Lewis histo-blood group \(\alpha1,3/\alpha1,4\) fucose residues may both mediate binding to GII.4 noroviruses

Waqas Nasir2, Martin Frank1,3, Chaitanya A K Koppisetty3,4, Göran Larson2, and Per-Georg Nyholm1,3,5

1Department of Clinical Chemistry and Transfusion Medicine, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; 2Department of Computer Science and Engineering, Chalmers University of Technology, 41296 Gothenburg, Sweden; 3Biognos AB. Generatorsgatan 1, P.O. Box 8963, 40274 Gothenburg, Sweden; 4Department of Medical Biochemistry and Cell Biology, University of Gothenburg, Gothenburg, Sweden

Received on October 21, 2011; revised on May 6, 2012; accepted on May 7, 2012

Human noroviruses cause recurrent epidemics of gastroenteritis known to be dominated by the clinically important GII.4 genotype which recognizes human Secretor gene-dependent ABH histo-blood group antigens (HBGAs) as attachment factors. There is increasing evidence that GII.4 noroviruses have undergone evolutionary changes to recognize Lewis antigens and non-Secretor saliva. In this study, we have investigated the possibilities of the Lewis \(\alpha1,3/\alpha1,4\) fucoses as mediators of binding of GII.4 noroviruses to Lewis antigens. The study was carried out using molecular dynamics simulations of Lewis type-1 and type-2 chain HBGAs in complex with VA387 P domain dimers in explicit water. Based on the computer simulations, we suggest the possibility of two receptor binding modes for Lewis HBGAs: the “Secretor pose” with the Secretor Fuc\(\alpha1,2\) in the binding site and the “Lewis pose” with the Lewis Fuc\(\alpha1,3/\alpha1,4\) residues in the binding site. This was further supported by an extensive GlyVicinity analysis of the Protein Data Bank with respect to the occurrence of the Lewis and Secretor poses in complexes of Lewis antigens with lectins and antibodies as well as GII norovirus strains. The Lewis pose can also explain the interactions of GII.4 norovirus strains with \(Le^a\) and \(SLe^a\) structures. Moreover, the present model suggests binding of complex branched polysaccharides, with the Lewis antigens at the nonreducing end, to P domain dimers of GII.4 strains. Our results are relevant for understanding the evolution of norovirus binding specificities and for in silico design of future antiviral therapeutics.

Keywords: ABO antigens / histo-blood group antigens / Lewis antigens / molecular dynamics / norovirus

Introduction

Noroviruses are known to cause major outbreaks of acute nonbacterial gastroenteritis worldwide (Fankhauser et al. 2002; Lopman et al. 2003; Hutson et al. 2004; Nguyen et al. 2007). Among human noroviruses, the GII.4 (genogroup II, genotype 4) noroviruses are responsible for 60–80% of the outbreaks (Lindesmith et al. 2008; Donaldson et al. 2010). Noroviruses are positive-sense, single-strand RNA viruses which belong to the Caliciviridae family. The virus envelope is made from the major capsid protein VP1, encoded by ORF2 of the viral genome, and consists of 180 monomers or 90 dimers arranged in \(T=3\) icosahedral symmetry (Prasad et al. 1999). Each monomer is comprised of a shell (S) domain, forming the core of the capsid protein, and the protruding (P) domain which extends from the shell domain and constitutes the outer surface of the virus capsid (Prasad et al. 1999). The virus–host interaction sites are located on the surface of the hyper-variable P domain of VP1 (Donaldson et al. 2010).

Noroviruses recognize the Secretor, Lewis and ABO families of human histo-blood group antigens (HBGAs) in a strain-specific manner (Marionneau et al. 2002; Huang et al. 2003, 2005; Harrington et al. 2004; Shirato et al. 2008). Outbreak studies and volunteer challenge studies have linked HBG A recognition of noroviruses to clinical infection and illness (Hutson et al. 2002, 2005; Lindesmith et al. 2003; Tan, Jin et al. 2008), typically in the case of Secretor (FUT2) gene-dependent HBGAs. Furthermore, binding studies using saliva and synthetic oligosaccharides have shown that most of the GII.4 norovirus strains recognize HBGAs (Huang et al. 2003, 2005) containing the Fuc\(\alpha1,2\)Gal\(\beta\) carbohydrate moiety. The structural details of the binding of GII.4 norovirus to HBGAs were shown by the crystal structure of VA387/B-trisaccharide complex (Cao et al. 2007). The crystal structure demonstrated that the \(\alpha1,2\)-linked fucose interacts strongly with the protein residues in the binding site through a number of hydrogen bonds as well as with hydrophobic interactions. These data were further confirmed by a mutagenesis study (Tan, Xia et al. 2008). We have previously explored the computational fit of a series of Secretor gene-dependent HBGAs with \(\alpha1,2\)-linked fucose in the binding site of VA387 (Koppisetty et al. 2010). The Fuc\(\alpha1,2\) residue was estimated to contribute...
at least half of the binding energy in HBGA binding to GII.4 noroviruses (Koppisetty et al. 2010). The role of the \(\alpha1,2\)-linked fucose residue in recognition of HBGAs by GII.4 noroviruses is thus well understood.

