Journal of Computer-Aided Molecular Design, **15:** 117–128, 2001.
KLUWER/ESCOM

Simulation of carbohydrate-protein interactions: Computer-aided design of a second generation GM1 mimic

Anna Bernardi[a,*], Marta Galgano[a], Laura Belvisi[a] & Giorgio Colombo[b,*]
[a]*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;* [b]*Istituto CNR di Biocatalisi e Riconoscimento Molecolare, via M. Bianco 9, 20133 Milano, Italy*

Received 20 March 2000; Accepted 7 August 2000

Key words: carbohydrate-protein interactions, conformational search, docking, enterotoxins, free energy perturbations, ganglioside GM1, glycomimetics, molecular dynamics, rational design of carbohydrate mimics

Summary

The oligosaccharide of ganglioside GM1 [Galβ1-3GalNAcβ1-4(NeuAcα2-3)Galβ1-4Glcβ1-1Cer] is the cellular target of two bacterial enterotoxins: the cholera toxin (CT) and the heat-labile toxin of *E.coli* (LT). We recently reported that the pseudosaccharide **2** [Galβ1-3GalNAcβ1-4(NeuAcα2-3)DCCHD] is a high-affinity ligand for CT, and thus a functional mimic of GM1 (Bernardi, A., Checchia, A., Brocca, P., Sonnino, S. and Zuccotto, F., J. Am. Chem. Soc., 121 (1999) 2032–2036). In this paper we describe the design of a second-generation mimic, formally obtained from **2** by inverting the configuration of a single stereocenter, thus transforming a N-acetyl galactosamine into a N-acetyl glucosamine. The design process involved modeling of the free ligand and its LT complex, followed by qualitative and quantitative comparison with the corresponding structures of **2**. The protocol employed relied on both conformational search and molecular dynamics methodologies to account for the flexibility of both the ligand and the protein receptor. The conformational search of the LT:inhibitor complex showed that, compared to **2**, the new compound can insert one more hydroxy group within the protein binding site. Molecular dynamics simulations showed that, in turn, this may trigger a series of rearrangements and reorientations of side chains and crystallographic water molecules in the toxin, leading to new H-bond contacts which may result in enhanced affinity of the new inhibitor. FEP calculations were performed by mutating the structure of **2** in solution and in the protein complex, and the prediction was made that the second-generation mimic should be a stronger binder than its parent compound.

Introduction

The inhibition of protein-carbohydrate interaction provides a powerful strategy for the treatment of a variety of diseases. In particular, the development of small-molecule mimics of carbohydrates associated with important recognition events has attracted a great deal of attention as a way to develop therapeutic agents with good stability and synthetic availability [1]. A major synthetic effort toward glycomimetic compounds has resulted, which has rarely been accompanied by a comparable effort in the rational design of such molecules. The rational design of carbohydrate mimics is indeed hampered by many fundamental difficulties. Despite the abundance of structural data on carbohydrate-protein complexes, the driving force for their formation is still much debated [2]. The accessible conformations of even moderately complex carbohydrates are notoriously difficult to determine [3], and not always conserved going from the solution to the protein complex [4–8]. Computational tools are stretched to their limits when required to deal with carbohydrates in water, let alone their proteic complexes [9]. Additionally, strongly bound water molecules,

*To whom correspondence should be addressed: E-mail: anna.bernardi@unimi.it, G.Colombo@chem.rug.nl Present address: Department of Biophysical Chemistry, University of Groningen, Nijenborg 4, 9747 AG Groningen, the Netherlands.

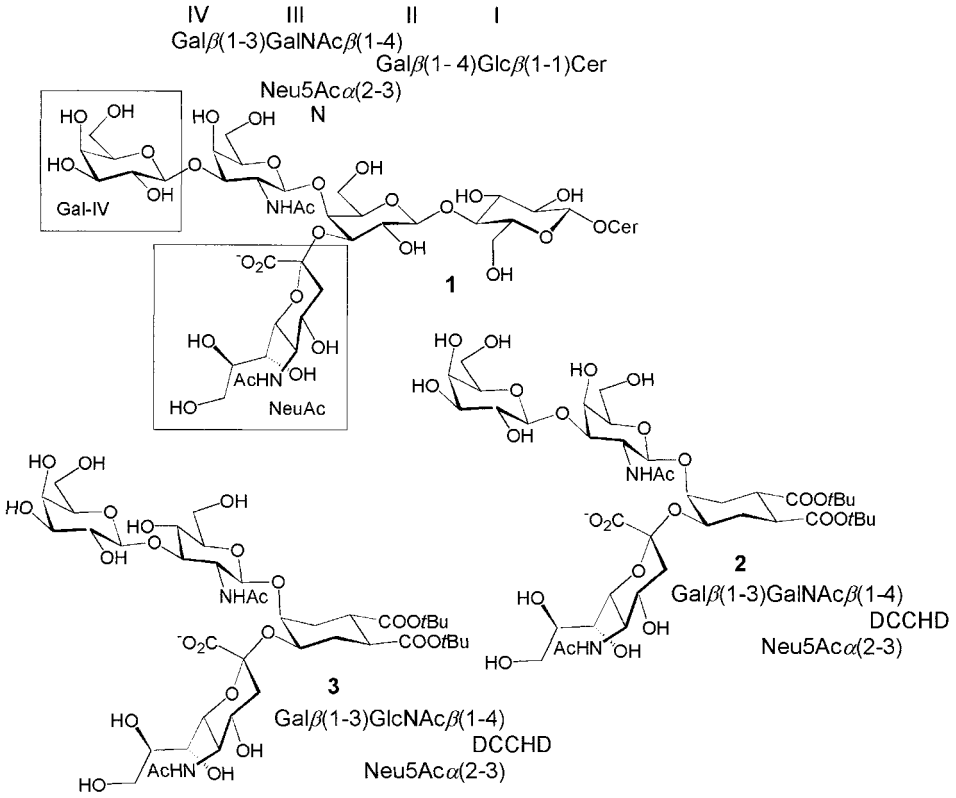


Figure 1. Structures of ganglioside GM1, **1**, its mimic **2**, and the GlcNAc derivative **3**. The binding determinants of **1** are highlighted in the box.

that are often found in carbohydrate binding sites, are involved in carbohydrate recognition, and must be dealt with during the design process [10–12].

