Available online at www.sciencedirect.com
ScienceDirect
Carbohydrate Research 342 (2007) 1651–1660
Carbohydrate RESEARCH

First round of a focused library of cholera toxin inhibitors

Črtomir Podlipnik,[a,†] Ingrid Velter,[b] Barbara La Ferla,[b] Gilles Marcou,[a,‡] Laura Belvisi,[a] Francesco Nicotra[b,*] and Anna Bernardi[a,*]

[a]*Dipartimento di Chimica Organica e Industriale, Universita' degli Studi di Milano, Via Venezian 21, I-20133 Milano, Italy*
[b]*Dipartimento di Biotechnologie e Bioscienze, Universita' degli Studi di Milano-Bicocca, Piazza della Scienza 2, I-20126 Milano, Italy*

Received 23 February 2007; received in revised form 29 May 2007; accepted 2 June 2007
Available online 9 June 2007

Abstract—*C*-Galactosides have been used as scaffolds to design a library of non-hydrolysable inhibitors of cholera toxin (CT). Test elements from the library were synthesized and found to inhibit CT binding to an asialofetuin-coated SPR chip with micromolar affinity. Preliminary results are reported.

Keywords: Carbohydrate mimics; Cholera toxin; Glycomimetics; *C*-Glycosides; GM1 ganglioside

1. Introduction

Vibrio cholerae cholera toxin (CT) is the causative agent of cholera, a disease that is responsible for the death of hundreds of thousands of people each year in developing countries.[1] CT is a member of the AB5 holotoxin family. It consists of a single catalytically active component, A, and a non-toxic receptor-binding component, a pentamer of B subunits. The B pentamer (CTB5) is responsible for binding to the cell surface and this function is retained even in the absence of the A subunit.[2,3]

The mechanism of CT action is initiated by binding of the B subunits to the oligosaccharide head-group of the GM1 ganglioside, β-Gal*p*-(1→3)-β-Gal*p*NAc-(1→4)-(α-Neu*p*Ac-(2→3)-β-Gal*p*-(1→4)-β-Glc–OH, **1** (o-GM1, Fig. 1), on intestinal epithelial cell membranes. The A subunit is then translocated into the host cell, where it mono-ADP-ribosylates Gsα, a guanine nucleotide-binding regulatory protein, resulting in persistent activation of adenylyl cyclase.[3] The rising level of cAMP leads to massive loss of fluids, which in turn can lead to dehydration and shock.[4,5]

Small molecules acting as decoys for the toxin's GM1 binding site could work as toxin inhibitors by competing with GM1, possibly saturating the toxin, which would prevent it from binding to the epithelial cells and causing the symptoms of cholera. Natural and synthetic antagonists of CT are known, and they are under investigation as potential agents for the treatment of cholera and other toxin-caused enteropathies.[6]

The interactions of o-GM1 with CTB5, as seen in the X-ray structure of the complex, are highlighted in Figure 1.[7,8] The terminal galactose stacks on top of Trp88 and is linked through hydrogen bonds to Asn90, Lys91, Glu51, and Gln61 (Fig. 1). The terminal sialic acid (NeuAc) is exposed to solvent on the α-face and it is in van der Waals contact with Tyr12 on the β-face. Hydrogen bonds are formed between the sialic acid polar groups with crystallographic water molecules and with the backbone of Glu11 and Tyr12. The carboxyl group of NeuAc establishes a key interaction with Trp88, mediated by a key water molecule. The outer part of the binding pocket, with Glu11 and Arg35, constitutes a dipole that orients the *N*-acetyl fragment of the sialic acid (Fig. 1).

The terminal galactose and sialic acid residues contribute the most to GM1 binding. Although the two

* Corresponding authors. Tel.: +39 02 64483457 (F.N.); tel.: +39 02 50314092; fax: +39 02 50314072 (A.B.); e-mail addresses: francesco.nicotra@unimib.it; anna.bernardi@unimi.it
† Present address: Faculty of Chemistry and Chemical Technology, Aškerčeva 6, 1000 Ljubljana, Slovenia.
‡ Present address: Institut de Chimie de Strasbourg, UMR7177,1, rue Blaise Pascal, 67000 Strasbourg, France.

0008-6215/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carres.2007.06.006

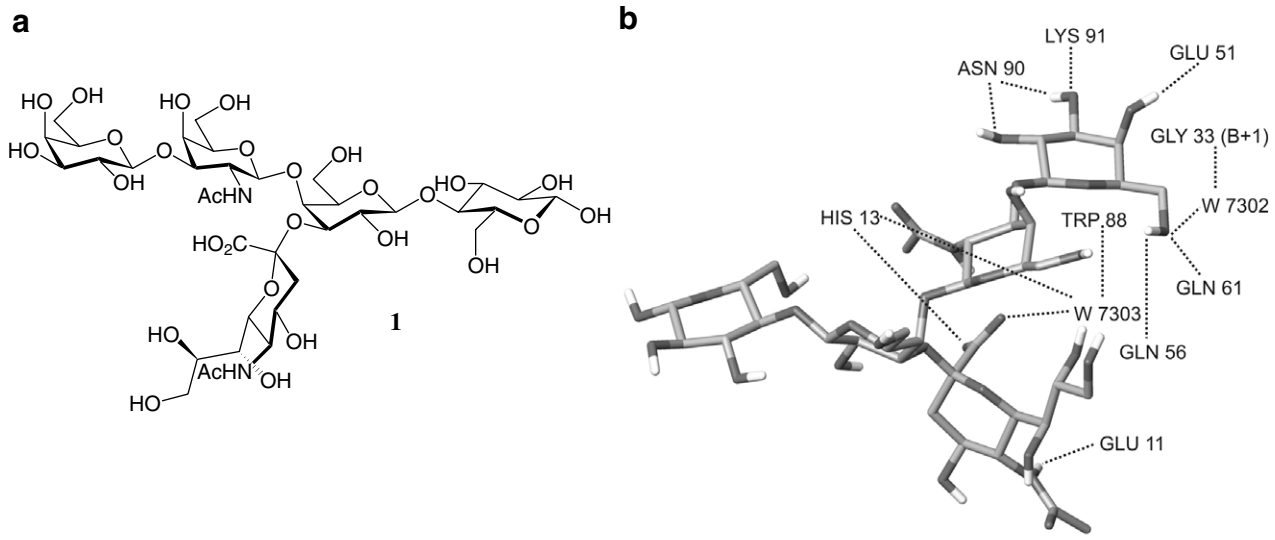


Figure 1. (a) The GM1 oligosaccharide o-GM1, and (b) its key interactions with CT binding site.

isolated monosaccharides show very low affinity for the protein (Gal 40 mM; NeuAc > 200 mM),[9] o-GM1 binds to CT with a dissociation constant $K_D = 43$ nM.[9] This appears to depend on the conformational preorganization of o-GM1,[9,10] which has been shown by NMR to adopt one major conformation[11] closely resembling the bound conformation observed in the CT complex.[7,12]

According to this interpretation, the pseudo-oligosaccharide **2** (Fig. 2), which reproduces the 3D structure of o-GM1, was found to display a similar affinity for CTB5.[10] A second group of pseudo-oligosaccharides rationally designed to replace the sialic acid moiety while retaining CT affinity have led to micromolar inhibitors of toxin (Fig. 2).[12–18] Similar results have been achieved by using galactose as a CT-binding anchor and developing several generations of galactoside libraries[3,6a,19–22] (Fig. 2). Some of the molecules discussed above display good affinity for CT and are structurally much simpler than the natural ligand GM1. However, they all are *O*-glycosides, and are unlikely to be metabolically stable to any significant extent. Furthermore, the synthetic methods used to connect the pharmacophoric sugar moieties are those of traditional carbohydrate chemistry, which are often laborious and not high yielding procedures.