A few studies have reported the binding of GII.4 noroviruses to HBGAs not containing a Fuc\(\alpha1,2\) residue. Rydell et al. (2009) reported binding of the Dijon strain to SLe\(^x\), whereas de Rougemont et al. (2011) reported binding of the Dijon, Osaka and DenHaag GII.4 strains to SLe\(^x\) as well as to Le\(^x\). Similarly, a recent report on the interactions of Ast0139 GII.4 strain with HBGAs by STD NMR demonstrated the binding to SLe\(^x\), Le\(^x\) and Le\(^x\) (Fiege et al. 2012). The Lewis a and b (type-1 chain structures) contain \(\alpha1,4\)-linked fucoses, whereas the Lewis x and y (type-2 chain) epitopes contain \(\alpha1,3\)-linked fucose residues. Although the binding of GII.4 noroviruses to these Lewis-positive HBGAs has been reported (Rydell et al. 2009; de Rougemont et al. 2011; Fiege et al. 2012), the exact binding mode in these cases is poorly understood due to lack of crystallographic data. During our previous work on the interaction of Secretor gene-dependent Fuc\(\alpha1,2\)Gal\(\beta\) containing ligands with the documented VA387 binding site (Koppisetty et al. 2010), we realized that, e.g. the \(\alpha1,3\)-linked fucose residue of the Lewis x epitope could interact with the fucose-binding site similarly as \(\alpha1,2\)-linked fucose. Such a binding mode was recently observed in the crystal structure of VA207 (GII.9 strain) P domain dimer in complex with SLe\(^x\) (Chen et al. 2011). In this study, we present results on the receptor-binding mode for Lewis-positive HBGAs in complex with the GII.4 norovirus VA387. The pose in which \(\alpha1,3/\alpha1,4\)-linked fucose fits in the binding site of VA387 in the same orientation as \(\alpha1,2\)-linked fucose is here termed the “Lewis pose”. Comparisons are made with the “Secretor pose”, which is observed in the VA387/B-trisaccharide crystal structure (Cao et al. 2007). The work is focused on VA387 due to the clinical importance of GII.4 noroviruses and the fact that the VA387 strain interactions are known from the crystal structure. The results are obtained using molecular dynamics (MD), molecular mechanics-generalized Born surface area (MM-GBSA) and Glide scoring techniques. The work reported herein is of clinical importance of GII.4 noroviruses and the fact that the Lewis-positive HBGAs containing a Fuc\(\alpha1,2\)Gal\(\beta\) portion of SLex. However, these interactions were rather dynamic during the production runs due to the higher flexibility of the terminal sugar residues (Supplementary data, Figures S2–S4). Interestingly, for SLe\(^x\), we observed conformational transitions of the \(\alpha2,3\)-linkage (Supplementary data, Figure S4).

The protein–carbohydrate interactions for the GlcNAc\(\beta\) residue were different for type-1 and type-2 chain HBGAs (Supplementary data) due to the \(\sim 180^\circ\) flipped orientation of the GlcNAc\(\beta\) residue. Hydrogen bonds and water bridges were observed between protein residues S441, Q390, D391 and G392 of the binding pocket and the GlcNAc\(\beta\) residue in type-1 chain HBGAs. For type-2 chain ligands, residues A346 and H347 of the binding pocket were instead observed to have water bridges with the GlcNAc\(\beta\) residue. Additionally, water bridges between Q390, D391, G392 and S441 amino acids and the GlcNAc\(\beta\) residue were observed in a small number of snapshots from the MD simulations. The Fuc\(\alpha1,2\) residue of the Secretor gene-dependent HBGAs in the Lewis pose was observed to have hydrogen bonds and solvent-mediated contacts with the residues Q390, D391 and

**Results**

During the building process of initial protein–ligand complexes, we found that all the ligands, except Le\(^x\), were easily accommodated in their low-energy conformation in both the Lewis and Secretor poses. The acetamido group of the GlcNAc\(\beta\) residues in type-2 chain ligands differs by \(\sim 180^\circ\) from its orientation in type-1 chain HBGAs. In the case of Le\(^x\), this causes minor clashing of the acetamido group of the GlcNAc\(\beta\) residue with the side chains of S441 in the Lewis pose and with G392 in the Secretor pose. These clashes were relieved on subsequent energy minimization.

