ARTICLE

OBC
www.rsc.org/obc

Mimics of ganglioside GM1 as cholera toxin ligands: replacement of the GalNAc residue †

Anna Bernardi,*[a] Daniela Arosio,[a] Leonardo Manzoni,[b] Diego Monti,[b] Helena Posteri,[a] Donatella Potenza,[a] Silvia Mari[a] and Jesús Jiménez-Barbero[c]

[a] *Universita' di Milano – Dipartimento di Chimica Organica e Industriale e Centro di Eccellenza CISI, via Venezian 21, 20133 Milano, Italy*

[b] *CNR-Istituto di Scienze e Tecnologie Molecolari (ISTM), Via Golgi 19, 20133 Milano, Italy*

[c] *Dept. de Estructura y Función de Proteínas, Centro de Investigaciones Biológicas CSIC, Velázquez 144, 28006 Madrid, Spain*

Received 28th October 2002, Accepted 13th December 2002
First published as an Advance Article on the web 5th February 2003

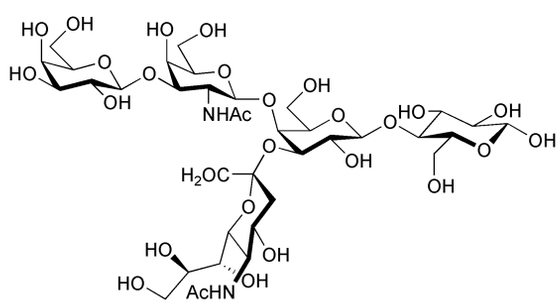
Two new cholera toxin (CT) ligands (**4** and **5**) are described. The new ligands were designed starting from the known GM1 mimics **2** and **3** by replacement of their GalNAc residue with the C4 isomer GlcNAc. As predicted by molecular modelling, the conformational properties of the equivalent pairs **2**–**4** and **3**–**5** are very similar and their affinity for CT is of the same order of magnitude. NMR experiments have also proved that **5** occupies the GM1-binding site of the toxin and have revealed its bound conformation.

Introduction

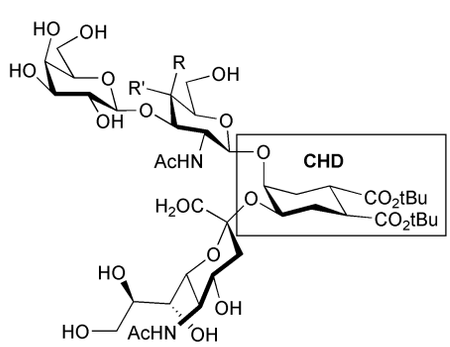
Carbohydrate–protein complexes are formed in the initial steps of a large number of physiological and pathological processes, which range from cell–pathogen interaction, to cell–cell recognition, to tumor metastasis, *etc.*[1] Interference with these recognition events could be used to modulate or alter signal transmission, or to prevent the onset of diseases. Therefore the synthesis of functional sugar mimics capable of antagonising oligosaccharides at the protein receptor level has attracted a great deal of attention as a way to develop drugs with good stability and synthetic availability.[2]

Our group is currently involved in an effort to devise a rational approach to the design and synthesis of glycomimetics by scaffold replacement.[3] In brief, we have proposed the use of conformationally stable cyclic diols[4] to replace non-pharmacophoric parts of bioactive oligosaccharides, while preserving the correct pharmacophore orientation. Computational tools can be used to predict the three-dimensional structures of the mimics and compare them with the structure of the natural ligand. When supported by adequate experimental work, molecular modelling also makes it possible to obtain at least qualitative predictions on the binding mode of new substrates, and to design further simplifications of the glycomimetic structures aimed at reducing the carbohydrate-likeness of the mimics and increasing their drug-like properties.

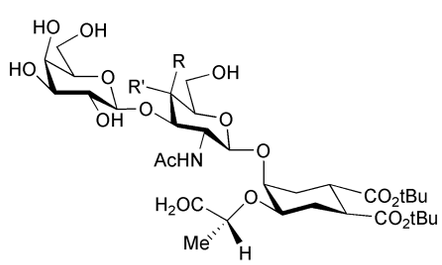
This approach has been validated[5] using as a model system the recognition pair composed of the head-group of ganglioside GM1 **1** (Chart 1) and the two bacterial enterotoxins (cholera toxin (CT) and heat-labile toxin of *E. coli* (LT)) that use it as their target on cell membranes. GM1 interacts with CT and LT using the Gal and NeuAc residues at the oligosaccharide non-reducing end, as has been shown by biochemical[6] and structural[7] studies. Not long ago, we described the rational design and the synthesis of the pseudo-oligosaccharide **2**[5,8] (Chart 1), a functional and structural mimic of ganglioside GM1 based on the use of a conformationally restricted cyclohexanediol (CHD)[3] to replace the reducing end of the ganglioside head-group, which is not involved in toxin binding. Mimic



1 GM1headgroup (oGM1)



R=OH, R'=H **2** (psGM1)
R=H, R'=OH **4**



R=OH, R'=H **3**
R=H, R'=OH **5**

Chart 1

† Electronic supplementary information (ESI) available: synthetic details, product characterisations and full NOE contact list. See http://www.rsc.org/suppdata/ob/b2/b210503a/

DOI: 10.1039/b210503a

2 was found to be as active as GM1 in binding to CT.[5] More recently, we reported three second-generation mimics[9] that were conceived by replacing the sialic acid (NeuAc) moiety of **2** with simple α-hydroxyacids. The most active compound of this series was the (*R*)-lactic acid-containing ligand **3** (Chart 1).

In an effort to further simplify the synthetic complexity of the structure we also examined the pseudo-tetrasaccharide **4** (Chart 1), which is derived from **2** by replacing the GalNAc residue with an *N*-acetylglucosamine (GlcNAc). Modification of the hexosamine was considered because it would simplify the synthesis of the artificial receptors (GalNAc is actually made from GlcNAc), while not negatively interfering with the formation of the toxin complex, since the GalNAc of GM1 interacts with the protein only *via* the *N*-acetyl group. Furthermore, the 4-hydroxy group of GalNAc in the experimental CT–GM1 complex[7] and in the computational LT–GM1[10] and LT–**2**[5] models is located outside the protein binding pocket and fully exposed to the solvent. Assuming that **4** will bind to CT and/or LT with the same general mode of GM1 and **2**, it appeared likely that inversion of configuration at C4 of the hexosamine could allow new H-bond contacts to be formed between the toxin and the substrate. In principle, these may or may not compensate for the loss in complex solvation due to the burying of the hydroxy group in the binding cavity. To evaluate the potential of **4** as a GM1 antagonist, computer simulations were used to model the free ligand **4** and its LT complex, followed by comparison with the corresponding structures of **2**.[11]

The following predictions were made:[11]

1. Changing the hexosamine in the pseudo GM1 structure should not modify the overall molecular conformation, and, in particular, should not alter the relative position of the Gal and NeuAc binding determinants.

