Mapping substrate interactions of the human membrane-associated neuraminidase, NEU3, using STD NMR

Amgad Albohy, Michele R Richards, and Christopher W Cairo

Alberta Glycomics Centre, Department of Chemistry, University of Alberta, Edmonton, AB, Canada T6G 2G2

1To whom correspondence should be addressed: Tel: +1-780-492-0377; Fax: +1-780-492-8231; e-mail: ccairo@ualberta.ca

Received 26 May 2014; Revised 30 September 2014; Accepted 1 October 2014

Abstract

Saturation transfer difference (STD) nuclear magnetic resonance (NMR) is a powerful technique which can be used to investigate interactions between proteins and their substrates. The method identifies specific sites of interaction found on a small molecule ligand when in complex with a protein. The ability of STD NMR to provide specific insight into binding interactions in the absence of other structural data is an attractive feature for its use with membrane proteins. We chose to employ STD NMR in our ongoing investigations of the human membrane-associated neuraminidase NEU3 and its interaction with glycolipid substrates (e.g., GM3). In order to identify critical substrate–enzyme interactions, we performed STD NMR with a catalytically inactive form of the enzyme, NEU3(Y370F), containing an N-terminal maltose-binding protein (MBP)-affinity tag. In the absence of crystallographic data on the enzyme, these data represent a critical experimental test of proposed homology models, as well as valuable new structural data. To aid interpretation of the STD NMR data, we compared the results with molecular dynamics (MD) simulations of the enzyme–substrate complexes. We find that the homology model is able to predict essential features of the experimental data, including close contact of the hydrophobic aglycone and the Neu5Ac residue with the enzyme. Additionally, the model and STD NMR data agree on the facial recognition of the galactose and glucose residues of the GM3-analog studied. We conclude that the homology model of NEU3 can be used to predict substrate recognition, but our data indicate that unstructured portions of the NEU3 model may require further refinement.

Key words: ganglioside, glycolipid, glycosyl hydrolase, neuraminidase, NMR

Introduction

The membrane-associated neuraminidase, NEU3, is one of four NEU isoenzymes identified in humans (Miyagi and Yamaguchi 2012). Previous work has provided evidence for NEU3 localization to the plasma membrane, endosome and the nuclear envelope (Zanchetti et al. 2007; Wang et al. 2009). As its membrane localization may suggest, NEU3 has been associated with roles including control of apoptosis, cell signaling and regulation of cell–cell interactions (Papini et al. 2004; Valaperta et al. 2006). Additionally, NEU3 is proposed to play important roles in the development of human cancer (Miyagi et al. 2003, 2004, 2007). NEU3 is a glycosyl hydrolase (GH; EC 3.2.1.18) enzyme and has an apparent preference for glycolipid substrates over glycoproteins (Kopitz et al. 1996; Monti et al. 2002; Ha et al. 2004). The enzyme is likely involved in the regulation of glycolipid composition of the membrane through consumption of specific ganglioside substrates (Li et al. 2001). Despite the importance of NEU3 in disease, there is limited information on the enzyme’s specificity and recognition of substrate.

NEU3 substrate specificity has primarily been tested against naturally occurring gangliosides. The enzyme has been reported to cleave GM3 (1), GD1a, GD1b and GT1b (Kopitz et al. 1996). Studies of membrane-bound NEU isolated from human brain tissue found
activity for an α(2→3) Neu able to cleave GM3, GM4 and GD1α; as well as an α(2→8) activity able to cleave GD3, GD2, GD1β, GT1β and GQ1b (Oehler et al. 2002). Relative to GM3, GD3 and GM4 were less active substrates (ca. 70%), whereas GD1α, GD1β and GD2 were significantly less active (ca. 5%) (Oehler et al. 2002). Notably, several studies have found that glycolipids with sialic acid (Sia) at an internal branch point are poor substrates. The most common examples of this group are GM1 and GM2, which both contain an α(2→3) Neu5Ac at an internal galactose (Gal) residue with a β(1→4) GlcNAc linkage (Kopitz et al. 1996; Oehler et al. 2002). Interestingly, lyso-gangliosides are known to be cleaved by membrane Neu, while O-acetylation at C9 of the Sia was found to inhibit activity (Oehler et al. 2002).