**MD simulations**

The snapshots of the protein–ligand complexes after 5 ns of MD production run are shown in Figures 1 and 2. For the ligands in the Secretor pose, protein–ligand interactions and carbohydrate torsion angles were similar to those reported earlier (Koppisetty et al. 2010). The results presented here, therefore, will focus on the Lewis pose unless otherwise specified.

The explicit solvent MD simulations of ligands in complex with VA387 P domain resulted in stable trajectories for each ligand in both the Lewis and Secretor poses. The root mean square deviation (RMSD) analyses of each residue in the sugar and the protein-binding site demonstrated that the Fuc\(\alpha\) residue in the fucose-binding site was almost immobile with an average RMSD of \(\sim 1\) Å from the starting structure for all ligands in both poses. Similarly, the amino acids T344, R345, A346, D374 and Y443 interacting with the Fuc\(\alpha\) residue in the site had minor deviations from the starting position (RMSD\(_{\text{max}}\) \(\sim 2\) Å). In contrast, the terminal sugar residues were more flexible in the site. All the ligands, particularly the type-2 chain HBGAs, were more or equally stable in terms of RMSD in the Lewis pose than in the Secretor pose (Supplementary data, Figure S1). The torsion angles of glycosidic linkages of bound ligands were in energetically favorable regions of conformational energy maps (Frank et al. 2007) for free oligosaccharides. Also, the carbohydrate conformations observed in MD simulations were in agreement with available crystallographic data for bound Lewis antigens (Supplementary data, Table S1).

The direct contacts of the fucose in the Secretor pose with residues T344, R345, D374 and Y443 of the P domain were also observed for the binding fucose in the Lewis pose despite no positional restraints were applied (Figure 1). The analyses of the MD trajectories (Supplementary data) indicate that the Lewis-positive HBGAs fit at least equally well in the Lewis pose as in the Secretor pose. Also, according to the modeling data reported here, the lack of a Fuc\(\alpha1,2\) residue in Le\(^x\) and SLe\(^x\) HBGAs does not influence the binding of the Fuc\(\alpha1,3\) residue.

In the Lewis pose, the protein residues N393, H395 and Y443 were involved in hydrogen bonds and water-bridging interactions with the terminal residue GalNAC\(\alpha\) of ALe\(^x\) and ALe\(^x\) and with the Neu5Ac\(\alpha\) residue of SLe\(^x\). However, these interactions were rather dynamic during the production runs due to the higher flexibility of the terminal sugar residues (Supplementary data, Figures S2–S4). Interestingly, for SLe\(^x\), we observed conformational transitions of the \(\alpha2,3\)-linkage (Supplementary data, Figure S4).

The protein–carbohydrate interactions for the GlcNAc\(\beta\) residue were different for type-1 and type-2 chain HBGAs (Supplementary data) due to the \(\sim 180^\circ\) flipped orientation of the GlcNAc\(\beta\) residue. Hydrogen bonds and water bridges were observed between protein residues S441, Q390, D391 and G392 of the binding pocket and the GlcNAc\(\beta\) residue in type-1 chain HBGAs. For type-2 chain ligands, residues A346 and H347 of the binding pocket were instead observed to have water bridges with the GlcNAc\(\beta\) residue. Additionally, water bridges between Q390, D391, G392 and S441 amino acids and the GlcNAc\(\beta\) residue were observed in a small number of snapshots from the MD simulations. The Fuc\(\alpha1,2\) residue of the Secretor gene-dependent HBGAs in the Lewis pose was observed to have hydrogen bonds and solvent-mediated contacts with the residues Q390, D391 and
It is also interesting to note that in the case of the Secretor pose of Le^b, Le^b and ALe^b, these interactions are instead maintained by the Lewis fucose (Supplementary data, Figures S2 and S3).

