The Role of Waters in Docking Strategies with Incremental Flexibility for Carbohydrate Derivatives: Heat-Labile Enterotoxin, a Multivalent Test Case

Wendy E. Minke,[†] David J. Diller,[†] Wim G. J. Hol,[†,‡] and Christophe L. M. J. Verlinde*[,†]

Department of Biological Structure and Howard Hughes Medical Institute, Biomolecular Structure Center, University of Washington, Box 357742, Seattle, Washington 98195

Received August 12, 1998

Molecular docking studies of carbohydrate derivatives in protein binding sites are often challenging because of water-mediated interactions and the inherent flexibility of the many terminal hydroxyl groups. Using the recognition process between heat-labile enterotoxin from *Escherichia coli* and ganglioside GM_1 as a paradigm, we developed a modeling protocol that includes incremental conformational flexibility of the ligand and predicted water interactions. The strategy employs a modified version of the Monte Carlo docking program AUTODOCK and water affinity potentials calculated with GRID. After calibration of the protocol on the basis of the known binding modes of galactose and lactose to the toxin, blind predictions were made for the binding modes of four galactose derivatives: lactulose, melibionic acid, thiodigalactoside, and *m*-nitrophenyl-α-galactoside. Subsequent crystal structure determinations have demonstrated that our docking strategy can predict the correct binding modes of carbohydrate derivatives within 1.0 Å from experiment. In addition, it is shown that repeating the docking simulations in each of the seemingly identical binding sites of the multivalent toxin increases the chance of finding the correct binding mode.

Introduction

Heat-labile enterotoxin (LT) and cholera toxin (CT) are the major virulence factors of two diarrheal diseases that result in the deaths of several hundred thousand people each year.[1,2] LT and CT are produced by enterotoxigenic *Escherichia coli* and *Vibrio cholerae*, respectively. Both toxins are oligomeric AB_5 proteins, share 80% sequence identity, and have a similar mode of action. In the host the toxins attach to the mucosal cells of the intestinal system via the B-subunits, which recognize ganglioside GM_1 as a receptor on the cell surface (GM_1 = Galβ1-3GalNAcβ1-4{NeuAcα2-3}Galβ1-4Glcβ1-ceramide).[3,4] This recognition process triggers a chain of intracellular trafficking events that eventually result in the translocation of the toxin's A-subunit into the cytosol. There it ADP-ribosylates $G_{s\alpha}$,[5] which triggers the massive loss of fluids and ions from the host cell.

A way to prevent the toxic action of the A-subunit is to inhibit entry of the toxin into the host cells by blocking the binding of the B-subunit to the GM_1 receptor. The residues involved are almost all conserved between heat-labile enterotoxin and cholera toxin, the only exception being residue 13 which is an arginine in porcine LT (pLT) and a histidine in CT.[6] This makes the GM_1 binding site of the B-subunit an important target for the design of prophylactic drugs acting as B-subunit antagonists.

The crystal structure of the B-pentamer of CT in complex with the oligosaccharide part of the natural receptor ganglioside GM_1 (oligo-GM_1)[7] reveals that each of the five B-subunits provides a binding surface for one molecule of oligo-GM_1. In addition, Gly33 from a neighboring subunit assists in receptor binding through a solvent-mediated hydrogen bond. The crystal structure shows that the two terminal sugars of GM_1, galactose and sialic acid, contribute most to the interaction with the toxin, with a smaller contribution from the *N*-acetyl galactosamine residue. The galactose residue is buried deepest in the binding pocket and its binding mode is identical in all crystal structures of complexes solved so far (Table 1). It stacks neatly on top of a tryptophan (Trp88) and takes part in an extensive hydrogen-bonding network with the protein and with protein bound waters, providing specificity to the binding. Therefore, galactose was chosen as a lead for B-subunit antagonist design.

Table 1. Three-Dimensional Structures of LT and CT and Complexes

toxin−oligomeric state	B-ligand	resolution (Å)	PDB code
pLT−AB_5		1.95	1LTS[a]
pLT−AB_5	lactose	2.3	1LTT[b]
pLT−AB_5	galactose	2.2	1LTA[c]
pLT−AB_5	D-Gal-β1,3-GalNAc	2.13	1LTI[d]
CT−B_5	oligo-GM_1	2.2	1CHB[e]
CT−B_5		2.4	1FGB[f]
CT−AB_5		2.4	1XTC[g]

[a] See ref 8. [b] See ref 9. [c] See ref 7. [d] See ref 10. [e] See ref 11. [f] See ref 12. [g] See ref 13.

A structural analysis of the terminal galactose in the ganglioside binding site revealed sufficient space to allow for binding of C1 and C2 epimer derivatives of galactose. After a substructure search in the ACD-3D 95.2 database (MDL), 31 galactose derivatives with various substitutions at the C1 or C2 positions were purchased and tested for their ability to inhibit binding of LT and CT to GD_{1b}. In the ELISA assay ganglioside

* To whom correspondence should be addressed.
[†] Department of Biological Structure.
[‡] Howard Hughes Medical Institute.

10.1021/jm980472c CCC: $18.00
Published on Web 05/01/1999

GD_{1b} was used rather than the natural receptor GM_1 since binding of the toxins to GM_1 is very strong.[14] LT and CT bind, respectively, 10 and 20 times weaker to GD_{1b} than to GM_1,[15] and therefore, with GD_{1b} it is possible to identify inhibitors with lower affinity. It appeared that 15 of the initial 31 galactose derivatives showed inhibition of GD_{1b} receptor binding for both LT and CT, with IC_{50} values of the same order as or superior to that of galactose, the terminal sugar of the natural toxin receptor.[16]

The next step in the drug design process is to design a new set of B-subunit antagonists by combining information from the detailed binding modes from this original set of galactose derivatives. However, experimental determination of the binding modes by X-ray crystallography can be very time-consuming. Therefore, a fast and reliable computational method for determining the binding mode of the ligands in the GM_1 binding site would be of major importance. This paper describes steps to achieve that objective.

Conformational flexibility is a key factor to be considered in docking strategies.[17] Algorithms in which both receptor and ligand are kept rigid are relatively fast.[18] "Semi-flexibility", in which the ligand is flexible but the receptor remains rigid, is approached by Monte Carlo techniques,[19,20,41] genetic algorithms,[21−23] and incremental docking algorithms.[24,25] Flexibility of the receptor was considered by Jones and co-workers,[26] who allowed for some flexibility in the hydroxyls of the receptor, and by Leach,[27] who described an algorithm that uses discrete protein side chain flexibility. Recently, Caflisch and co-workers developed a strategy in which both the ligand and the receptor are completely flexible.[28] A rational decision as to the level of flexibility needed for our problem was reached as follows. The B-subunits of the many structures of LT and CT (Table 1) are very similar; for example the RMS deviation between the 515 B-subunit Cα's of pLT (porcine heat-labile enterotoxin) and that of the pLT:galactose complex is 0.4 Å.[29] Thus, protein flexibility can be ignored. On the other hand, sugars are very flexible ligands with respect to their many hydroxyls and require a thorough search of their spatial as well as conformational space. Therefore, a "semi-flexible" docking program is very suitable for attacking our problem. Many of these available "semi-flexible" docking programs generate multiple conformations of the ligand and subsequently use a rigid docking method. Because of the many hydroxyls in the carbohydrates, the number of stable conformations is too large to be considered in such an approach. Therefore we used another method.