To address these shortcomings, we are currently working toward the development of CT ligands starting from simple *C*-galactosides. *C*-Glycosides are well known metabolically stable analogs of *O*-glycosides.[23–29] Furthermore, some of us have recently shown that the functionalized *C*-galactosides **7** and **8** (Fig. 3) can be easily synthesized in a few steps directly from galact-

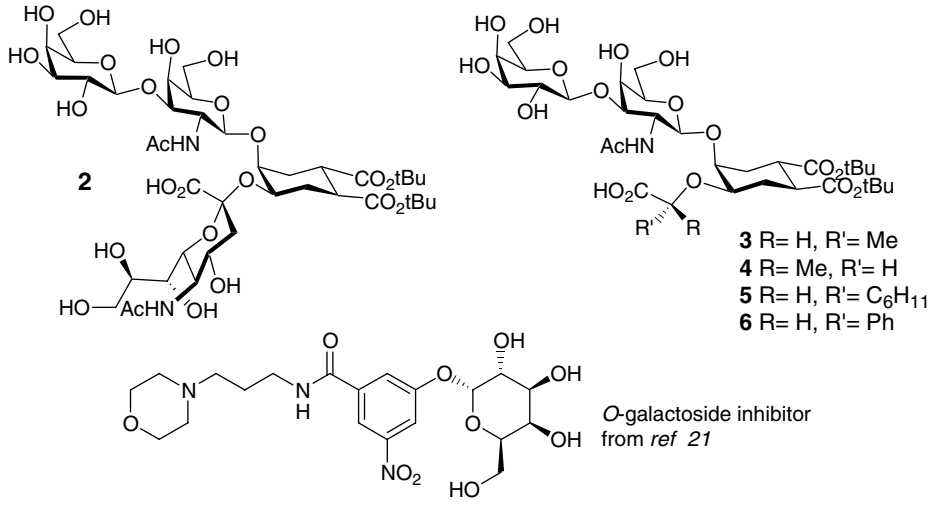


Figure 2. Some currently known CT inhibitors.

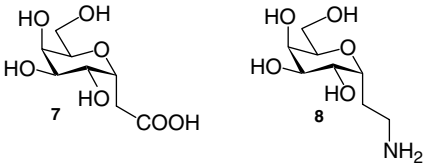


Figure 3. The *C*-galactoside anchors used for the development of metabolically stable CT-ligands.

ose avoiding the need for protecting groups.[30] The approach we have adopted to identify CT inhibitors involves the following steps: (a) the development and validation of a docking/scoring protocol based on a set of known pseudo-GM1 ligands; (b) the design of a focused library of *C*-galactosides starting from **7** and **8**; (c) the synthesis and affinity evaluation (by SPR) of selected elements of the library. We report here on our initial results, which have led to the identification of simple *C*-galactosides with micromolar affinity for CT.

Table 1. Computed (Autodock/Glide) and measured activity of CT ligands

Entry	Compound	Glide G-score[a]	pIC_{50} measured	pIC_{50} computed
1	**1**	−12.38	5.80[b]	5.65
2	**2**	−11.79	5.89[b]	5.18
3	**3**	−10.27	3.70[b]	3.97
4	**4**	−9.65	3.00[b]	3.48
5	**5**	−11.47	4.43[c]	4.92
6	**6**	−11.66	5.00[c]	5.08
7	**9**	−8.03	2.69[d,e]	2.20
8	**10**	−9.45	2.45[e]	3.32
9	**11**	−10.08	3.90[e]	3.82
10	**12**	−10.42	3.63[e]	4.09
11	**13**	−9.38	3.51[e]	3.27
12	**14**	−11.58	3.22[e]	5.01

[a] G-score obtained starting from Autodock poses following the procedure described in Section 3 (Docking experiments).
[b] Fluorescence measurement, from Ref. 14.
[c] Fluorescence measurement, from Ref. 15.
[d] Fluorescence measurement, from Ref. 46.
[e] This paper (SPR measurement).

2. Results and discussion

2.1. Validation of the docking protocol

The validation of the docking protocol was attempted using a set of seven known CT ligands, which included o-GM1 **1** (Fig. 1), the pseudo-GM1 **2**–**6** (Fig. 2) and methyl α-D-galactopyranoside **9** (Fig. 4). The CT complex of all these compounds was at least partially defined either from crystallographic[7,19] or NMR studies[13,16] and IC_{50} from fluorescence titrations were available[14,15] and covered a good range of different affinities (Table 1, entries 1–7).

Autodock[31] was previously reported to reproduce the bound pose of CT-ligands with good fidelity.[32] Indeed, we obtained very good results using Autodock when generating poses for o-GM1 and its derivatives, but the ranking based on Autodock score was poor. Better results were obtained by clustering the Autodock poses, minimizing the cluster leaders in the toxin's cavity and ranking the poses by in situ scoring with Glide.[33,34] The Glide score (G-Score) obtained ranged from −12.38 for the best ligand **1**, to −8.03 for methyl α-D-galactopyranoside **9**, a millimolar CT ligand (Table 1). The seven test compounds were ranked correctly relative to their affinity (Table 1, Fig. 4). Visual inspection of the best ranking pose for each ligand allowed us to recognize the features expected based on the available experimental data.

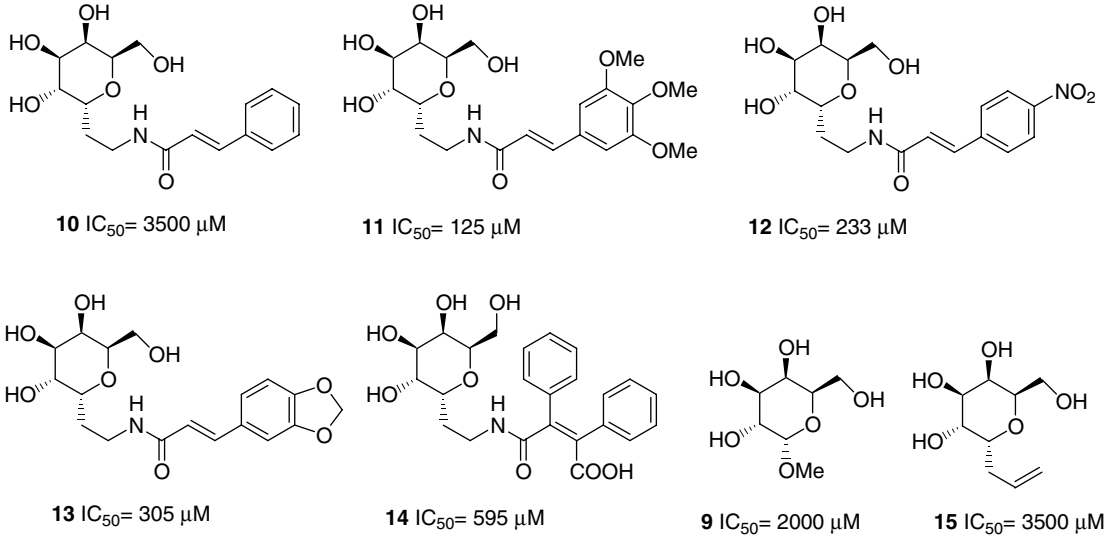


Figure 4. Structure and CT inhibition data (SPR) of ligands **9**–**15**.