Specific binding between bacterial enterotoxins and oligosaccharide receptors on the host cell membrane is a paradigm for protein-sugar interaction. One of the best characterized recognition pairs is formed by ganglioside GM1 [Galβ1-3GalNAcβ1-4(NeuAcα2-3)Galβ1-4Glcβ1-1Cer] and the bacterial enterotoxins cholera toxin (CT) and heat-labile toxin of *E. coli* (LT) [13]. CT and LT are 80% homologous AB5 proteins sharing a common mechanism of action: after the B5 pentamer has docked to the cell by interacting with GM1 headgroups on its surface, a catalytic fragment of the A subunit is translocated through the membrane and actual cell intoxication is initiated. Biochemical [14–19] and structural [20–22] data have shown that the oligosaccharide headgroup of GM1 (o-GM1) binds the B5 pentamer using the two terminal sugars at its non-reducing end, Gal-IV and NeuAc (see Figure 1). The intrinsic dissociation constant for CT binding to o-GM1 is 0,95 μM [18], one of the strongest known values for sugar-protein complexes. Last but not least, o-GM1 has been shown to exist mainly in a single conformation in solution [23], which closely resembles the bound conformation observed by X-ray crystallography [20, 21].

Given the amount of available data, and the characteristic of its structure and function, GM1 appeared to be ideally suited as a subject for rational mimic design. As a first step, a computational model of the LT:GM1 complex which was a fair reproduction of the CT:GM1 X-ray structure was obtained using a Monte Carlo/Energy Minimization (MC/EM) conformational search of the sugar within the toxin binding pocket [10]. Then, the same computational protocol was used to design a functional mimic of ganglioside GM1, the pseudosugar **2** [Galβ1-3GalNAcβ1-4(NeuAcα2-3)DCCHD]. The central residue of GM1, a 3,4-disubstituted galactose unit, was recognized as the ganglioside scaffold element and substituted with a conformationally locked cyclohexanediol (DCCHD). The binding determinants of the natural receptor were retained in **2**. Conformational analysis, both in solu-

tion and in the toxin binding site, showed that DCCHD is a good functional mimic of a 3,4-branched galactose, and predicted that **2** would preserve the same orientation of the binding determinants as observed in **1**. Indeed, the solution structure of **2** and its binding ability to CT were found to be analogous to those of the GM1 oligosaccharide [24].

In an effort to further simplify the molecular and synthetic complexity of the structure, second generation mimics are presently studied, and among these the pseudotetrasaccharide **3** (Figure 1), which is derived from **2** by replacing the GalNAc residue with a N-acetylglucosamine (GlcNAc). Modification of the hexosamine was considered for the following reasons: i) it would greatly simplify the synthesis of the artificial receptors (GalNAc is actually made from GlcNAc); ii) it should not interfere negatively with the formation of the toxin complex, since the hexosamine residue interacts with the protein only via the N-acetyl group, which is conserved in **3**. Furthermore, the 4-hydroxy group of GalNAc in the experimental CT:GM1 complex and in the computational LT:GM1 and LT:**2** models is located outside of the protein binding pocket and fully exposed to the solvent. Assuming that **3** will bind to CT and/or LT with the same general mode of GM1 and **2**, it appears likely that inversion of configuration at C4 of the hexosamine may allow new H-bond contacts to be formed between the toxin and the substrate. These may (or may not) compensate for the loss in complex solvation due to the burying of the hydroxy group in the binding cavity. Thus the study of the new artificial receptor **3** appears to be of great interest also because it may provide some insights about the fundamentals of protein-sugar complex formation.

The effects of the hexosamine replacement on the structure of the ligand and of the LT complex were modeled in a series of steps. Initially, the three-dimensional structure and conformational behavior of **3** were assessed by performing conformational searches of the ligand in (GB/SA) water solution and within the LT:**3** complex. These were performed using MacroModel's MC/EM search protocol and GB/SA solvation model, as previously established for **2** [24]. In this phase, the protein binding site was essentially constrained to the crystallographic coordinates, and only the conformational freedom of **3** was probed. The results showed that, like GM1, **3** is significantly conformationally restricted and that its solution lowest energy conformation is conserved in the LT complex. The dynamic behavior of the complex was then studied by molecular dynamics. In this phase, the effects due to flexibility of the side chains in the receptor binding pocket and to the presence and behavior of solvent molecules were uncovered. Since subtle solvation effects involving the hexosamine 4-hydroxy groups were expected to be crucial to the definition of the differential binding energy of **2** and **3**, an explicit solvation model (TIP3P) was included in this phase of the study. The same model was used to evaluate the differential hydration of the two free ligands in water solution. Finally, the qualitative observations collected so far were put on a quantitative basis using Free Energy Perturbation (FEP) methodologies.

The computational procedure outlined above efficiently addresses the problem of ligand flexibility by using the MC/EM conformational search phase to allow a fast and thorough exploration of the conformational space available to the substrate within the protein binding pocket. Thus, compared to studies based on molecular dynamics only [12], this protocol should be particularly useful as a tool in the design of inhibitors of carbohydrate-protein interactions.

The results have led us to conclude that **3** should be a better LT binder than **2**, and its synthesis is currently in progress in our laboratories.

Computational methods

*Conformational search (MC/EM) of isolated **3***

The calculations were performed using the MacroModel/Batchmin [25] package (version 5.5) and the AMBER* force field, augmented by MNDO derived sialic acid parameters [26]. Bulk water solvation was simulated using MacroModel's generalized Born GB/SA continuum solvent model [27]. 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 (a van der Waals cutoff of 8.0 Å and an electrostatic cutoff of 20.0 Å) were used. The conformational searches were carried out using 30,000 steps of the usage directed MC/EM procedure. Over the last 5,000 steps, no new minima were located within the first 2 kcal/mol. Following previously established protocols [10, 24], all the extraannular bonds of the sugars, except the C_7-C_8 bond of NeuAc, were used as torsional variables. The united atom version of AMBER* was used for the initial search. All the minima within 50 kJ/mol from the global minimum were saved and

energy minimized with the all atom version of the force field, after addition of explicit H atoms [10].

*Conformational search (MC/EM) of the LT:**3** complex*

The conformational searches were carried out using the usage directed MC/EM procedure, with the same protocol used in the study of the LT:psGM1 complex [24]. In brief, 14,000 MC/EM steps were performed, using the torsional variables described above. Translation (1.5 Å maximum) and rotation (180°) of the substrate within the binding cavity were allowed during the MC steps. Bulk water solvation was simulated using the GB/SA model. Five crystallographic water molecules were retained, as previously described [10, 24]. All calculations were carried out on a B2 (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 following force constants were used: 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; 400 kJ/Å^2 for the atoms within 4–5 Å. The periphery 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.