AUTODOCK[30,31,19] performs automated docking of the whole ligand with user-specified dihedral flexibility within a rigid protein binding site. The program uses a Monte Carlo simulated annealing technique for configurational and translational exploration with a rapid energy evaluation and does not require subsequent energy minimization. Energies are evaluated via precalculated grids with molecular affinity potentials. Here we report on the use of AUTODOCK in two test cases, i.e., docking of galactose and lactose into their binding sites of pLT. In addition, AUTODOCK was used in four cases where the answer was not yet known when the calculations where carried out (their chemical structures

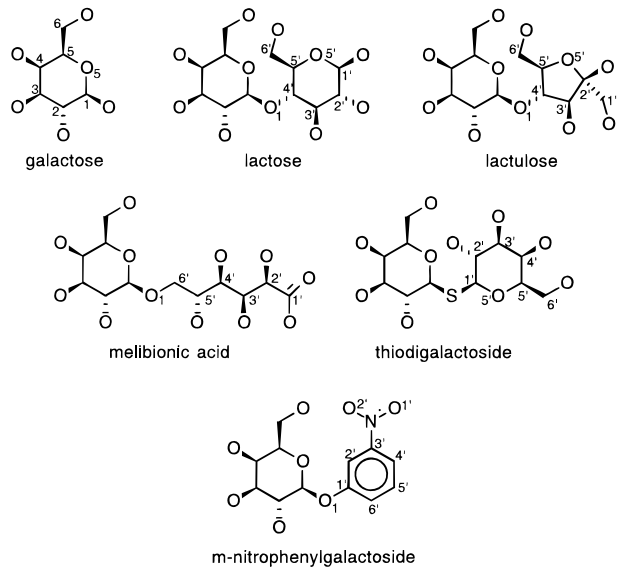

Figure 1. Galactose and galactose derivatives.

Table 2. IC_{50} Values of D-Galactose and Derivatives in the LT GD_{1b} ELISA[16]

D-galactose and derivatives	linkage	galactose substituent	LT-IC_{50} (mM)
galactose			57 (8)
lactose	1β−4	D-glucosyl	45 (10)
lactulose	1β−4	D-fructosyl	15 (3)
melibionic acid	1α−6	D-gluconic acid	7 (1)
thiodigalactoside	1β−1	*S*-galactosyl	4 (1)
m-nitrophenyl-α-galactoside	1α	*m*-nitrophenyl	0.6 (2)

are depicted in Figure 1, and the IC_{50} values for these compounds are listed in Table 2). Subsequently, cocrystallization experiments with these galactose derivatives were performed so that a critical evaluation of the docking strategy was possible.[32] The major finding of our computational experiments is that successful docking of carbohydrates and their derivatives critically depends on including water molecules.

Results and Discussion

Understanding of this section requires the definition of AUTODOCK terminology. Usually multiple Monte Carlo *runs* are performed in each AUTODOCK docking experiment. Each run provides one *solution*, i.e., a predicted binding mode. At the end of a docking job with multiple runs AUTODOCK performs cluster analysis: (1) solutions that are within 1 Å of each other are put in the same *cluster* and are referred to as *members* of that cluster; (2) in addition, clusters are *ranked* by the lowest energy representative of each cluster. Correct solutions are referred to as *hits*, which means that their RMS deviations from the experimental structure are under 1.0 Å. It should be noted that the AUTODOCK energies as reported in this paper are not directly related to true binding constants,[42] rather they are just a measure of the scoring function used.

Test Case 1: Galactose. Since the structure of pLT complexed with galactose was known at the start of this study, it was used as a test case for our docking strategy. The binding sites are very similar in all five B-subunits of the pLT:galactose complex. This is reflected in the RMS deviations of the atoms residing within 10 Å of

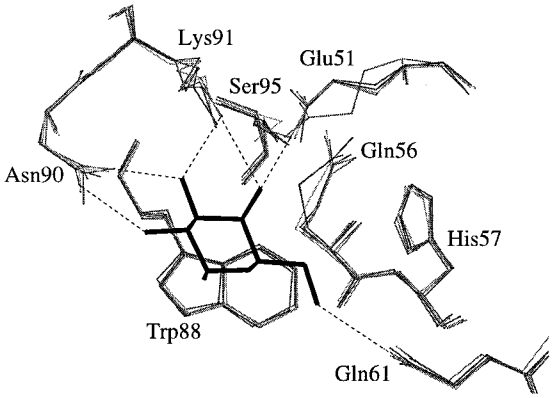

Figure 2. Superposition of galactose binding sites 1 through 5 of pLT:galactose. Sites 1, 3, 4, and 5 were superpositioned onto binding site 2 using the residues that reside within 5 Å of galactose. The galactose shown is from binding site 2. Figures 2−5 were produced with MOLSCRIPT.[33]

Table 3. pLT:Galactose Docking Results in the Five GM_1 Binding Sites in the Absence of Waters

binding site	no. of clusters[a]	rank of cluster	members in cluster[a]	lowest energy in cluster (kcal/mol)[b]	RMSD (Å)[c]
1	56	1	27	−57.87	0.64
		2	3	−54.65	2.87
2	69	1	10	−55.46	1.10
		2	15	−54.17	0.52
3	64	1	7	−50.94	1.34
		2	7	−50.85	0.76
4	63	1	2	−56.56	1.14
		2	23	−56.44	0.33
5	70	1	13	−56.26	0.40
		2	7	−55.26	1.15

[a] Clustering of a total of 128 runs. [b] The energies listed are as determined by AUTODOCK. [c] RMS deviation between the docked and the crystallographically determined galactose.

galactose, which range between 0.5 and 1.0 Å (Figure 2). In addition, the RMS deviations between the galactoses in the different subunits range only from 0.27 to 0.56 Å after superpositioning as described in the methods section. Also, the hydrogen bond pattern between the sugars and LT is similar in all subunits.[7] Finally, the galactose binding site accommodates five crystallographic water molecules that make at least two interactions to the protein or associated waters. The positions of these water molecules and their interactions are similar in each of the five binding sites, and three water molecules are involved in direct hydrogen bonds with galactose.[7] The 5-fold degeneracy of the pLT:galactose complex structure provides an excellent opportunity to investigate the effect of small differences in atomic positions on the results of ligand docking procedures.

Galactose was docked in all five binding sites in the absence of the crystallographic waters. The results differ substantially for each of the five binding sites (Table 3) which is remarkable in view of the small RMS deviations between the different binding sites and hints toward a rather steep scoring function. Binding sites 2, 3, and 4 yielded hits as second ranked solutions, while the first ranked solutions differed only slightly from the crystallographic binding mode. In contrast, in binding sites 1 and 5 the first ranked binding mode of galactose as suggested by AUTODOCK corresponded to the crystallographic one. Binding site 1 had the most members in the first ranked, correct solution and in addition had the largest energy difference between the first and second ranked solution. This suggested that the grid created around binding site 1 contained the best information to mimic real interactions and therefore all further docking experiments were done in binding site 1.

Before we discuss docking of galactose derivatives the ring conformation of galactose deserves some comments. In all docking experiments we kept galactose in the $^4C_1$ ring conformation for the following three reasons. One, this particular conformation was the only one observed in the crystal structures of pLT or CT with various galactose derivatives (galactose, lactose, T-antigen, GM1). Moreover, each of these structures allowed for five independent observations of the ring conformation because the toxin possesses five binding sites which are not related by crystallographic symmetry. Two, various experts on molecular conformation point out that it has been established that hexopyranoses clearly prefer the $^4C_1$ ring conformation, e.g., Allinger.[44] Three, our own survey of 1-substituted galactose derivatives in the Cambridge Crystallographic Database only revealed the $^4C_1$ conformation (entries: ADGALA01, ADGALA03, ADGALA10, BDGLOS01, BDGLOS10, CAGALA10, HCEREB, JEYDAS, NIBYUS). Hence, it seemed entirely justified to assume that the galactose ring conformation would be identical in all our docking studies.

Test Case 2: Lactose. Lactose was chosen as a more demanding test than galactose for docking strategies because of the higher degree of flexibility. Lactose is a disaccharide (Galβ1-4Glc) which is not a part of GM_1 but for which the experimental binding mode to pLT is known.[9] For lactose the protocol followed for galactose was modified in the approach toward flexibility of the hydroxyls. The large number of hydroxyls in lactose leads to 12 flexible bonds. Docking of ligands with more than eight flexible bonds rapidly becomes intractable[42] with Monte Carlo simulated annealing. Therefore, we decided to use an incremental strategy in order to reduce the conformational search space: (1) the hydroxyl dihedrals of the galactose moiety of lactose were selected to make optimal hydrogen bonds to the protein and kept rigid during the docking runs, (2) the bridging dihedrals between the two sugars were allowed to vary, and (3) the dihedrals of the glucose moiety were allowed to vary. This reduced the number of rotatable bonds during docking of lactose to seven. It was realized that using this approach docking of the galactose moiety is directed to its experimentally frequently observed binding mode. However, in all experimental structures of LT or CT with galactose containing ligands[7,9−11] the galactose moiety is observed in an identical binding mode.