Direct docking (and scoring) with Glide yielded poses of lower quality and some of the compounds, including the natural ligand, were especially problematic, possibly because they are more flexible and larger than ideal for virtual screening software. However, roughly the same ranking of the test compounds was obtained, indicating that direct screening with Glide, which is orders of magnitude faster than the protocol described above, could be used for the initial analysis of larger libraries.

2.2. Design, synthesis and testing of the test library

Based on these results, a virtual screening of the *C*-galactoside library was attempted. A library of 309 compounds was built by conjugating scaffolds **7** and **8** with a group of selected amines or acids, respectively. The compounds were designed to contain either an acid function, which could substitute for the carboxyl group of sialic acid, or lipophilic groups, which would be able to exploit the shallow lipophilic surface surrounding the CT galactose binding site, or both.

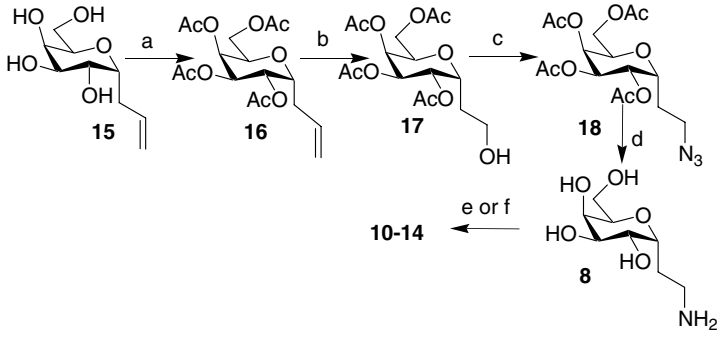
The virtual library was built by in-house software as described in Section 3.3 and screened using Glide 4.0. Analysis of the G-score showed that none of the screened compounds were ranked above the natural ligand o-GM1, but that several were ranked in a G-score range associated with ligands of micromolar affinity in the validation set.

From the sublibrary of cinnamic acid derivatives obtained starting from scaffold **8**, the five amides **10**–**14** (Fig. 4) were selected for synthesis using five commercially available cinnamic acids. These five compounds were re-docked in CT binding site using the Autodock/Glide protocol described above, and the G-score obtained is reported in Table 1 (entries 8–12). All these compounds are α-*C*-glycosides, as opposed to native GM1, which contains a β-galactosyl residue. According to the calculated poses, this has no influence on the binding mode of the galactose anchor, which retains all the known interactions in the binding site (see Fig. 1b).

The compounds were synthesized starting from known *C*-allyl galactose **15** (Scheme 1).[35,36] Compound **15** was acetylated and subjected to Lemieux oxidation followed by reduction to afford alcohol **17**. Tosylation and treatment with NaN_3 afforded the azido derivative **18**, which was reduced and deprotected providing amine **8** in good overall yields. Coupling of **8** with diphenylmaleic anhydride or freshly prepared cinnamoyl chlorides provided **10**–**14** in 50–80% yield.

The inhibition properties of compounds **10**–**14** were analyzed in a surface plasmon resonance (SPR) competition assay, using the cholera toxin B pentamer CTB5 as the target and a SPR chip coated with asialofetuin (ASF), as previously described.[37] A series of measurements with increasing CTB5 concentrations yielded a binding isotherm and a K_{chip} of 70 μM.[38] Inhibition studies were then performed using ligands **9**–**15** and CTB5 at 70 μM. First, the methyl α-D-galactopyranoside **9** (Fig. 4) was measured and an IC_{50} of 2 mM was determined, consistent with the known CTB5 affinity of galactose.[9,19] No affinity improvement was shown by the cinnamic acid derivative **10** ($IC_{50} = 3500$ μM) over **9** or the α-*C*-allyl galactoside **15**. On the contrary, the inhibition potency of the ligands increased by one order of magnitude for compounds **11**–**14**, which displayed additional functional groups. Amide **11**, bearing three methoxy groups on the aromatic ring, is the most active of the set ($IC_{50} = 125$ μM). Inspection of the Autodock poses (Fig. 5) does not suggest specific H-bonding interactions for the three methoxy groups, rather they appear to provide additional hydrophobic interactions at the bottom of the binding site. Furthermore, the three oxygen atoms of **11**, which are largely exposed to the aqueous solvent, may improve the solvation of the complex relative to the unsubstituted analog **10**.

Furthermore, we could use the G-scores and the experimental pIC_{50} for the seven test compounds to generate a simple model for predicting IC_{50} for CT inhibitors. The model is expressed by Eq. 1, which was then used to predict activity values for the set of the newly



Scheme 1. Synthesis of ligands **10**–**14**. Reagents and conditions: (a) Ac_2O, pyridine; (b) OsO_4, $NaIO_4$, 0 °C→rt, then $NaBH_4$; (c) TsCl, pyridine, CH_2Cl_2, 0 °C→rt then NaN_3, DMF; (d) H_2 Pd–C, MeOH then MeONa, IRA-120H$^+$; (e) ArCOCl, toluene, reflux, 12 h; (f) diphenylmaleic anhydride, MeOH, K_2CO_3, 2 h.

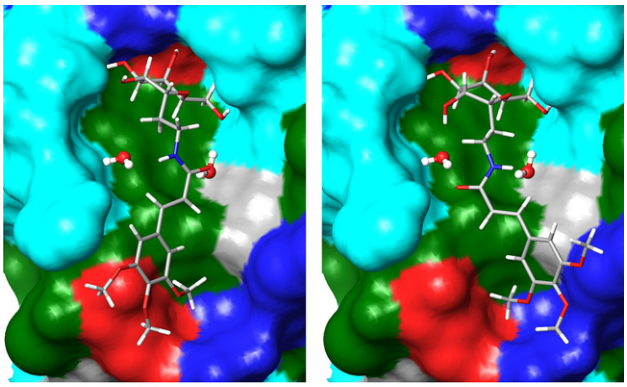


Figure 5. Poses obtained with Autodock/Glide procedure for compound **11**.

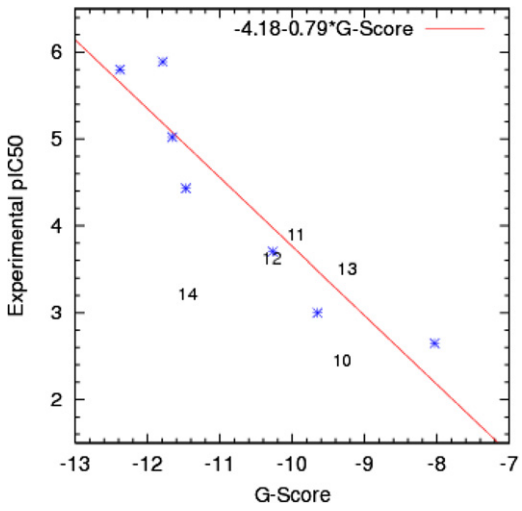


Figure 6. Correlation between Glide score and affinity for CT binders. Stars represent entries 1–7 in Table 1, which are included in the model described by Eq. 1. The numbers represent compounds **10**–**14**.