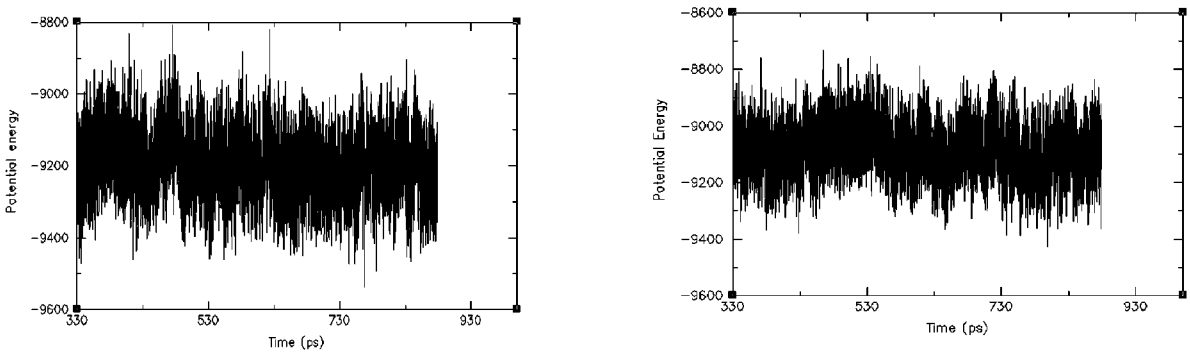
Molecular Dynamics studies with TIP3P water

The starting structures for Molecular Dynamics simulations of the toxin complexes of **2** and **3** with explicit TIP3P water [28] solvation were the global minimum structures obtained from previous MC/EM conformational searches. We retained the five water molecules in both the inhibitor-protein complexes [10, 24], and added a TIP3P water cap with a radius of 22 Å from the center of mass of the hexosamine in the docked pseudooligosaccharide, to ensure that the inhibitor and the surface region of the protein near the binding site were fully solvated. All residues within 15 Å radius of the center of mass of the two hexosamines, as well as the ligand and all the cap water molecules were allowed to move during the dynamics.

The appropriate charges on the atoms of the inhibitors were obtained by applying a divide and Conquer ESP fitting methodology, as implemented in the program DIVCON [29], to the uncomplexed molecules. The AM1 Hamiltonian was chosen in this part of the calculation as it has been shown that it could very well reproduce the geometric and electrostatic parameters obtained with higher level calculations [30]. The total charge on the ligand was -1, and the partial charges on the carboxylate oxygen atoms were calculated to be -0.62 and -0.55. The AMBER94 [31] all atom force field augmented with parameters obtained from Glycam_93 [30] for glycoproteins was used for all the explicit water simulations.

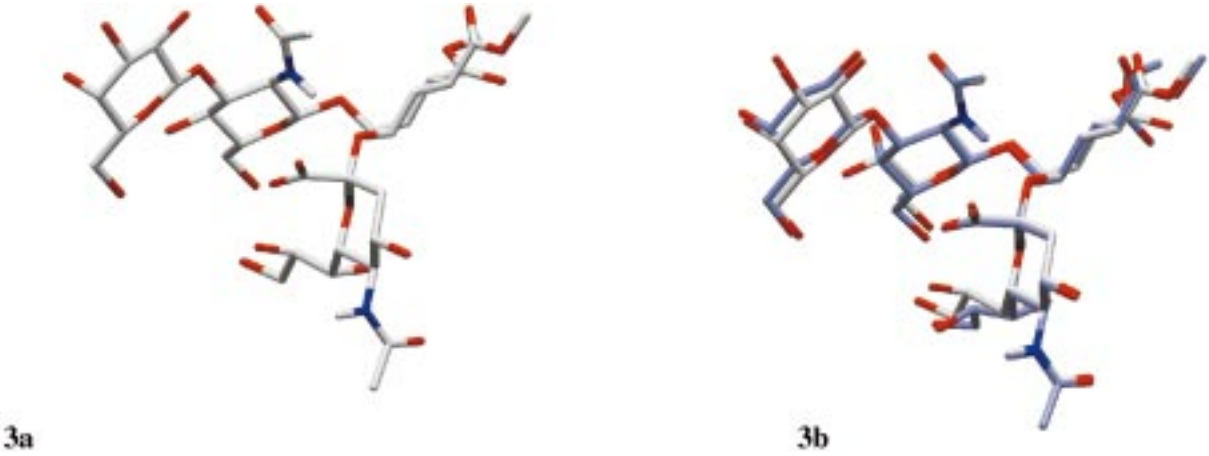
Both capped systems were minimized for 1500 steps using steepest descent energy minimization for the first 200 steps and conjugated gradient for the remaining part. The minimized configurations were used as starting points for molecular dynamics (MD) simulations. MD simulations were carried out using a residue based cutoff of 10 Å and a time step of 1.5 fs. In both cases, the bond lengths were constrained to their equilibrium values using the SHAKE [32] algorithm with a tolerance of 0.0005 Å. Both models were equilibrated by gradually increasing the temperature in stages (3 ps at 100 K, 3 ps at 200 K) to the target temperature of 310 K (3 more ps) by coupling to a thermal bath [33]. After 300 ps of MD simulation, the energy of the two systems had stabilized (Figure 2). From this point onwards, atomic coordinates were collected every 50 steps for 600 ps for the two systems, yielding a total of 900 ps of simulations for both complexes. The last 300 ps of each simulation were used for analysis on structures saved every 50 steps.

The starting configurations for uncomplexed **2** and **3** were the minimum structures found with previous MC/EM conformational search approaches. They were solvated by a periodic box of TIP3P water molecules. Both molecules were minimized according to the minimization protocol described above, and slowly heated from 0 to 310 K by a 9 ps MD. In this case the temperature and pressure were maintained constant at 310 K and 1 atm using the Nosè Hoover Chain (NHC) [34] algorithm for temperature and pressure coupling. Ewald sums and long range Van der Waals corrections were used for the accurate treatment of long-range electrostatic and Van der Waals interactions, respectively [35]. A *K*max value of 8 was used along with a nonbonded cutoff of 10 Å, and the pairlist was updated every 50 steps. SHAKE was applied to all bonds with a tolerance of 0.0005 Å. A time step of 1.5 fs was used



2a **2b**

Figure 2. Time evolution of the potential energy (kcal/mol) during the dynamic simulations. (a) LT:**2** complex; (b) LT:**3** complex



3a **3b**

Figure 3. (a) Minimum energy conformation of **3**, as calculated by MC/EM. (b) Superimposition of **2** (light blue) and **3** (white).

