work[8] and were further refined by explicitly including two crystallographic water molecules that were found to be tightly associated with the recognition surface of the protein.[7]

Computational Methods

All calculations were performed using the MacroModel/Batchmin package[14] (version 4.0 or 4.5) and the AMBER* force field, augmented by MNDO-derived sialic acid parameters.[11] To simplify the computational problem, all calculations on GM1 were performed on the methyl derivative **1b**. The X-ray coordinates for the LT:lactose complex (2.3 Å resolution) were kindly made available by Prof. Hol.[6,15] In order to reduce the complexity of the structure while maintaining the integrity of the sugar-binding cavity, all calculations were carried out on a B_2 (B + B(+1)) dimer. Only the ligand and a shell of residues surrounding the binding site of LT were subjected to energy minimization. All the residues within 5 Å of the sugars were included in the shell. The ligand and all binding site polar hydroxy and amino hydrogens were unconstrained during energy minimization. All other atoms that belonged to the substructure being minimized were constrained to their crystallographic coordinates by parabolic restraining potentials that increased with the distance from the sugar substrate. The force constants used are as follows: 100 kJ/Å^2 for the atoms within 0−3 Å of any atom of the ligand, 200 kJ/Å^2 for the atoms within 3−4 Å, and 400 kJ/Å^2 for the atoms within 4−5 Å. The perifery of the restrained structure was checked with the EdgeD command of MacroModel, and isolated atoms were included to avoid incomplete functional groups. All other atoms were ignored.

The calculations were carried out either with a dielectric constant of 4 or using the GB/SA solvation model[16] of MacroModel. This model treats the solvent as an analytical continuum starting near the van der Waals surface of the solute and uses a dielectric constant of 78 for the bulk water and 1 for the molecule. Extended nonbonded cutoff distances were used. Thus all calculations were run with a van der Waals cutoff of 8.0 Å and an electrostatic cutoff of 20.0 Å.

The conformational searches were carried out using either a usage-directed MC/EM procedure[17] or its pseudosystematic variant (SUMM).[18] The extraannular bonds of the sugars that can undergo free rotation were used as torsional variables in the MC steps. Previous studies on GM1[11] had shown that AMBER* in GB/SA water overestimates the stability of *gauche* conformations for the C_7-C_8 diol in the NeuAc side chain, compared to experimental data which are consistent with an essentially *anti* conformation.[10,19] It was found that a good solution to this problem consists in starting the MC search for GM1 with a 7,8-*anti* conformer and not including the C_7-C_8 bond in the variable list. Conformational filtering (TORC) was also used to screen out *gauche* rotamers during the search. Translation (1.5 Å maximum) and rotation (180° maximum) of the substrate within the binding cavity were allowed during the MC steps. The united atom version of AMBER* was used for the initial search. Energy minimization was performed using the truncated Newton conjugate gradient (TNCG) procedure,[20] and was terminated either after 500 iterations or when the energy gradient rms fell below 0.1 kJ/mol Å. All conformers that differed from the global minimum energy conformation by no more than 100 kJ/mol were saved. After addition of explicit H atoms on the sugar, they were subjected to further energy minimization to reduce the energy gradient rms to 0.01 kJ/mol Å.

Results and Discussion

As we mentioned, at the beginning of this project no X-ray data were available for any GM1:toxin complex. However, the LT:lactose structure had just been determined by Hol and co-workers.[6] This structure showed that the disaccharide was interacting with the protein mainly *via* its galactose unit, whereas the glucose at the reducing end was free in the cavity and loosely solvated by a few crystallographic water molecules. A schematic representation of the binding interactions between the Gal unit and LT is reported in Figure 2. Every hydrogen bond donor and acceptor of galactose (except the ether oxygen) are hydrogen bonding to LT either directly or with the mediation of the three well-defined water molecules at solvation sites 2, 3, and 5. The protein residues involved are Glu-51, Lys-91, Gln-61, and Asn-90. Solvent site 2 mediates a hydrogen-bonding interaction with the Gly-33 of the adjacent B(+1) monomer. Extensive van der Waals contacts are found between the hydrophobic α face of galactose and Trp-88.

These data and previous biochemical observations led to the hypothesis that a galactose-binding site exists in CT and LT and is the same for lactose and GM1. If this is the case, a meaningful model of the LT:GM1 complex could be generated by superimposing the Gal-IV residue of GM1 (see Figure 1) to the galactose of bound lactose in the LT:lactose crystal structure. After

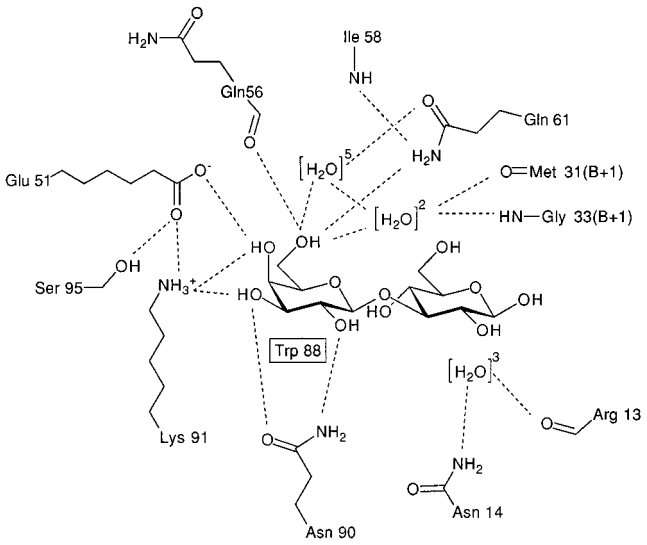


Figure 2. Interactions of lactose with LT, as seen in the X-ray structure.