synthesized amides **10**–**14**. The predicted values are collected in Table 1 (column 5) and compared to the experimental values derived from the SPR measurements (column 4). A reasonable correlation between experimental pIC_{50} and computed Glide scores for the CT inhibitors can be observed in Figure 6. The largest deviation observed is for the predicted IC_{50} value of compounds **14**, which is two orders of magnitude higher than that measured by SPR experiment.

$$pIC_{50} = -4.18(\pm 1.48) - 0.79(\pm 0.14) * \text{G Score} \quad (n = 7,\ r^2 = 0.87,\ F = 34.03,\ s^2 = 0.26,\ q^2 = 0.73) \quad (1)$$

3. Conclusions

In conclusion, we have shown that simple *C*-galactosides can be used as anchors for non-hydrolysable ligands capable of CT inhibition. The small test library that we have described here is by no means optimized. Nevertheless, four of the five non-hydrolysable compounds prepared displayed an inhibition potency that compares well to previously described artificial ligands that are more complex and/or hydrolysable, such as those depicted in Figure 4.[14,15,21] The library design was supported by in silico docking protocols which were validated using a training set of known CT ligands. It is well known that binding mode prediction and accurate prediction of binding affinity are different tasks, which are accomplished with very different efficiency by currently available docking programs.[41] In the case at hand, we have tested two different programs and found that one (Autodock) produced good ligand poses (as judged by comparison with the known X-ray or NMR structures) and poor affinity ranking and the second one (Glide) gave a better reproduction of the binding affinity and poor ligand poses. Eventually we have validated a method that relies on the initial rapid library screening with Glide followed by redocking with Autodock and in situ scoring with Glide. Based on this procedure we have built a predictive model to estimate a ligands' inhibition potency from virtual screening scores. With these tools in hand, we will continue our studies toward the design and synthesis of more potent analogs.

4. Experimental

4.1. Synthesis

4.1.1. General methods. All solvents were dried with molecular sieves for at least 24 h prior to use. Thin layer chromatography (TLC) was performed on silica gel 60 F254 plates (Merck) with detection using UV light when possible, or by charring with a solution of concd H_2SO_4–EtOH–H_2O (5:45:45) or a solution of $(NH_4)_6Mo_7O_{24}$ (21 g), $Ce(SO_4)_2$ (1 g), concd H_2SO_4 (31 mL) in water (500 mL). Flash column chromatography was performed on silica gel 230–400 mesh (Merck). ^1H and ^{13}C NMR spectra were recorded at 25 °C with a Varian Mercury 400 MHz instrument using $CDCl_3$ as the solvent unless otherwise stated. Chemical shift assignments, reported in ppm, are referenced to the corresponding solvent peaks. Mass spectra were recorded with a MALDI2 Kompakt Kratos instrument, using α-cyano-4-hydroxycinnamic acid (CHCA) or gentisic acid (DHB) as the matrix; HRMS were recorded on a QSTAR elite LC/MS/MS system with a nanospray ion source. Optical rotations were measured at rt using a

Krüss P3002 electronic polarimeter and are reported in units of 10^{-1} deg cm^2 g^{-1}.

4.1.2. 3-(2,3,4,6-Tetra-*O*-acetyl-α-D-galactopyranosyl)-propene (16). To a solution of the *C*-allyl-galactoside **15** (2.50 g, 12.25 mmol) in pyridine (18.0 mL), acetic anhydride (9.0 mL, 98.04 mmol) was added at 0 °C. The reaction was stirred for 4 h at rt, diluted with ethyl acetate (50 mL), and then washed with water (1 × 30 mL), 10% $CuSO_4$ (3 × 30 mL), water (1 × 30 mL), 5% HCl (1 × 30 mL), water (1 × 30 mL), and brine (1 × 30 mL). The organic layer was dried (Na_2SO_4), filtered and the filtrate concentrated under vacuum. The resulting residue was purified by flash column chromatography over silica gel (petroleum ether–EtOAc, 2.5:1) to afford **16** (4.35 g, 95%); yellow oil, $R_f = 0.6$ (petroleum ether–EtOAc, 1.5:1); $[\alpha]_D$ +87 (*c* 1, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$): δ 5.80–5.69 (m, 1H, C*H*=CH_2), 5.42–5.40 (m, 1H, H-4), 5.28–5.25 (dd, $J = 4.8$, 9.3 Hz, 1H, H-2), 5.22–5.21 (dd, $J = 3.2$, 9.3 Hz, 1H, H-3), 5.15–5.08 (m, 2H, CH=CH_2), 4.31–4.28 (q, $J = 4.8$, 10.0 Hz, 1H, H-1), 4.22–4.17 (dd, $J = 8.8$, 12.8 Hz, 1H, H-6), 4.12–4.06 (m, 2H, H-5, 6), 2.52–2.41 (m, 1H, H-1′), 2.32–2.22 (m, 1H, H-1′), 2.12 (s, 3H, CH_3), 2.07 (s, 3H, CH_3), 2.04 (s, 3H, CH_3), 2.03 (s, 3H, CH_3); ^{13}C NMR (100 MHz, $CDCl_3$): δ 168.1 (C), 167.6 (C), 167.5 (C), 167.3 (C), 130.8 (CH), 115.3 (CH_2), 69.1 (CH), 65.9 (CH), 65.6 (CH), 65.3 (CH), 59.2 (CH_2), 28.7 (CH_2), 18.7 (CH_3), 18.6 (CH_3), 18.5 (CH_3).

4.1.3. 2-(2,3,4,6-Tetra-*O*-acetyl-α-D-galactopyranosyl)-ethanol (17). A solution of **16** (2.08 g, 5.6 mmol) in H_2O–acetone–*t*-BuOH 2:1:1 (40 mL) was cooled to 0 °C and OsO_4 (0.01 equiv) was added and the reaction mixture was stirred for 15 min prior to the addition of solid $NaIO_4$ (3.64 g, 17.02 mmol). After 3 h, another portion of solid $NaIO_4$ (2.40 g, 11.22 mmol) was added and the reaction mixture was stirred overnight at rt. The suspension was then filtered through a Celite pad and the filtrate was concentrated. The residue was dissolved in EtOAc, filtered again through a Celite pad, and the filtrate was dried (Na_2SO_4), filtered and concentrated. The resulting aldehyde (1.73 g) was submitted to the next step without further purification.

To a stirred solution of the crude aldehyde (1.73 g, 4.63 mmol) in MeOH (20 mL), at 0 °C, was added sodium borohydride (70 mg, 1.85 mmol) in three portions. After 30 min, the pH was adjusted to 3 by the addition of AcOH and the mixture was evaporated. The residue was then diluted with CH_2Cl_2 (15 mL), washed with aqueous HCl, with brine (2 × 10 mL), dried (Na_2SO_4) filtered and concentrated under reduced pressure. The residue was purified by flash column chromatography over silica gel (petroleum ether–EtOAc, 6:4) to provide alcohol **17** in 61% yield over the two steps (1.28 g, 2.41 mmol); $[\alpha]_D$ +104.3 (*c* 1, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$): δ 5.41 (t, $J = 3.2$ Hz, 1H, H-4), 5.22–5.16 (m, 2H, H-2, 3), 4.53 (br dd, $J = 11.8$, 9.2 Hz, 1H, H-6a), 4.43 (dt, $J = 11.5$, 3.1 Hz, 1H, H-1), 4.20–4.16 (m, 1H, H-5), 4.02 (dd, $J = 11.8$, 3.9 Hz, 1H, H-6b), 3.74–3.73 (m, 2H, CH_2CH_2OH), 2.11 (s, 3H, CH_3C=O), 2.10 (s, 3H, CH_3C=O), 2.09 (s, 3H, CH_3C=O), 2.06 (s, 3H, CH_3C=O), 1.95–1.86 (m, 1H, H-1′a), 1.69–1.61 (m, 1H, H-1′b); ^{13}C NMR (100 MHz, $CDCl_3$): δ 171.0 (C), 170.0 (C), 169.9 (C), 169.8 (C), 69.99 (CH), 69.84 (CH), 69.00 (CH), 67.94 (CH), 67.32 (CH), 61.37 (CH_2), 60.06 (CH_2), 21.14–21.03 (CH_3); MALDI-TOFMS (CHCA) 377 [M+1], 398 [M+23].

