Note

An *in silico* virtual screening study for the design of norovirus inhibitors: Fragment-based molecular docking and binding free energy calculations‡

Magnus Lundborg,[1] Eunus Ali,[2] and Göran Widmalm*

Department of Organic Chemistry, Arrhenius Laboratory, Stockholm University, S-106 91 Stockholm, Sweden

*Corresponding author. Tel. +46 8 16 37 42. E-mail address: gw@organ.su.se

‡Dedicated to the memories of Professor Lennart Kenne and Professor Malcolm Perry.

[1]Present address: Science for Life Laboratory, KISP (Karolinska Institutet Science Park), Tomtebodavägen 23A, S-171 65 Solna, Sweden

[2]Present address: School of Medicine, Faculty of Health Sciences, Flinders University, GPO Box 2100, Adelaide 5001, South Australia, Australia

ABSTRACT

Gastrointestinal infections caused by noroviruses may be prevented by inhibition of their binding to histo-blood group carbohydrate antigens. A fragment-based virtual screening approach was used, employing docking followed by molecular dynamics simulations in order to enable binding free energy calculations using the linear interaction energy method. The resulting structures, composed of high-affinity fragments, can be a good starting point for lead optimizations and four molecules that pass both REOS and SYLVIA filters, which can remove known toxic features and assess the synthetic accessibility, respectively, are proposed as inhibitors.

Keywords: Carbohydrate antigen, Influenza, *Caliciviridae*, *In silico* screening, Scoring function, Ro3

Noroviruses (NVs) are RNA viruses belonging to the *Caliciviridae* family, causing acute gastrointestinal infections, such as the 'winter vomiting disease'.[1,2,3,4,5] They are considered responsible for over two thirds of foodborne illnesses due to known agents, but only 7% of deaths caused by these diseases.[6] However, there are currently no vaccines or drugs against NVs,[7] although significant effort is presently devoted to the understanding of the interaction between NVs, experimentally studied in the form of viruslike particles, and glycosphingolipids[8] as well as galactoceramides.[9] Many viruses bind carbohydrates, of different epitopes, on the cell surface in order to attach to their host cells.[1] Histo-blood group antigens (HBGAs) act as receptors to NVs, and separate NV strains have different binding patterns.[7,10,11] HBGAs are glycans linked to glycoproteins or glycolipids and can be found on red blood cells, in saliva and in the gut.[12] The backbone structures of the HBGAs are important for presenting the oligosaccharide antigen, which is what is recognized in the interaction,[13] and inhibition of the association of NVs to HBGAs may stop the infection or prevent its spreading.[7] An alternative approach to the inhibition of NVs was recently presented based on bisulfite adducts of transition state inhibitors to its 3C-like protease,[14] which is essential for virus replication. The GI.1 (Norwalk) virus was the first norovirus discovered and it has been shown that its capsid is formed by 90 protein dimers.[1,13] The capsid protein can be divided into two domains, the shell domain (S) and the protruding domain (P), which in turn can be divided into the subdomains P1 and P2.[13] GII.4 (VA387) is an NV strain, binding strongly to the HBGA epitopes A, B and H.[15] It was used in this study since there are crystal structures available of the capsid P domain interacting with HBGA trisaccharides, through direct as well as water mediated interactions (Figure 1), but also because of its broad specificity.

Herein virtual screening[16] by fragment-based[17] docking, using AutoDock Vina,[18] has been performed, followed by molecular dynamics (MD) simulations[19] and binding free energy calculations employing the linear interaction energy (LIE)[20] method. The more accurate binding free energy (BFE) calculations enable more confident pose prediction, which is important when combining high-affinity fragments into potential inhibitors, whose binding modes and affinities are assessed in the same way. The HBGA A trisaccharide in PDB entry 2OBS is incorrect, as has been reported before,[21] and therefore, it was decided to focus on the HBGA B trisaccharide in the comparisons between docking and crystal structures. The protein residues contributing most to the binding affinities of the HBGA B trisaccharide were identified (see Table 1 for interaction energies) since these ones should be key residues for inhibitors to bind to.