this simple docking operation optimization of the LT:GM1 complex could be carried out using MC/EM conformational searches. This method has been shown to be effective at reproducing both enzyme−inhibitor complexes[13] and the solution structure of GM1.[11] The implementation of the foregoing strategy presented two main problems:

1. Due to the difficulties associated with the computational chemistry of sugars and to the complexity of the case at hand, the procedure chosen had to be validated by showing that it could reproduce at least the main features of experimentally known structures.[21] This could not be done directly on the GM1:LT complex, for which experimental structures were not available. Therefore the definition of the computational protocol and its validation had to be sought using the known LT:lactose structure as benchmark.

2. An additional point of concern was represented by the water molecules found by the X-ray analysis within the LT binding cavity in the complex. In the LT:lactose complex there are nine such molecules within 5 Å of the sugar. Some of them appear to be tightly bound to the galactose unit and to mediate its interaction with the protein (see Figure 2); others appear to simply fill the cavity around glucose. Since lactose is a much smaller substrate than GM1, the point could be made that at least some of the solvent molecules observed in the crystal structure may be in the cavity to *replace* parts of the ganglioside, and thus they should be deleted to study the LT:GM1 complex. Ideally, an implicit solvation model such as GB/SA could be used to substitute for explicit solvation, and therefore it should be possible to ignore the crystallographic water molecules and still obtain reasonable structures. However, this is not necessarily true, particularly for those molecules that are deeply buried in the cavity (such areas may not be seen as solvent accessible by the solvation algorithm) or for those that mediate specific interactions between the toxin and the substrate. With this problem in mind, the conformational analysis of the LT:lactose complex was carried out under various conditions of solvation and dielectric constant, with the aim of determining to which extent the GB/SA model could replace explicit solvation in this particular context.

Table 1. Minimization of LT:Lactose[a]

	entry		
	1	2	3
elect treatment/solvation	$\epsilon = 4$	GB/SA	GB/SA
cryst H_2O retained[b]	12	12	3
rmsd (Å)[c] total	0.264	0.583	1.050
Gal	0.172	0.418	0.468
Glc	0.353	0.748	1.660

[a] From the X-ray structure, using TNCG algorithm and AMBER*, to an energy gradient rms of 0.01 kJ/Å mol. [b] See text. [c] Measured between equivalent atoms of the sugar after superimposition with the X-ray structure.

Table 2. MC/EM Conformational Search of LT:Lactose

	entry		
	1	2	3
elect treatment/solvation	$\epsilon = 4$	GB/SA	GB/SA
cryst H_2O retained[a]	12	12	3
MC/EM steps	3000	3000	5000
no. of low-energy structures[b]	24	9	5 (+19)[a]
glob min rmsd (Å)[c] total	0.453	2.067	2.74
Gal	0.201	0.312	0.689
Glc	0.726	3.983	4.898

[a] See text. [b] Structures within 20 kJ/mol from the global minimum. [c] Measured between equivalent atoms of the sugar after superimposition with the X-ray structure.

The LT:Lactose Complex. The substructure that was actually taken into account was a shell of 5 Å surrounding the sugar, which included nine crystallographically determined water molecules. Simple minimization of this structure using a dielectric constant $\epsilon = 4$ essentially duplicates the X-ray coordinates. The value of 4 is generally accepted as the best option for the dielectric constant in molecular mechanics reproduction of crystal structures.[22] When the GB/SA solvation model was used ($\epsilon_{bulk} = 78$, $\epsilon_{mol} = 1$) for the same minimization, the lactose molecule translated inside the cavity toward the carboxylate of Glu-51 (see Figure 2). In the crystal structure this residue appears to be hydrogen bonded to the 3-hydroxy group of the sugar. Apparently, when the carboxylate charge is not reduced by a polar dielectric constant, such interaction is overestimated by the program. The GB/SA model should provide the required shielding effect by simulating water solvation, but Glu-51 is located too deeply in the cavity to appear solvent accessible to the GB/SA algorithm. On the other hand, in the X-ray structure the carboxylate of Glu-51 is found to be specifically solvated by three crystallographic water molecules that were not initially included in the calculation because they are farther apart than 5 Å from the sugar. When these molecules were included in the substructure being minimized[23] (adding to a total of 12 explicit water molecules), the X-ray structure could be reproduced using either $\epsilon = 4$ (Table 1, entry 1) or the GB/SA model (Table 1, entry 2) with a root mean square deviation (rmsd)[24] of 0.264 and 0.583 Å, respectively. Removing all crystallographic water molecules but the three solvating Glu-51 carboxylate (Table 1, entry 3), the rmsd between the GB/SA-minimized structure and the X-ray structure increases to 1.050 Å, still well below the X-ray resolution. It is interesting to note that in all minimizations reported in Table 1 the largest deviation is observed for the glucose residue, which is indeed less strongly bound in the cavity than galactose.