4.1.4. 1-Azido-2-(2,3,4,6-tetra-*O*-acetyl-α-D-galactopyranosyl)ethane (18). To a stirred solution of alcohol **17** (1 g, 2.67 mmol), in pyridine (5 mL), at 0 °C, solid tosyl chloride (5.60 mg, 2.92 mol) was added. The reaction mixture was warmed to rt and stirred overnight, quenched by the addition of H_2O (1 mL) and stirred for another 30 min. The reaction was then diluted with $CHCl_3$ and washed with H_2O, then brine (1 × 50 mL), dried (Na_2SO_4), filtered and concentrated under reduced pressure. The residue was submitted to the next step without further purification.

To a stirred solution of the crude tosylate (1.25 g) in DMF (5 mL), sodium azide (NaN_3) (229 mg, 3.53 mmol) was added and the resulting mixture stirred at rt for 12 h and was then concentrated under reduced pressure. The residue was purified using flash column chromatography over silica gel (petroleum ether–EtOAc, 3:1) to provide **18** (852 mg, 80% over the two steps); ^{1}H NMR (400 MHz, $CDCl_3$): δ 5.41 (t, $J = 3.2$ Hz, 1H, H-4), 5.24 (dd, $J = 8.6$, $J = 4.7$ Hz, 1H, H-2), 5.17 (dd, $J = 8.6$, $J = 3.22$ Hz, 1H, H-3), 4.35–4.29 (m, 2H, H-1, 6a), 4.10–4.06 (m, 2H, H-5, 6b), 3.45–3.38 (m, 2H, CH_2N_3), 2.11 (s, 3H, CH_3C=O), 2.09 (s, 3H, CH_3C=O), 2.07 (s, 3H, CH_3C=O), 2.05 (s, 3H, CH_3C=O), 1.96–1.89 (m, 1H, C*H*CH_2N_3), 1.74–1.66 (m, 1H, C*H*CH_2N_3); ^{13}C NMR (100 MHz, $CDCl_3$): δ 170.6 (C), 169.9 (C), 169.8 (C), 169.7 (C), 69.16 (CH), 69.05 (CH), 68.44 (CH), 68.04 (CH), 67.40 (CH), 61.39 (CH_2), 47.96 (CH_2), 26.25 (CH_2), 21.14-21.03 (CH_3); Maldi-Tof MS (CHCA) 374 [M+1−28].

4.1.5. 2-(α-D-Galactopyranosyl)ethylamine (8). To a solution of azide **18** (490 mg, 1.22 mmol) in MeOH (3 mL), at rt, was added $Pd(OH)_2$–C (100 mg, 10% by weight). The reaction mixture was then placed under an atmosphere of H_2 and stirred for 16 h. The mixture was then filtered through Celite and the filtrate evaporated under reduced pressure. The residue contained a partially deacylated product, so it was dissolved in dry MeOH (3 mL) and NaOMe (1.2 mL, 0.1 M) was added. After 15 min the reaction was neutralized with resin IR-

120 H+, filtered and the filtrate evaporated under reduced pressure to afford **8** (237 mg, 94%); amorphous white solid; ^{1}H NMR (400 MHz, CD_3OD): δ 4.04–3.98 (m, 1H, H-1), 3.85–3.80 (m, 2H, H-2, 4), 3.68–3.52 (m, 4H, H-3, 5, 6a, 6b), 3.05–2.95 (m, 2H, CH_2N), 2.01–1.85 (m, 2H, CH_2CH$_2$N); ^{13}C NMR (100 MHz, CD_3OD): δ 76.14 (CH), 74.99 (CH), 72.31 (CH), 71.53 (CH), 70.46 (CH), 66.36 (CH_2), 63.86 (CH_2), 33.9 (CH_2).

4.1.6. General procedure for amides 10–13. To a solution of the corresponding cinnamic acid derivative (0.5 mmol) in THF (1 mL), at rt, oxalyl chloride (0.2 mL/mmol) was added. The reaction was stirred for 2.5 h, then the volatiles were evaporated under reduced pressure. The residue was dissolved in toluene and added to a mixture of amine and Na_2CO_3 and heated at reflux for 2.5 h. The solvent was evaporated and the residue purified by flash column chromatography on silica gel, using a gradient EtOAc–MeOH, 10:0→8:2.

Compound **10**. Yield 86%; $[\alpha]_D$ +0.183 (*c* 0.5); ^{1}H NMR (400 MHz, D_2O): δ 7.59–7.52 (m, 2H, *H*Ar), 7.42–7.35 (m, 4H, *H*Ar, C*H*=), 6.55 (d, $J = 15.4$ Hz, 1H, C*H*=), 4.15–4.08 (m, 1H, H-1), 3.91–3.87 (m, 2H, H-2, 4), 3.70–3.60 (m, 4H, H-3, 5, 6a, 6b), 3.33–3.26 (m, 2H, CH_2N), 1.91–1.85 (m, 1H, C*H*CH$_2$N), 1.81–1.72 (m, 1H, C*H*CH$_2$N); ^{13}C NMR (100 MHz, D_2O): δ 171.4 (C), 143.7 (CH), 137.1 (C), 132.9 (CH), 131.7 (CH), 130.6 (CH), 122.8 (CH), 76.35 (CH), 74.63 (CH), 72.34 (CH), 71.70 (CH), 70.69 (CH), 63.87 (CH_2), 39.26 (CH_2), 25.93 (CH_2); HRMS calcd for $[M+H]^+$: 338.1604, $[M+Na]^+$: 360.1423. Found: 338.1765, 360.1302.

Compound **11**. Yield 70%; ^{1}H NMR (400 MHz, D_2O): δ 7.31 (d, $J = 15.7$ Hz, 1 H, C*H*=), 6.82 (s, 2H, *H*Ar), 6.45 (d, $J = 15.7$ Hz, 1H, C*H*=), 4.13–4.07 (m, 1H, H-1), 3.98–3.90 (m, 2H, H-2, 4), 3.85–3.65 (m, 13H, H-3, 5, 6a, 6b, 3OCH_3), 3.38 (br t, 2H, CH_2N), 1.98–1.83 (m, 2H, CH_2CH$_2$N); ^{13}C NMR (100 MHz, D_2O): δ 180.4 (C), 155.2 (C), 143.3 (CH), 133.6 (C), 122.6 (CH), 108.0 (CH), 74.63 (CH), 72.33 (CH), 71.66 (CH), 70.67 (CH), 63.84 (CH_2), 63.55 (CH), 58.63 (CH_3), 39.20 (CH_2), 23.63 (CH_2); HRMS calcd for $[M+H]^+$: 428.1921. Found: 428.1967.