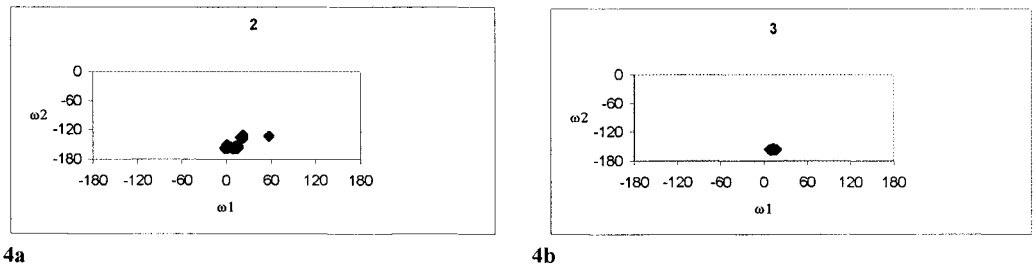
and the simulations were carried on for 900 ps. The last 300 ps of each simulation were used for analysis on structures saved every 50 steps.

FEP simulations

FEP simulations were carried out using the GIBBS module of the AMBER 5.0 [31] suite of programs. We have employed the 'slow growth' approach for the determination of the free energy. This method has been applied successfully in several previous applications [36, 37]. The relative free energy of binding $\Delta\Delta G_{3-2}$ (Scheme 1) was obtained by mutating the OH4 hydroxy group of the hexosamine from the equatorial (as in **3**) to the axial (as in **2**) position both in the sugar-protein complexes (ΔG_4) and in the free, water-solvated carbohydrate mimics (ΔG_1).

The partial charges on the starting and final points of these simulations were obtained through a coupled QM/MM ESP fitting procedure [38]. The procedure was applied to the ligands in aqueous solution and complexed to the protein. Atomic point charges for the substrates could have been determined in the gas phase as it is typical practice [39, 40]. However, this would lead to an unbalanced model between the simulations in water and in the toxin binding cavity. In the protein case, the charges calculated by ESP fitting do include polarization arising from the surrounding environment, thus it is also important to include these effects in the solution case. Hence, the partial charges on each atom of the two mimics were calculated by applying the ESP procedure on structures obtained after 300 ps of MD simulations.

The FEP simulations were run using a timescale of 600 ps and a timestep of 1.5 fs. To investigate the effects of sampling on the computed free energy values, four different simulations were carried out starting from temporally separated starting con-



4a 4b

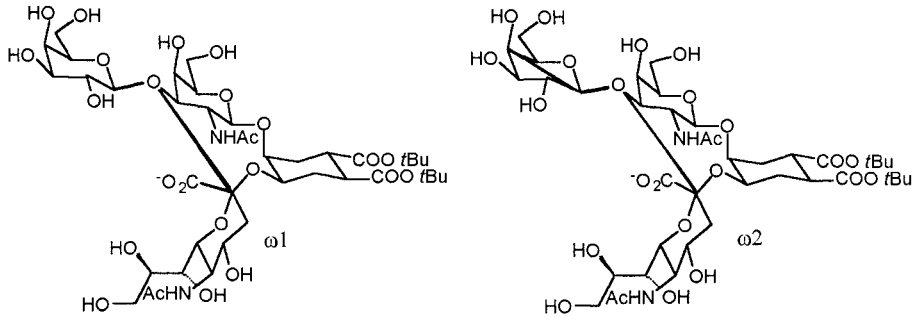


Figure 4. Internal coordinate plots for the MC/EM analysis of **2** (a) and **3** (b).

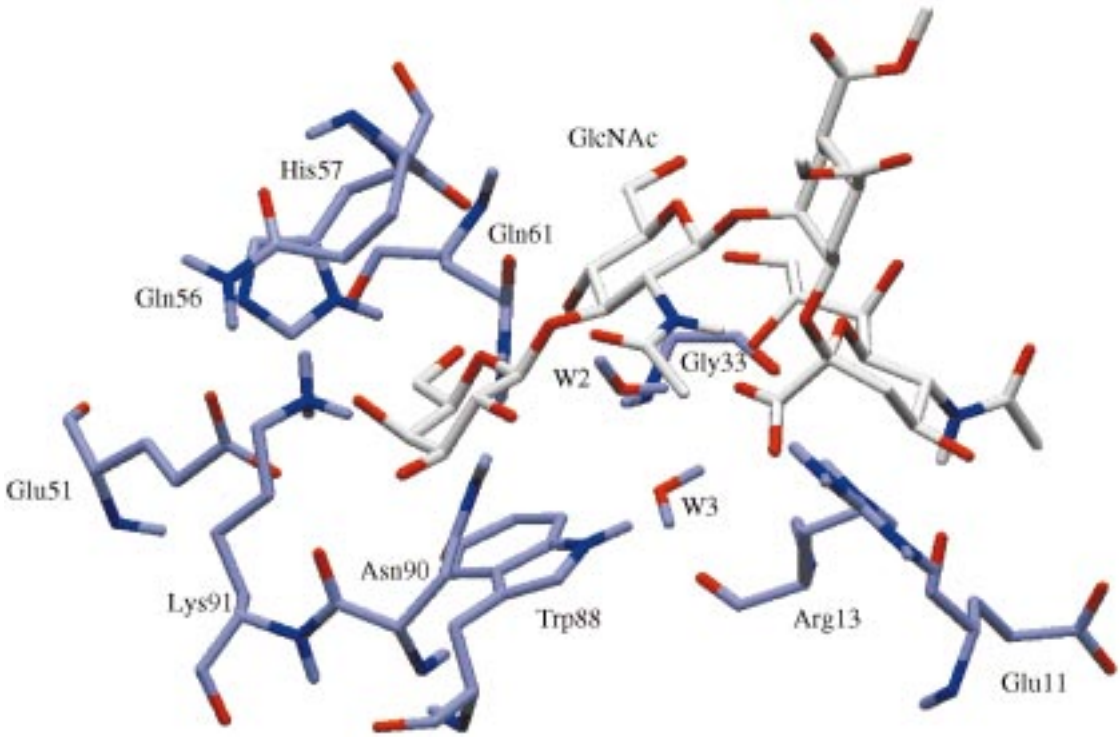


Figure 5. LT:**3** complex, as calculated by MC/EM.

figurations. These were obtained by equilibrating the starting structure for 270, 360, 450 and 900 ps. All simulations (both for the complex and the solvated ligands) were run by mutating **3** into **2**. Each simulation required ca. 120 CPU h on an SGI R10000 Origin machine.

Results and discussion

MC/EM conformational searches

Conformational analysis of **3** was performed both for the isolated molecule and in the binding pocket of LT, and the predicted three - dimensional structures were compared to those of **2**.