Recombinantly expressed Neu3 has been used to test enzyme specificity showing similar results with commonly occurring gangliosides (Ha et al. 2004). Investigations of Neu3 with synthetic GM3 analogs have found that the enzyme does not tolerate modifications to the N5 side chain, while the C9 site can be modified with preservation of some activity (Sandbhör et al. 2011). Importantly, this study found that the hydrophobic aglycone of the ganglioside substrate was critical for activity, yet could be substituted by a simple octyl chain. Additionally, a specific recognition site for the hydrophobic aglycone was proposed based on these findings. Although a range of substrate activity has been demonstrated for Neu3, there are still limited data on the structure of the active site or its recognition of specific substrates.

Although Neu2 has been studied by crystallography, there are currently no experimentally determined structures of the membrane-associated Neu3. Homology models of Neu3 have been developed (Magesh et al. 2006; Albohy et al. 2010). The model of Neu3 we previously reported has been tested extensively by site-directed mutagenesis and substrate studies (Albohy et al. 2010; Sandbhör et al. 2011). We have used the homology model of Neu3 in the rational design of inhibitors for human neuraminidases (hNeu) (Zou et al. 2010; Albohy et al. 2013). Although the homology model of Neu3 has been a useful tool, we recognized that the model is likely incomplete. For example, a significant portion of our homology model does not overlap with the template Neu2 structure used to develop it, suggesting some portions of the protein may adopt a different fold than seen in Neu2 (Albohy et al. 2010). As a result, we set out to obtain more detailed structural information regarding substrate interactions with the Neu3 active site and to identify portions of our model that require refinement. We turned to saturation transfer difference (STD) nuclear magnetic resonance (NMR) as a method that provides details of substrate interactions with the enzyme.

Mayer and Meyer introduced STD NMR as a method to map the interaction of substrate epitopes in close contact with a protein-binding site (Mayer and Meyer 1999, 2001). STD NMR has also been used to study the binding conformation of ligands in their active sites (Viegas et al. 2011; Mohabbi et al. 2012). Carbohydrate–lectin interactions are often studied by this method due to their moderate binding affinity and substrate size (Roldós et al. 2011). Von Itzstein and coworkers have used STD NMR to study the binding of free Sia and some of its derivatives to the Vibrio cholerae Neu, revealing that the enzyme has a preference for the α-anomer (Haselhorst et al. 2004). STD NMR has also been used to analyze carbohydrate–ligand interactions with membrane proteins (Claassen et al. 2005; Mari et al. 2005). Thus, we expected that STD NMR could provide crucial insight into the active site of Neu3 and its interaction with the enzyme’s most active ligand, GM3.

To investigate the molecular features of the Neu3–GM3 enzyme–substrate complex, we needed to tailor both the enzyme and its ligand. First, the native enzyme would rapidly hydrolyze the Neu5Ac glycosidic linkage of the GM3 trisaccharide, making it impossible to study the intact complex. To deal with this issue, we used a catalytically inactive form of the enzyme, Neu3(Y370F), which we have previously characterized (Albohy et al. 2010). The native glycosphin-golipid substrate, GM3, is also known to have a high propensity for micelle formation (Sonnino et al. 1990, 1994), which could interfere with the use of STD NMR. To avoid this problem, we chose to use a glycolipid analog of GM3 with an octyl chain as the aglycone, a substrate that has been previously shown to have comparable activity with GM3 (Sandbhör et al. 2011). Using STD NMR and molecular dynamics (MD), we report here the epitope mapping of the GM3-octyl derivative (Neu5Ac-α(2→3)-Gal-β(1→4)-Glc-β-octyl; 2) with the active site of Neu3 (Figure 1). We compared these data to the STD NMR of the octyl-lactoside (3) as a weak-binding control. In addition, the epitope mapping of the commonly employed fluorogenic substrate, 2‘-(4-methylumbelliferyl)-α-D-N-acetylneuraminic acid (4MU-NANA; 4), is reported. We compare the experimentally obtained epitope maps to the predictions of the homology model using MD simulations of the complexes.