Compound **12**. Yield 75%; ^{1}H NMR (400 MHz, CD_3OD): δ 8.29–8.20 (m, 2H, *H*Ar), 7.80–7.75 (m, 2H, *H*Ar), 7.59 (d, $J = 15.7$ Hz, 1H, C*H*=), 6.81 (d, $J = 15.7$ Hz, 1H, C*H*=), 4.12–4.04 (m, 1H, H-1), 3.96–3.82 (m, 3H, H-2, 4, 6a), 3.80–3.76 (m, 1H, H-5), 3.72–3.65 (m, 2H, H-3, 6b), 3.59–3.50 (m, 1H, C*H*N), 3.48–3.362 (m, 1H, C*H*N), 1.97–1.80 (m, 2H, CH_2CH$_2$N); ^{13}C NMR (100 MHz, CD_3OD): δ 170.6 (C), 152.3 (C), 145.5 (C), 141.9 (CH), 132.6 (CH), 129.0 (CH), 127.9 (CH), 77.22 (CH), 76.91 (CH), 74.65 (CH), 72.93 (CH), 72.93 (CH), 65.07 (CH_2), 40.92 (CH_2), 28.61 (CH_2); HRMS calcd for $[M+H]^+$: 405.1274, $[M+Na]^+$: 360.1423. Found: 383.1613, 405.1408.

Compound **13**. Yield 72%; ^{1}H NMR (400 MHz, D_2O): δ 7.42 (d, $J = 15.7$ Hz, 1H, C*H*=), 7.17 (d, $J = 1.7$ Hz, 1H, *H*Ar), 7.11 (dd, $J = 1.7$, $J = 8.0$ Hz, 1H, *H*Ar), 6.92 (d, $J = 8.0$ Hz, 1H, *H*Ar), 6.45 (d, $J = 15.7$ Hz, 1H, C*H*=), 6.02 (s, 2H, OCH_2O), 4.38–4.21 (m, 1H, H-1), 4.06–3.92 (m, 2H, H-2, 4), 3.84–3.70 (m, 4H, H-3, 5, 6a, 6b), 3.43 (br t, 2H, CH_2N), 2.04–1.90 (m, 1H, C*H*CH$_2$N), 1.84–1.65 (m, 1H, C*H*CH$_2$N); ^{13}C NMR (100 MHz, CD_3OD): δ 178.5 (C), 150.4 (2C), 143.4 (CH), 131.9 (C), 126.4 (CH), 124.8 (CH), 111.3 (CH), 109.1 (CH), 104.2 (CH_2), 76.8 (CH), 74.7 (CH), 72.4 (CH), 71.6 (CH), 70.6 (CH), 64.0 (CH_2), 39.2 (CH_2), 26.2 (CH_2); HRMS calcd for $[M+Na]^+$: 404.1321. Found 404.1257.

Compound **14**. A solution of galactose amine **8** (16 mg, 0.078 mmol), diphenyl maleic anhydride (24 mg, 0.097 mmol), Na_2CO_3 (33 mg, 0.31 mmol) and H_2O (100 μL) in MeOH and THF (1 mL, 1:1, vol:vol) was stirred at rt for 1.5 h. The mixture was filtered, the solvents were evaporated under reduced pressure and the residue was purified using a solvent system gradient EtOAc–MeOH, 10:0→6:4 to yield 10 mg of compound **14** (28%); ^{1}H NMR (400 MHz, D_2O): δ 7.30–7.21 (m, 6H, *H*Ar), 7.18–7.04 (m, 4H, *H*Ar), 4.18–4.05 (m, 1H, H-1), 4.00–3.92 (m, 2H, H-2, 4), 3.82–3.61 (m, 4H, H-3, 5, 6a, 6b), 3.44–3.31 (m, 2H, CH_2N), 2.02–1.90 (m, 1H, C*H*CH$_2$N), 1.89–1.80 (m, 1H, C*H*CH$_2$N); ^{13}C NMR (100 MHz, CD_3OD): δ 174.7 (C), 146.8 (C), 139.4 (C), 138.6 (C), 136.5 (C), 133.0 (CH), 132.6 (CH), 132.3 (CH), 131.9 (CH), 131.6 (CH), 131.1 (CH), 130.9 (CH), 130.7 (CH), 130.6 (CH), 130.5 (CH), 74.8 (CH), 72.4 (2 × CH), 71.7 (CH), 70.9 (CH), 63.9 (CH_2), 39.4 (CH_2), 26.1 (CH_2); HRMS calcd for $[M+Na-H_2O]^+$: 462.1529. Found: 462.1719.

4.2. SPR measurements

The SPR measurements were performed on a double channel BIOCORE instrument at 25 °C, in HBS buffer. The glycoprotein asialofetuin (ASF) was covalently attached to a dextran functionalized gold SPR chip via NHS active ester intermediacy following a protocol previously established.[37] A series of measurements with increasing CTB5 concentrations yielded a binding isotherm and a K_{chip} of 70 μM.[38] After each binding experiment the surface was regenerated with 10 mM NaOH for 2.5 min. In a typical competition experiment, CTB5 (70 μM in HBS buffer) was mixed with variable concentrations of the ligands and allowed to reach equilibrium for 5 min at rt. The equilibrated samples were then introduced to the chip-immobilized protein. These measurements generated a new series of binding isotherms, from which equilibrium constants (Req) were

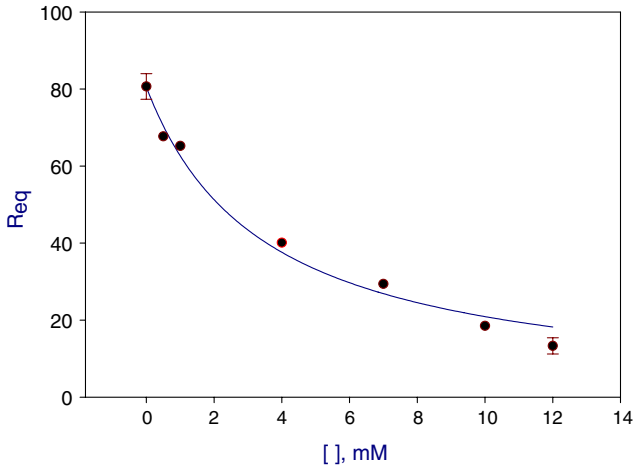


Figure 7. SPR experiments: hyperbolic fitting for compound **10**.

obtained using SPR instrument software. A hyperbolic fitting ($y = \mathrm{Rmax}\ \mathrm{IC}_{50}/(\mathrm{IC}_{50} + x)$) was then used in order to extrapolate the ligand's affinity to the protein. In Figure 7, the hyperbolic fitting of **10** is shown.

4.3. Modeling section

4.3.1. Model preparation. We started from the X-ray CT–GM1 complex (PDB code:3CHB).[8] It is a conformation of the B_5 subunit of the Cholera Toxin with five pentasaccharides GM1 bound to the five binding sites. The binding sites are formed by pairs of B units that are in contact with one another. So two domains on the five that constitute the B_5 subunit of the Cholera Toxin are necessary to model the binding site. We may have chosen any of those pairs, but in the sites involving the chains D and F lattice interactions in the crystal form have prevented the position of the terminal glucose residue to be determined. Therefore they are missing in the PDB file. So we have chosen the chains G and H as models. A sulfur atom of unknown origin and a molecule of 2-(*N*-morpholino)-ethane-sulfonic acid coming from the well solution used for crystallization were removed.