2. Compared to **2**, the new compound **4** will indeed insert one more hydroxy group within the protein binding site. Molecular dynamics simulations suggested 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.

3. Free Energy Perturbation (FEP) calculations performed by carrying out the **2** ⟶ **4** mutation in solution and in the protein complex suggested that the GlcNAc mimic **4** should be a stronger binder than its parent compound **2**.

We now report the experimental data relative to the GlcNAc-containing mimics **4** and **5**, the latter being the GlcNAc-containing version of the second generation binder **3**. The ligands have been synthesised and their solution structure has been studied by NMR spectroscopy, confirming the computational predictions.[11] The complex formed by **5** with the recognition element of CT, CTB5, has been studied by TR-NOE NMR, and the bound conformation of the ligand has been determined. Competition experiments, carried out by adding oGM1 to a solution of the CTB5–**5** complex, could also be monitored by NMR, and have confirmed that the GlcNAc-containing ligand and the ganglioside are indeed competing for the same binding site. Affinity constants for CTB5 have been obtained for **4** and **5** by fluorescence titrations, and found to be similar to those of the parent compounds **2** and **3**, respectively. Thus, the computational predictions on the GlcNAc binders were revealed to be qualitatively correct, but the FEP calculations were quantitatively incorrect.

Results

Synthesis of the ligands

The GlcNAc-containing ligands **4** and **5** were prepared following the synthetic scheme adopted for the parent compounds **2**[8] and **3**[9] and reported in Scheme 1.

The full synthetic sequence and product characterisations are reported in the Supplementary Information. † Briefly, the known acceptors **6**[8] and **7**[9] (Scheme 1) were glycosylated at the axial hydroxy group using the trichloroacetimidate **8** as a Galβ-1,3-GlcNAc donor, and TfOH or TMSOTf as the promoter. The adducts were routinely deprotected to yield the free ligands.

NMR studies of the free ligands 4 and 5

Two-dimensional 400-MHz NOESY and ROESY spectra of **4** and **5** were obtained in D_2O. The relevant inter-residual contacts are collected in Table 1, and compared with the data reported in the literature for **1**,[12] **2**[5] and **3**[13] in D_2O. Complete spectral assignments and full contact lists are collected in the Supplementary Information. †

The NMR data obtained for **4** (Table 1, column **4**) show the characteristic set of NOE cross-peaks (Fig. 1b) which is also observed for **1**[12] and **2**[5] (Table 1, columns **1** and **2**; Fig. 1a). The data are consistent with one major conformation of the oligosaccharide framework featuring the Galβ-1,3-GlcNAcβ-1,4-

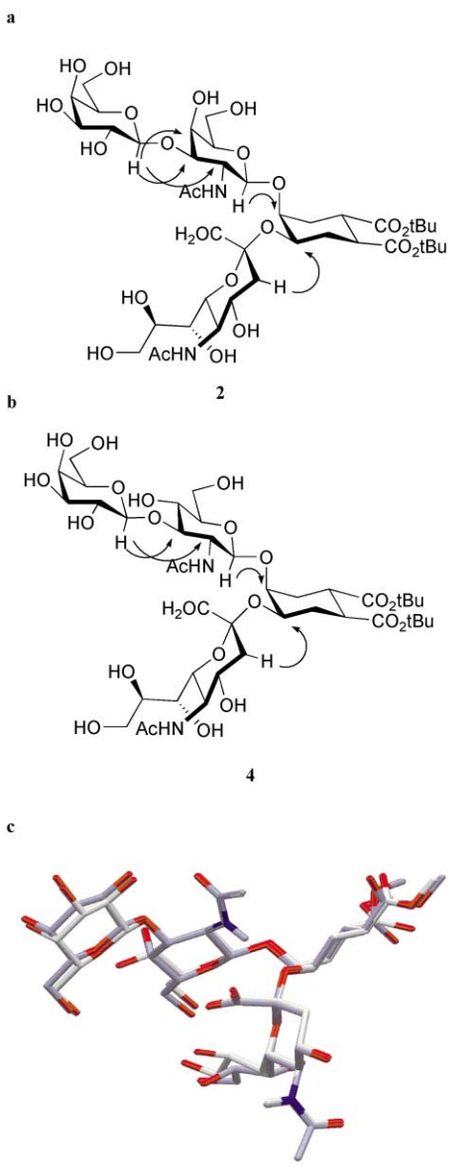
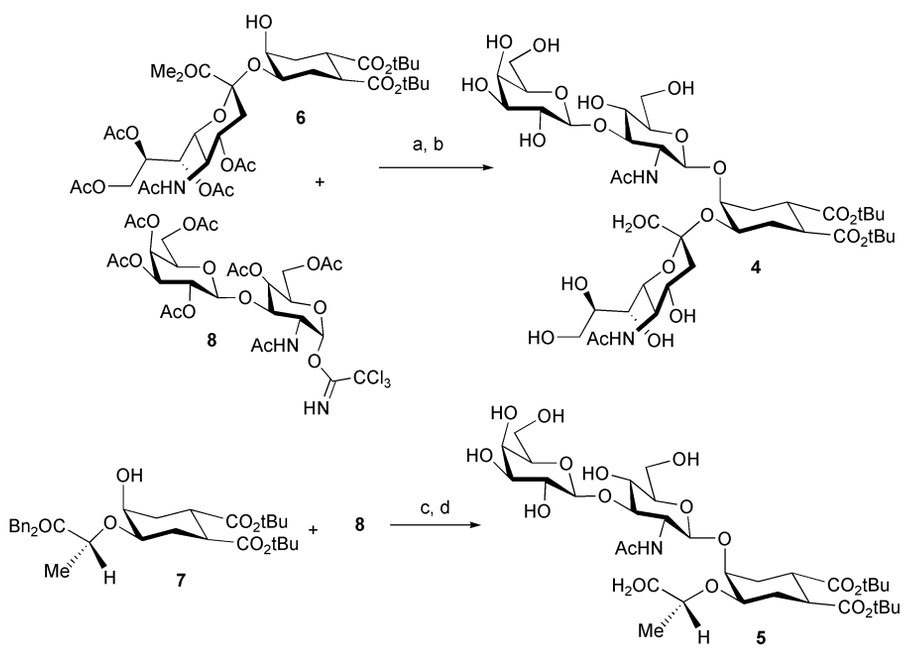


Fig. 1 **a**. Observed NOE contacts in the NMR spectra of **2**. **b.** Observed NOE contacts in the NMR spectra of **4**. **c.** Overlap of the calculated structures of **2** (light blue) and **4** (white). For the Galβ-1,3-GalNAcβ-1,4-CHD pseudo-trisaccharide the glycosidic angles ϕ and ψ are defined as H1–C1–O–Cn and C1–O1–Cn–Hn, respectively. For the NeuAcα-2,3-CHD fragment ϕ and ψ are defined as (O=)C1–C2–O–Cn and C2–O–Cn–Hn, respectively.