Results

1D 1H NMR assignment

To conduct STD NMR experiments, we first needed to determine a complete assignment for the proton signals in the substrates to be used. Standard 1D and 2D experiments (1D 1H NMR, 1D 13C, COSY, HMQC, HMBC and CSSF-TOCSY) (Sandbhör et al. 2011) were used to make assignments for 1H signals of 2, 3 and 4. For example, CSSF-TOCSY (Robinson et al. 2004; Duncan et al. 2007) experiments were used to assign some overlapping signals (see Supplementary data).

Protein expression

Recombinant Neu3, or the catalytically inactive mutant Neu3 (Y370F) protein containing an N-terminal MBP-affinity tag, was expressed and purified as previously reported (Albohy et al. 2010). Aliquots of the protein were stored at −80°C and used immediately after thaw for experiments. To prepare the enzyme sample for STD NMR experiments, the buffer was rapidly exchanged using a spin column (MWCO 30 kDa). The Neu3(Y370F) mutant was found to have only 0.1% Neu activity when compared with the wild type using 4 as a substrate. In addition, MBP alone was expressed and purified using identical methods and used for control experiments.

STD NMR of Neu3(Y370F)–substrate complexes

To confirm that the substrate and enzyme could be used for STD studies, we conducted control experiments to exclude possible artifacts (Supplementary data, Figure S12): STD experiments with Neu3 alone, 2 alone, or the MBP-affinity tag alone showed only minor signals. STD experiments run with 1 showed a large signal in the aliphatic region, which we attribute to micelle formation (Sonnino et al. 1990, 1994). GM3 is known to have a very low critical micelle concentration, and our experiments required a high concentration of the substrate (2.5 mM) (Palestini et al. 1995). Adding DMSO (up to 8% final concentration) did not disrupt the aliphatic STD NMR signals. Attempts to eliminate the aliphatic signals of GM3 by modification of the on-frequency resonance were also unsuccessful (see Supplementary data). As a result, we concluded that the use of GM3 for STD NMR experiments would obscure protein–aglycone interactions due to micelle formation. We then tested 2, an analog of GM3 known to
be a good substrate for the enzyme (Sandbhor et al. 2011). When used in STD NMR experiments alone, spectra of 2 did not contain any signals in the aliphatic region (Supplementary data, Figure SI2b). Based on the known substrate activity for 2 and our STD NMR control experiments, we chose to use this ligand as a proxy for the native substrate.

The mutant enzyme used here also contains an MBP-affinity tag at the N-terminus of NEU3. In previous studies, we have found that the fusion tag is important for maintaining the stability of recombinant NEU3. Thus, the affinity tag was retained for this work; however, we needed to confirm that any binding interactions observed by STD NMR were the result of NEU3–substrate interactions alone (Albohy et al. 2010). To test for interactions of MBP with 2, we expressed and purified MBP alone, and performed STD NMR experiments with the protein alone and with 2. These experiments showed no significant signals in either the aliphatic or carbohydrate regions of the spectrum, confirming the lack of interactions between 2 and MBP. Furthermore, to test the importance of Sia for binding of 2 to NEU3 we tested a complex of NEU3 and 3. These experiments found a very weak STD signal in the aliphatic (0.9–1.3 ppm) and carbohydrate regions (3.5–4.5 ppm) of the spectrum (Supplementary data, Figure SI2c). The STD NMR spectrum for this interaction was too noisy to make reasonable assignments of interactions, likely due to the low affinity of the ligand.

In addition to binding of 2 and 3, we also tested 4 and 2-deoxy-2,3-didehydro-N-acetylmuramic acid (DANA) as ligands for NEU3(Y370F). The STD NMR spectrum of the 4–NEU3(Y370F) complex had good signal to noise and could be interpreted (vide infra). The NEU3(Y370F)–DANA complex gave poor STD signal under several conditions and could not be easily interpreted. We also tested DANA with NEU3 at different temperatures (25, 32 and 37°C) and at different times of observation (3, 24, and 48 h). Unfortunately, none of these conditions resulted in significant improvements to the signal-to-noise ratio of the STD NMR spectra (data not shown). To provide additional insight into this observation, we tested the ability of DANA to disrupt the NEU3(Y370F)–2 and NEU3(Y370F)–4 complexes in competition experiments (Supplementary data, Figure SI10). We found that high concentrations of DANA (10–15 mM) were able to compete the binding of compounds 2 and 4 to NEU3(Y370F) in separate experiments. These data confirm that DANA is binding to the protein. Additionally, these data support that all three ligands are binding to the same site and help to rule out non-specific ligand–protein interactions.