Two small-molecule libraries were selected for this study. The Maybridge library is a diverse and small library which has been used successfully before.[22] This was complemented by a selection from the ZINC database, which was designed to retrieve slightly smaller compounds in order to

improve the efficiency of the ligands, whilst still covering a large chemical space. The differences in binding free energies, from AutoDock Vina, of the docked fragments to the active site lacking water molecules were small. Considering the general lack of accuracy of scoring functions used in docking programs[23] it is indeed difficult to select fragments to be used as scaffolds. By approximating the affinity, using the more accurate LIE method on the top-ranked structures from docking, it was possible to combine promising fragments into candidate molecules for inhibition of the binding of the norovirus capsid protein to HBGAs. The docking and MD simulations with six water molecules retained did not provide any additional information, resulting in less favorable affinities of the HBGA trisaccharides and comparable results for the fragments of the libraries (data not shown). Thus, it was decided to focus on the calculations without water in the active site bridging between protein and fragment ligands that were to be the basis of inhibitors to be formed. The rationale for this was that an additional entropy gain[24] may be obtained for the proposed structures if water molecules will be released upon ligand binding to the protein.

To ensure that convergence was reached during the equilibration phase of the MD simulations, the interaction energies during the first and second half of the production phase were compared. If the average van der Waals or electrostatics interaction energies differed by more than 1 kcal·mol^{-1} between the two parts, the trajectories were not deemed stable, and that production run was instead considered a prolonged equilibration. A new production run was started instead – continuing from the end of the last equilibration trajectory using the same parameters. If convergence was still not reached, the interaction energies from the last set of simulations were used anyhow as an estimate of the BFE. The reported binding free energy for each ligand was the average over the production phase of the respective simulation. The calculated binding free energies using the LIE method of the two HBGA trisaccharides and the top ten molecules from the Maybridge fragment library (all 500 structures are listed in the supporting information of ref. 22) as well as the top ten fragments of the ZINC database are compiled in Table 2.

During the course of this study Rademacher et al. reported results of a competitive NMR screening targeting norovirus,[25] focused on the L-Fuc binding region, using the Maybridge fragment library. They selected fragment number 160 for further developments of a multivalent inhibitor. Using our methodology, that compound was ranked very low by AutoDock Vina (position 397), but the calculated LIE of structure 160 was −1.3±0.0 kcal·mol^{-1} (the uncertainty in the BFE calculation is the difference of the BFE of the first and second half of the MD simulation), placing it in the middle region of the LIE results. This clearly shows that using only the top-20 ligands from docking may not be enough to capture all binders, especially not when the overall difference in docking score is small.

Four molecular structures are proposed from high scoring fragments binding in different parts of the binding site, close enough to be linked together. In order not to proceed with and optimize unsuitable compounds a REOS (Rapid Elimination of Swill)[26,27] filter procedure was used in Canvas (version 1.3, Schrödinger, LLC, New York, NY, 2010). This filter can remove compounds that contain, for example, known toxic features. The four proposed compounds **23** – **26** (Scheme 1) passed the tests by REOS, showing that they would be good candidates for further optimizations. The synthetic accessibility was assessed using SYLVIA (http://www.molecular-networks.com),[28] giving scores ranging from 3.85 to 5.61, where 1 is easy and 10 is difficult to synthesize. This means that they are all appraised to be of intermediate difficulty to synthesize using commercially available starting materials. These four molecules all have high calculated affinities for the norovirus capsid protein (legend of Scheme 1) and may be effective inhibitors of the HBGA binding. They have in common planar aromatic rings and one or several nitrogen atoms. Compound **23** contains a sugar residue, compounds **24** and **25** a carboxyl-hydrazide functional group and compound **26** a trifluoromethyl group. The best poses of these compounds, as well as HBGA B as reference, are shown in Figure 2. The docking of the HBGA B trisaccharide to the protein where six water molecules were retained showed a smaller RMSD < 1 Å (Figure 2a) whereas when all water molecules were removed a larger RMSD > 2 Å (Figure 2b) was observed, as could be anticipated, and supports that the docking procedure works well. Interestingly, the four potential inhibitors **23** – **26** that are independently proposed herein all cover the region where the α-L-Fuc group of the trisaccharide resides (Figure 2c – 2f), that is, the same region targeted by Rademacher et al. in their recent study.[25] Furthermore, during evolution over the last two decades the fucose-binding amino acid residues have stayed strictly conserved in GII.4 viruses, whereas residues interacting with the substituted galactose sugar have varied,[29] supporting that aiming to inhibit the 6-deoxy-sugar region should be a suitable approach. Additionally, the computational interaction studies of ABO oligosaccharides with norovirus GII.4 showed that the binding strength could be improved with larger oligosaccharides, but it was still the α-(1→2)-linked fucosyl residue that was most important for the contact to the protein.[21] This line of reasoning is further supported by the fact that humans having a G428A mutation in the *FUT2* gene, which encodes for an α-(1→2)-fucosyl transferase, are strongly protected from infection since the enzyme thereby becomes inactivated and cannot form the pertinent fucosyl-containing oligosaccharide structure(s) necessary for infection by the virus.[30]