Table 1 Relevant NOE contacts in ligands **1**–**5**[a]

Residue		**1**[b]	**2**[c]	**3**[d]	**4**[e]	**5**[e] free state	**5**[e] CTB5 complex
NeuAc [N]	N-3ax	GII-3 (s)	CHD3 (s/m)	—	CHD3 (s/m)	—	—
	N8	GN1 (m)[f]	[g]		[g]		
			CHD4 (w)				
(*R*)-lactic acid	H_L	—	—	CHD3 (s)	—	CHD3 (s)	CHD3 (m)
				CHD4 (s)		CHD4 (s)	—
				CHD2eq (m)		CHD2eq (m/w)	CHD2eq (s)
				GN1 (w)		GN1 (w)	—
	Me			CHD3 (s)		CHD3 (s)	CHD3 (m)
				GN1 (m)		GN1 (m)	—
GalNAc [GN]	GN1	GII-4 (s)	CHD4 (s)	CHD4 (s)	—	—	—
GlcNAc [GN]	GN1	—	—	—	CHD4 (s)	CHD4 (s)	CHD4 (s)
Gal [G]	G1	GN3 (s)[f]	GN3 (s)	GN3 (s)	GN3 (s)	GN3 (s)	GN3 (s)
		GN2 (w)	GN2 (w)	GN2 (w)	[h]	GN2 (w)	
		GN4 (w)[f]	GN4 (w)	GN4 (w)			

[a] ROESY cross-peaks in D_2O solution. In parentheses the observed intensities; s: strong, m: medium, w: weak. [b] From ref. 12. [c] From ref. 5. [d] From ref. 13. [e] This work. [f] Data from DMSO-d6 solution (ref. 31); not seen in D_2O, due to signal overlap. [g] N8–GN1 not measurable, due to signal overlap. [h] GN2 overlaps with GN3.



a. TMSOTf (0.4 eq), CH_2Cl_2, reflux (13%); b. 0.002 M MeONa in MeOH, then H_2O (92%); c. TfOH (0.2 eq), CH_2Cl_2, r.t., then reflux (20%); d. H_2, Pd-C, then b (70% over the two steps).

Scheme 1 Synthesis of **4** and **5**.

CHD pseudo-trisaccharide and the NeuAcα-2,3-CHD fragment exclusively in the *syn*, *syn* (ϕ, ψ 55°, 0°) and the *anti*, *syn* (ϕ, ψ −170°, −30°) conformations, respectively (for the definition of the ϕ, ψ inter-glycosidic torsion angles, see the caption of Fig. 1). A marked conformational restriction is a distinctive feature of the oGM1 pentasaccharide and its mimics. In general, all experiments find that the ganglioside head-group can be broken down into two areas: a so-called "core trisaccharide", the GalNAcβ1-4(Neu5Acα2-3)Gal trisaccharide which is often described as "rigid", and more mobile regions corresponding to the external sugars at both ends of the head-group. However, also the mobility of the Galβ1-3GalNAc fragment at the non reducing end is actually limited to ample oscillations around well-defined average glycosidic torsion angles. Dynamic simulations of the oligosaccharide suggest that the NMR data showing two weak and equivalent NOE contacts between Gal-H1 and GalNAc-H2 and H4 (Table 1, column 1; see Fig. 1 for the equivalent contacts in **2**) are accounted for by a model whereby the Galβ1–3GalNAc bond populates a single, broad energy minimum, rather than two individual conformations.[14] The energy well is centered at ϕ, ψ 55°, −5°, which is also the conformation observed in the CTB5–oGM1 X-ray structure (ϕ, ψ 55° ± 10°, 0° ± 10°).[7] Therefore it can safely be asserted that oGM1 exists largely as one main conformer. The same behaviour has been observed for **2**, which assumes the conformation depicted in Fig. 1c (light blue structure).[5] In **4**, the flexibility of the Galβ-1,3-GlcNAcβ-1,4-CHD fragment is further reduced by the H-bond interaction of GlcNAc-OH4 with the Gal oxygen, which constrains the amplitude of the oscillations closer to the observed bound conformation. In fact, the conformational analysis of **4**[11] led to 18 conformers within 10 kJ mol^{-1} from the global minimum that all shared a common conformation of the core fragment. This conformation is completely superimposable with the NMR solution structure of **2**,

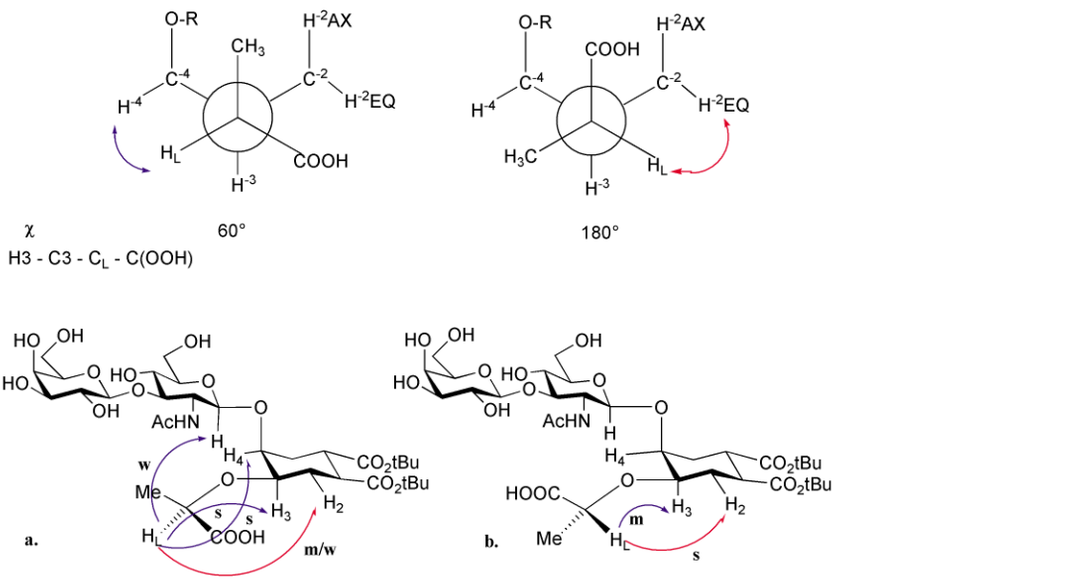


Fig. 2 Experimentally observed side-chain conformations in the (*R*)-lactic acid-containing ligand **5**. The intensity of the observed NOE cross-peaks is indicated: w:weak; s:strong; m/w:medium/weak. **a**. Free state. The mutually exclusive NOE contacts H_L/H4 and H_L/H2eq reveal the existence of two different conformations in water solution. The conformation depicted is the prevalent one. **b**. Bound state. The single conformation selected by the toxin is not the most abundant one in the free state.