An important observation for ligands in the Lewis pose is the orientation of the O1 atom of the GlcNAcβ residue. According to our modeling data, the chain continuation points for the two ligands, accommodated in the binding sites of the VA387 binding site with the Lewis and Secretor poses of difucosylated Lewis antigens. The snapshots after 5 ns of MD simulations in explicit waters are shown. The figure was made using Open Source Pymol (Schrodinger 2010). Ligands are shown as stick models and the surface of the protein is color-coded (Chain A: light red; Chain B: light green) according to the distance from the ligand atoms with the threshold of 4 Å. The HBGAs are color-coded (fucose: red; galactose: cyan; N-acetyl-glucosamine: yellow; N-acetyl-galactosamine: blue; downstream methoxy group: white, hydrogen bonds are depicted by stroked yellow lines and interacting amino acids are annotated. Water molecules were removed for clarity. The Lewis fucose in the Lewis pose retains the same interactions as reported in the crystal structure for α1,2-linked fucose in the VA387/B-trisaccharide complex (Cao et al. 2007).
VA387 P dimer with Secretor pose, are \( \sim 40 \text{ Å} \) apart in contrast to \( \sim 26 \text{ Å} \) for the Lewis pose. Furthermore, the O1 atoms of the GlcNAc\( \beta \) residues point away from each other in the Secretor pose but not in the Lewis pose (Figure 3).

Glide scoring and MM-GBSA calculations
Glide scoring and the MM-GBSA approach were used to estimate interaction energies for the protein–ligand complexes in the Lewis and Secretor poses (Figure 4). The predicted scores for all the ligands were in the same range as those for the VA387/B-trisaccharide MD-simulated complex.

For Le\( ^b \) and Le\( ^s \), the scores of the Lewis pose were similar to those of the Secretor pose which can be explained by the protein–ligand interactions observed in each pose (see above). Glide scores for ALe\( ^b \) and ALe\( ^s \) indicate significant differences between the Lewis pose and the corresponding Secretor pose. This difference in predicted binding energy is due to differences in the interaction profile of the terminal GalNAc\( \alpha \) residue in the two poses. The binding pocket defined by amino acid residues Q331, A346, H347, K348 and I389, adjacent to the fucose-binding site, has multiple hydrogen bonds with the GalNAc\( \alpha \) residue in the Secretor pose for these ligands. These interactions can explain the better score of the Secretor pose in the case of ligands with a terminal GalNAc\( \alpha \) residue. In the Lewis pose, the terminal GalNAc\( \alpha \) residue interacts with amino acid residues N393, H395 and Y443 of the protein dimer. These interactions are more superficial which can explain a lower score. Since ALe\( ^b \) and ALe\( ^s \) have higher scores than the known binders Le\( ^b \) and Le\( ^s \), the former are predicted to be good binders in both poses. Furthermore, since the chain continuation point for HBGAs in the Lewis pose is in the binding pocket defined by residues Q331, A346, H347, K348 and I389 (Tan, Xia et al. 2008), one can expect more interactions to arise in this region due to elongation of the chain.

SLex in “gauche” conformation is scored two standard deviations (SDs) higher than the “trans” conformation (Figure 4) according to Glide. This difference in the scores can be explained by different sets of interactions of the Neu5Ac\( \alpha \) residue with protein in the two conformations. In the gauche conformation, Neu5Ac\( \alpha \) residue has been observed to have multiple hydrogen bonds with G392, N393 and H395 of the protein dimer, some of which are lost in the trans conformation because the sialic acid residue is pointing more towards the bulk solvent. Despite this difference in the interaction profile for the two conformations, SLex\( ^s \) is scored slightly higher than Le\( ^b \), Le\( ^s \) and Le\( ^s \) according to Glide. Finally, Le\( ^s \), despite lacking the Fuc\( 1,2 \) residue, had similar Glide score to those of the known binders Le\( ^b \) and Le\( ^s \) in the Lewis pose. For MM-GBSA calculations, similar trends in relative binding energies were observed (Figure 4).

Analysis of fucose binding sites in the Protein Data Bank
An analysis of the Protein Data Bank with the GlyVicinity software (Lutteke et al. 2005) revealed that difucosylated Lewis structures make primary contacts with the protein through either the Secretor gene-dependent fucose or the Lewis fucose (Table I). Based on the major contacts, the studied complexes have been classified as having a Secretor pose or a Lewis pose. However, the low number of studied cases does not allow us to make any statements on the preference of either of these poses. Furthermore, the analysis shows that the methyl group of the fucose residues in all the studied complexes makes hydrophobic contacts with an aromatic residue of the protein and that the OH groups of the fucose are involved in hydrogen bonds with 2–3 polar amino acids (frequently arginine and serine).