The crystal structure of the pLT:lactose complex is known at 2.3 Å resolution.[9] The RMS deviations between the lactoses in the different binding sites range from 0.74 to 1.86 Å. A comparison with the lactose docking results was disappointing. The RMS deviation between the first ranked solution and the crystal structure was 3.49 Å (Table 4, Figure 3A), which is much larger than the RMS deviations between the lactoses in the different binding sites in the crystal structure. The first solution approximating the crystal

Table 4. pLT:Galactose Derivative Docking Results

galactose derivative	no. of waters	no. of clusters[a]	rank of cluster	members in cluster[a]	lowest energy in cluster (kcal/mol)[b]	RMSD (Å)[c]
lactose	0	85	1	8	−75.44	3.49
			2	2	−74.21	3.95
	4	94	1	10	−69.82	0.83
			2	5	−68.23	0.97
melibionic acid	0	128	1	1	−76.20	*d*
			2	1	−75.76	*d*
	2	127	1	1	−73.97	*d*
			2	1	−72.53	*d*
	4	128	1	1	−76.74	*d*
			2	1	−75.56	*d*
lactulose	0	91	1	3	−80.61	3.19
			2	4	−77.97	3.05
	2	100	1	2	−76.70	3.25
			2	2	−72.53	1.64
	4	94	1	2	−73.96	2.20
			2	6	−73.80	1.61
thiodigalactose	0	91	1	6	−66.81	2.67
			2	7	−65.77	3.70
	2	92	1	4	−65.29	2.51
			2	6	−65.05	1.71
	4	105	1	8	−64.07	0.74
			2	3	−61.93	1.39
m-nitrophenylgalactose	0	39	1	22	−72.63	2.60
			2	30	−72.31	0.67
	2	52	1	42	−69.03	1.51
			2	3	−63.66	5.34
	4	76	1	17	−67.24	1.61
			2	19	−65.20	3.36

[a] Clustering of a total of 128 runs. [b] The energies listed are as determined by AUTODOCK. [c] The RMSD values between the top ranked docked and the X-ray ligands are obtained after superpositioning the Cα's of two subunits of the corresponding crystal structure (B4B5 of pLT:lactose,[9] B1B2 of pLTI232T/Y233H:lactulose,[32] B3B4 of pLTB:thiodigalactose,[32] B4B5 of pLTB:nitrophenylgalactose[32]) onto B2B3 of pLT:galactose.[7] [d] The electron density for melibionic acid in pLTC236ins:melibionic acid[32] was not unambiguous, and the model was not refined.[32]

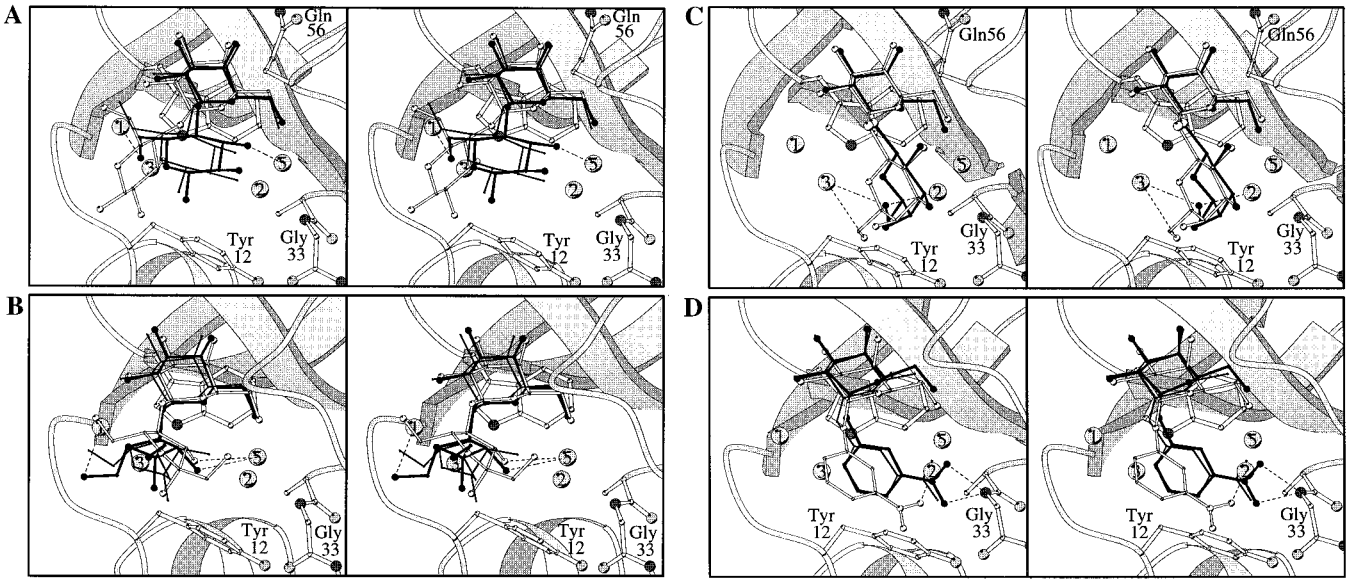

Figure 3. Superposition of experimental and docked structures of complexed LT. In all panels the ligand in open sticks is the proposed docked structure; the ligand in thick black sticks is the crystallographically determined structure; the ligand in thin black sticks is the docked structure that best resembles the experimental structure. Numbering of the waters is as suggested by Merritt and co-workers.[7] (A) Lactose in open sticks is the first ranked solution from the run without waters; lactose in thin black sticks is the first ranked solution from the run with four waters. (B) Lactulose in open sticks is the third ranked solution from the run with four waters; lactulose in thin black sticks is the fourth ranked solution from the run with four waters. (C) Thiodigalactoside in open sticks is the first ranked solution from the run with four waters, which is the correct binding mode. (D) Nitrophenyl-α-galactose in open sticks is the first ranked solution from the run with four waters; nitrophenylgalactose in thin black sticks is the second ranked solution from the run without waters.

structure was only ranked 17th and had a large energy difference of 11.51 kcal/mol with the first ranked solution (RMS deviation = 0.87 Å, two members in that cluster, $E = -64.93$ kcal/mol). Analysis of the first ranked, but wrong, binding mode revealed that the galactose residue was positioned reasonably well (RMS deviation = 1.07 Å), but the position of the docked glucose residue differed substantially from the crystal

Table 5. Properties of the Crystallographically Observed Water Molecules in the GM_1 Binding Site of LT

water molecule[a]	occurrences (max is 25)[b]	consensus hydrogen bonds[c]	avg B-factor (Å^2)[b]
1	14	Asn14: Oδ1/Nδ2[d] water 3	40
2	23	Gly33:N water 5	25
3	19	Arg13:O Trp88:Nε1 water 1	37
5	13	Gln56:O Gln61:Oε1 water 2	40

[a] Numbering of the waters is as suggested by Merritt and coworkers.[7] [b] In each of the five subunits of the following five structures: pLT,[8] pLT:lactose,[9] pLT:galactose,[7] pLT:T-antigen,[10] and CTB:oligo-GM_1.[11] [c] See Figure 6. [d] The hydrogen bond pattern does not allow for an unambiguous identification of these atoms. Either is a possibility.

structure (RMS deviation = 4.73 Å, Figure 3A). The difficulty in docking the glucose moiety correctly was traced to the absence of waters in the simulation, since (1) the docked lactose occupies the position of two crystallographically observed waters and (2) in the crystal structure O3′ and O6′of lactose are hydrogen-bonded to waters (Figure 3A). Therefore, we studied the effect of including waters into docking.

Test Case 3: Lactose and Waters. AUTODOCK docking simulations of lactose to binding site 1 were successful after adding the crystallographically determined waters to the protein coordinate file as described in the Methods section. The RMS deviation between the first ranked docked lactose and the crystal structure was 0.83 Å (Table 4, Figure 3A). Thus, docking of lactose is improved dramatically upon inclusion of crystallographic waters in the simulation.