STD NMR of NEU3(Y370F)–2 complex

We next conducted STD NMR experiments of 2 with NEU3(Y370F) (Figure 2; Supplementary data, Figure SI3). STD NMR spectra were
analyzed, and the relative integrations of assigned peaks were used to construct an epitope map of the substrate (Figure 3; Supplementary data, Table SI1). In the case of overlapped signals, peaks were treated together as an average.

In the epitope map of 2, the strongest STD signal was found for the aliphatic aglycone hydrogens, consistent with the presence of a hydrophobic-binding site that recognizes the lipid aglycone, as previously proposed (Sandbhor et al. 2011). All signals were normalized to the strongest carbohydrate signal as reference (H1′, 100%). Although some of the octyl protons showed stronger signals than H1′, we chose to focus on relative differences among the oligosaccharide contacts since these should be more critical for substrate specificity. The strongest STD signals within the glucose (Glc) residue were seen at H1′ (100%) and H2′ (83%), suggesting that this residue, distal from the glycosidic linkage cleaved by the enzyme, was important for recognition. The remaining protons of Glc showed only moderate STD signals.

The intervening Gal residue may play a less important role in recognition by NEU3. STD signals for Gal were all moderate, ranging from 37 to 50%. Among these, the strongest STD signals for Gal were seen for H2′ (50%) and H5′ (46%); which were not completely resolved in the NMR spectrum. It is notable that both A and B faces of Gal and Glc residues show significant contact with the protein, suggesting that the glycan is not complexed in a shallow-binding site that interacts with only one face of each residue.

The Neu5Ac residue of 2 showed strong STD signals for several positions. The strongest signal for the Neu5Ac residue was found for the glycerol side chain H8′ (73%) and H9′ (46%). Moderate signals were also observed for portions of the pyranose ring (H3′(eq), H4′, H6′) and H7′, H9′, while signals for H3′(ax) and the N5-Ac were relatively weak. In summary, these results are consistent with the aglycone and Glc residues acting as an anchor for the binding of 2. The Gal residue does not appear to have any critical participation in binding, consistent with the weak binding observed for the lactoside alone (vide supra). The presence of the Neu5Ac residue provides additional contacts, most notably within the glycerol side chain of the residue and likely the C1′ carboxylate (although this second contact cannot be observed in the experiment).
STD NMR of NEU3(Y370F)–4 complex

We next turned to fluorogenic NEU3 substrate, 4, which lacks the intervening lactoside and contains an aromatic aglycone. STD NMR of 4 with NEU3(Y370F) was performed under the same conditions used for 2 as the substrate (Supplementary data, Figure S14 and Table S12), and an epitope map was constructed (Figure 4). The strongest signals found were in the Sia residue and were then used as reference (H4′ and H8′). Signals for H8 and the C4-methyl (100%) and H3 (76%) showed the highest relative intensities within the aglycone. Once again, the glycerol side chain of Sia showed strong interactions at H7′ and H8′. Within the pyranose ring, H4′, H5′ and H6′ all showed relatively strong contacts with the protein. In 4, we observed only a small difference between the H3′(ax) and H3′(eq) signals, with the H3′(eq) signal the stronger of the two (44% vs. 49%). We have previously proposed that the glycerol side chain of Neu5Ac may be particularly critical for recognition of substrates and inhibitors by NEU3, and these data support that this portion of both substrates is an important enzyme–substrate contact (Albohy et al. 2011).