The protein was prepared for the virtual screening experiments following the protocol recommended by Schrödinger.[39] The first phase was to suppress all water molecules except those that are relevant for the binding. Two water molecules in the binding site (W7302 and W7303 in the PDB file) have been identified as playing an important role for CT binding events[40] and were retained. Hydrogens were added to the ligand automatically via the Maestro interface, leaving no lone pair and using an explicit all-atom model. Multiple bonds and formal charges were checked and corrected for the carboxyl group of the sialic acid in GM1. After separation of the complex in separate files for the protein, the ligand, and the water molecules, the Schrödinger, Inc script *Pprep* was run on protein in the presence of bound ligand. This script neutralizes all residues of the protein except those that are close to the ligand or those that form salt bridges. No H-bonding conflict was found. We then proceeded to add the hydrogens to the protein and to the water molecules. One water molecule was reoriented so as to form hydrogen bond with the carboxyl group of sialic acid, as it is suggested by analysis of the X-ray.[8] We then achieved structure preparation by using the Schrödinger, Inc script *premin*. This script performs a minimisation in three steps in order to allow the complex obtained so far to relax without losing the original structure. The minimization is limited in number of steps and is not intended to attain convergence. Instead, the process is stopped as soon as the RMS relative to the original structure exceeds 0.4 Å.

For the full minimization of the complexes, we used the Amber* force field with the Still-Senderowitz pyranose parameters.[42] The Polak-Ribiere Conjugate Gradient (PRCG) algorithm was used with a limit of 5000 steps, unless the energy gradient converged below 0.05 kJ mol^{-1} before the limit was reached. The minimisation used several shells so as to allow the ligand to relax in its new environment and the protein to be flexible enough to allow this relaxation. The first shell consisted of the ligand and the polar hydrogens within 5 Å of the ligand. The second shell consisted of all the atoms within 5 Å of the ligand that did not belong to the first shell. A constraint of 200 kJ mol^{-1} $\mathrm{Å}^{-2}$ was applied to maintain these atoms in place. The third shell consisted of all atoms within 15 Å of the ligand that were not already included. Those atoms were completely frozen, though they participated in the force field. All other atoms were ignored.

4.3.2. Docking experiments. The minimized complex CTB5:GM1 was used for generation of Autodock's and Glide's grids. The grids were needed for flexible docking with Autodock and subsequent scoring in situ with Glide. All single bonds in ligands are treated as freely rotatable in the Autodock procedure, except the C3–OH, C4–OH, C5–C6 and C6–OH bonds of the galactose anchor.

4.3.2.1. Autodock setup. Autodock 3.0 was used, starting from the complex structure prepared as described in the Modeling section and generating the grid from the GM1 ligand using the standard Autodock procedure (from Autodock tutorial). The grid spacing was standard, and the size is set to include all the CT binding sites. We used standard setup of the software, which makes use of a Lamarckian genetic algorithm. Each docking experiment was repeated 100 times. The population size was set to 50, the number of energy evaluations per run was set to 10^6, and the maximum number of generations to 54,000. All other parameters

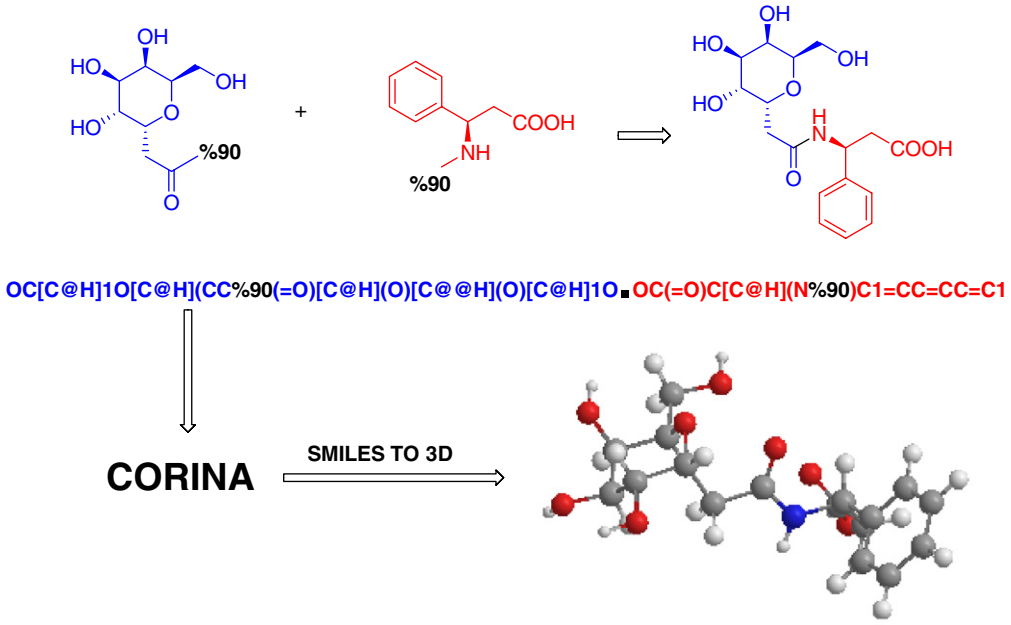


Figure 8. Virtual chemical reaction using a combination of SMILES strings.

used default values. In particular, the rms tolerance for clustering was set to 1 Å. The poses obtained by this procedure were minimized using the same protocol described above for the CT–GM1 complex and scored in situ with Glide. The result of a docking experiment following this procedure is reproducible, that is, even if Autodock uses a genetic algorithm, the final set of docked poses is always the same as that of the GlideScore.

4.3.2.2. Glide setup. Glide uses a number of grids to explore the conformational space of ligands in the binding site and scores each of them. We have used the complex formed by the GH chains of the B_5 subunit of the Cholera Toxin and GM1 to help Glide define the ligand site. All the grids are calculated in an enclosing box. The size of this box has been fitted to GM1 as it was the largest ligand we wanted to dock. Note that the grids were defined for a given conformation of the binding site. This conformation changed slightly each time we minimized a ligand in the binding pocket because the orientation of the polar hydrogen was unconstrained. Those differences were not sufficient nonetheless to justify recalculations of the grid for each minimized complex.

4.4. Focused virtual library

The virtual library was generated using an in-house software by systematically combining a set of fragments on a set of scaffolds. The user defines attachment points for addition of the fragments to the scaffolds. Thus, we were always sure to generate libraries of compounds that were synthetically accessible.

A virtual combinatorial library was prepared by combining SMILES strings.[43] Attaching points on both reactants have been assigned by prefacing the same two-digit number with percent sign (%). The new compounds are created by combining SMILES string of two reactants into the product using dot (.) operator. Example of such virtual chemical reaction is shown in Figure 8.

The final structure represented by SMILES string was then converted to 3D structures using CORINA.[44,45]

Acknowledgments

We thank the European Union for financial support (under contract HPRN-CT-2002-00173, Glycidic Scaffolds Network) and the Interdepartmental Center for Large Instrumentation (CIGA) at the University of Milan for use of the SPR instrumentation.

References

1. Cholera 1999, *WHO Weekly Epidemiolog. Rev.* Vol. 75, August 4, 2000; pp 249–256.
2. Lencer, W. I.; Slaslwsky, D. *Biochim. Biophys. Acta* **2005**, *1746*, 314–321.
3. Fan, E.; O'Neil, C. J.; Mithchell, D. D.; Robien, M. A.; Zhang, Z.; Pickens, J. C.; Tan, X.-J.; Korotkov, K.; Roach, C.; Krumm, B.; Verlinde, C. L. M. J.; Merritt, E. A. *Int. J. Med. Micr.* **2004**, *294*, 217–223.
4. Moss, J.; Vaughan, M. *J. Biol. Chem.* **1977**, *252*, 2455–2457.
5. Cassel, D.; Pfeuffer, T. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 2669–2673.