Docking with and without Waters. Waters can play an important role in ligand binding. In the case of lactose binding to LT the waters direct the orientation of the glucose moiety through hydrogen bonds. However, for other ligands a docking strategy in which all crystallographic waters are kept could fail because certain waters might be displaced by functional groups. In the galactose binding sites of the pLT:galactose complex four crystallographically observed waters are conserved (numbering of the waters is as suggested by Merritt and coworkers[7]). To investigate whether it can be predicted which of the waters can be displaced from the GM_1 binding site of pLT, a statistical analysis based on the crystal structures was performed. The number of occurrences of each of the four waters in the five B-subunits of several LT structures was determined (Table 5). Waters 2 and 3 had the highest occurrence in the crystal structures and it was predicted that these waters were the most tightly bound in the GM_1 binding site and the hardest to be displaced by functional groups. Therefore, docking jobs with the galactose derivatives were performed in the absence of waters, in the presence of two (waters 2 and 3), and in the presence of all four water molecules.

Docking Galactose Derivatives. The information obtained from the galactose and lactose docking experiments was used to set up the docking simulations for the galactose derivatives with unknown binding modes. This means that (1) dihedrals in phenyls or saccharide rings were kept rigid during docking, (2) the dihedrals of the galactose moiety were optimized to make perfect hydrogen bonds to the protein and were kept rigid, (3) all other acyclic dihedrals were allowed to vary, and (4) the derivatives were docked in galactose binding site 1 of the pLT:galactose structure without waters, with waters 2 and 3, and with four waters. In Table 4 the results are presented for the four galactose derivatives studied: melibionic acid, lactulose, thiodigalactoside, and *m*-nitrophenylgalactoside. Assuming that one of the many solutions obtained from the docking runs represents the correct binding mode, the next challenge resides in its identification. We devised a protocol that may distinguish this true binding mode.

Table 6. pLT:Galactose Derivative Docking Results with GRID Derived Waters in the Protein

galactose derivative	no. of clusters[a]	rank of cluster	members in cluster	lowest energy in cluster (kcal/mol)[b]	RMSD (Å)[c]
melibionic acid	128	1	1	−79.61	*d*
		2	1	−79.35	*d*
lactulose	87	1	10	−73.08	1.62
		2	3	−69.64	3.08
thiodigalactose	84	1	9	−68.74	2.62
		2	5	−64.97	2.55
m-nitrophenyl-galactose	40	1	37	−77.31	0.56
		2	31	−69.42	2.19

[a] Clustering of a total of 128 runs. [b] The energies listed are as determined by AUTODOCK. [c] The RMSD values between the top ranked docked and the X-ray ligands are obtained after superpositioning the Cα's of two subunits of the corresponding crystal structure (B1B2 of pLT*I*232T/Y233H:lactulose,[32] B3B4 of pLTB:thiodigalactose,[32] B4B5 of pLTB:nitrophenylgalactose[32]) onto B2B3 of pLT:galactose.[7] [d] The density for melibionic acid in pLTC236ins: melibionic acid was ambiguous, and the model was not refined.[32]

Foremost, the previously described test cases showed that not only the first ranked binding modes should be considered as possible candidates. Rather, any solution with an energy close to the first ranked binding mode may represent the correct binding mode. Therefore, we decided to examine every solution within 5 kcal/mol of the first ranked solutions in all three docking experiments (without waters, with two waters, and with four waters).

Then, the results from the simulations with all waters were compared with the results from the experiments with missing waters. In the absence of waters the ligand can be positioned by AUTODOCK in a space that was occupied by waters in the crystal structure. Two steps are suggested to compare the results from the docking simulations with different waters. As a first step, their AUTODOCK energies were compared. We only considered solutions from the runs with missing waters as potential candidates when their energy was lower than the energy of the first ranked solution from the simulation with all waters. As a second step, the solutions were analyzed visually to examine whether the ligand was positioned in a space that was occupied by waters in the crystal structure. Therefore, only those binding modes were considered as possible candidates (1) in which the ligand did not occupy space that favorably holds a water or (2) in which the ligand replaced a water and in compensation formed good hydrogen bonds.

Finally, we focused on unsatisfied hydrogen bond donors and acceptors. AUTODOCK does not include a

Table 7. Flexible Dihedrals during the AUTODOCK Docking Studies[a]

inhibitor	rotatable bonds
galactose	C1−O1, C2−O2, C3−O3, C4−O4, C5−C6, C6−O6
lactose	C1−O1, O1−C4′, C5′−C6′, C6′−O6′, C1′−O1′, C2′−O2′, C3′−O3′
melibionic acid	C1−O1, O1−C6′, C6′−C5′, C5′−O5′, C5′−C4′, C4′−O4′, C4′−C3′, C3′−O3′, C3′−C2′, C2′−O2′, C2′−O1′
lactulose	C1−O1, O1−C4′, C5′−C6′, C6′−O6′, C2′−O2′, C2′−C1′, C1′−O1′, C3′−O3′
thiodigalactose	C1−S, S−C1′, C2′−O2′, C3′−O3′, C4′−O4′, C5′−C6′, C6′−O6′
m-nitrophenylgalactose	C1−O1, O1−C1′

[a] See Figure 1 for the assignment of atom names.

penalty for burying unsatisfied hydrogen bond donors or acceptors, which can lead to false positives. Therefore, after visual inspection, suggested binding modes were discarded if they involved burying of an unsatisfied hydrogen bond donor or acceptor.

After the docking studies were finished, the crystal structures were solved of complexes of pLT with melibionic acid, lactulose, thiodigalactoside, and *m*-nitrophenylgalactoside,[32] making a critical evaluation of the docking strategy possible. The crystallographically determined binding modes together with the proposed docked binding modes are depicted in Figure 3. The results of the docking experiments of each of the four galactose derivatives are summarized in Table 4, and a brief discussion of each of the four complexes is given in the following paragraphs.

Melibionic Acid. Melibionic acid is a disaccharide which is much more flexible than lactose: the gluconic acid is in the free carboxylic acid form at neutral pH rather than in a closed, lactone form and has therefore many degrees of freedom (Table 7). AUTODOCK did not find a single preferred solution. Most clusters only had one member in any of the three docking runs (without, with two, or with four waters), and none of them had more than two members. Most of the solutions had galactose positioned as observed in the crystal structures, but due to the diversity of proposed binding modes for the gluconic acid moiety, it was impossible to suggest a reasonable or single binding mode for melibionic acid.

The crystal structure of the pLT:melibionic acid complex was solved at 2.8 Å resolution.[32] The electron density for the galactose moiety was unambiguous. However, the density for the gluconic acid moiety was insufficient to resolve its conformation. These results suggest that the gluconic acid moiety of melibionic acid has multiple binding modes in pLT, which is also supported by fluorescence spectroscopy studies which indicate the existence of at least two binding modes.[16] Therefore, the results of the docking study with melibionic acid correspond with the experimental data: no single binding mode of melibionic acid in pLT is preferred.

Lactulose. Lactulose is a disaccharide composed of galactose linked via a β1−4 linkage to fructose. The fructose moiety can undergo mutarotation in solution, which results in a mixture of β-furanose, α-furanose, and β-pyranose forms. In the solid state this mixture corresponds to a ratio of 75:10:15, respectively,[34] whereas the crystalline trihydrate is purely in the β-furanose conformation.[35] Therefore, lactulose was docked in the latter conformation, as was later confirmed by the crystallographically determined structure. The solutions that passed the 5 kcal/mol cutoff requirement (16 solutions in total) were analyzed as described above. In most of the retained binding modes the fructose moiety did not make any specific interactions with the protein or the waters. However, the three top ranked binding modes of the docking run with four waters did have multiple, similar interactions between lactulose and the protein with its waters. In all three top ranked binding modes the fructose ring sits against Cα and Cβ of Gln56. The AUTODOCK energies of these three binding modes are really close to one another (−73.96, −73.80, and −73.36 kcal/mol). The second and third ranked binding modes each have six members, while the first ranked binding mode only has two members. The second ranked binding mode has one hydrogen bond of length 2.27 Å with an angle of 170°, between O1′ and the Ser55 backbone carbonyl. The third ranked binding mode has a slightly stronger hydrogen bond of length 2.10 Å with an angle of 167°, between O3′ and water 5. Therefore the third ranked binding mode of the docking job with four waters was proposed to be the correct binding mode (Figure 3B).