In conclusion, using molecular docking and MD simulations we have proposed four new molecules that may act as high-affinity inhibitors of noroviruses in the process of preventing its binding to histo-blood group antigens. Since these potential inhibitors should be relatively straightforward to synthesize by apt organic chemists and the fact that biophysical techniques such as surface plasmon

resonance and NMR spectroscopy are being used to evaluate protein-ligand binding of inhibitors[31] we envision a combined approach along these lines as the next step in the development of efficient inhibitors of norovirus infection.

1. Methods

1.1 Model setup

The biological unit (dimer) of the crystal structure of the GII.4 (VA387) capsid P domain binding the HBGA B trisaccharide (PDB entry 2OBT)[10] was used as target for the docking. The protein was prepared for docking using the Protein Preparation Wizard in Maestro (Maestro, Schrödinger, 2010), including setting protonation states of histidines. Two different protein set-ups were used: discarding all water molecules as well as retaining waters 576, 586, 590, 645, 648 and 724, which were deemed potentially important for HBGA B interactions (cf. Figure 1). Protein pdbqt files were prepared using AutoDock Tools,[32] which included addition of partial charges.[33]

1.2 Fragment libraries

Two fragment libraries were used as ligand sources for docking. The first was the Maybride Ro3 library of 500 fragments (Thermo Fisher Scientific Inc.) and the second was molecules from the ZINC database[34] (without subset restrictions) retrieved with the following search parameters: $-2 \leq$ charge ≤ 2, xLogP ≤ 3, number of rotatable bonds ≤ 4, number of hydrogen bond donors ≤ 4, number of hydrogen bond acceptors ≤ 4 and $125 \leq$ molecular weight ≤ 225. This query gave 6423 small molecules from ZINC (September 10, 2010). In addition to this the trisaccharides of HBGA A and B were prepared. The HBGA B trisaccharide from the PDB file 2OBT was used, whereas the structure of the HBGA A trisaccharide in 2OBS contains errors, as reported previously,[21] and was therefore modeled based on the HBGA B trisaccharide. The two trisaccharides were energy minimized using the MMFF94s force field[35] in Avogadro[36] prior to docking.

1.3 Docking

All molecules were converted to pdbqt files, that is, partial charges and active torsions set, using AutoDock Tools. Three separate docking simulations were performed: the HBGA trisaccharides, the Maybridge library and the fragments from the ZINC database. AutoDock Vina[18] was used to perform the dockings. The box used for docking was 20 Å × 20 Å × 20 Å centered close to the α-L-Fuc*p*-(1→2)-β-D-Gal*p* glycosidic linkage, in order to encompass as much of the binding site as possible. Ten binding modes were kept from the docking; the best three binding modes were analyzed by MD simulations. The 20 top-ranked structures from the Maybridge and ZINC libraries

as well as the two trisaccharides were analyzed using MD simulations for LIE calculations. The docking procedures were performed on an Intel(R) Core(TM) 2 Quad CPU Q6600 2.40 GHz processor with 4 GB RAM running Ubuntu 10.10 64-bit operating system. The four molecules built from fragments with high affinity were treated in the same way as the trisaccharide structures (partial charges assigned followed by energy minimization).