Discussion
The difucosylated ligands considered in the present study are the products of fucosyl transferases encoded by \( FUT2 \).
(Secretor) and FUT3 (Lewis) genes for type-1 chain structures, and FUT1 and FUT3, FUT4, FUT5, FUT6 or FUT9 genes for the type-2 chain structures. SLe\(^\alpha\) (dependent on FUT7 and FUT4) and Le\(^\alpha\) are the two exceptions which depend on neither FUT1 nor FUT2 since they lack a Fuc\(_{\alpha}1,2\) residue in the sugar chain. The binding of SLe\(^\alpha\) neoglycoprotein to GII.4 and GII.3 norovirus strains was shown in 2009 (Ryddell et al. 2009), whereas the first report on the recognition of Le\(^\alpha\) by GII.4 noroviruses came in a recent study, which reported binding of GII.4 noroviruses to a Le\(^\alpha\) neoglycoconjugate (de Rougemont et al. 2011). Furthermore, recent STD NMR data on the interactions of HBGA with a GII.4 norovirus show that the Lewis fucose of Le\(^\alpha\) and SLe\(^\alpha\) saccharides interact with the protein (Fiege et al. 2012). Moreover, recent reports on the binding pattern of the norovirus strains Den Haag and Osaka show that non-Secretor gene-dependent antigens may also bind to GII.4 noroviruses (de Rougemont et al. 2011). These findings raised the intriguing question whether the Lewis fucose could bind in the fucose-binding site similarly to the \(\alpha 1,2\)-linked fucose.

In this study, we have explored the possibility of the Lewis fucose mediating the binding of the Lewis antigens to GII.4 noroviruses using molecular modeling techniques. We used the same MD methodology as was used in our previous study on the binding of the Fuc\(_{\alpha}1,2\)Gal\(\beta\) in the fucose binding site (Koppisetty et al. 2010), where the contacts of the Secretor fucose were well retained when compared with the crystal structure of the VA387/B-trisaccharide complex with an RMSD of <0.8 Å. Similar results were obtained for the Lewis antigens in the present study with very little deviation from the starting pose of the fucose residue bound in the site.

The present study shows the detailed interactions of the Lewis antigens in the known fucose-binding site of VA387 (Figures 1 and 2). The results obtained for SLe\(^\alpha\) and Le\(^\alpha\) are particularly interesting due to the lack of a Fuc\(_{\alpha}1,2\) residue in these antigens. The Glide score and MM-GBSA energy for Le\(^\alpha\) are similar to those of Le\(^\alpha\) and Le\(^b\), which have previously been shown to bind GII.4 noroviruses (Huang et al. 2003, 2005). This is in agreement with the experimental results.
reportedly recently on the recognition of \( \text{Le}^b \) by GII.4 strains (de Rougemont et al. 2011). The authors used repetitive \( \text{Le}^b \) with three Fuc1,3 residues (JL Pendu, personal communication) for the study. The fact that several assay systems have failed to detect \( \text{Le}^b \) binding to GII.4 noroviruses (Huang et al. 2005; Rydell et al. 2009) might be explained by presentation effects. Therefore, the use of trimeric \( \text{Le}^b \) might expose the Lewis fucose in a better way for the antigen to interact with the known fucose-binding site. For monomeric \( \text{Le}^b \), there is a need for more sensitive assays such as STD NMR for the detection of binding (Fiege et al. 2012).

For \( \text{SL}^b \), the present results show slightly better Glide and MM-GBSA scores than for \( \text{Le}^b \), suggesting that the sialic acid residue may play a role in increasing the binding affinity of the complex. In fact, the present model suggests that \( \text{SL}^b \) is recognized by GII.4 noroviruses at least as strongly as the \( \text{A} \text{Le}^b \) and \( \text{AL} \text{Le}^b \) antigens. Furthermore, the analyses of the trajectories from the explicit solvent MD for the VA387/SLex complex show that sialic acid residue makes several water-mediated contacts with the protein (Supplementary data, Figure S2 and Table S1). The sialic acid residue in the crystal structure which agrees with our finding of the gauche conformation in parts of our MD simulations (Figure 5). The crystal structure shows the Lewis fucose in a pose which is in very good agreement with the MD simulations (Chen et al. 2011; Hansman et al. 2011; Shanker et al. 2011) and therefore results in structurally distinct conformations in contrast to the conserved fucose binding pocket (Figure 5). Also, crystallographic data (Cao et al. 2007; Shanker et al. 2011) and modeling studies (Supplementary data, Figure S1) show that this loop region is highly flexible. Due to this flexibility in the loop region and the possible conformations of the \( \alpha 2,3 \text{-linkage} \) in \( \text{SL}^b \), one can expect a variation in the interactions of \( \text{SL}^b \) with the binding site residues in the loop.