Two conformational searches were then performed using either $\epsilon = 4$ (Table 2, entry 1) or the GB/SA model

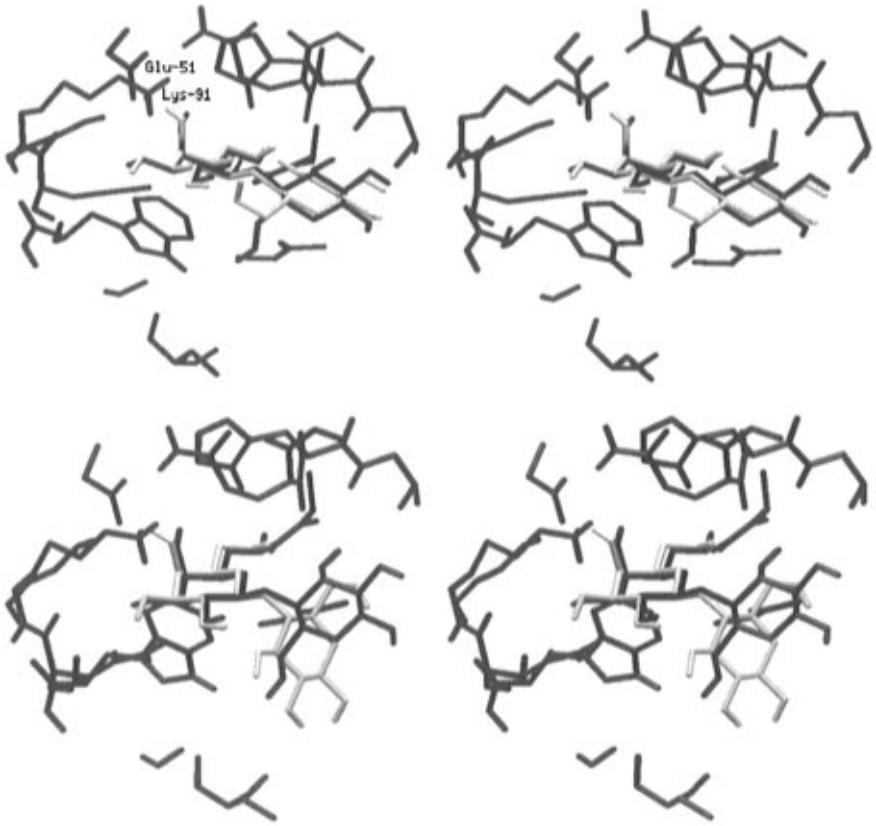


Figure 3. Superimpositions of the experimental (black) and calculated (gray) structures of LT:lactose. The calculated structures are the global minima found retaining all the 12 crystallographic water molecules (Table 2, entries 1 and 2): (top) calculated using $\epsilon = 4$ (Table 2, entry 1) and (bottom) calculated using GB/SA (Table 2, entry 2).

(Table 2, entry 2). The 12 crystallographic water molecules were retained, and the torsional degrees of freedom of lactose and its position relative to the protein were varied, for a total of five MC variables. The searches (3000 steps each) yielded 24 and 9 low-energy conformations (within 20 kJ/mol from the global minimum), respectively. The superimpositions between the global minima and the X-ray structure are reported in Figure 3 (water molecules deleted for clarity). Figure 4 shows a superimposition of the calculated low-energy conformations of the sugar (protein deleted for clarity). The minimum energy structures are quite closely related to the one derived crystallographically. The major deviation from the X-ray coordinates occurs for the GB/SA minimum (Table 2, entry 2; Figure 3, bottom) and concerns the position of glucose. We have already observed that this residue is not really bound by the protein.[6] On the contrary, the primary binding interactions due to the terminal galactose and the position of this residue in the cavity are reproduced rather accurately by the global minima (Figure 3) and by *all* the low-energy conformations (Figure 4).

Finally, all crystallographic water molecules were removed, except the three solvating Glu-51, and a 5000 step MC/EM search was performed using the GB/SA solvation model (Table 2, entry 3). Quite reassuringly, the global minimum was still very similar to the X-ray structure (Figure 5), and four additional low-energy conformers with the same characteristics were found. As usual, the position of the galactose residue is accurately reproduced by the calculation. However, in this case 19 of the conformers found within 20 kJ/mol of the minimum locate the lactose molecule *outside* the protein binding cavity.[25] The main driving force for this