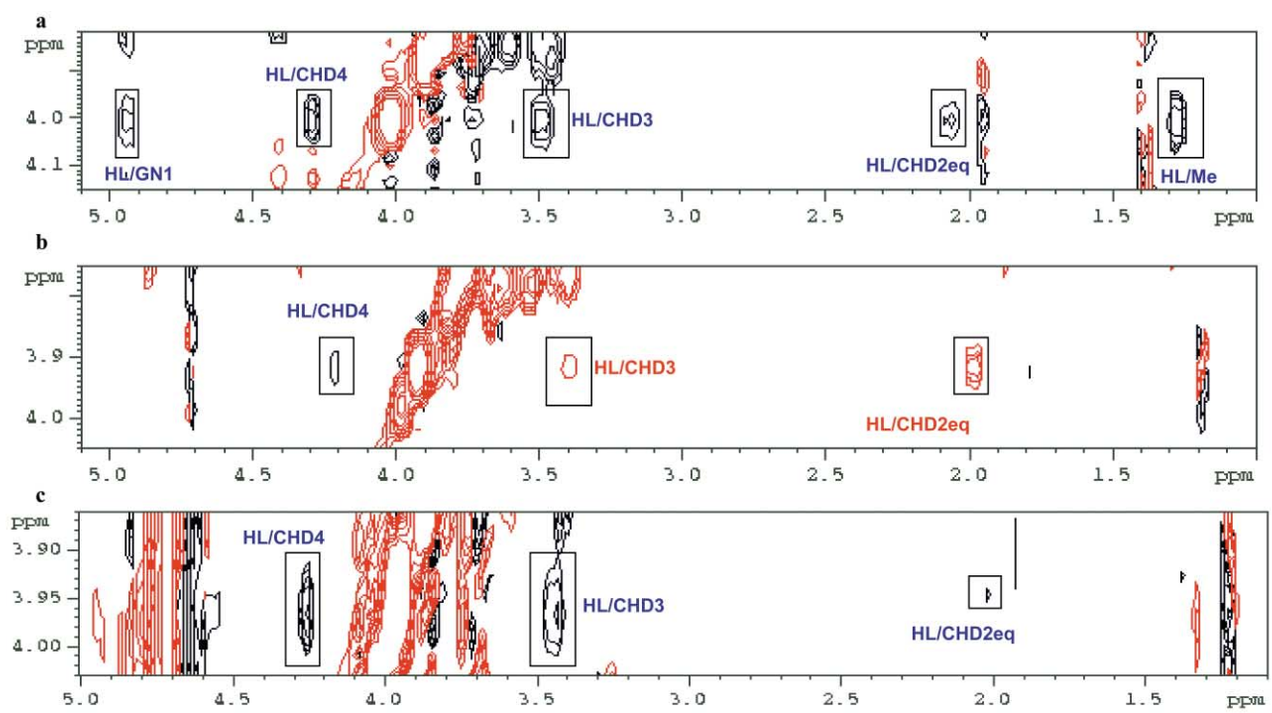


Fig. 3 NMR spectra of **5**. **a**) NOESY spectrum of **5** in the free state. **b**) TR-NOESY spectrum of **5** in the CTB5 complex. **c**)TR-NOESY spectrum of the CTB5–**5** complex after adding oGM1. Blue cross-peaks are due to the free state form, red cross-peaks are TR-NOE signals arising from the bound state of **5**.

which also corresponds to the bound structure observed in the CT–GM1 complex (Fig. 1c). The experimental data that are now available for **4** confirm the computational prediction by showing exclusively the expected pattern of NOE contacts (Fig. 1b).

For the (*R*)-lactic acid-containing ligand **5** the NMR data (Table 1, column **5**, free state) are once more consistent with a *syn*, *syn* disposition of the Galβ1-3GlcNAcβ-1,4-CHD pseudotrisaccharide, but show a substantial flexibility of the (*R*)-lactic acid side-chain. At least two conformations are represented with χ 60° (major, Fig. 2a) and χ 180° (minor), where χ is the improper dihedral angle H3–C3–C_L–C(=O) (see Fig. 2) defining the position of the carboxylate relative to the cyclohexanediol ring. This is inferred by the presence of two mutually exclusive NOE contacts from H_L to CHD4 and CHD2eq (Fig. 3a), which should arise from the χ 60° and χ 180° conformations (Fig. 2), respectively. The same pattern of NOE contacts was also observed for the GalNAc derivative **3**[13] (Table 1, column **3**).

Thus, the NMR experiments for the unbound ligands suggest that the GalNAc-containing and the GlcNAc-containing molecules are indeed adopting identical conformations, regardless of the nature of the hexosamine.

NMR studies of the CT–5 complex

Information about the conformation of complexed ligands can be derived from transferred nuclear Overhauser effect (TR-NOE) studies,[15] provided that the exchange between the

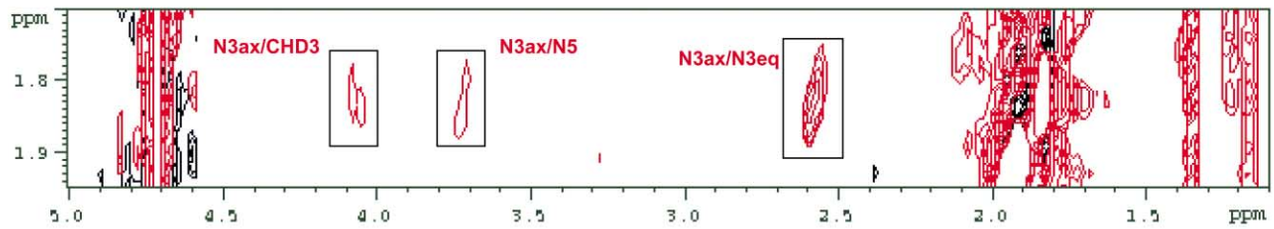


Fig. 4 TR-NOESY spectrum of the CTB5–**5** complex after addition of the GM1 oligosaccharide **1** (oGM1). In-phase (red) TR-NOESY cross-peaks belonging to the H3ax of the NeuAc residue in oGM1 (N3ax) show that oGM1 is bound to CT.