The crystal structure of the pLTI232T/Y233H:lactulose complex was solved at 2.65 Å resolution.[32] The average RMS deviation between the lactuloses in the different B-subunits is 1.0 Å (0.6 Å for the galactose moiety and 1.2 Å for the fructose moiety) and reflects a superposition of several different orientations of the furanose ring with respect to the protein molecule. Binding site 1 had the clearest electron density, and the lactulose model from this site was used for comparison with the modeled ligand. The crystal structure showed that the C1−O1 dihedrals were similar within 10° but the O1−C4′ dihedrals differed by about 50° between the modeled and experimental lactulose binding modes, which is reflected in the RMS deviation of 1.68 Å (Figure 3B). In the crystal structure the hydrogen bond between O3′ and water 5 is observed as it was predicted in the model. However the fructose ring is rotated in the crystal structure so it does not have any hydrophobic interactions but forms an additional hydrogen bond between fructose O6′ and water 1. This observed binding mode was found by AUTODOCK as the fourth ranked binding mode in the docking job with four waters present ($E = -72.43$ kcal/mol, members in cluster = 4, RMSD = 0.74 Å, Figure 3B). Thus, AUTODOCK was able to find the correct binding mode with an energy value just 1.5 kcal/mol higher than the energy value of the first ranked binding mode.

Thiodigalactoside. Thiodigalactoside is a symmetric molecule: the β-anomeric C1's of two galactose moieties are linked via a sulfur atom (Figure 1). This disaccharide represents an interesting case since our docking strategy includes freezing of the dihedrals of the galactose anchor while leaving the dihedrals of the C1 substituent flexible. Therefore, in this case AUTODOCK can position the galactose moiety with the flexible dihedrals into the galactose pocket, while leaving the

C1 substituent galactose with the fixed hydroxyl dihedrals to be positioned elsewhere. The galactose moiety with the fixed dihedrals appeared to be positioned in the galactose pocket for all 14 solutions that passed the 5 kcal/mol cutoff requirement. The binding modes were analyzed as previously described, and only one binding mode was left in which the flexible galactose moiety made distinct interactions with the protein or its bound waters: the first ranked binding mode from the docking job with four waters (Figure 3C). The sulfur interacts with Cβ and Cγ of Gln56, and O6′ forms a hydrogen bond with water 3. In addition this binding mode clearly is a representative of the largest cluster and has a much lower energy than the second ranked solution. Hence the first ranked binding mode from the runs with four waters present was predicted to be the correct binding mode.

The 1.7 Å resolution crystal structure of the pLT: thiodigalactoside complex showed very clear density for both galactose moieties,[32] with RMS deviations ranging from 0.15 to 0.38 Å between the thiodigalactosides in the different B-subunits. For comparison with the modeled binding mode we used thiodigalactoside from binding site 3, since it has the lowest average B-factor. The RMS deviation between the experimental and the proposed binding mode was only 0.74 Å (Figure 3C). Thus, docking of thiodigalactoside was very successful.

Nitrophenylgalactoside. Since *m*-nitrophenyl-α-galactoside has the lowest IC_{50} of the galactose derivatives studied in this paper (Table 2), we expected that docking of this compound would be relatively straightforward. Again, solutions that were within the 5 kcal/mol cutoff were analyzed, a total of 11 binding modes. After initial filtering as described previously, the remaining binding modes were compared. The only binding mode in which the nitrophenyl ring shows interaction with the protein or its associated waters is the first ranked binding mode from the docking run with four waters. In this binding mode, the nitro group forms a hydrogen bond with water 2 (Figure 3D). However, this hydrogen bond is the only interaction of the nitrophenyl group, and therefore this proposed binding mode does not explain the 100-fold improvement in IC_{50} over galactose.

The 2.2 Å crystal structure showed that the nitro group of *m*-nitrophenylgalactoside replaced water 2 and was hydrogen-bonded to Gly33[32] (Figure 3D, RMS deviations ranging from 0.31 to 0.66 Å between the *m*-nitrophenylgalactosides in the different B-subunits). The distance between the nitro oxygen and the backbone nitrogen of Gly33 is 2.8 Å and the N−O···N angle is 113°. This hydrogen bond is likely to be relatively strong, since the nitro−O···H distance is 0.35 Å shorter than the average hydrogen bond distance of 2.30 (1) Å for nitro−O···H bonds.[43]

For comparison with the modeled binding mode we used *m*-nitrophenylgalactose from binding site 4, since it has the lowest average B-factor. The RMS deviation between the experimental and the proposed binding mode was 1.51 Å. Contrary to what was expected, the second ranked solution from the docking job without any waters was a hit. This second ranked, correct solution was not considered to be a possible candidate for the following reasons. First, the replaced water 2 has a relatively high number of occurrences in other complex structures (Table 5), which suggested this water to be bound very tightly. Second, in this second ranked binding mode the hydrogen bond to Gly33 was rather poor, with a distance of 2.3 Å between the backbone amide hydrogen and the nitro oxygen.

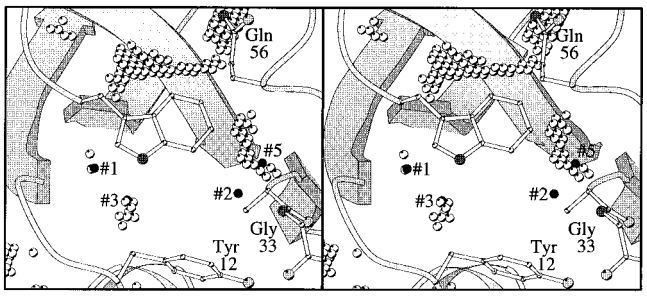

Figure 4. Results from GRID using a water molecule as the probe atom on a grid with 0.5 Å spacing. White balls represent gridpoints with GRID energies lower than −8 kcal/mol. Black balls represent water molecules observed experimentally in the pLT:galactose complex.

Docking with Predicted Waters. Further analysis of the waters was carried out since the *m*-nitrophenylgalactoside results were very unexpected, especially the displacement of water 2. Previously, only the number of occurrences of each of the four waters was considered a measure of the tightness of binding. Now, the following study approach was taken in order to explain why only water 2 was replaced by a functional group (Table 5). First, the hydrogen bonds each water molecule makes were tabulated. Second, the average B-factor of each of the waters in the five B-subunits of several structures was determined. Third, GRID energies[36,37] were determined for each gridpoint on a 0.5 Å grid in the GM_1 binding site of pLT:galactose in the absence of galactose. The gridpoints with energies lower than −8 kcal/mol are depicted in Figure 4. The experimental observations suggest that water 2 is least likely to be displaced: it has the highest number of occurrences. Whether the fact that it has the lowest average B-factor of all five water molecules in the binding site is a good predictor for ease of displacement is debatable because, in principle, B-factors do not reflect the depth of the potential but the width. However, the results obtained with GRID show that water 2 is not in a strongly preferred water binding site, in contrast with the other water molecules. Therefore, it appears that the wrong waters were chosen as candidates to be displaced by a functional group of a ligand.

The GRID results suggested that an AUTODOCK docking experiment should be carried out in the GM_1 binding site with waters obtained using the GRID program. A general protocol for obtaining GRID-derived waters is described in the Methods section. Using this strategy 12 GRID waters were included in the docking runs, with two waters directly in the GM_1 binding site (Figure 5): at the positions of waters 1 and 3. Indeed, an *m*-nitrophenylgalactoside AUTODOCK docking job with GRID waters in the binding site yielded the correct binding mode as the first ranked solution (Table 6). In addition, the energy for this binding mode is more than 10 kcal/mol better than the energy of the first ranked cluster in the *m*-nitrophenylgalactoside docking job with four observed waters. This large energy gain clearly

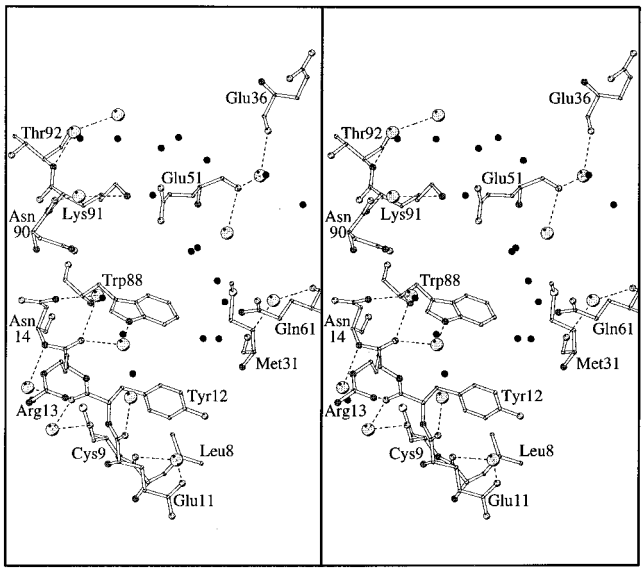

Figure 5. The galactose binding site with GRID waters. The GRID waters are depicted as light gray spheres. Their generation is described in the Methods section. For comparison, the experimental waters are shown as black spheres.

compensates for the enthalpy loss accompanying the release of water 2. This final docking study with *m*-nitrophenylgalactoside clearly showed that AUTODOCK is very capable of predicting the correct binding mode when the right information concerning waters is implemented in the docking process.