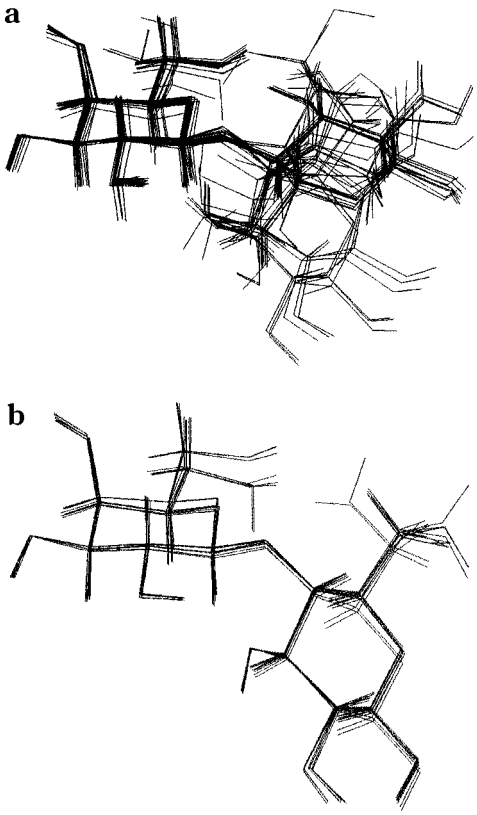


Figure 4. Superimpositions of the LT:lactose low-energy conformations (within 20 kJ/mol of the global minimum) calculated retaining all the 12 crystallographic water molecules (Table 2, entries 1 and 2). Only the lactose molecule is shown: (a) the 24 conformers found using $\epsilon = 4$ (Table 2, entry 1) and (b) the nine conformers found using GB/SA (Table 2, entry 2).

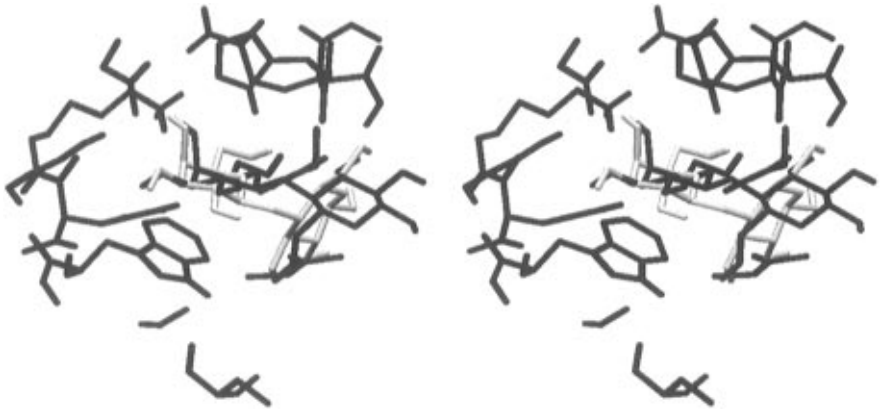


Figure 5. Superimposition of the experimental (black) and calculated (gray) structures of LT:lactose. The calculated structure is the global minimum found retaining only three crystallographic water molecules (Table 2, entry 3).

dramatic movement of the substrate appears to be the strong hydrogen bond network created with the three water molecules retained, which compensates for the lost solvation of glucose. This problem is likely to be reduced when calculating the LT:GM1 complex, due to the larger number of contacts between the sugar and the protein, which ought to stabilize the position of the substrate.

The fact that even under the extreme circumstances of total water deletion the global minimum and four additional low-energy structures were found as reasonable reproductions of the galactose-binding disposition appeared to be encouraging enough to move on to the next step, *i.e.*, the determination of the LT:GM1 complex structure. In doing so, we were warned that major local distorsions of the structure could occur in the absence of crucial water molecules.

The LT:GM1 Complex. The previously calculated lowest energy solution conformation of GM1[11] was docked into the toxin binding site by superimposing its Gal-IV residue to the galactose coordinates in LT: lactose. After removing all crystallographic water molecules but the three solvating Glu-51, the complex was minimized by taking into account the usual 5 Å shell. The resulting structure was the starting point for a 8600 step MC/EM conformational search (search 1, Table 3, entry 1) which yielded 51 low-energy conformations within 20 kJ/mol from the global minimum.

At this stage the X-ray structure of the CT:GM1 complex at 2.2 Å resolution became available[8] and allowed direct comparison of our LT:GM1 model with experimental data. In the crystal structure there are five nonidentical copies of GM1 pentasaccharide bound to each of the five B subunits of CT. The quality of the X-ray data allowed unambiguous determination of a model for all five copies of the first two sugar residues (Gal-IV and GalNAc) and for four of the NeuAc residues. The entire pentasaccharide model could be built only for subunit B5, but it appears to be consistent with residual electron density in the other four binding sites. The binding was found to resemble a two-fingered grip, with only two residues of GM1, Gal-IV and NeuAc, interacting directly or *via* solvent-mediated hydrogen bonds with CT. A map of the main interactions between sugars and toxin is reported in Figure 6.

Comparison with the calculated structure was achieved by superimposing the side chains of three conserved binding site residues (Lys-91, Trp-88, and Glu-51) of the calculated and X-ray complex. Figure 7 shows the result of such superimposition between the global minimum of search 1 and the CT:GM1 crystal structure (LT deleted for clarity).