For difucosylated Lewis antigens without the terminal \( \alpha 1,3 \text{-linkage} \) in SLex, one can expect a variation in the interactions between the Neu5Ac residue with the protein. Extended MD simulations are underway in order to investigate the flexibility of the \( \alpha 2,3 \text{-linkage} \) and the interactions of the Neu5Ac residue in the bound state in more detail. The other ligands do not show major conformational transitions in any of the glycosidic linkages and therefore we assumed that 5 ns should be long enough to sufficiently sample their binding properties.

This report describes the detailed interactions of \( \text{SL}^b \) with a norovirus GII.4 strain. It is of great interest to compare the present model with a crystal structure that was recently reported for a GII.9 strain in complex with \( \text{SL}^b \) (Chen et al. 2011). The crystal structure shows the Lewis fucose in a pose which is in very good agreement with the MD simulations (Figure 2). The \( \alpha 2,3 \text{-linkage} \) is in the gauche conformation in the crystal structure which agrees with our finding of the gauche conformation in parts of our MD simulations (Supplementary data, Figure S2 and Table S1). The sialic acid residue in the crystal structure however, does not show contacts with the binding site which can be explained by a different loop conformation in the site. This loop region is highly flexible in sequence among different GII norovirus strains (Chen et al. 2011; Hansman et al. 2011; Shanker et al. 2011) and therefore results in structurally distinct conformations in contrast to the conserved fucose binding pocket (Figure 5).
Since the fucose binding sites are essentially conserved between GII.4 and GII.9 strains, these data support our results that, even in the presence of the $\alpha_{1,2}$-linked fucose, binding of the difucosylated Lewis antigens to GII.4 noroviruses can occur in the Lewis pose. It is noteworthy that the Lewis and Secretor poses of difucosylated Lewis HBGAs have been observed also for the binding of these ligands to lectins and antibodies (Table I). For instance, the Lewis pose was observed for the binding of Lectin IV of *Griffonia simplicifolia* with Leb and for antibody MBR96 binding with Ley. On the other hand, Lectinolysin in crystalline complex with Le b and Ley is an example of binding in Secretor pose (Table I). Also, it is well known that a range of anti-Le$^b$Y antibodies consistently cross-react with Le$^b$ and/or H type-1/2 antigens (e.g. 2-25LE, BR55, F3) (Manimala et al. 2007). This cross-reactivity suggests that H and Lewis antigens can bind in different poses (corresponding to Secretor and Lewis poses of noroviruses) to the same site. Since the fucose binding pockets in GII noroviruses share similar features to those present in lectins and antibodies (Table I), the observation of the Lewis and Secretor poses in lectins and antibodies strengthens our conclusion that noroviruses exhibit the same switch in binding specificity for the Lewis antigens.

Another interesting finding of this study is the orientation of the reducing ends of the HBGAs in the Secretor and Lewis poses for the Lewis antigens in complexes with GII.4 noroviruses (Figure 3). Appropriate spacing of epitopes could favor the binding of branched sugar structures to the same P dimer. The binding of branched Lewis-positive ligands in the Lewis pose might enhance the binding affinity to GII.4 noroviruses. The question of epitope orientation may become crucial for a computational design of dendrimeric multivalent carbohydrate mimetics for inhibition of binding to GII.4 noroviruses.

The results reported here are in full agreement with the binding data and crystallographic data for GII noroviruses (Chen et al. 2011; de Rougemont et al. 2011; Hansman et al. 2011; Shanker et al. 2011; Fiege et al. 2012). The observation that two of the more recent GII.4 strains, Den Haag and Osaka, recognize non-Secretor saliva is of special interest (de Rougemont et al. 2011). These strains, like most of the post-2002 GII.4 strains, have an insertion in the flexible and highly variable loop region in the binding site. Preliminary modeling data suggest that this insertion does not influence the HBG A binding to GII.4 noroviruses in the Lewis pose. Thus, the Lewis pose should be considered for an understanding of the GII.4 norovirus pathogenesis and the evolutionary steps that the virus takes to infect phenotypically broader populations.

**Materials and methods**

**Nomenclature**

The gauche conformation of the $\alpha_{2,3}$-linkage in SLex is defined as $\Phi(C1-C2-O3-C3) \approx -70^\circ$ and the trans conformation as $\Phi(C1-C2-O3-C3) \approx 190^\circ$.