complexed and uncomplexed states is sufficiently fast.[16] When a ligand interacts with a macromolecule, the cross-relaxation rates of its bound states become opposite in sign to those of the free ligand and generate negative NOEs. So, in the TR-NOESY experiment the cross-peaks which are due to bound species are in phase with the diagonal signals, whereas unbound species give rise to cross-peaks which are out of phase. The bound form in-phase signals can be exploited to assess the conformational properties of the bound state. The TR-NOE technique allows the study in solution of how a ligand fits into a protein binding site. The phase behaviour of NOESY cross-peaks can be used to analyze the affinity of individual ligands within mixtures of compounds,[17] and can also be exploited for competition experiments.[18] Our group has recently used it to study the complexes formed by CTB5 and a group of GM1 mimics, including **3**.[13]

The affinities of **1**,[19] **2**[5] and **4** (see below) for CTB5 are on the upper limit of what can be revealed by the TR-NOE technique,[20] but the study of the CTB5–**5** complex yielded much important information. No difference is observed between the free-state and bound-state conformations of the Gal-GlcNAc and GlcNAc-CHD fragments of **5**, which are always found to be *syn* (ϕ, ψ 55°, 0°; see Table 1, column **5**, CTB5 complex). In contrast, a clearly different set of cross-peaks was observed for the NOEs involving the hydroxyacid protons H_L and CH_3 (see Table 1) upon passing from the free to the bound state. In particular, the cross-peaks observed for H_L in both free and bound states are shown in Fig. 3a and 3b, respectively. The out-of-phase NOESY cross-peaks relative to the free state of **5** are depicted in blue, whereas the in-phase TR-NOESY peaks arising from its bound state are shown in red, like the diagonal signals. The free state spectrum (Fig. 3a) shows the set of NOE contacts (blue cross-peaks) discussed above, and sketched in Fig. 2a (CHD3 and CHD4 strong, CHD2eq m/w). The TR-NOESY spectrum obtained upon addition of CTB5 to the solution is shown in Fig. 3b (see also Table 1). The strong H_L/CHD-H4 cross-peak found in the free state NOESY of **5** (Fig. 3a) almost disappears in the TR-NOESY spectrum of the CTB5–**5** complex, and the small residual signal is still out-of-phase (blue), therefore it still belongs to the free state of the ligand. The in-phase (red) cross-peaks belonging to the bound state of **5** show a close H_L/CHD-H2eq contact, much stronger than in the free state. This implies a conformational transition of the hydroxyacid side-chain, with the χ 180° conformer (Fig. 2b) being selected in the bound state preferentially with respect to the most abundant χ 60° conformer (Fig. 2a). The behaviour of **5** described here closely parallels the observations already reported for **3**,[13] and is strongly suggestive of a common binding mode for the two ligands in the toxin.

Finally, a competition experiment was performed by adding oGM1 to the CTB5–**5** complex (Fig. 3c). oGM1 has a much higher affinity for CT than **5** (see below), therefore in a competition experiment it is expected to bind selectively to the toxin and displace **5** from its binding site. Indeed, the cross-peaks of **5** which are in phase (red) in the TR-NOESY spectrum of its CT complex (*e.g.* H_L/CHD3 and H_L/CHD2eq in Fig. 3b) become out-of-phase (blue) upon addition of oGM1 to the mixture (Fig. 3c), and the relative intensity of the H_L/CHD4 and H_L/CHD2eq cross-peaks typical of the free state of **5** (CHD4 s, CHD2eq m/w) is also restored. Thus, it clearly appears that **5** is no longer bound to the toxin. Furthermore, the NeuAc-H3eq/Gal-H3 NOE contact typical of GM1 (see Table 1) is observed (Fig. 4) as an in-phase (red) cross-peak in the same spectrum, proving that GM1 has indeed displaced **5** from the toxin binding site. This experiment clearly establishes that oGM1 and **5** are competing for the same binding site in CTB5, as expected on the basis of the computational results.

Binding studies

The interaction of cholera toxin B5 pentamer (CTB5) with Gal-containing ligands can be studied using the intrinsic fluorescence of the Trp88 residue in the toxin binding site.[21] Binding of **1**–**3** to CTB5 induces bathochromic shifts in the emission spectrum whose extent depends on the ligand.[9] oGM1 binding is also accompanied by an increase in the intensity of fluorescence emission, whereas a variable decrease in fluorescence is seen upon binding of **2** (small decrease) and **3** (larger decrease). Surprisingly, when 0.5 μM CTB5 was titrated with the GlcNAc-containing ligands **4** and **5** no bathochromic shift occurred. However, a small decrease in fluorescence was observed with **4** and a larger one with **5** (see Supplementary Information, Figs. S1 and S2). † The binding isotherms of GalNAc-containing and GlcNAc-containing ligands were therefore compared by analysing the intensity data. This requires a relatively high CTB5 concentration (0.5 μM). However, wavelength titrations performed for **1** and **2** at lower toxin concentrations (0.1 μM) gave comparable results (See Supplementary Information, Fig. S3). †

Binding of **1** to CTB is known to occur cooperatively.[19] The observed concentration of the ligand at 50% saturation (IC_{50}) in our titration (0.5 μM CTB5, see Fig. 5) is 1.5 μM. Under the same conditions, the intrinsic association constant, determined by calorimetric titration, is 1.05×10^6 M^{-1}.[19] The first generation mimic **2** was also found to bind cooperatively and with comparable affinity (IC_{50} 1.2 μM, see Fig. 5).[9] Binding of the

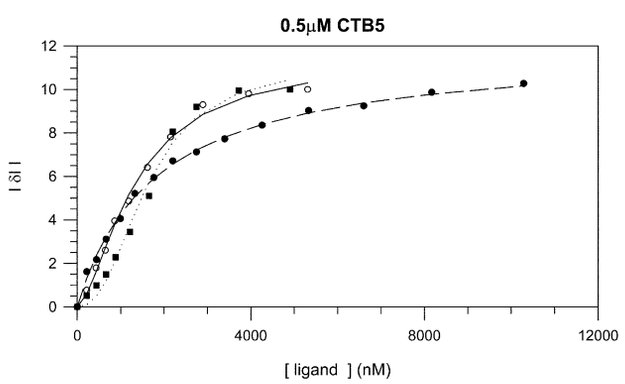


Fig. 5 Fluorescence intensity titrations of CTB5 (0.5 μM) with o-GM1 **1** (black square, dotted line); ps-GM1 **2** (empty circles, solid line); the GlcNAc-containing analogue **4** (black circles, broken line). The absolute values of the variation of fluorescence intensity emission at 350 nm (|δI|) have been normalised to 10, and plotted against the total concentration of the ligands (nM).

GlcNAc analogue **4** fits a simple 1 : 1 isotherm and doesn't show any cooperativity (Fig. 5), the IC_{50} is 1.3 μM and the dissociation constant, determined by nonlinear regression analysis, is 1.8 μM. Although quantitative comparison is complicated by the different cooperativity behaviour, the relative binding affinities of **1**, **2**, and **4** toward CTB5 can be appreciated by comparing the corresponding binding isotherms, collected in Fig. 5 and showing that the three ligands clearly appear to be of comparable potency.