Globally, the similarity of the two structures is rather satisfactory. Both the sugar residues which do contribute to binding are roughly located in the same position with respect to the protein. The major deviation from the experimental data is observed for Glc-I, which in

Table 3. MC/EM Conformational Search of LT:GM1[a]

		entry		
		1	2	
cryst H_2O retained[b]		3	5	
MC/EM steps		8600	15000	
no. of low-energy structures[c]		51[d]	8[e]	
glob min rmsd (Å)[f]	Gal-IV	0.844	0.376	(0.417)
	GalNAc	0.806	0.196	(0.493)
	Gal-II	1.214	1.007	nd[g]
	Glc-I	5.700	3.839	nd[g]
	NeuAc	1.376	0.898	(0.641)
NeuAc side chain	C7	1.751	0.995	(0.763)
	C8	2.164	0.737	(0.712)
	C9	2.733	0.897	(0.824)
	O7	2.382	1.077	(0.923)
	O8	1.968	0.600	(0.486)
	O9	3.137	0.754	(0.686)

[a] All searches were performed using GB/SA water solvation and u.a. AMBER*. [b] See text. [c] Structures within 20 kJ/mol from the global minimum. [d] Minimized to 0.01 kJ/mol Å using u.a. AMBER*. [e] Minimized to 0.01 kJ/mol Å with explicit hydrogens on the sugar. [f] Measured between sugar ring centroids and NeuAc side chain atoms after superimposition with the X-ray structure. In parentheses is given the rmsd between the two best defined crystallographic pentasaccharides. [g] Only one copy of this residue was experimentally determined.

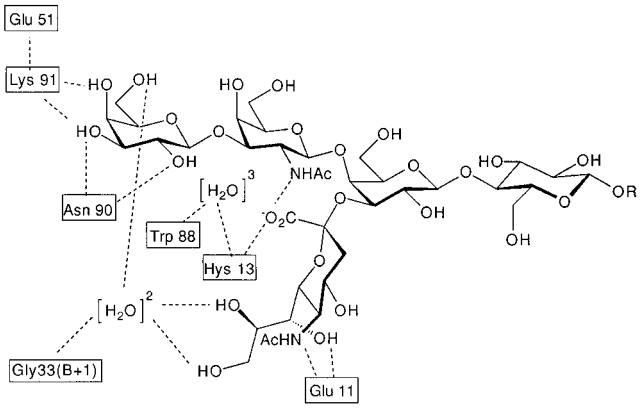


Figure 6. Interactions of GM1 with CT, as seen in the X-ray structure.

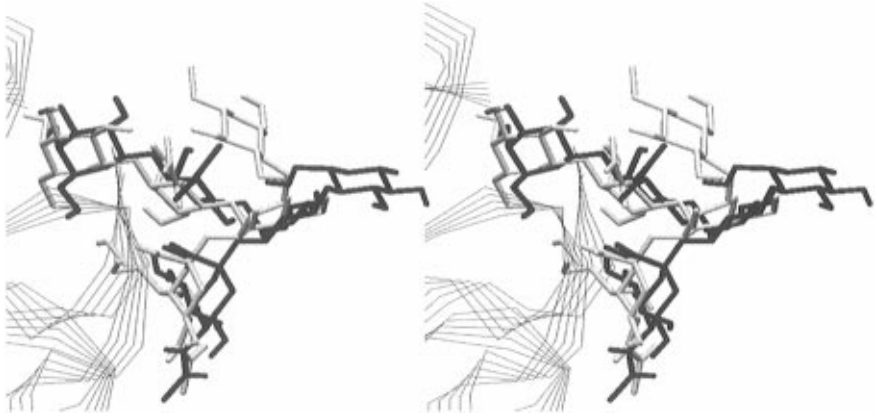


Figure 7. Superimposition between the X-ray structure of CT:GM1 (black) and the global minimum of search 1 in Table 3 for the LT:GM1 complex (gray) (LT deleted).

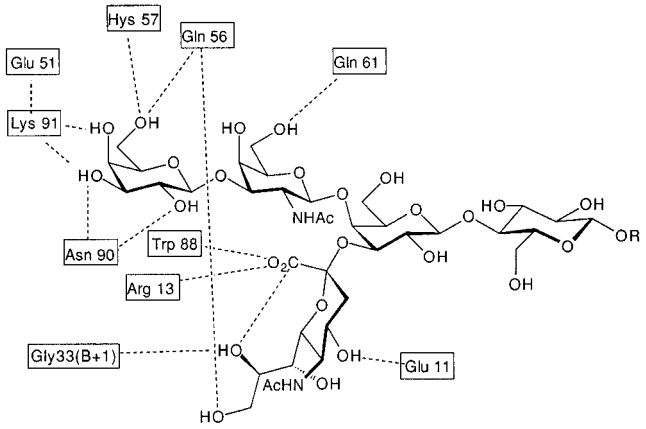


Figure 8. Interactions of GM1 with LT, as calculated by search 1 in Table 3.