**Preparation of the initial protein–ligand complexes**

The Lewis-positive HBGAs, A Lewis b (ALeb), A Lewis y (ALEy), Lewis b (Leb), Lewis y (Ley), Lewis x (Lex) and sialyl Lewis x (SLex), are all characterized by the presence of either a Lewis Fuc$z$1,3 or a Lewis Fuc$z$1,4 residue in the sugar chain (Table II). The structures of the ligands considered were prepared using the LEAP utility of AMBER tools...
Table II. Carbohydrate ligands considered in the present study

<table>
<thead>
<tr>
<th>Ligand abbreviation (name)</th>
<th>Ligand structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Le^t^ (Lewis b)</td>
<td>Fucα2Galβ3(Fucα2)GlcNAcβ-O-CH₃</td>
</tr>
<tr>
<td>ALe^t^ (A Lewis b)</td>
<td>GalNAcβ3(Fucα2)Galβ3(Fucα4) GlcNAcβ-O-CH₃</td>
</tr>
<tr>
<td>Le^t^ (Lewis y)</td>
<td>Fucα2Galβ4(Fucα3)GlcNAcβ-O-CH₃</td>
</tr>
<tr>
<td>ALe^t^ (A Lewis y)</td>
<td>GalNAcα3(Fucα2)Galβ4(Fucα3) GlcNAcβ-O-CH₃</td>
</tr>
<tr>
<td>Le^t^ (Lewis x)</td>
<td>Galβ4(Fucα3)GlcNAcβ-O-CH₃</td>
</tr>
<tr>
<td>SLe^t^ (Sialyl Lewis x)</td>
<td>Neu5Acα2,3Galβ4(Fucα3)GlcNAcβ-O-CH₃</td>
</tr>
<tr>
<td>B-Tri (B-Trisaccharide)</td>
<td>Galβ3(Fucα2)Galβ-O-CH₃</td>
</tr>
</tbody>
</table>

1.4 or the SWEET2 server (Bohne et al. 1999) based on information on preferred torsion values for the linkages derived from the Protein Data Bank using GlyTorsion (Lutke et al. 2005). Structural optimization was performed with the molecular mechanics force field MM3 using the TINKER program (Ponder 2010).

Two binding poses for the ligands Le^t^, ALe^t^, Le^t^ and ALe^t^ in the VA387 site were produced using superimposition of the Secretor and the Lewis fucose residues on the α1,2-linked fucose of the B-trisaccharide/VA387 crystal structure complex: the Secretor pose with α1,2-linked fucose in the site and the Lewis pose with α3,1/α4,1-linked fucose in the site. Using the same procedure, the Lewis pose was produced for Le^t^ and SLe^t^.

The initial protein–ligand complexes were processed for MD simulation using LEAP (AMBER software package, version 8.0; University of California, San Francisco, CA). All histidine residues were assumed to be neutral and were protonated at the epsilon-position. Hydrogen atoms were added to the protein, and 14 sodium ions were added to counterbalance the overall negative charge of the system. For periodic boundary condition simulations, the system was solvated in a box of TIP3P water with approximate dimensions of 80 Å × 90 Å × 1170 Å (about 14,000 water molecules). GLYCAM06 force field with version 8.0; University of California, San Francisco, CA) All parameters and charges (Kirschner et al. 2008) were assigned for the carbohydrates and the ff99 force field parameters (Wang et al. 2000) for the protein.

MD simulations
We performed explicit solvent MD simulations for at least 5 ns with TIP3P explicit water molecules and periodic boundaries to explore the computational fit of the Lewis ligands reported herein and to study the protein carbohydrate interactions of these ligands. For the VA387/SLe^t^ complex, two MD simulations were performed for ~5 ns each. A two-step energy minimization was performed before running the production dynamics. In the first minimization step, comprising 1000 minimization cycles, mild position restraints (5 kcal mol⁻¹ Å⁻²) were applied on protein backbone atoms, whereas the atoms in protein side chains, sugars and explicit water molecules were kept unrestrained. Steepest descent algorithm and conjugate gradient algorithm were used to calculate 500 cycles each. The second minimization stage was comprised of complete unrestrained minimization of the whole system for 2500 minimization cycles. The steepest descent algorithm was used for the first 1000 cycles followed by conjugate gradient minimization for 1500 cycles. A 100 ps equilibration was then performed for the minimized protein–ligand complexes. For the first 20 ps, constant volume (NVT) equilibration was carried out using position restraints (5 kcal mol⁻¹ Å⁻²) applied on the protein backbone atoms, whereas the remaining atoms in the system were free to move. The temperature was gradually adjusted from 0 to 300K during this time. This was followed by 80 ps of constant pressure (NPT) equilibration in which all the atoms of the system were kept unrestrained. The equilibration process was then followed by production dynamics of at least 5 ns performed at constant temperature and pressure of 310 K and 1 atm, respectively. All atoms in the system were kept unrestrained in the MD production runs. During both the MD equilibration and production runs, the temperature was regulated using Berendsen temperature bath with a temperature coupling parameter of 1.0 ps. Pressure was kept constant with isotropic position scaling and 2.0 ps of pressure relaxation time. A cutoff of 10 Å was applied for calculating nonbonded interactions: 1–4 electrostatic and nonbonded interactions were scaled by the default values of 1/1.2 and 1/ 2.0, respectively. The SHAKE algorithm was employed to constrain solute bonds involving a hydrogen atom allowing a time step of 2 fs for the simulations. The trajectory files were written every 1 ps.