Finally, docking studies with these GRID waters were carried out with the other three ligands: melibionic acid, lactulose, and thiodigalactose (Table 6). For melibionic acid, the results are similar to those in any of the other docking jobs: no specific binding mode was proposed by AUTODOCK, suggesting that melibionic acid has multiple binding modes. For lactulose the energy of the first ranked binding mode in the docking job with the GRID waters is even higher than the energy of the first ranked binding mode in the docking job with four waters, suggesting that this is not a very likely binding mode, which is confirmed by an RMS deviation with the crystal structure of 2.20 Å. For thiodigalactoside the energy of the first ranked binding mode of the docking job with GRID waters is only 4.67 kcal/mol lower than the energy of the docking job with four observed waters. In addition, in this binding mode a potential hydrogen bond acceptor is buried (Oϵ1 of Gln61), and therefore the binding mode as suggested by AUTODOCK for thiodigalactoside in the docking job with GRID waters is not a potential binding mode. Thus, the use of GRID-derived waters in docking using AUTODOCK yields the expected results for the four galactose derivatives.

Conclusions

Our studies showed that AUTODOCK with our modification for waters can be a very powerful program in docking carbohydrates into crystallographically determined protein structures. The docking strategy described in this paper seems to indicate that useful information can be obtained in other cases when the following points are taken into consideration.

1. Whenever possible, the use of multiple binding sites obtained from multiple experimental target protein structures seems beneficial. AUTODOCK considers the protein to be rigid and does not even allow slight variations in atomic positions. This can result in failure of the docking procedure, especially because hydrogen atoms are added to the protein without experimental data. However, with the use of multiple target sites, chances are increased to have a proper representation of the binding site. In our case with the five binding sites it was shown that the scoring grid created around binding site 1 clearly contained the best information for AUTODOCK to find the correct binding mode for galactose.

2. A consequence of the multiple hydroxyls of carbohydrates is the high flexibility of the compound, leading quickly to a too large number of degrees of freedom. Therefore, we chose to use an incremental approach. First, one moiety of the ligand was docked while it was kept completely flexible. The best binding mode for this moiety was selected and the dihedrals were optimized for that binding mode. Subsequently a docking experiment was performed with the second moiety attached to the first, while the dihedrals of the first moiety were kept rigid and the dihedrals of the second moiety were allowed to rotate. In this way binding of the first moiety will be directed to its preferred binding pocket, but at the same time this binding pocket is not excluded for exploration by the second moiety.

3. Most importantly, due to the multiple hydroxyl functions of the carbohydrates, possible hydrogen bonds to water molecules should not be neglected since they can be crucial for carbohydrate binding, as was seen for lactose and thiodigalactoside. Thus, experimentally determined waters should be implemented in the docking strategy. On the other hand, waters can also be displaced by functional groups of a ligand. Therefore, it is probably best to perform docking studies in the presence of all water molecules, without any waters, and in the presence of a select group of waters. This select group of waters can be chosen by using the program GRID. In our case it appeared a posteriori that GRID was able to predict which waters have to be selected in order to propose a correct binding mode for *m*-nitrophenylgalactoside.

Multiple docking experiments, with and without waters, lead to a very large number of suggested binding modes. Therefore, the following strategy is suggested for selecting a subset of solutions to be analyzed: first, all solutions within 5 kcal/mol from the first ranked binding mode were selected for all simulations, and second, for the simulations without waters or with two waters the only solutions to be selected were the ones that have an energy below the energy of the first ranked binding mode of the simulations with four waters.

4. Visual analysis of the results is of great importance, most importantly perhaps, because the scoring function used in AUTODOCK does not include a penalty for burying unsatisfied hydrogen bond donors or acceptors. Also, in the case that a functional group of a ligand displaces a water molecule, the replacing protein−ligand interaction has to be similar to the replaced protein−water interaction. If this would not be the case, the protein would rather have a water than the ligand in the suggested binding mode.

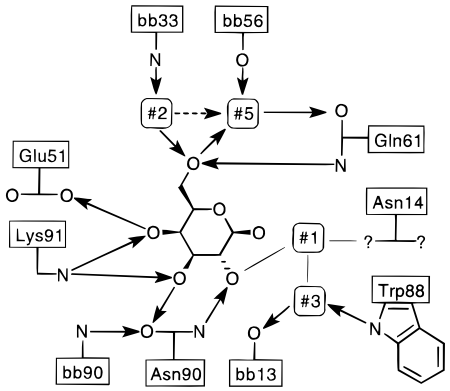

Figure 6. Consensus hydrogen-bonding pattern for the galactose binding pocket as obtained by summarizing the hydrogen bonds in the five binding sites of pLT:galactose and of CT:oligo-GM_1. Arrows denote hydrogen bonds of which the direction could be determined unambiguously; lines denote hydrogen bonds of which the direction is ambiguous. Solid lines/arrows denote atom to atom distances below 3.1 Å; dotted lines/arrows denote atom to atom distances between 3.1 Å and 3.5 Å. Question marks denote that the identity of Oδ1 and Nδ2 could not be determined unambiguously.

In conclusion, we believe that the general incorporation in modeling protocols of our four recommendations may lead to more successful and efficient predictions of carbohydrate binding modes.

Computational Methods

Protein Model. The coordinates of pLT complexed to galactose[7] were taken from the Brookhaven Protein Data Bank (PDB code 1LTA). Water molecules not located in the GM_1 binding site were discarded, leaving only water molecules 1, 2, 3, and 5 to be part of the protein (using the numbering as proposed by Merritt and co-workers[7]). With INSIGHTII (version 95.0, Biosym/MSI) hydrogens were added, and potentials and charges were assigned using the CVFF force field.[38]

To perform a docking experiment in the presence of waters, careful attention has to be paid to the positioning of the hydrogens on the waters. The hydrogen-bonding network of the four conserved waters in the galactose binding sites is not immediately obvious. Therefore the hydrogen bond patterns in 10 binding sites were studied, five each in the pLT:galactose complex[7] and the CT:oligo-GM_1 complex,[11] and a consensus hydrogen-bonding pattern was found for waters 2, 3, and 5 (Figure 6). Hydrogens were added to the waters and positioned to conform to the consensus hydrogen-bonding pattern. The hydrogens of water 1 were positioned to optimize interactions with the protein, which was different for each binding site. Then, the hydrogen positions were energy-minimized. The protein, galactose-heteroatoms and the water oxygens were positionally constrained, and a conjugate gradient algorithm was used with a distance-dependent dielectric constant ($\epsilon = 4r$) until the RMS derivative was smaller than 0.001 kcal mol^{-1} $Å^{-1}$.

Superpositioning of the GM_1 binding sites of the toxin is required for exploring virtually the same space around the five binding sites, as well as for RMS deviation calculations between ligands. Two subunits are involved in forming each GM_1 binding site; for example, subunits B1 and B2 form binding site 1. Therefore, the Cα's of the two subunits forming a particular binding site are superimposed onto the Cα's of subunits B2 and B3 of the pLT:galactose structure, using the superimpose routine in INSIGHTII.