Figure 9. Superimposition of the first two LT:GM1 low-energy conformations, as calculated by search 2 in Table 3 (LT deleted).

the calculated complex folds in toward the toxin to pick up hydrogen bonds from the rest of the sugar and the protein. This can be due to insufficient sampling or to a poor reproduction of bulk water solvation by the GB/SA model. However, it should be noted that the X-ray structure features a gluconolactone instead of a glucose at the sugar reducing end and that this residue is connected to Gal-II with an unusual (−46,−7) (ϕ,ψ) anomeric linkage, which may be forced by crystal packing. Furthermore, four of the five GM1 copies in the X-ray structure exhibit a lack of good electron density for this residue, which is consistent with flexibility.[8]

Closer inspection of the computed and crystal structure of the sugar, as reported in Figure 7, reveals more subtle differences. Most notably, GM1 is calculated to stick deeper in the binding pocket than it is experimentally observed. The rmsd of the single-sugar residues in the computed and experimental structures are reported in Table 3 (entry 1) and compared with the deviation between the two best-defined crystallographic pentasaccharides (in parentheses). The major deviation is seen for the NeuAc residue and particularly for its side chain. In the computed structure the sialic acid carboxylate is found at hydrogen-bonding distance from the indolic NH of Trp-88 (see Figure 8).

Apparently, the different orientation of the NeuAc side chain in the computed structure finds its origin in the absence of water molecules in sites 2 and 3 (see Figure 6). In fact, in the crystal structure those water molecules mediate the interactions between GM1 and both Trp-88 of the B unit and Gly-33 of the adjacent B(+1) residue. From biochemical studies, these interactions were indeed found to be essential for GM1 binding.[4] In the computed structure, site 2 is actually filled by NeuAc OH-8, which directly binds to Gly-33. The absence of site 3 water is compensated by thrusting forward the NeuAc carboxylate by about 1.5 Å. As a consequence of the different conformation of the NeuAc side chain, the terminal hydroxy OH-9 is at hydrogen bond distance from the hydroxymethylene group of Gal-IV. Thus, from the results of search 1, it clearly appears that explicit water molecules are needed in sites 2 and 3, since they are involved in very specific interactions which cannot be accounted for by the GB/SA water model. This is in agreement with a recent study, which focused on the water molecules found in the proximity of the galactose-binding site in a set of five different X-ray structures of LT/CT of their sugar complexes.[7] It was concluded that "solvent sites 2 and 3 are consistently occupied in the ensemble of structures and may be regarded as constituting a part of the recognition surface of the protein, *i.e.*, the specific binding site 'seen' by the GM1 receptor consists of the toxin B pentamer with solvents already bound at sites 2 and 3.".[7]

With this in mind a new search was performed (search 2, Table 3, entry 2), which included the two crystallographically determined water molecules at sites 2 and 3. The search was carried out for 15 000 steps using u.a. AMBER*. Eighty structures were found

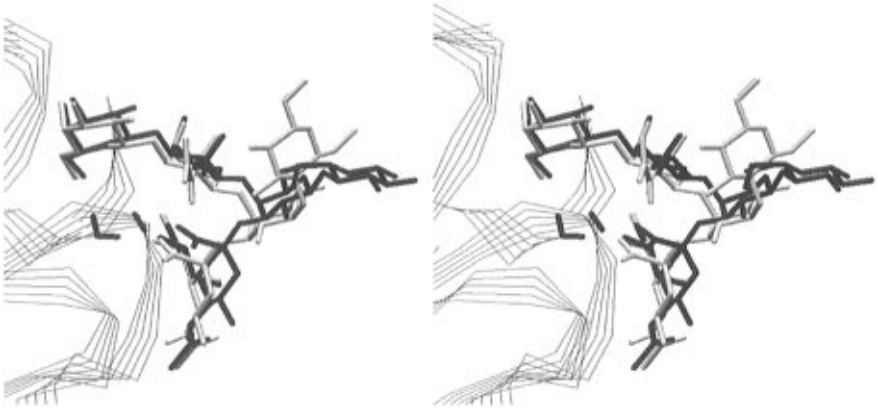


Figure 10. Superimposition between the X-ray structure of CT:GM1 (black) and the global minimum of search 2 in Table 3 for the LT:GM1 complex (gray) (LT deleted).

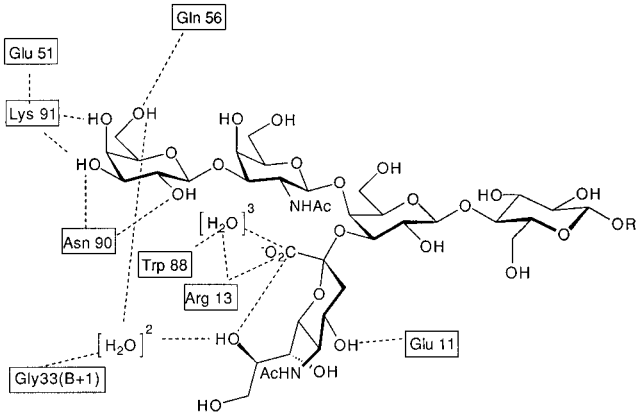


Figure 11. Interactions of GM1 with LT, as calculated by search 2 in Table 3.

within 42 kJ/mol from the global minimum. They were reminimized after adding explicit hydrogens to the sugar, to yield two low-energy conformations within 12 kJ/mol and eight conformations within 20 kJ/mol from the global minimum. Superimposition of the first two structures is reported in Figure 9 and is representative of the differences found among all the eight low-energy conformations. The positions of Gal-IV and GalNAc do not vary in any of the structures, which mainly differ in the coordinates of the terminal glucose and the NeuAc side chain.