Glide scoring
For each protein–ligand complex, Glide scoring was performed for the last 10 snapshots with a difference of 100 ps from the MD trajectory. For VA387/SLe^t^ complex, an additional set of Glide scores, from the second MD simulation, was calculated for selected frames showing the trans conformation of α2,3-linkage of sialic acid residue in the bound state. After removing the water molecules and countions from the snapshots, a restrained minimization of the protein side chains in the presence of the sugar was performed in OPLS-2005 force field. The resulting protein–carbohydrate complexes were scored using Glide XP scoring function with the “Refine” option of Glide docking. The GlideScore XP 5.0 parameters were used during the refinement.

MM-GBSA
MM-GBSA calculations have been shown to perform well when it comes to estimating the relative free binding energies of protein–ligand complexes (Hou et al. 2011). In the present study, the MM-GBSA approach was used as a complement to the Glide scoring technique. The calculations were performed on the same MD trajectory snapshots as used for Glide. The calculations were preceded by implicit solvent GB (Onufriev et al. 2004) minimization with position restraints of 5 kcal mol⁻¹ on protein main chainatoms. GLYCAM06 (Kirschner et al. 2008) and ff99 (Wang et al. 2000) force fields were used for carbohydrates and protein, respectively. A total of 1000 minimization cycles were performed using steepest descent method followed by conjugate gradient algorithms with 500 steps each. Separate receptor, ligand and complex pdb files were prepared for processing. MM-GBSA calculations were then performed on minimized trajectories using python script mmpbsa.py in AMBER 11. The GB model by

1170
Onufriev et al. (2004) was used in calculations. The internal dielectric constant of 1.0 was used for solutes, whereas an external dielectric constant of 78.5 was used to produce solvent effects with a salt concentration of 100 mM.

Trajectory analysis
The analysis of MD simulations was performed using the Conformational Analysis Tools (CAT) software (www.md-simulations.de/CAT) together with the ptraj module of AMBER tools 1.4, which was used for the superimposition of the trajectory frames. The analyses included RMSD, hydrogen bond analysis and bridging water analysis. These analyses were performed on each frame of the MD production runs. For the VA387/SLe5 complex, the simulation showing only the gauche conformation was analyzed for protein–carbohydrate interactions. The second MD simulation of the VA387/SLe5 complex was used to study the transitions of sialic acid residue between gauche and trans conformations. The resulting information from all the analyses is provided in the form of figures in the Supplementary data.

Analysis of lectins and antibodies containing difucosylated ligands
GlyVicinity analysis tool (Lutteke et al. 2005) was used to obtain information about the most abundant protein residues in fucose-binding sites from the crystal structure data. The distance threshold was set to 4.0 Å. In a total number of 98 pdb files, 379 aromatic protein residues (W, Y, H, F) out of the total 1645 were identified to be within the specified distance threshold which suggests hydrophobic contacts with the methyl group of fucose. The classification criterion was then set based on the total number of observed hydrogen bonds and hydrophobic contacts between fucose and protein residues in difucosylated protein–ligand complexes. The crystal structures of antibodies and lectins containing Le$^b$ were downloaded from Protein Data Bank and analyzed manually for contacts. Hydrogen bonding cutoff was set to 3.2 Å.

Note
During the revision of this manuscript, Fiege et al. published an STD NMR study, which supports our original conclusions and is now included in the list of references.

Supplementary data
Supplementary data for this article is available online at http://glycob.oxfordjournals.org/.

Acknowledgements
Computational resources at Biognos AB and the Department of Medical Biochemistry at the University of Gothenburg are gratefully acknowledged. The authors thank Prof. Fredrik Höök, Division of Biological Physics, Chalmers University of Technology, Sweden, for fruitful discussions regarding the kinetics and relative binding affinities of the protein–ligand complexes and for reading the manuscript.

Abbreviations
AMBER, assisted model building with energy refinement; CAT, conformational analysis tools; HBGAs, histo-blood group antigens; MD, molecular dynamics; MM-GBSA, molecular mechanics-generalized Born surface area; RMSD, root mean square deviation; SD, standard deviation; STD NMR, saturation transfer difference nuclear magnetic resonance; ORF, open reading frame.

References