Galactose Derivative Models. The galactose derivatives were built with the biopolymer module of INSIGHTII in conjunction with the AMBER force field modified for polysaccharides.[39] The newly constructed small molecules were energy-minimized using conjugate gradient with a distance-dependent dielectric constant ($\epsilon = 4r$) until the RMS derivative was smaller than 0.001 kcal mol^{-1} $Å^{-1}$.

The charges obtained from the AMBER force field were adequate for building the molecules but were not for docking, because charges for certain functional groups were missing; for example, the charges for the nitro group in *m*-nitrophenylgalactoside. New charges were obtained by semiempirical quantum mechanical calculations using the AM1 Hamiltonian in the program AMSOL.[40] Because AUTODOCK uses a united atom force field, the nonpolar hydrogen atoms were deleted from the coordinate file and their charges were merged with the charges of the carbons to which they are bonded (Table 8 in the Supporting Information).

AUTODOCK Docking Studies. All docking studies were performed with the program AUTODOCK (version 2.2.).[30] The target in each docking run was constructed of two neighboring B-subunits of pLT, including all four, two (waters 2 and 3), or no water molecules in the GM_1 binding site. Affinity grid files were generated using the auxiliary program AUTOGRID. C1 of galactose in binding site 2 of the crystal structure was chosen as the center of the grids, and the dimensions of the grid were $23 \times 20 \times 20$ $Å^3$ with grid points separated by 0.25 Å. Of the two sets of parameters supplied with the program, the original Lennard-Jones and hydrogen-bonding potentials of version 1.0 were used.

For docking in the presence of waters it was essential to modify the AUTOGRID program supplied with the AUTODOCK 2.2 package. The only oxygen type that AUTOGRID considers being a potential hydrogen bond acceptor is an oxygen bound to a carbon atom. Therefore, AUTOGRID was changed so it would also consider an oxygen bound to two hydrogen atoms as a potential hydrogen bond acceptor, allowing for proper treatment of water molecules.

The starting position of the small molecule was randomized as follows. First, the galactosyl moiety was superimposed onto its crystallographic counterpart of the pLT:galactose complex. Then the ligand was translated by at least 4 Å. Finally, the ligand was rotated 175° around the $x, y, z = 1, 1, 1$ axis. It was assumed that the galactose moiety of the derivatives would bind in a fashion similar to that of the galactose moieties observed in all pLT or CT complexes, since in these experimental complexes the galactose was always positioned in the same Trp88 pocket making identical interactions. Therefore, the dihedrals of the galactose moiety of the new ligand were adjusted in such a way that they could make the same optimal hydrogen bonds with the protein as interpreted from the crystal structures. This resulted in the following dihedral angles: C1−C2−O2−HO2 = 60°; C2−C3−O3−HO3 = −65°; C3−C4−O4−HO4 = −117°; C4−C5−C6−O6 = 158°; C5−C6−O6−HO6 = −55°. During docking these dihedrals were kept rigid, whereas other dihedrals were allowed to rotate. Table 7 gives an overview of the rotatable dihedrals.

The parameters for the AUTODOCK Monte Carlo simulated annealing configurational search were identical for all docking jobs. Each simulated annealing run lasted 110 cycles with a starting temperature of 403 K in the first cycle and a temperature reduction factor of 0.95 $cycle^{-1}$. The maximum number of accepted or rejected steps per cycle was 30000. The maximum translation step was kept constant at 0.2 Å for every cycle, but the maximum quaternion rotation step and dihedral rotation step had a reduction factor of 0.99 $cycle^{-1}$ starting from a maximum rotation of 5.0° in the first cycle. In order for the search to be extensive, the program uses a multiple start approach in combination with a time-dependent random number generator. The maximum allowed number of runs was used (128).

After docking, the 128 solutions were clustered in groups with RMS deviations lower than 1.0 Å. The clusters were ranked by the lowest energy representative of each cluster. Unsuspecting users should be warned about the default method for clustering in AUTODOCK version 2.2, based on atom similarity. This means that the distance between an atom of the first structure and the nearest atom of identical atom type of the second structure is used to determine the

RMS deviation, rather than the distance between identical atoms. In most cases the default clustering leads to nonsense results and should be turned off, especially for sugars which are substituted with identical functional groups all over.

The docking experiments were performed on a DEC 4100/400 with three processors and 512 MB of memory. The CPU time varied from 8 h and 6 min for docking galactose to 16 h and 43 min for docking melibionic acid.

GRID Waters. The program GRID version 15[36,37] was used for two different purposes. First, GRID was used to determine the interaction energies between crystallographically observed waters and the protein. Second, GRID was used to find energetically favorable water binding sites which were used in the AUTODOCK studies.

To find these favorable water binding sites, we started by calculating the GRID energies for a water probe on a 23 × 20 × 20 Å^3 grid with 0.5 Å spacing around binding site 1 of the pLT:galactose structure. Figure 4 shows that the gridpoints with energies below −8 kcal/mol form many distinct groups; a group is defined as all gridpoints with energies below −8 kcal/mol that are within 1.4 Å of one another (1.4 Å is the van der Waals radius for a water probe used by GRID). Any group that has a radius larger than 2.8 Å (2 times the van der Waals radius of water) represents a volume that does not contain a very localized water and is therefore not considered to contain a crucial water binding site for docking. For each group with a radius smaller than 2.8 Å, we selected the gridpoint with lowest energy. If those lowest energy gridpoints were within 2.8 Å of one another, the highest energy gridpoint was discarded. The remaining gridpoints were considered to be water sites, and hydrogens were added to the water oxygen. The hydrogen positions were optimized in five cycles of molecular dynamics and subsequent energy minimization. During these cycles only the water hydrogens were allowed to move, using a distance-dependent dielectric constant ($\epsilon = 4r$) and the CVFF force field. Each molecular dynamics run consisted of 100 equilibration cycles and subsequently 1000 1 fs steps at 300 K, with the Verlet velocity algorithm. For each energy minimization the conjugate gradient algorithm was used until the RMS derivative was below 0.001 kcal mol^{-1} Å^{-1}. For each water molecule the lowest energy conformation of the hydrogens was selected after the five cycles of molecular dynamics and energy minimization. Using this procedure twelve GRID waters were obtained that were implemented in docking (Figure 5).

Acknowledgment. We are very grateful to Arthur Olson and his colleagues at the Scripps Research Institute for providing AUTODOCK and to Peter Goodford for providing GRID. Also we thank Ethan Merritt for access to unpublished crystallographic data. This research received financial support from the National Institutes of Health through a grant to C.L.M.J.V. (GM54618). We are indebted to the Murdoch Charitable Trust for a major equipment grant to the BMSC.

Supporting Information Available: Charges obtained by the program Amsol[40] and used for the AUTODOCK docking studies for galactose, lactose, melibioinic acid, lactulose, thiodigalactoside, and *m*-nitrophenyl-α-galactoside. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Black, R. E. The epidemiology of cholera and enterotoxigenic *E. coli* diarrheal disease. In *Developments of Vaccines and Drugs Against Diarrhea*, 11th Nobel Conference, Stockholm, 1985; Holmgren, J., Lindberg, A., Möllby, R., Eds.; Studentlitteratur: Lund, 1986; pp 23−32.

(2) Wanke, C. A.; Lima, A. A.; Guerrant, R. L. Infectious diarrhea in tropical and subtropical regions. *Baillieres Clin. Gastroenterol.* **1987**, *1*, 335−359.

(3) Van Heyningen, W. A.; Carpenter, C. C. J.; Pierce, N. F.; Greenough, W. B., III. Deactivation of cholera toxin by ganglioside. *J. Infect. Dis.* **1971**, *124*, 415−418.

(4) Cuatrecasas, P. Interaction of *Vibrio cholerae* enterotoxin with cell membranes. *Biochemistry* **1973**, *12*, 3547−3558.

(5) Cassel, D.; Pfeuffer, T. Mechanism of cholera toxin action: Covalent modification of the guanyl nucleotide-binding protein of the adenylate cyclase system. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 2669−2673.

(6) Domenighini, M.; Pizza, M.; Jobling, M. G.; Holmes, R. K.; Rappuoli, R. Identification of errors among database sequence entries and comparison of correct amino acid sequences for the heat-labile enterotoxins of *Escherichia coli* and *Vibrio cholerae*. *Mol. Microbiol.* **1995**, *15*, 1165−1167.