The two lactic acid-containing compounds **3** and **5**, both binding CTB5 without any cooperativity effect, allow an easier estimation of relative affinity. The binding isotherms obtained in a titration of 0.5 μM CTB5 are reported in Fig. 6. The dissociation constants calculated by non-linear regression analysis are 190 and 450 μM for **3** and **5**, respectively.

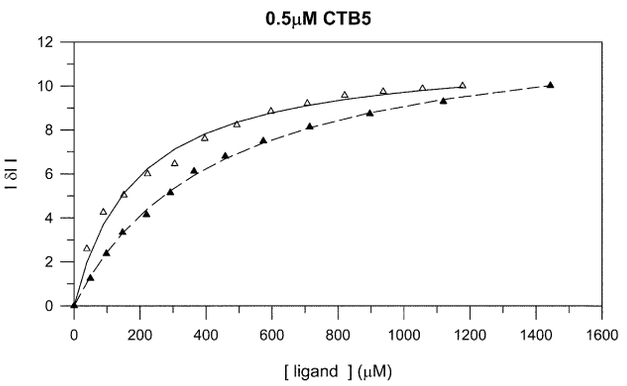


Fig. 6 Titrations of CTB5 0.5 μM with **3** (empty triangles, solid line) and **5** (black triangles, broken line). The absolute values of the variation of fluorescence intensity emission at 350 nm have been normalised to 10, and plotted against the total concentration of the ligands (μM). The dissociation constants calculated by non-linear regression analysis are 190 and 450 μM, respectively.

Thus, it appears that, compared to the parent ganglioside and pseudo-gangliosides, the GlcNAc compounds **4** and **5** are indeed showing a comparable affinity for CT, as expected based on the conformational analysis and molecular dynamic simulation of the toxin–**4** complex.[11] However, contrary to the FEP predictions,[11] **4** is not more active than **2**, nor does **5** appear to be any more active than **3**.

Discussion

The complex formed by ganglioside GM1 [Galβ1-3GalNAcβ1-4(NeuAcα2-3)Galβ1-4Glcβ1-1Cer] and the two AB5 enterotoxins CT and LT is one of the best characterised protein–carbohydrate pairs. Both toxins recognise the pentasaccharide head-group of ganglioside GM1 (o-GM1, **1**) on the host cell epithelial surface using the B5 pentamer. The pentamer has a characteristic, doughnut-like shape, with a central pore. Once the toxin is attached to the membrane, a fragment of the A monomer moves through the pore and is finally translocated through the host cell membrane into the cytosol. Biochemical[6] and structural[17] data have shown that there are five binding sites in the toxin, and that the ganglioside binds to them using the two terminal sugars at its non-reducing end, a galactose (Gal) and a sialic acid (NeuAc). Binding of oGM1 to CT displays positive cooperativity, as determined by calorimetric titrations,[19] and in biological settings the interaction between the B5 pentamer and several membrane-bound GM1 molecules is a text-book case of high-affinity multivalent interaction.[22]

We have shown that it is possible to replace the reducing end lactose of oGM1 with an appropriate diol as in psGM1 **2** with no loss of affinity or cooperativity effect.[5] The sialic acid can also be replaced with various hydroxyacids, as in the lactic acid derivative **3**.[9] In this case, the amount of affinity retained varies with the nature of the hydroxyacid side-chain,[23] but in all the analogues we have studied so far cooperativity is lost.[9,23]

Much less is known about the role played by the hexosamine. In the X-ray structure of the CTB5–oGM1 complex the GalNAc residue does not exhibit any directed or mediated H-bond interaction with the protein, but the acetamide methyl group makes a van der Waals contact with the Cβ of His13.[7] There is no natural glycolipid which contains a Galβ1,3-GlcNAcβ-1,4-(NeuAcα-2,3-)Gal fragment, and, to the best of our knowledge, no binding data are available for this oligosaccharide or for larger entities that contain it. In a previous paper,[11] we evaluated the potential of **4**, a GlcNAc analogue of psGM1 **2**, to behave as a GM1 antagonist using computer simulations to model the free ligand **4** and its LT complex. On the basis of this computational work it was expected that exchanging GalNAc with GlcNAc would yield a new molecule with the same overall molecular conformation of **2**, capable of fitting the GM1-binding site of CT and displaying a higher affinity than psGM1.

The experimental results described above concerning **4** and its lactic acid analogue **5** have now shown that the prediction was qualitatively correct, in particular:

1. NMR experiments have shown that the conformation of GlcNAc-containing ligands **4** and **5** in the free state and in the CTB5–**5** complex do not differ from those observed for the GalNAc substrates **2**[5] and **3**[13] (free state) and CTB5–**3**[13] complex.

2. Competition experiments using the TR-NOESY technique have unequivocally shown that **5** binds into the GM1-binding site of CTB5, from which it can be displaced by the higher-affinity natural ligand oGM1.

3. Fluorescence emission titrations of CTB5 have shown that the affinities of the GlcNAc-containing ligands **4** and **5** are of the same order of magnitude as those measured for the corresponding GalNAc-containing compounds **2** and **3**. However, in contrast with the behaviour of **2**, binding of **4** to CTB5 is not cooperative. Furthermore, the increase of affinity predicted by FEP calculations[11] in the **2** to **4** mutation is not borne out by experiment.

The structural reasons leading to the cooperative behaviour of GM1 binding to CT are unclear, thus it is very difficult to understand what leads to the loss of cooperativity in some of the psGM1 analogues. Computational analysis (let alone prediction) of cooperativity in systems of this size is currently unfeasible, so at this time we can only try a speculative interpretation of the available experimental data. Based on X-ray structures of bound and unbound CT and LT, the main structural effect that sugar binding has on these toxins is a tightening of the 51–60 loop of the B subunits, a loop that connects beta strand β4 to helix α2 of the B monomer.[24] The position of the α2 helix controls the size of the pore in the B5 doughnut,[25] and thus it may regulate translocation of the A fragment through the pore and ultimately through the host cell membrane. Through the α2 helix, carbohydrate binding to one site could in principle be signalled to the adjacent B monomer. However, this same tightening is elicited by sugars that are cooperative binders, such as GM1, and sugars that are non-cooperative binders, such as galactose.[24] When we first determined the affinity of **3** and other analogues obtained by sialic acid replacement,[9] we found that none of them displayed cooperative binding and speculated that the communications between CTB monomers elicited by **1** and **2** may be mediated by the NeuAc residue. The NeuAc side-chain, in fact, is in contact with the Gly-33 residue of the B+1 monomer through one water molecule (W2) located at crystallographic site two in the CT–GM1 complex.[7] This molecule is not seen in the known X-ray structures of isolated enterotoxins,[24] but it is present in many[24] (but not all)[26] of the known toxin–ligand structures. Neither of the new GlcNAc-containing mimics **4** and **5** appears to exhibit cooperative behaviour in binding to CTB5,