1.4 Molecular dynamics simulations

The program Q was used for the molecular dynamics (MD) simulations.[19] Two types of simulations were carried out for each ligand. Firstly, the free state started from an energy minimized conformation, obtained using a MacroModel (MacroModel, Schrödinger, 2010) Conformational Search (ConfGen), with 25 steps performed during the conformational search, employing the thorough search mode and a maximum of 5 iterations for post energy minimization of generated structures using the optimal minimization method. The conformation of lowest potential energy was kept for the MD simulations without any protein present. Secondly, the top three poses from the docking results were used as alternative bound states. The ligand atom types were assigned by making an OPLS AA VEGA ATD template[37] and the ligand OPLS all-atom[38] partial charges were assigned using Maestro.[39] In both cases the systems were solvated by generating a TIP3P water[40] sphere with a radius of 20 Å, centered on the atom closest to the geometrical center of the ligand. All water molecules closer than 2.4 Å to any solute heavy atoms were removed. OPLS all-atom[38] force field parameters were used. All charged protein residues that were outside the water sphere, or with the atom carrying the formal charge of the residue within 4 Å of the water sphere edge, were set to their neutral state. If there still was a net charge in the protein, charged residues were made neutral, starting with residues furthest away from the water sphere center until the system was neutral. An interaction energy correction for transforming remote charged residues into their uncharged state was calculated for each ligand pose before the MD simulations were started.[41] This was done by iterating through all atoms in the ligand and calculating the electrostatic interaction energy between it and the atom carrying the formal charge (of the charged state) for each neutralized protein residue using a dielectric constant of 80.

In the protein-ligand simulations all atoms outside the water sphere were tightly restrained to their initial coordinates and nonbonded interactions between restrained pairs of atoms were not calculated. In the simulations without protein a 10.0 $kcal \cdot mol^{-1} \cdot Å^{-2}$ harmonic restraint was applied at the center of the molecule to prevent it from approaching the edge of the water sphere. The SHAKE algorithm[42] was applied to all bonds to hydrogen atoms in order to allow for longer time steps. The water molecules at the sphere surface were subjected to radial and polarization restraints, similar to the SCAAS model.[43] During the equilibration phase the nonbonded cutoff was set to 10 Å

for all atoms except for the ligand, to which no cutoff was applied. In the production phase (also referred to as the dynamics phase), when results were collected, the nonbonded cutoff of the non-ligand atoms was increased to 13 Å regarding solvent-solvent and solute-solvent interactions. The nonbonded pair lists were updated every 15 and 10 time steps for the equilibration and production phases, respectively. Long-range electrostatic interactions were treated with the local reaction field approximation,[44] except for the ligand. Equilibration simulations were run for almost 400 ps in the unbound state and 420 ps in the bound state. During the equilibration simulations the temperature and time steps were increased gradually and the relatively short equilibration stages also allowed analyzing the atom positions at regular intervals, based on the restart files, without saving the whole trajectories. The production runs lasted for 1 ns for unbound and 600 ps for protein ligand-bound simulations. The ligand-surrounding energies were sampled every 100 fs during the production runs. Three complete simulations (except the first equilibration stage, which was assumed not to have a large impact) were carried out for each of the three docked poses, as well as the ligand-only state, using different random seeds. The total simulation time of the production phase in the free state of each ligand was 3.0 ns (3 runs of 1.0 ns) and for each of the three top poses from the docking of every ligand the total simulation time of the production phase was 1.8 ns (3 runs of 0.6 ns). The MD simulations of each ligand was performed on one node, consisting of two quad-core Intel Harpertown 2.66 GHz CPUs (E5430) with 8 GB RAM, running the free state and the three bound conformations on two CPU cores each. At most 20 simulation nodes were running in parallel, that is, MD simulations of 20 ligands.

LIE binding free energies were calculated from the MD simulations using the following equation:[45] $\Delta G_{bind}^{LIE} = \alpha\Delta\left\langle V_{l-s}^{vdW}\right\rangle + \beta\Delta\left\langle V_{l-s}^{el}\right\rangle + \gamma$, where α is the van der Waals interaction coefficient, β is the electrostatic coefficient and γ is a constant to correct for systematic deviations. The difference between the average van der Waals interaction energies, of the ligand to its surroundings, in the free and bound state, is represented by $\Delta\left\langle V_{l-s}^{vdW}\right\rangle$ whereas $\Delta\left\langle V_{l-s}^{el}\right\rangle$ is the difference between the average intermolecular electrostatic interaction energies of the ligand in the two states. The factor α was set to 0.18 and β to either 0.33, 0.37, 0.43 or 0.50 depending on the ligand charge and the number of hydroxyl groups.[45] Since the absolute binding free energy was not important to calculate in this study γ was not optimized and was set to zero instead.