The conformational search of isolated **3** was run for 30,000 steps simulating water solvation with the GB/SA model, and yielded 18 conformers within 10 kJ mol^{-1} from the global minimum. They all shared the common conformation of the branched pseudo-trisaccharide fragment shown in Figure 3a. Comparison of the global minimum with the NMR solution structure of **2** [24] (Figure 3b) clearly shows that the two structures are completely superimposable, but for the 4-hydroxy group of the hexosamines. A more extensive comparison can be obtained by mapping the improper dihedral angles $\omega 1$ and $\omega 2$, which describe the relative position of the sialic acid carboxy group and of the galactose residue (Figure 4), for all the accessible conformations of **2** and **3** within 10 kJ/mol from the global minimum. The figure clearly shows that changing the hexosamine in the pseudo GM1 structure does not alter the relative position of the binding determinants. The reduced flexibility of **3** compared to **2** may result from H-bonding of the hexosamine 4-OH group to the oxygen bridge of the anomeric linkage in **3**.

The MC/EM analysis of the LT:**3** complex was also very promising. Within the first 12.5 kJ/mol 16 conformations were found, all featuring $\omega 1$ $-1° \pm 2°$ and $\omega 2$ $-175°$. The buried surface area of the protein binding-site in the lowest-energy conformation of the LT:**3** complex was calculated to be 333 $Å^2$. This compares well with the value of 347 $Å^2$ previously calculated for the LT:**2** complex [24] and for the experimental value of 403 ± 10 $Å^2$ for the CT:GM1 complex [21]. The global minimum of the MC/EM search is depicted in Figure 5, and a map of the interactions of **3** with LT, as seen in the global minimum, is reported in Figure 6. The corresponding map for LT:**2** (from ref 24) is reported in Figure 7. All the calculated low energy conformations were found to feature the expected substrate-protein contacts [20–22, 24]. Furthermore, the 4-OH group of GlcNAc was found to make H-bond contacts to the crystallographic water molecule at site 2 (W2 in Figure 5) in all conformations, and to Gln56 backbone carbonyl in conformers 3 (+1.1 kJ/mol) and 4 (+1.4 kJ/mol). Thus it appears that the C4 inversion of the hexosamine is actually improving the ability of the artificial binder to interact with the protein, by inserting the corresponding hydroxy group within the binding site and at interaction distance with it. However, compared to the LT:**2** complex, this hydroxy group is subtracted to the interaction with the external solvent, which may reflect negatively on the complex solvation. To further explore this issue and

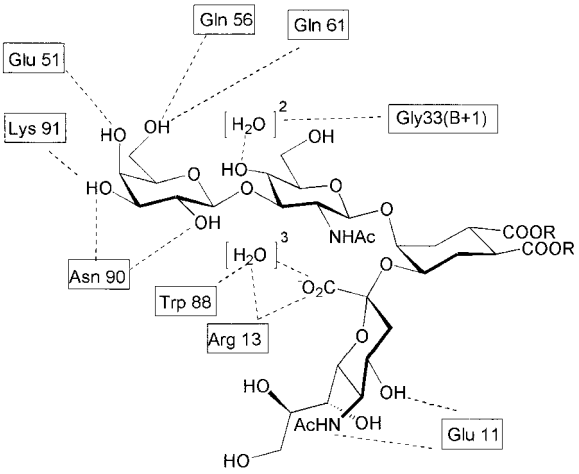


Figure 6. Protein substrate contacts in the global minimum of LT:**3**, as calculated by MC/EM.

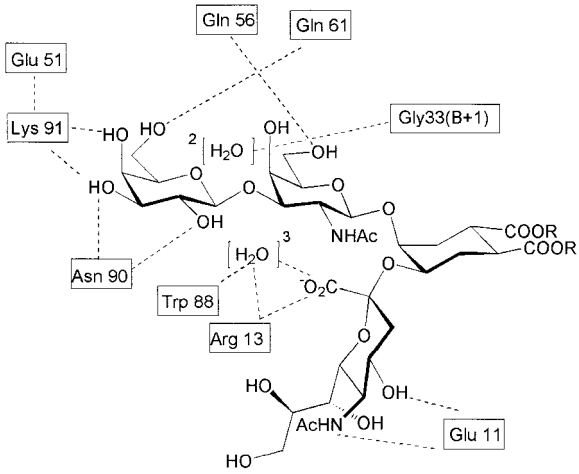


Figure 7. Interaction of **2** with LT (from ref 24).

to get a detailed picture of the solvent environment of the hexosamine in both receptors, we switched to an explicit solvent model, using TIP3P water, and performing molecular dynamics simulations of the isolated pseudosugars and of their toxin complexes.

Molecular dynamics simulations

Molecular dynamics simulations were run for 900 ps starting from the MC/EM global minimum conformation for both complexes, and using TIP3P water solvation as described in the method section. Both the inhibitors, docked in the binding pocket, experience a van der Waals contact and a stacking interaction between the cyclohexyl ring of the terminal galactose and the indol ring of Trp 88. This kind of hydrophobic interaction has proven to be one the major factors in determining recognition in a number of cases [41, 42] and our simulations seem to confirm these findings: the terminal galactose ring remains stacked on Trp 88 and in close contact with the side chains of Lys 99 and Ser 95 of the protein during the whole simulation.

In the LT:**2** complex (Figure 8a), the interactions found at the end of the MD simulations are not different from those found in the global minimum conformation for the MC/EM conformational search [24] and reported in Figure 7. The side chain amide group of Gln 61 is hydrogen bonded to the Gal-OH6 group and to the crystallographic water at site 2, W2. This water molecule, is also H-bonded to the NH group of Gly 33(B+1). The crystallographic water at site 3, W3, also remains in its starting position where it shares hydrogen bonding interactions with the side chains of NeuAc, Trp 88 and Arg 13 (not shown in Figure 8a). The axial 4-hydroxy group of the GalNAc residue of **2** remains exposed to the solvent where it is solvated and stabilized by three water molecules. The backbone carbonyl oxygen of Gln 56 is stabilized by a hydrogen bond to GalNAc-OH6. The side chain of Gln 56 resides in its original crystal structure position, where it is partially exposed to the water solvent and the side chain amide hydrogens are in contact with the backbone carbonyl oxygen of Val 52. We will show in the next part of this section that Gln 56 is involved in a rearrangement process of the recognition pocket upon binding of **3**.