6. (a) Fan, E.; Merritt, E. A.; Verlinde, C. L. M. J.; Hol, W. G. J. *Curr. Opin. Chem. Biol.* **2000**, *10*, 680–686; (b) Velter, I. A.; Politi, M.; Podlipnik, Č.; Nicotra, F. *Minirev. Med. Chem.* **2007**, *7*, 159–170.
7. Merritt, E. A.; Kuhn, P.; Sarfaty, S.; Erbe, J. L.; Holmes, R. K.; Hol, W. G. J. *J. Mol. Biol.* **1998**, *282*, 1043–1059.
8. Merritt, E. A.; Sarfaty, S.; van den Akker, F.; L'Hoir, C.; Martial, J. A.; Hol, W. G. J. *Protein Sci.* **1994**, *3*, 166–175.
9. Turnbull, W. B.; Precious, B. L.; Homans, S. W. *J. Am. Chem. Soc.* **2004**, *126*, 1047–1054.
10. Bernardi, A.; Checchia, A.; Brocca, P.; Sonnino, S.; Zuccotto, F. *J. Am. Chem. Soc.* **1999**, *121*, 2032–2036.
11. Acquotti, D.; Poppe, L.; Dabrowski, J.; van der Lieth, C.-W.; Sonnino, S.; Tettamanti, G. *J. Am. Chem. Soc.* **1990**, *112*, 7772–7778.
12. Bernardi, A.; Podlipnik, Č.; Jiménez-Barbero, J. GM1 Glycomimetics and Bacterial Enterotoxins. In *Protein–Carbohydrate Interactions in Infectious Diseases*; RSC: Bewley, CA, 2006; pp 73–91.
13. Bernardi, A.; Arosio, D.; Potenza, D.; Sanchez-Medina, I.; Mari, S.; Cañada, F. J.; Jiménez-Barbero, J. *Chem. Eur. J.* **2004**, *10*, 4395–4406.
14. Arosio, D.; Baretti, S.; Cattaldo, S.; Potenza, D.; Bernardi, A. *Bioorg. Med. Chem. Lett.* **2003**, 3831–3834.
15. Bernardi, A.; Carrettoni, V.; Grosso Ciponte, V.; Monti, D.; Sonnino, S. *Bioorg. Med. Chem. Lett.* **2000**, *13*, 2197–2200.
16. Bernardi, A.; Potenza, D.; Capelli, A. M.; Garcìa-Herrero, A.; Cañada, F. J.; Jiménez-Barbero, J. *Chem. Eur. J.* **2002**, *8*, 4598–4612.
17. Bernardi, A.; Arosio, D.; Sonnino, S. *Neurochem. Res.* **2002**, *27*, 539–545.
18. Bernardi, A.; Arosio, D.; Manzoni, L.; Monti, D.; Posteri, H.; Potenza, D.; Mari, S.; Jiménez-Barbero, J. *Org. Biomol. Chem.* **2003**, *1*, 1–9.
19. Minke, W. E.; Roach, C.; Hol, W. G.; Verlinde, C. L. *Biochem.* **1999**, *38*, 5684–5692.
20. Minke, W. E.; Hong, F.; Verlinde, C. L.; Hol, W. G.; Fan, E. *J. Biol. Chem.* **1999**, *274*, 33469–33473.
21. Pickens, J. C.; Merritt, E. A.; Ahn, M.; Verlinde, C. L. M. J.; Hol, W. G. J.; Fan, E. *Chem. Biol.* **2002**, *9*, 215–224.
22. Mitchell, D. D.; Pickens, J. C.; Korotkov, K.; Fan, E.; Hol, W. G. J. *Bioorg. Med. Chem.* **2004**, *12*, 907–920.
23. Postema, M. H. D. *C*-Glycoside Synthesis; CRC Press: Boca Raton, 1995.
24. Levy, D. E.; Tang, C. *The Chemistry of C-Glycosides.* Tetrahedron Organic Chemistry Series; Pergamon: New York, 1995; Vol. 13.
25. Nicotra, F. Synthesis of *C*-glycosides of Biological Interest. In *Topics in Current Chemistry*; Springer: Berlin Heidelberg, 1997; Vol. 187, pp 56–83.
26. Nicotra, F. *Modified Carbohydrates and Carbohydrate Analogs*; Chapman & Hall, Blackie Academic & Professional: London, 1997, pp 384–429.
27. Du, Y.; Linhardt, J. R. *Tetrahedron* **1998**, *54*, 9913–9959.
28. Bililign, T. G. B. R.; Thorson, J. S. *Nat. Prod. Rep.* **2005**, *22*, 742–760.
29. Yuan, Y.; Linhardt, J. R. *Curr. Top. Med. Chem.* **2005**, *5*, 1393–1430.
30. La Ferla, B.; Cardona, F.; Perdigão, I.; Nicotra, F. *Synlett* **2005**, *17*, 2641–2642.
31. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Comput. Chem.* **1998**, *19*, 1639–1662.
32. Minke, W. E.; Diller, D. J.; Hol, W. G. J.; Verlinde, C. L. *J. Med. Chem.* **1999**, *42*, 1778–1788.
33. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shaw, D. E.; Shelley, M.; Perry, J. K.; Francis, P.; Shenkin, P. S. *J. Med. Chem.* **2004**, *47*, 1739–1749.
34. Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. *J. Med. Chem.* **2004**, *47*, 1750–1759.
35. Bennek, J. A.; Gray, G. A. *J. Org. Chem.* **1987**, *52*, 892–897.
36. Broxterman, H. J. G.; Kooreman, P. A.; van der Elst, H.; Roelen, H. C. P. F.; van der Marel, G. A.; van Boom, J. H. *Recl. Trav. Chim. Pays-Bas* **1990**, *109*, 583.
37. Arosio, D.; Vrasidas, I.; Valentini, P.; Liskamp, R. M. J.; Pieters, R. J.; Bernardi, A. *Org. Biomol. Chem.* **2004**, *2*, 2113–2124.
38. de Mol, N. J.; Plomp, E.; Fischer, M. J. E.; Ruijtenbeek, R. *Anal. Biochem.* **2000**, *279*, 61–70.
39. Schrodinger. FirstDiscovery 2.5, User Manual, March **2003**.
40. Bernardi, A.; Raimondi, L.; Zuccotto, F. *J. Med. Chem.* **1997**, *40*, 1855–1862.
41. Warren, G. L.; Andrews, C. W.; Capelli, A.-M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. *J. Med. Chem.* **2006**, *49*, 5912–5931.
42. Senderowitz, H.; Parish, C.; Still, W. C. *J. Am. Chem. Soc.* **1996**, *118*, 2078–2086.
43. Weininger, D. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31–36.
44. Sadowski, J.; Gasteiger, J. *Chem. Rev.* **1993**, *93*, 2567–2581.
45. Sadowski, J.; Gasteiger, J.; Klebe, G. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1000–1008.
46. Mertz, J. A.; McCann, J. A.; Picking, W. D. *Biochem. Biophys. Res. Commun.* **1996**, *226*, 140–144.