The superimposition between the global minimum and the X-ray structure was carried out as described above, and the result is reported in Figure 10. Fitting between the calculated and observed structure has dramatically improved, and all the sugar/toxin contacts are now reproduced by the computation (compare Figure 6 and Figure 11).[26]

The foregoing results suggest that the substructure here identified, which includes the protein shell within 5 Å from GM1 plus the three water molecules solvating Glu-51 carboxylate plus the two water molecules at crystallographic sites 2 and 3, can be taken as a realistic representation of the toxin binding surface. Calculation of the interactions between such surface and potential sugar or sugar-like substrates should yield reasonable models of the complexes and allow at least a qualitative comparison of the binding ability of different substrates. Therefore it should be possible to use this method as a predictive tool in designing artificial LT and/or CT binders and thus potential anticholera drugs. Studies are in progress in our group to test this point.

Acknowledgment. This work was supported by grants from MURST, CNR, and Centro di Modellistica Computazionale del CILEA (6C007). We thank Prof. Hol for kindly providing the coordinates of the LT: lactose complex and Dr. Zanferrari for technical advice.

References

(1) Masserini, M.; Freire, E.; Palestini, P.; Calappi, E.; Tettamanti, G. Fuc-GM1 Ganglioside Mimics the Receptor Function of GM1 for Cholera Toxin. *Biochemistry* **1992**, *31*, 2422−2426.
(2) De Wolf, M. J. S.; Fridkin, M.; Kohn, L. D. Tryptophan Residues of Cholera Toxin and its A Protomers and B Protomers. Intrinsic Fluorescence and Solute Quenching upon Interacting with Ganglioside GM1, Oligo-GM1, or Dansylated Oligo-GM1. *J. Biol. Chem.* **1981**, *256*, 5489−5496.
(3) Merritt, E. A.; Hol, W. G. J. AB(5) Toxins. *Curr. Opin. Struct. Biol.* **1995**, *5*, 165−171.
(4) Ludwig, D. S.; Holmes, R. K.; Schoolnik, G. K. Chemical and Immunochemical Studies on the Receptor Binding Domain of Cholera Toxin B Subunit. *J. Biol. Chem.* **1985**, *260*, 12528−12534. Tsuji, T.; Honda, T.; Miwatani, T.; Wakabayashi, S.; Matsubara, H. Analysis of Receptor Binding Site in *E. Coli* Enterotoxin. *J. Biol. Chem.* **1985**, *260*, 8552−8558. Jobling, M. G.; Holmes, R. K. Analysis of Structure and Function of the B Subunit of Cholera Toxin by the Use of Site-directed Mutagenesis. *Mol. Microbiol.* **1991**, *5*, 1755−1767.
(5) Sixma, T. K.; Kalk, K. H.; v. Zanten, B. A. M.; Dauter, Z.; Kingma, J.; Witholt, B.; Hol, W. G. J. Refined Structure of *E. Coli* Heat Labile Enterotoxin, a Close Relative of Cholera Toxin. *J. Mol. Biol.* **1993**, *230*, 890−918. Sixma, T. K.; Pronk, S. E.; Kalk, K. H.; Wartna, E. S.; v. Zanten, B. A. M.; Witholt, B.; Hol, W. G. J. Crystal Structure of a Cholera Toxin-Related Heat Labile Enterotoxin from *E. Coli*. *Nature* **1991**, *351*, 371−377.
(6) Sixma, T. K.; Pronk, S. E.; Kalk, K. H.; Wartna, E. S.; v. Zanten, B. A. M.; Berghuis, A. M.; Hol, W. G. J. Lactose Binding to Heat Labile Enterotoxin Revealed by X-Ray Crystallography. *Nature* **1992**, *355*, 561−564.
(7) Merritt, E. A.; Sixma, T. K.; Kalk, K. H.; v. Zanten, B. A. M.; Hol, W. G. J. Galactose Binding-site in *E. Coli* Heat Labile Enterotoxin (LT) and Cholera Toxin (CT). *Mol. Microbiol.* **1994**, *13*, 745−753.
(8) Merritt, E. A.; Sarfaty, S.; v. d. Akker, F.; L'Hoir, C.; Martial, J. A.; Hol, W. G. J. 2.2 Å Crystal Structure of Cholera Toxin B_5 Pentamer Bound to Receptor GM1 Pentasaccharide. *Protein Sci.* **1994**, *3*, 166−175.
(9) Lanne, B.; Schierbeck, B.; Karlsson, K. A. On the Role of the Carboxyl Group of Sialic Acid in Binding of Cholera Toxin to Receptor Glycosphingolipid, GM1. *J. Biochem.* **1994**, *16*, 1269. Fukuta, S.; Magnani, J. L.; Twiddy, E. M.; Holmes, R. K.; Ginsburg, V. Comparison of the Carbohydrate Binding Specificities of Cholera Toxin and *E. Coli* Heat Labile Enterotoxins LTH-1, LT-1a, and LT-1b. *Infect. Immun.* **1988**, *56*, 1748−1753. Ångstrøm, J.; Tenenberg, S.; Karlsson, K.-A. Delineation and Comparison of Ganglioside-Binding Epitopes for the Toxins of *Vibrio Cholerae*, *E. Coli*, and *Clostridium Tetani*: Evidence for Overlapping Epitopes. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 11859−11863. Schengrund, C.-L.; Ringler, N. J. Binding of *Vibrio Cholerae* Toxin and the Heat Labile Enterotoxin of *E. Coli* to GM1 Derivatives of GM1, and Nonlipid Oligosaccharide Polivalent Ligands. *J. Biol. Chem.* **1989**, *264*, 13233−13237.