The presence of a GlcNAc in the new synthetic inhibitor **3** complexed to the protein appears to cause a series of rearrangements in the binding pocket, which eventually determine a complex network of stabilizing hydrogen bonds. The side chain of Gln 56, in fact, moves from its original crystallographic position to form a hydrogen bond with the equatorial O4 of GlcNAc (Figure 8b). The time series of GlcNAc-O4 Gln56-HE2 distances collected over the last 600 ps of LT:**3** simulation (Figure 9b) clearly shows the transition to a new configuration, which does not appear in the corresponding LT:**2** series (Figure 9a). In the new position, the side chain of Gln 56 is stabilized by two water-mediated hydrogen bonding interactions with the HD atom of the imidazole ring of His 57 and with the O atom of the C6 hydroxyl group of GlcNAc (Figure 8b). Another hydrogen bond connects the amide hydrogens of the Gln 56 in the new orientation to the O4 of the terminal Gal. Furthermore, the hydrogen bond between the Gal-OH6 and the side chain of Gln 61 is conserved in the simulation of the LT:**3** complex, as it was in the previous case. During the simulation the crystallographic water W2 moves out of its original binding pocket towards the hydroxyl proton of the GlcNAc OH4 group. This fact is extremely interesting: W2 in fact, loses its favorable interaction with Gly 33(B+1), but becomes part of a second network of hydrogen bonds involving the 4-hydroxy group of GlcNAc, the carbonyl group of Gln 61 and two hydroxy groups of the NeuAc side chain. Thus, the whole complex gains an energetic stabilization due to the presence of the new, more extensive hydrogen bonding pattern. Moreover, there may be a synergic effect associated with the rearrangement of the Gln 56 previously described. All the other stabilizing interactions are not lost during the whole MD simulation. It should be pointed out that the crystallographic water at site 2 (W2) is not in a strongly preferred water binding site, according to GRID calculations [11], and it was found to be displaced by the ligand in the recently determined X-ray structure of the LT:m-nitrophenyl α-galactoside complex [43].

Thus the dynamic simulations confirm that the axial 4-hydroxy group of the hexosamine in **2** is indeed contributing to the solvation of the LT:**2** complex, by interacting with three water molecules. On the contrary, the presence of an equatorial 4-hydroxy group in the hexosamine of **3** appears to trigger a series of rearrangements and reorientations of the protein groups which can make the binding between the toxin and its inhibitor energetically more favorable.

In order to gain further information on the energetics of the binding process, the different solvation of the free inhibitors **2** and **3** was also taken into account. Molecular dynamics simulations were performed on the free inhibitors in a water box, and their hydration

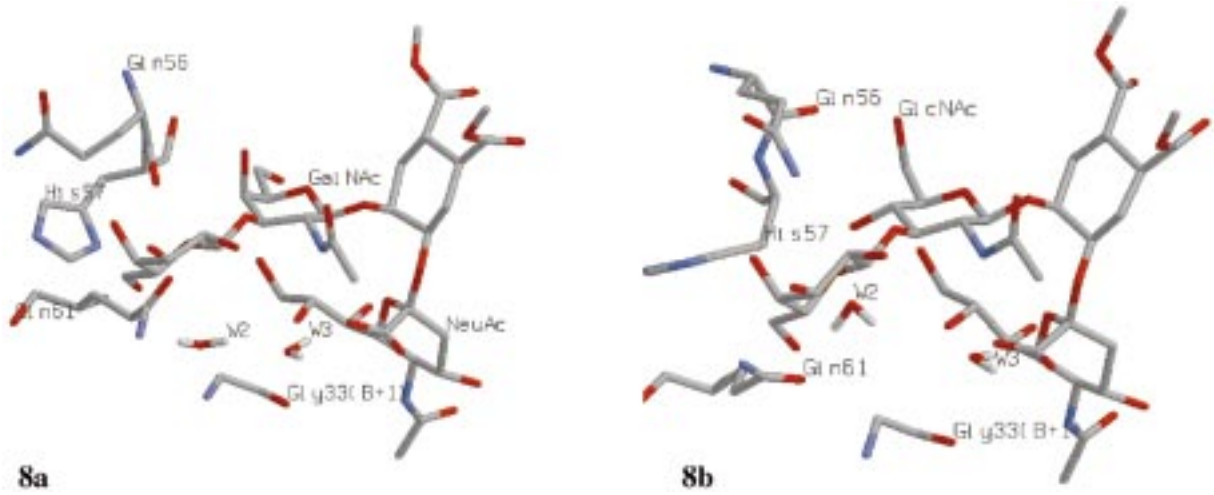


Figure 8. LT:ligand complexes as calculated by molecular dynamics. (a) LT:**2** complex; (b) LT:**3** complex.

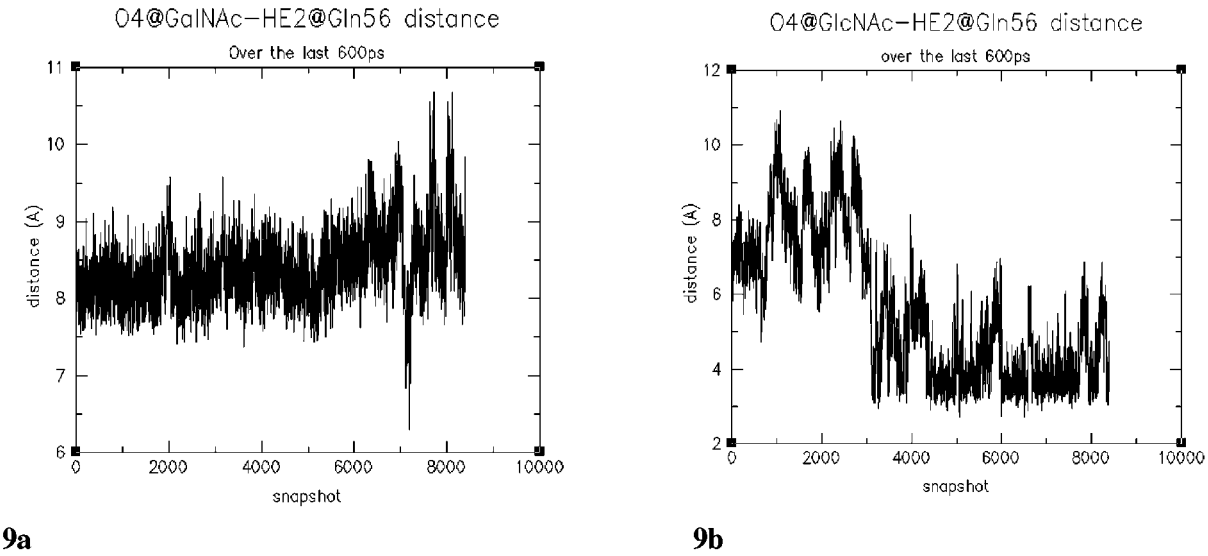
Table 1. Free energy perturbation results

Run	Initial config.[a]	ΔG_1 (kcal/mol)	ΔG_4 (kcal/mol)	$\Delta\Delta G_{3-2}$ (kcal/mol)
1	270	−2.01	+2.29	−4.30
2	360	−0.42	+3.28	−3.70
3	450	−1.54	+1.44	−2.98
4	900	−2.79	+1.60	−4.39
Average		-1.69 ± 0.99	$+2.15 \pm 0.88$	-3.84 ± 1.87