Acknowledgements

This work was supported by a grant from the Swedish Research Council. Computing resources were kindly provided by the Center for Parallel Computers (PDC), Stockholm, Sweden.

Supplementary data

Supporting information for this article is available on-line. It contains additional information on the MD simulations, binding free energies and schematic structures of trisaccharides and the twenty top-scored fragments of each of the two libraries.

References

1. Hutson, A. M.; Atmar, R. L.; Estes, M. K. *Trends Microbiol.* **2004**, *12*, 279-287.
2. Zheng, D.-P.; Ando, T.; Frankhauser, R. L.; Beard, R. S.; Glass, R. I.; Monroe, S. S. *Virology* **2006**, *346*, 312-323.
3. Nordgren, J.; Kindberg, E.; Lindgren, P.-E.; Matussek, A.; Svensson, L. *Emerg. Infect. Dis.* **2010**, *16*, 81-87.
4. Shanker, S.; Choi, J.-M.; Sankaran, B.; Atmar, R. L.; Estes, M. K.; Prasad, B. V. V. *J. Virol.* **2011**, *85*, 8635-8645.
5. Shirato, H. *Front. Microbiol.* **2012**, *3*, 1-3.
6. Mead, P. S.; Slutsker, L.; Dietz, V.; McCaig, L. F.; Bresee, J. S.; Shapiro, C.; Griffin, P. M.; Tauxe, R. V. *Emerg. Infect. Dis.* **1999**, *5*, 607-625.
7. Feng, X.; Jiang, X. *Antimicrob. Agents Chemother.* **2007**, *51*, 324-331.
8. Bally, M.; Gunnarsson, A.; Svensson, L.; Larson, G.; Zhdanov, V. P.; Höök, F. *Phys. Rev. Lett.* **2011**, *107*, 188103.
9. Bally, M.; Rydell, G. E.; Zahn, R.; Nasir, W.; Eggeling, C.; Breimer, M. E.; Svensson, L.; Höök, F.; Larson, G. *Angew. Chem. Int. Ed.* **2012**, *51*, 12020-12024.
10. Cao, S.; Lou, Z.; Tan, M.; Chen, Y.; Liu, Y.; Zhang, Z.; Zhang, X. C.; Jiang, X.; Li, X.; Rao, Z. *J. Virol.* **2007**, *81*, 5949-5957.
11. Tan, M.; Hegde, R. S.; Jiang, X. *J. Virol.* **2004**, *78*, 6233-6242.
12. Marionneau, S.; Cailleau-Thomas, A.; Rocher, J.; Le Moullac-Vaidye, B.; Ruvoën, N.; Clément, M.; Le Pendu, J. *Biochimie* **2001**, *83*, 565-573.
13. Prasad, B. V.; Hardy, M. E.; Dokland, T.; Bella, J.; Rossmann, M. G.; Estes, M. K. *Science* **1999**, *286*, 287-290.