independent of the presence of the sialic acid moiety (see Fig. 6), apparently suggesting that NeuAc is not the cooperativity determinant. However, it should be noted that during the dynamics simulations of the LT–**4** complex[11] a displacement of the water molecule W2 was observed from the position it occupies in CT–**1**[7] and LT–**2**[11] and toward the 4-hydroxy group of the GlcNAc residue of **4**. Thus, in the LT–**4** complex W2 loses its interaction with Gly-33(B+1) and participates in a different H-bond network.[11] On the basis of this dynamics simulation and of the above speculation on inter-monomer communication, it is still possible that the cooperativity behaviour observed for **1** and **2** is indeed determined by the sialic acid side-chain *via* the W2 water molecule. The latter, however, is displaced from its normal position by the GlcNAc-containing ligands and cannot transmit its signal to the adjacent (B+1) monomer through the Gly-33 residue. If this is not the case, more subtle effects must be operating in defining the cooperativity behaviour of ganglioside binding to CT.

Free Energy Perturbation (FEP) calculations have recently been systematically applied to a series of antibody–sugar complexes using GLYCAM, TIP3P water, RESP charges and stepwise perturbations.[27] The simulations were shown to reproduce reasonably the known geometries of ligand–protein complexes, while the calculated values of $\Delta\Delta G$ of binding were found to be qualitatively reproduced and to depend heavily on the choices made about the protonation state of an His located in the vicinity of the binding site. Our estimate of the $\Delta\Delta G$ of binding between **2** and **4** was obtained[11] with a similar set of parameters (GLYCAM, TIP3P water, ESP-derived charges on the carbohydrate atoms) and the mutation was performed in the **4** to **2** sense using the slow-growth algorithm rather than stepwise perturbations of the ligand. The energetic gain calculated for **4** (3.8 ± 1.9 kcal mol^{-1}) resulted from both a better solvation of the GalNAc ligand **2** in the free state, and a more favourable interaction of **4** with the protein. Since, as we have discussed above, the pseudo-ganglioside oligosaccharides experimentally display a limited internal flexibility,[5,13,14] incomplete sampling of the free ligands doesn't appear to be a likely explanation of the rather large error in the relative free energies computed. However this might not hold in the case of the protein-complex, where sampling of the protein's degrees of freedom is also important (*vide infra*). The use of LT rather than CT in the calculation was also taken into account as a possible source of error in the calculation. However, preliminary fluorescence titrations[28] obtained with a sample of the entire LT toxin[29] showed the same trends observed with CT, *i.e.* equivalent pairs of ligands (**3** and **5**, and **2** and **4**) have similar affinity for LT as well as for CT.

The most likely explanation for the error in the computed relative free energies lies in the computational treatment of the pseudo-ganglioside–toxin complex part of the FEP. First, incomplete sampling of the complex simulation cannot be ruled out, particularly considering that a substantial amount of side-chain rearrangements occur in the protein binding site on passing from the LT–**4** to the LT–**2** complex. Furthermore, the state of protonation of His-57, one of the amino acids most involved in the rearrangement, was not addressed at all during the FEP simulation.

Although achieving a quantitative prediction of relative binding energies for oligosaccharide–protein complexes may well require a lot more effort, this work shows that computational tools can be used with success to design new inhibitors of carbohydrate–protein interaction, and that they yield qualitatively useful results. Indeed, at the start of this project, nothing could suggest that inversion of the hexosamine C4 in psGM1 **2** would yield a ligand with good CT affinity. We have already noted that, to the best of our knowledge, the natural counterpart of **4** has never been described, and certainly has never been tested as a ligand for bacterial enterotoxins. The calculations correctly predicted that inversion of the hexosamine C4 in **2** would yield a new molecule with overall shape and conformational properties very similar to psGM1. It also allowed a prediction that the new molecule **4** would be able to interact with the cholera toxin in a way similar to its GalNAc analogue. The question remains open of how to reliably achieve quantitative prediction of relative binding affinities in this and similar systems.

In conclusion, two new artificial ligands of the cholera toxin, the pseudo sugars **4** and **5** are described. The new ligands were designed starting from the known GM1 mimics **2** and **3** by replacement of their GalNAc residue with the C4 isomer GlcNAc. Such substitution had been suggested by inspection of the CT–GM1 complex, and supported by computational predictions, which suggested that the three-dimensional shape of the new ligands and their mode of interaction with CT would be similar to those of the starting structures. These predictions are now confirmed by the experimental results showing that the conformational properties of the equivalent pairs **2**–**4** and **3**–**5** are indeed very similar and that their affinity for CT is of the same order of magnitude. NMR experiments have also allowed the gathering of information on the structure of the CT–**5** complex showing that **5** occupies the GM1-binding site of the toxin and that it binds with a conformation similar to the one adopted by **3** in the CT–**3** complex. Since GalNAc is normally synthesised by C4 inversion of GlcNAc, the GlcNAc ligands constitute a new class of GM1 mimics with improved synthetic accessibility.

Experimental section

Synthesis

The synthesis of compound **6** was described in ref. 8. The full synthetic sequence leading to **4** and **5** and the product characterisations are reported in the Supplementary Information.

NMR

NMR spectra of **4** and **5** were recorded at 25–30 °C in D_2O, on Varian Unity 500 MHz or Bruker AVANCE 400 MHz spectrometers. For the experiments with the free ligands, the corresponding compound (1–1.5 mg) was dissolved in D_2O and the solution was degassed by passing N_2. COSY, TOCSY and HSQC experiments were performed using standard sequences at temperatures between 298 and 310 K. NOESY experiments were performed with mixing times of 500, 700, 1000 ms (**4**) and 400, 600, 700, 800 ms (**5**). ROESY experiments[30] were performed with mixing times of 50, 100, 150, 200, 300 ms (**4**) and 100, 200, 250, 300 ms (**5**).

The cholera toxin CTB pentamer (CTB5) was purchased from List Biological Laboratories Inc. The commercial sample was ultrafiltered to remove EDTA and tris salt, redissolved in phosphate buffer and subjected to two cycles of freeze-drying with D_2O to remove traces of H_2O. The sample was then dissolved with D_2O, and the solution transferred to the NMR tube to give a final concentration *ca.* 0.1 mM. TR-NOESY experiments were performed with mixing times of 100, 200 and 300 ms, for a *ca.* 50 : 1 molar ratio of **5** : lectin. TR-ROESY experiments were also performed with mixing times of 100, 200, 250, 300 ms. In the competition experiment, 1.5 mg of oGM1 **1** (gift from Professor Sandro Sonnino, Dipartimento di Chimica e Biochimica Medica, Universita' di Milano) were added to the same solution and the TR-NOESY spectra were recorded as described above.