(10) Acquotti, D.; Poppe, L.; Dabrowski, J.; v. d. Lieth, C.-W.; Sonnino, S.; Tettamanti, G. 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 in Native Glycolipid in Dimethylsulfoxide and in Water-Dodecylphosphocoline Solutions. *J. Am. Chem. Soc.* **1990**, *112*, 7772−7778.
(11) Bernardi, A.; Raimondi, L. Conformational Analysis of GM1 in Water Solution with a New Set of Parameters for the Neu-5-Ac Moiety. *J. Org. Chem.* **1995**, *60*, 3370−3377.
(12) Lee, Y. C.; Lee, R. T. Carbohydrate-Protein Interactions: Basis of Glycobiology. *Acc. Chem. Res.* **1995**, *28*, 321−327.
(13) Guida, W. C.; Bohacek, R. S.; Erion, M. D. Probing the Conformational Space Available to Inhibitors in the Thermolysin Active Site Using Monte Carlo/Energy Minimization Techniques. *J. Comput. Chem.* **1992**, *13*, 214−228.
(14) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. MacroModel. An Integrated Software System for Modeling Organic and Bioorganic Molecules using Molecular Mechanics. *J. Comput. Chem.* **1990**, *11*, 440−467.
(15) The PDB entry for this structure is p11tt.
(16) Still, W. C.; Tempzyk, A.; Hawley, R.; Hendrickson, T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(17) Chang, G.; Guida, W. C.; Still, W. C. An Internal Coordinate Monte Carlo Method for Searching Conformational Space. *J. Am. Chem. Soc.* **1989**, *111*, 4379−4386.
(18) Goodman, J. M.; Still, W. C. An Unbounded Systematic Search of Conformational Space. *J. Comput. Chem.* **1991**, *12*, 1110−1117.
(19) Brown, E. B.; Brey, W. S.; Weltner, W., Jr. Cell-Surface Carbohydrates and their Interactions . I. NMR of N-Acetyl Neuraminic Acid. *Biochim. Biophys. Acta* **1975**, *399*, 124−130. Sabesan, S.; Block, K.; Lemieux, R. U. The Conformational Properties of the Gangliosides GM2 and GM1 Based on ^{1}H and ^{13}C Nuclear Magnetic Resonance Studies. *Can. J. Chem.* **1984**, *62*, 1034−1045.
(20) Ponder, J. W.; Richards, F. M. An Efficent Newton-like Method for Molecular Mechanics Energy Minimization of Large Molecules. *J. Comput. Chem.* **1987**, *8*, 1016−1024.
(21) It is assumed that exact duplication of the X-ray structure would be unachievable, since the X-ray structure represents a model of an average structure rather than an energy minimum.
(22) French, D. A.; Rowland, R. S.; Allinger, N. L. Modeling of Glucopyranose: the Flexible Monomer of Amylose. In *Computer Modeling of Carbohydrate Molecules*; French, D. A., Brady, J. W., Eds.; ACS Symposium Series 430; American Chemical Society: Washington, DC, 1990; pp 120−140.
(23) In the following calculations the positions of these water molecules were generally fixed by restraining the oxygen atoms with a parabolic potential with a force constant of 400 kJ/Å^2. However, these molecules appear to be so strongly associated with the carboxy group that their positions do not change significantly upon total optimization or including their position in the MC variables in short trial runs.
(24) The rms deviation was measured by superimposing the two structures and measuring the distances between equivalent atoms of the sugar.
(25) Most of these structures could easily be recognized as physically meaningless structures by visual inspection. In fact, they put the sugar in areas occupied by protein atoms which have been ignored by the calculation since they are external to the chosen substructure.
(26) The X-ray structure of the complex shows a hydrogen bond between the carboxylate of the sialic acid and the NH of GalNAc (ref 8), which is also suggested by the behavior of free GM1 in DMSO solution (ref 10). In the computed structures these two groups are indeed proximal (ca. 3 Å), but the GalNAc amido group is tilted by ca. 50° compared to the crystal structure, and the geometrical constraints for hydrogen bond are better satisfied by an interaction with the NeuAc anomeric oxygen.

JM9606812