(7) Merritt, E. A.; Sixma, T. K.; Kalk, K. H.; Van Zanten, B. A. M.; Hol, W. G. J. Galactose-binding site in *Escherichia coli* heat-labile enterotoxin (LT) and cholera toxin (CT). *Mol. Microbiol.* **1994**, *13*, 745−753.

(8) Sixma, T. K.; Kalk, K. H.; Van Zanten, B. A. M.; Hol, W. G. J. Refined structure of *Escherichia coli* heat-labile enterotoxin, a close relative of cholera toxin. *J. Mol. Biol.* **1993**, *230*, 890−918.

(9) Sixma, T. K.; Pronk, S. E.; Kalk, K. H.; Van 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.

(10) Van den Akker, F.; Steensma, E.; Hol, W. G. J. Tumor marker disaccharide D-Gal-β1,3-GalNAc complexed to heat-labile enterotoxin from *Escherichia coli*. *Protein Sci.* **1996**, *5*, 1184−1188.

(11) Merritt, E. A.; Sarfaty, S.; Van den Akker, F.; L'Hoir, C.; Martial, J. A.; Hol, W. G. J. Crystal structure of cholera toxin B-pentamer bound to receptor GM_1 pentasaccharide. *Protein Sci.* **1994**, *3*, 166−175.

(12) Zhang, R.-G.; Westbrook, M. L.; Westbrook, E. M.; Scott, D. L.; Otwinowski, Z.; Maulik, P. R.; Reed, R. A.; Shipley, G. G.; The 2.4 Å crystal structure of cholera toxin B subunit pentamer: choleragenoid. *J. Mol. Biol.* **1995**, *251*, 550−562.

(13) Zhang, R.-G.; Scott, D. L.; Westbrook, M. L.; Nance, S.; Spangler, B. D.; Shipley, G. G.; Westbrook, E. M. The three-dimensional structure of choleratoxin. *J. Mol. Biol.* **1995**, *251*, 563−573.

(14) Schengrund, C. L.; Ringler, N. J. Binding of *Vibrio cholerae* toxin and the heat-labile enterotoxin of *Escherichia coli* to GM_1, derivatives of GM_1, and nonlipid oligosaccharide polyvalent ligands. *J. Biol. Chem.* **1989**, *264*, 13233−13237.

(15) Sugii, S.; Tsuji, T. Binding specificities of heat-labile enterotoxins isolated from porcine and human enterotoxygenic *Escherichia coli* for different gangliosides. *Can. J. Microbiol.* **1989**, *35*, 670−673.

(16) Minke, W. E.; Roach, C.; Hol, W. G. J.; Verlinde, C. L. M. J. Structure-based exploration of the ganglioside GM1 binding sites of *E. coli* heat-labile enterotoxin and cholera toxin for the discovery of receptor antagonists. *Biochemistry* **1999**, *38*, 5684−5692.

(17) Verlinde, C. L. M. J.; Hol, W. G. J. Structure-based drug design: progress, results and challenges. *Structure* **1994**, *2*, 577−587.

(18) Kuntz, I. D.; Meng, E. C.; Shoichet, B. K. Structure-based molecular design. *Acc. Chem. Res.* **1994**, *27*, 117−123.

(19) Goodsell, D. S.; Morris, G. M.; Olson, A. J. Docking of flexible ligands. Applications of AUTODOCK. *J. Mol. Recognit.* **1996**, *9*, 1−5.

(20) Hart, T. N.; Read, R. J. Multiple-start Monte Carlo docking of flexible ligands. In *The protein folding problem and tertiary structure prediction*; Merz, K. M., Jr., Le Grand, S. M., Eds.; Birkhäuser: Boston, 1994; pp 71−108.

(21) Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Herman, C. J.; Fogel, D. B.; Fogel, L. J.; Freer, S. T. Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. *Chem. Biol.* **1995**, *2*, 317−324.

(22) Judson, R. S.; Jaeger E. P.; Treasurywala A. M. A genetic algorithm based method for docking flexible molecules. *J. Mol. Struct.* **1994**, *308*, 191−206.

(23) Oshiro, C. M.; Kuntz, I. D.; Dixon, J. S. Flexible ligand docking using a genetic algorithm. *J. Comput. Aided Mol. Des.* **1995**, *9*, 113−130.

(24) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470−489.

(25) Welch, W.; Ruppert, J.; Jain, A. N. Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites. *Chem. Biol.* **1996**, *3*, 449−462.

(26) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727−748.

(27) Leach, A. R. Ligand docking to proteins with discrete side-chain flexibility. *J. Mol. Biol.* **1994**, *235*, 345−356.

(28) Apostolakis, J.; Plückthun, A.; Caflisch, A. Docking small ligands in flexible binding sites. *J. Comput. Chem.* **1998**, *19*, 21−37.

(29) Hol, W. G. J.; Sixma, T. K.; Merritt, E. A. Structure and function of *E. coli* heat-labile enterotoxin and cholera toxin B pentamer. In *Bacterial toxins and virulence factors in disease. Handbook of natural toxins, Volume 8*; Moss, J., Iglewski, B., Vaughan, M., Tu, A. T., Eds.; Marcel Dekker Inc., 1995; pp 185−223.
(30) Goodsell, D. S.; Olson, A. J. Automated docking of substrates to proteins by simulated annealing. *Proteins: Structure, Function and Genetics* **1990**, *8*, 195−202.
(31) Goodsell, D. S.; Lauble, H.; Stout, C. D.; Olson, A. J. Automated docking in crystallography: Analysis of the substrates of aconitase. *Proteins* **1993**, *17*, 1−10.
(32) Merritt, E. A.; Sarfaty, S.; Feil, I. K.; Hol, W. G. J. Structural foundation for the design of receptor antagonists targeting *Escherichia coli* heat-labile enterotoxin. *Structure* **1997**, *5*, 1485−1499.
(33) Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **1991**, *24*, 946−950.
(34) Jeffrey, G. A.; Wood, R. A.; Pfeffer, P. E.; Hicks, K. B. Crystal structure and solid-state NMR analysis of lactulose. *J. Am. Chem. Soc.* **1983**, *105*, 2128−2133.
(35) Jeffrey, G. A.; Huang, D. B.; Pfeffer, P. E.; Dudley, R. L.; Hicks, K. B.; Nitsch, E. Crystal structure and NMR analysis of lactulose trihydrate. *Carbohydr. Res.* **1992**, *226*, 29−42.
(36) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.
(37) Wade, R. C.; Goodford, P. J. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 2. Ligand probe groups with the ability to form more than two hydrogen bonds. *J. Med. Chem.* **1993**, *36*, 148−156.
(38) Dauber-Osguthorpe, P.; Roberts, V. A.; Osguthorpe, D. J.; Wolff, J.; Genest, M.; Hagler, A. T. Structure and energetics of ligand binding to proteins: *E.coli* dihydrofolate reductase-trimethoprim, a drug-receptor system. *Proteins: Struct., Funct., Genet.* **1988**, *4*, 31−47.
(39) Homans, S. W. A molecular mechanical force field for the conformational analysis of oligosaccharides: comparison of theoretical and crystal structures of Manα1−3Manβ1−4GlcNAc. *Biochemistry* **1990**, *29*, 9110−9118.
(40) Cramer, C. J.; Truhlar, D. G. An SCF solvation model for the hydrophobic effect and absolute free energies of aqueous solvation. *Science* **1992**, *256*, 213−217.
(41) McMartin, C.; Bohacek, R. S. QXP: powerful, rapid algorithms for structure-based design. *J. Comput. Aided Mol. Des.* **1997**, *11*, 333−344.
(42) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639−1662.
(43) Allen, F. H.; Baalham, C. A.; Lommerse, J. P. M.; Raithby, P. R.; Sparr, E. Hydrogen-bond acceptor properties of nitro-O atoms: A combined crystallographic database and ab initio molecular orbital study. *Acta Crystallogr.* **1997**, *B53*, 1017−1024.
(44) Burkert, U.; Allinger, N. L. *Molecular Mechanics*; ACS Monograph 117; American Chemical Society: Washington, DC, 1982.

JM980472C