[a]Obtained by equilibrating the starting structure for 270, 360, 450 and 900 ps.

was evaluated from the analysis of the pair distribution functions (pdfs) [44]. Pdfs allow to obtain insights into the extent of water penetration and solvent ordering around atoms of particular interest, in this case GalNAc-O4 in **2** and GlcNAc-O4 in **3**. Our pdfs for the O4 of GalNAc in **2** and the oxygens of the surrounding water molecules show a high peak for the first solvation shell at 2.9 Å (solid line Figure 10), while in the case of **3** the first solvation peak shows a much lower intensity. This strongly suggests that the more exposed axial GalNAc-OH4 group of **2** is better solvated in water solution. The GlcNAc-OH4 group of **3** is not so exposed to the solvent because of its equatorial orientation and because it can have an intramolecular hydrogen bonding interaction with the oxygen atom on the anomeric linkage. The analysis of these solvation data suggests that, upon binding to the toxin, **2** may lose more solvation stabilizing interactions compared to the less solvated **3**, and this may make **2** a worse binder than **3**. Thus binding of **3** to the toxin should be favored compared to **2** not only because it gives rise to a larger number of stabilizing interactions through side chain rearrangement, but also because it involves a lower loss in solvation enthalpy, as seen with the pdfs analysis.

The simulations reported so far have provided us with qualitative, structural insight into the different binding modes of the two synthetic inhibitors. In order to obtain a quantitative prediction of the $\Delta\Delta G$ of binding, Free Energy Perturbation calculations were performed. The relative free energy of binding $\Delta\Delta G_{3-2}$ (Scheme 1) was obtained by calculating the free energies ΔG_1 and ΔG_4 for the non-physical processes which mutate ligand **3** into ligand **2** both in water solution (ΔG_1) and in the toxin complex (ΔG_4). Four separate simulations were run starting from four different starting configurations (see Methods). The relative free energy of binding calculated in each run are reported in Table 1 and lead to a calculated average free energy of binding ($\Delta\Delta G_{3-2}$) of -3.8 kcal/mol. Hence it appears that the presence of a hydroxyl group in the equatorial position of the hexosamine of **3** trig-



9a **9b**

Figure 9. Time series of the Gln56-HE2/hexosamine-O4 distance. (a) simulation of the LT-**2** complex; (b) simulation of the LT:**3** complex.

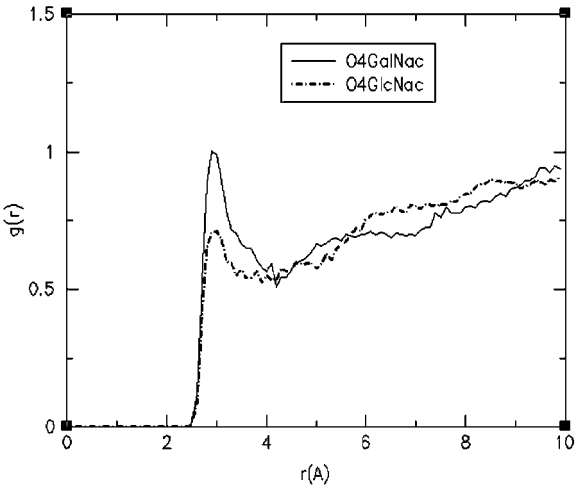
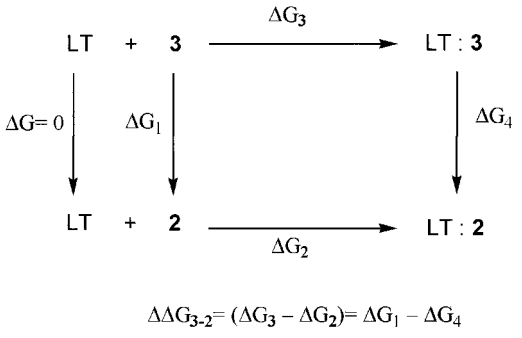


Figure 10. Pair distribution functions obtained from the dynamic simulations of **2** and **3** in TIP3P water.



$$\Delta\Delta G_{3\text{-}2} = (\Delta G_3 - \Delta G_2) = \Delta G_1 - \Delta G_4$$

Scheme 1. Thermodynamic cycle used for the FEP calculations.

gers a series of rearrangements and reorientations in the protein groups which can make the binding between the toxin and its inhibitor energetically more favorable. Exploiting this aspect of the recognition, together with keeping track of other important aspects like hydrophobic contacts, can help us design new and hopefully more powerful analogs of the natural receptors.

Conclusions

Computer-aided simulations of the interactions between the heat-labile toxin of *E. coli* LT and the two artificial receptors **2** and **3** has allowed a quantitative prediction of the relative affinity of the two ligands. Ligand **2** is a known GM1 mimic, which binds cholera toxin as efficiently as the natural receptor, ligand **3** was derived from **2** by inverting the C4 configuration of the hexosamine. The calculations predict that this inversion will improve the affinity of the ligand to the toxin by both increasing the number of binding interactions and reducing the solvent stabilization of the isolated ligand. Molecular dynamics simulations also predict that the binding of **3** to LT will occur together with side-chain rearrangement in the protein binding pocket [45]. The synthesis of **3** is currently in progress in our laboratories, and results of binding experiments with CT and LT will be reported in due course.

With the simulations described above we also introduce a promising method to simulate the docking of

flexible ligands to flexible receptors [46, 47]. Initially, the lowest energy conformations of the complex are located by sampling the conformational space of the ligand in the receptor through MC/EM analysis. The complex is then subjected to molecular dynamics simulations, and the effects due to the flexibility of side chains in the receptor binding pocket, the presence of solvent molecules, and the possible translational motions of small molecules (such as crystallographic water molecules) are accounted for. Compared to approaches based on molecular dynamics only [12], this protocol allows a more thorough exploration of the conformational space available to the substrate within the protein binding pocket, and hence should be particularly useful for carbohydrates epitopes, which are known to be often complexed in conformations which are not the most represented in solution [4–8].

Acknowledgements

This paper was supported by MURST, CNR and CILEA – Centro di Modellistica Computazionale.