14. Mandadapu, S. R.; Gunnam, M. R.; Tiew, K.-C.; Uy, R. A. S.; Prior, A. M.; Alliston, K. R.; Hua, D. H.; Kim, Y.; Chang, K.-O.; Groutas, W. C. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 62-65.
15. Huang, P.; Farkas, T.; Zhong, W.; Tan, M.; Thornton, S.; Morrow, A. L.; Jiang, X. *J. Virol.* **2005**, *79*, 6714-6722.
16. Waszkowycz, B.; Clark, D. E.; Gancia, E. *WIREs Comput. Mol. Sci.* **2011**, *1*, 229-259.
17. Boyd, S. M.; Turnbull, A. P.; Walse, B. *WIREs Comput. Mol. Sci.* **2012**, *2*, 868-885.
18. Trott, O.; Olson, A. J. *J. Comput. Chem.* **2010**, *31*, 455-461.
19. Marelius, J.; Kolmodin, K.; Feierberg, I.; Åqvist, J. Q. *J. Mol. Graph. Model.* **1998**, *16*, 213-225.
20. Åqvist, J.; Medina, C.; Samuelsson, J. E. *Protein Eng.* **1994**, *7*, 385-391.
21. Koppisetty, C. A. K.; Nasir, W.; Strino, F.; Rydell, G. E.; Larsson, G.; Nyholm, P.-G. *J. Comput. Aided Mol. Des.* **2010**, *24*, 423-431.
22. Rademacher, C.; Landström, J.; Sindhuwinata, N.; Palcic, M. M.; Widmalm, G.; Peters, T. *Glycoconjugate J.* **2010** , *27*, 349-358.
23. Scior, T.; Bender, A.; Tresadern, G.; Medina-Franco, J.-L.; Martínez-Mayorga, K.; Langer, T.; Cuanalo-Contreras, K.; Agrafiotis, D. K. *J. Chem. Inf. Model.* **2012**, *52*, 867-881.
24. Binder, F. P. C.; Lemme, K.; Preston, R. C. Ernst, B. *Angew. Chem. Int. Ed.* **2012**, *51*, 7327-7331.
25. Rademacher, C.; Guiard, J.; Kitov, P. I.; Fiege, B.; Dalton, K. P.; Parra, F.; Bundle, D. R.; Peters, T. *Chem. Eur. J.* **2011**, *17*, 7442-7453.
26. Walters, W. P.; Stahl, M. T.; Murcko, M. A. *Drug Discov. Today* **1998**, *3*, 160-178.
27. Walters, W. P.; Murcko, M. A. *Adv. Drug Deliv. Rev.* **2002**, *54*, 255-271.
28. Boda, K.; Seidel, T.; Gasteiger, J. *J. Comput. Aided Mol. Des.* **2007**, *21*, 311-325.
29. Lindesmith, L. C.; Donaldson, E. F.; LoBue, A. D.; Cannon, J. L.; Zheng, D.-P.; Vinje, J.; Baric, R. S. *PLoS Med.* **2008**, *5*, e31.
30. Rydell, G. E.; Kindberg, E.; Larsson, G.; Svensson, L. *Rev. Med. Virol.* **2011**, *21*, 370-382.
31. Schaefer, K.; Albers, J.; Sindhuwinata, N.; Peters, T.; Meyer, B. *ChemBioChem* **2012**, *13*, 443-450.
32. Sanner, M. F. *J. Mol. Graphics Mod.* **1999**, *17*, 57-61.
33. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765-784.
34. Irwin, J. J.; Shoichet, B. K. *J. Chem. Inf. Model.* **2005**, *45*, 177-182.
35. Halgren, T. A. *J. Comput. Chem.* **1999**, *20*, 720-729.

36. Avogadro: an open-source molecular builder and visualization tool. Version 1.0. http://avogadro.openmolecules.net/
37. Pedretti, A.; Villa, L.; Vistoli, G. *Theor. Chem. Acc.* **2003**, *109*, 229-232.
38. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225-11236.
39. Maestro, version 9.1, Schrödinger, LLC, New York, NY, 2010.
40. Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R. W.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926-935.
41. Åqvist, J. *J. Comput. Chem.* **1996**, *17*, 1587-1597.
42. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327-341.
43. King, G.; Warshel, A. *J. Chem. Phys.* **1989**, *91*, 3647-3661.
44. Lee, F. S.; Warshel, A. *J. Chem. Phys.* **1992**, *97*, 3100-3107.
45. Hansson, T.; Marelius, J.; Åqvist, J. *J. Comput. Aided Mol. Des.* **1998**, *12*, 27-35.
46. Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. *Prot. Eng.* **1995**, *8*, 127-134.