Titrations

Fluorescence titrations were performed with an LS50 Perkin Elmer fluorimeter, using a pH 7.5 tris buffer to dissolve CTB5 (0.5 μM) and ligands. Fluorescence from TRp-88 was measured with an excitation of 280 nm and an emission from 300 to 450

nm. The emission spectra are reported as Supplementary Information. The K_d reported in the text were determined by non-linear regression analysis, using Sigma Plot 2.0 (Jandel Corporation).

Acknowledgements

We thank Professor Sandro Sonnino (Dipartimento di Chimica e Biochimica Medica, Universita' di Milano) for a generous gift of oGM1, and for helpful discussions throughout this project. We thank Giorgio Colombo (IBMR, Milano) for carefully reading this manuscript and providing his comments. Funding was provided by MIUR, CNR, COST D-13/012, Azioni Integrate.

References

1 (*a*) H. Lis and N. Sharon, *Chem. Rev.*, 1998, **98**, 637–674; (*b*) R. A. Dwek, *Chem. Rev.*, 1996, **96**, 683–720; (*c*) A. Varki, *Glycobiology*, 1993, **3**, 97–130.
2 P. Sears and C. H. Wong, *Angew. Chem., Int. Ed.*, 1999, **111**, 2446–2471; P. Sears and C. H. Wong, *Angew. Chem., Int. Ed.*, 1999, **38**, 2300–2324 and references therein.
3 A. Bernardi, D. Arosio, L. Manzoni, F. Micheli, S. Pasquarello and P. Seneci, *J. Org. Chem.*, 2001, **66**, 6209–6216.
4 *trans*-Cyclohexane-1,2-diol has been used by many authors as an effective replacement for 3,4-disubstituted *N*-acetylglucosamine in the synthesis of sialyl Lewis X mimics. For a review, see: E. E. Simanek, G. J. McGarvey, J. A. Jablonowski and C.-H. Wong, *Chem. Rev.*, 1998, **98**, 833–862.
5 A. Bernardi, A. Checchia, P. Brocca, S. Sonnino and F. Zuccotto, *J. Am. Chem. Soc.*, 1999, **121**, 2032–2036.
6 C.-L. Schengrund and N. J. Ringler, *J. Biol. Chem.*, 1989, **264**, 13233–13237 and references therein.
7 (*a*) E. A. Merritt, S. Sarfaty, F. van den Akker, C. L'Hoir, J. A. Martial and W. G. J. Hol, *Protein Sci.*, 1994, **3**, 166–175; (*b*) E. A. Merritt, P. Kuhn, S. Sarfaty, J. L. Erbe, R. K. Holmes and W. G. J. Hol, *J. Mol. Biol.*, 1998, **282**, 1043–1059.
8 A. Bernardi, G. Boschin, A. Checchia, M. Lattanzio, L. Manzoni, D. Potenza and C. Scolastico, *Eur. J. Org. Chem.*, 1999, **6**, 1311–1317.
9 A. Bernardi, L. Carrettoni, A. Grosso Ciponte, D. Monti and S. Sonnino, *Bioorg. Med. Chem. Lett.*, 2000, **10**, 2197–2200.
10 A. Bernardi, L. Raimondi and F. Zuccotto, *J. Med. Chem.*, 1997, **40**, 1855–1862.
11 A. Bernardi, M. Galgano, L. Belvisi and G. Colombo, *J. Comput.-Aided Mol. Des.*, 2001, **15**, 117–128.
12 P. Brocca, P. Berthault and S. Sonnino, *Biophys. J.*, 1998, **74**, 309–318.
13 A. Bernardi, D. Potenza, A. M. Capelli, A. García-Herrero, F. J. Cañada and J. Jiménez-Barbero, *Chem.-Eur. J.*, 2002, **8**, 4597–4612.
14 P. Brocca, A. Bernardi, L. Raimondi and S. Sonnino, *Glycoconjugate J.*, 2000, **17**, 283–299.
15 A. A. Bothner-By and R. Gassend, *Ann. N. Y. Acad. Sci.*, 1973, **222**, 668–676.
16 P. L. Jackson, H. N. Moseley and N. R. Krishna, *J. Magn. Reson., Ser. B*, 1995, **107**, 289–292; V. L. Bevilacqua, D. S. Thomson and J. H. Prestegard, *Biochemistry*, 1990, **29**, 5529–5537; V. L. Bevilacqua, Y. Kim and J. H. Prestegard, *Biochemistry*, 1992, **31**, 9339–9349.
17 M. Vogtherr and T. Peters, *J. Am. Chem. Soc.*, 2000, **122**, 6093–6099.
18 M. Mayer and B. Meyer, *J. Med. Chem.*, 2000, **43**, 2093–2099.
19 A. Schoen and E. Freire, *Biochemistry*, 1989, **28**, 5019–5024.
20 An accurate NMR study of these complexes is in progress, and will be reported in due course.
21 J. A. Mertz, J. A. McCann and W. D. Picking, *Biochem. Biophys. Res. Commun.*, 1996, **226**, 140–144.
22 M. Mammen, S.-K. Chio and G. M. Whitesides, *Angew. Chem., Int. Ed.*, 1998, **37**, 2755–2794.
23 S. Cattaldo Tesi di Laurea, Universita' di Milano a.a. 2000–2001.
24 E. A. Merritt, T. K. Sixma, K. H. Kalk, B. A. M. van Zanten and W. G. J. Hol, *Mol. Microbiol.*, 1994, **13**, 745–753.
25 A. T. Amn, S. Fraser, E. A. Merritt, C. Rodrigherio, M. Kenny, M. Ahn, W. G. J. Hol, N. A. Williams, W. I. Lencer and T. R. Hirst, *Proc. Natl. Acad. Sci. USA*, 2001, **98**, 8536–8541.
26 W. E. Minke, C. Roach, W. G. J. Hol and C. L. M. J. Verlinde, *Biochemistry*, 1999, **38**, 5684–5692.
27 A. Pathiaseril and R. J. Woods, *J. Am. Chem. Soc.*, 2000, **122**, 331–338.
28 D. Arosio, unpublished.
29 Generous gift of Mariagrazia Pizza at Chiron Vaccines, Siena.
30 J. Dabrowski, T. Kozár, H. Grosskurth and N. E. Nifant'ev, *J. Am. Chem. Soc.*, 1995, **117**, 5534–5539.
31 D. Acquotti, L. Poppe, J. Dabrowski, C.-W. V.d. Lieth, S. Sonnino and G. Tettamanti, *J. Am. Chem. Soc.*, 1990, **112**, 7772–7778.