References

1. Sears, P. and Wong, C.-H., Angew. Chem. Int. Ed., 38 (1999) 2300–2324, and references therein.
2. Davis, A.P. and Wareham, R.S., Angew. Chem. Int. Ed., 38 (1999) 2978–2996, and references therein.
3. For a discussion, see Rubinstenn, G., Sinay, P. and Berthault, P., J. Phys. Chem. A, 101 (1997) 2536–2540.
4. Cooke, R.M., Hale, R.S., Lister, S.G., Shah, G. and Weir, M.P., Biochemistry, 33 (1994) 10591–10596.
5. Scheffler, K., Ernst, B., Katopodis, A., Magnani, J.L., Wang, W.T., Weisemann, R. and Peters, T., Angew. Chem. Int. Ed., 34 (1995) 1841–1844.
6. Poppe, L., Brown, G.S., Philo, J.S., Nikrad, P.V. and Shah, B.H., J. Am. Chem. Soc., 119 (1997) 1727–1736.
7. Harris, R., Kiddle, G.R., Field, R.A., Milton, M.J., Ernst, B., Magnani, J.L. and Homans, S.W., J. Am. Chem. Soc., 121 (1999) 2546–2551.
8. Bundle, D.R. and Milton, M.J., J. Am. Chem. Soc., 120 (1998) 10547–10548.
9. Woods, R.J., Glycoconj. J., 15 (1998) 209–216.
10. Bernardi, A., Raimondi, L. and Zuccotto, F., J. Med. Chem., 40 (1997) 1855–1862.
11. Minke, W.E., Diller, D.J., Hol, W.G.J. and Verlinde, C.L.M.J., J. Med. Chem., 42 (1999) 1778–1788.
12. Pathiaseril, A. and Woods, R.J., J. Am. Chem. Soc. 122 (2000) 331–338.
13. Spangler, B.D., Microbiol. Rev., 56 (1992) 622–647, and references therein.
14. Lanne, B., Schierbeck, B. and Ångström, J., J. Biochem., 126 (1999) 226–234.
15. Lanne, B., Schierbeck, B. and Karlsson, K.A., J. Biochem., 116 (1994) 1269–1274.
16. Ångström, J., Tenenberg, S. and Karlsson, K.A., Proc. Natl. Acad. Sci. USA, 91 (1994) 11859–11863.
17. Schengrund, C.-L. and Ringler, N.J., J. Biol. Chem., 264 (1989) 13233–13237.
18. Schön, A. and Freire, E., Biochem., 28 (1989) 5019–5024.
19. Fukuta, S., Magnani, J.L., Twiddy, E.M., Holmes, R.K. and Ginsburg, V., Infect.Immun., 56 (1988) 1748–1753.
20. Merritt, E.A., Sarfaty, S., v.d.Akker, F., L'Hoir, C., Martial, J.A. and Hol, W.G.J., Protein Sci., 3 (1994) 166–175.
21. Merritt, E.A., Sarfaty, S., Jobling, M.G., Chang, T., Holmes, R.K., Hirst, T.R. and Hol, W.G.J., Protein Sci., 6 (1997) 1516–1528.
22. Merritt, E.A. and Hol, W.G.J., Curr. Opin. Struct. Biol., 5 (1995) 165–171, and references therein.
23. Acquotti, D., Poppe, L., Dabrowski, J., v.d.Lieth, C.-W., Sonnino, S. and Tettamanti, G., J. Am. Chem. Soc., 112 (1990) 7772–7778.
24. Bernardi, A., Checchia, A., Brocca, P., Sonnino, S. and Zuccotto, F., J. Am. Chem. Soc., 121 (1999) 2032–2036.
25. Mohamadi, F., Richards, N.G.J., Guida, W.C., Liskamp, R., Lipton, M., Caufield, C., Chang, G., Hendrickson, T. and Still, W.C., J. Comp. Chem., 11 (1990) 440–467.
26. Bernardi, A. and Raimondi, L., J. Org. Chem., 60 (1995) 3370–3377.
27. Still, W.C., Tempzyk, A., Hawley, R. and Hendrickson, T., J. Am. Chem. Soc., 112 (1990) 6127–6129.
28. Jorgensen, W.L., Chandrasekhar, J., Madura, J., Impey, R.W. and Klein, M.L., J. Chem. Phys., 79 (1983) 926.
29. Dixon, S.L. and Merz, K.M. Jr., J. Chem. Phys., 107 (1997) 879–893.
30. Woods, R.J., Dwek, R.A., Edge, C.J. and Frasier-Reid, B.J.P.C., J. Phys. Chem., 99 (1995) 3832–3846.
31. Case, D.A., Pearlman, D.A., Caldwell, J.C., Cheatham, T.E.I., Ross, W.S., Simmerling, C.L., Darden, T.A., Merz, K.M. Jr., Stanton, R.V., Cheng, A.L., Vincent, J.J., Crowley, M., Ferguson, D.M., Radmer, R.J., Seibel, G.L., Singh, U.C., Weiner, P. and Kollman, P.A., AMBER 5.0; University of California, San Francisco, (1997).
32. van Gunsteren, W.F. and Berendsen, H.J.C., Mol. Phys., 34 (1977) 1311.
33. Berendsen, H.J.C., Potsma, J.P.M., van Gunsteren, W.F., DiNola, A.D. and Haak, J.R., J. Chem. Phys., 81 (1984) 3684–3690.
34. Cheng, A. and Merz, K.M. Jr., J. Phys. Chem., 100 (1996) 1927–1937.
35. Smith, P.E. and Pettitt, B.M., J. Chem. Phys., 105 (1996) 4289–4293.
36. Merz, K.M. Jr., Murcko, M.A. and Kollman, P.A., J. Am. Chem. Soc., 113 (1991) 4484–4490.
37. Colombo, G., Toba, S. and Merz, K.M. Jr., J. Am. Chem. Soc., 121 (1999) 3486–3493.
38. Toba, S., Damodaran, K.V. and Merz, K.M.Jr, J.Med.Chem. 42, (1999), 1225–1234.
39. Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A., J. Am. Chem. Soc., 117 (1995) 5179–5197.
40. Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S. and Weiner, P., J. Am. Chem. Soc., 106 (1984) 765–784.
41. Massova, I. and Kollman, P.A., J. Am. Chem. Soc., 121 (1999) 8133–8143.
42. Singh, S.B. and Kollman, P.A., J. Am. Chem. Soc., 121 (1999) 3267–3271.
43. Minke, W.E., Roach, C., Hol, W.G.J. and Verlinde, C.L.M.J., Biochemistry, 38 (1999) 5684–5692.

44. Colombo, G. and Merz, K.M. Jr., J. Am. Chem. Soc., 121 (1999) 6895–6903.
45. Different binding modes to the same protein have been calculated for diastereomeric monosaccharides binding to the mannose-binding protein: Liang, G., Schmidt, R.K., Yu, H.-A., Cumming, D.A. and Brady, J.W., J. Phys. Chem., 100 (1996) 2528–2534.
46. Mangoni, M., Roccatano, D. and Di Nola, A., Proteins, 35, (1999) 153–162.
47. Wang, J., Kollman, P.A. and Kuntz, I.D., Proteins, 35, (1999) 1–19.