Legends to figures and scheme

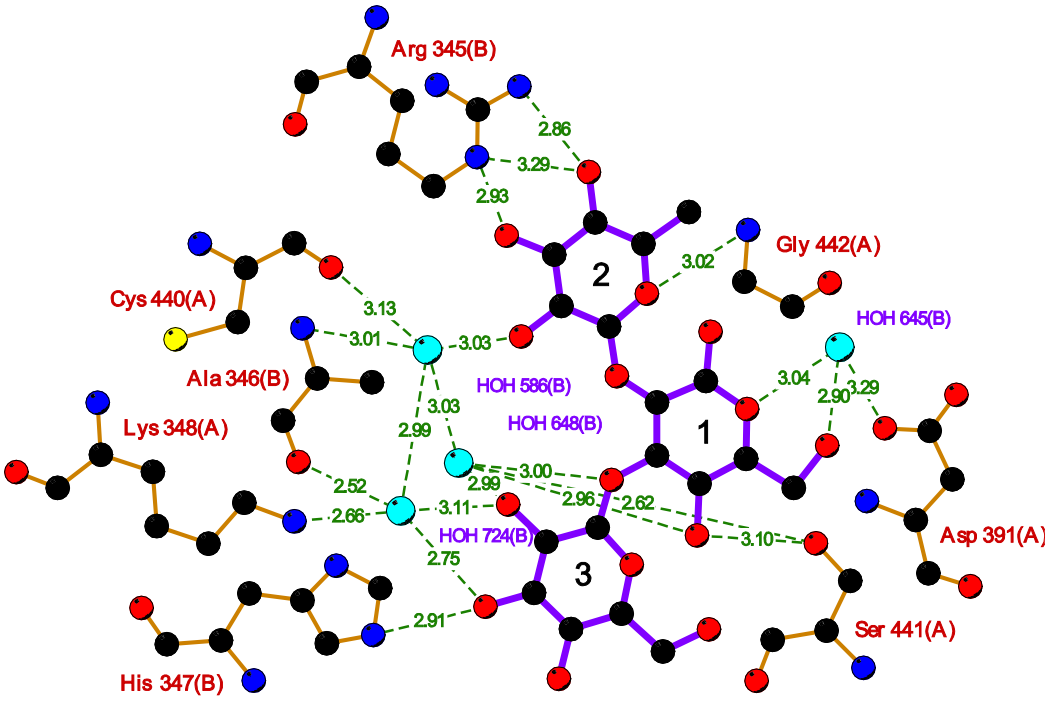
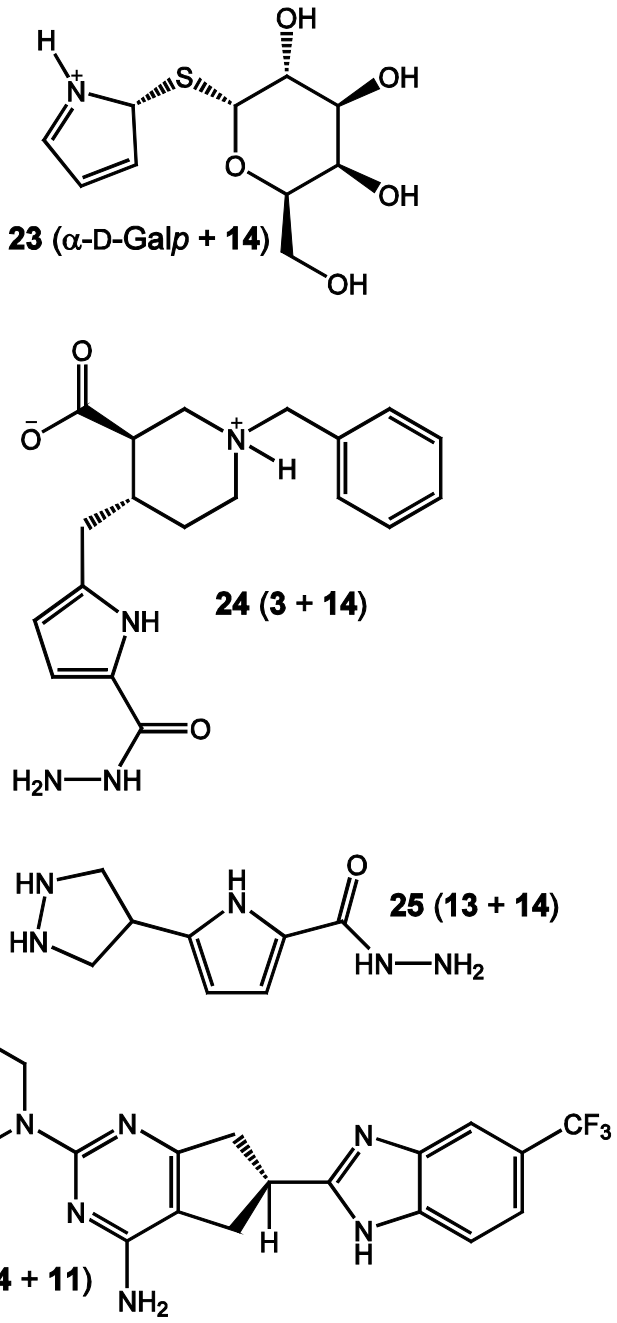


Figure 1. The HBGA B trisaccharide binds to the norovirus GII.4 capsid P domain through an intricate network of interactions. The structure is from PDB entry 2OBT and was generated using LIGPLOT.[10,46] The two protein monomers are denoted A and B. The residues in the trisaccharide denoted 1, 2 and 3 are β-D-Gal, α-L-Fuc and α-D-Gal, respectively.



Scheme 1. Proposed inhibitors (**23** – **26**) based on high-scoring fragments, indicated in parenthesis. The calculated BFE from LIE were –12.8±0.9, –12.1±0.1, –9.8±0.1, and –7.6±0.1 kcal·mol^{-1} and from AutoDock Vina (for reference) –5.1, –6.5, –5.8, and –7.9 kcal·mol^{-1} for compounds **23** – **26**, respectively. For compound **23** the protein-ligand simulation was not fully converged during the additional run and uncertainty span should therefore be regarded as an estimate.

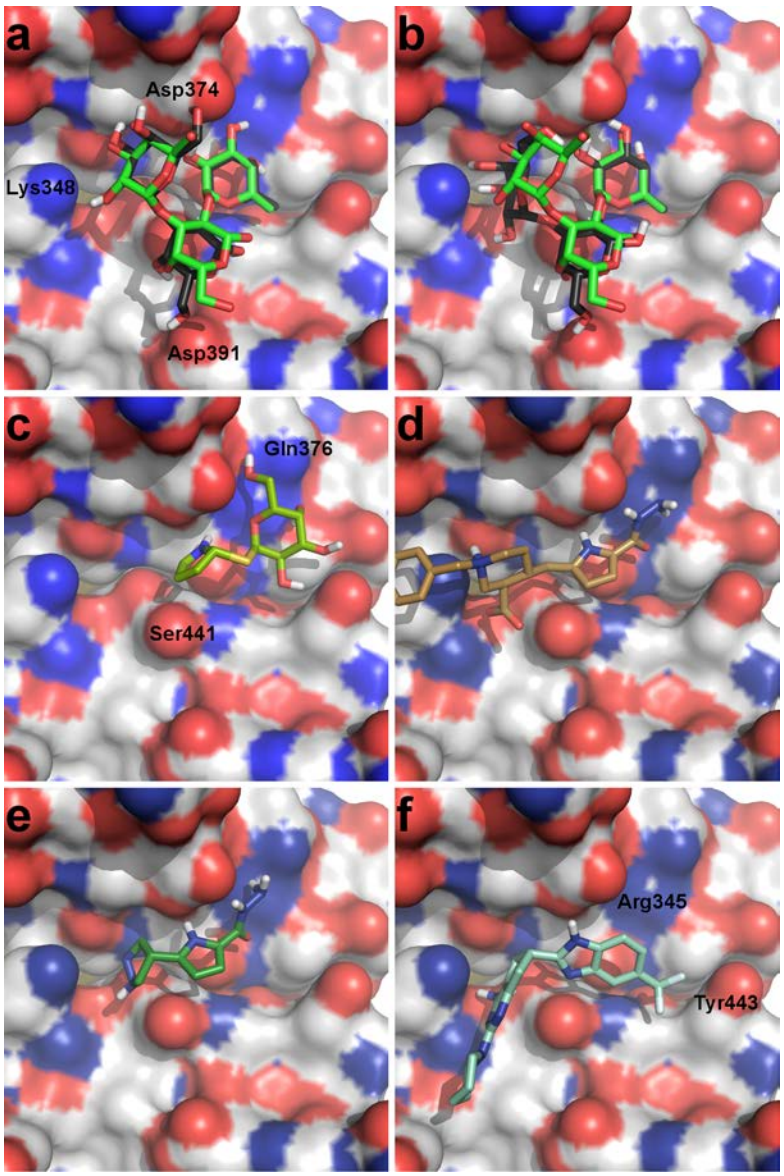


Figure 2. Examples of docked poses of high-affinity structures. Panel (a) shows HBGA B from the crystal structure (green) and docked (black) with water molecules retained; heavy atom RMSD was 0.9 Å between the two poses. Panel (b) shows HBGA B from the crystal structure (green) and docked (black) without any water molecules retained; heavy atom RMSD was 2.2 Å between the poses. Panels (c) – (f) show the docked poses, which subsequently were the starting points for the MD simulations, with highest affinity (according to LIE) of compounds **23**, **24**, **25** and **26**, respectively.