MD simulations of NEU3–substrate interactions

To develop improved molecular models of the NEU3–substrate complexes, we conducted MD simulations of each substrate bound to a homology model of NEU3 (Albohy et al. 2010). For both 2 and 4, we first docked the substrate into the active site. We then prepared for a production MD run by the addition of explicit water (TIP3P) (Jorgensen et al. 1983), minimization of the solvent, followed by minimization of the entire system. Additionally, the system was equilibrated by gradual heating and then cooling followed by a short dynamics cycle (0.1 ns), during which the system was brought to the production temperature (300 K). For these simulations, we used AMBER 10 (Case et al. 2005) with the GLYCAM06 force field (Kirschner et al. 2008). The resulting complexes were then subjected to 10 ns MD simulations to convergence at constant temperature and pressure (300 K; 1 atm; Supplementary data, Figures SI5 and SI6).

We analyzed the MD run of 4 by clustering 50,000 frames into 20 clusters using a means algorithm based on the RMSD of the ligand residues (Shao et al. 2007). The resulting clusters showed that the Sia residue maintained the expected key interactions throughout (Supplementary data, Figure S18). Two residues of the arginine triad (R25 and R340) maintain almost constant contact with the C1 carboxylate, while the third (R245) shows some switching between the carboxylate and the glycerol side chain oxygen O9′. We observed that O4′ of Neu5Ac maintained a stable hydrogen bond with R45. Additional contacts between the glycerol side chain and the protein included hydrogen bonds with E113, Y181 and R245. The orientation of the 4-methylumbelliferone aglycone was less conserved, and it was able to form non-polar contacts with the methylene groups of the R49 side chain.

Analysis of MD simulations on the NEU3–2 complex showed similar results to the model with 4, but provided additional information on the interaction of the Glc and Gal residues of the oligosaccharide (Figure 5). Simulations of the Neu5Ac residue exhibited the same

---

**Fig. 4.** STD epitope map of 4 (top) and comparison of MD and STD NMR results (bottom). The MD results are represented as $1/(r^6)$ based on the average distance to closest contacts in the protein, while the STD data are represented as a percentage of the effect.

**Fig. 5.** Molecular model of the NEU3–2 complex. (A) Interactions between 2 and NEU3. The more predominant position of the octyl side chain is shown in the hydrophobic pocket composed of V222, V224, P198 and P247 (some residues are omitted for clarity). The O4 of the Gal residue shows an H-bond contact with R48 found on the D50 loop. (B) An electrostatic potential surface representation of NEU3 with 2 bound in the active site in the same pose as shown in (A).
contacts in the active site observed for both 2 and 4. The C1″ carbonylate of 2 formed salt bridges with R25 (average distance of 2.97 Å). Additional salt bridges were formed between the C1″ carbonylate and R340 (average distance of 2.90 Å). The third member of the arginine triad, R245, was far from the carbonylate group, with an average distance of 6.54 Å. The average distance between the nucleophile oxygen of Y370 and C2″ of Neu5Ac was 5.33 Å, which is slightly larger than the distance observed for complexes of NEU2 with DANA by crystallography (Chavas et al. 2005). Meanwhile, E225 (the proposed general base) maintained a <3 Å distance to the nucleophilic tyrosine OH. The position of the octyl side chain exhibited switching between two different positions on the enzyme.

The aglycone of 2 spent most of its time in a hydrophobic pocket formed by V222, V224, P198, P247, and the backbone portions of H277 and R197 (Albohy et al. 2010). The octyl group was also found to occupy a second hydrophobic patch composed of I117, V118 and the backbone of R114 (ca. 20% of the MD run; Supplementary data, Figure S17). Examination of different clusters showed that the positions of the Neu5Ac and Gal residues were conserved throughout the run, while the Glc residue and the aglycone were more variable.

To compare the results of our STD NMR studies with those of the MD simulations, we analyzed the MD trajectories to determine average distances between specific ligand sites and H atoms of the protein. The MD trajectory was analyzed to identify 20 clusters that were then examined to choose potential close interactions between protons of the substrate and enzyme (Roe and Cheatham 2013). These distances were monitored over all frames of the MD simulation. The three shortest distances between a ligand site and potential close contacts were selected from each frame, and the resulting distances were averaged over the whole run. These average distances are summarized in Supplementary data, Table S1 for 2, and Supplementary data, Table S2 for 4. Note that these distances are not an absolute measure, but they were used as a relative measure of expected contacts for individual H atoms of the substrate with the enzyme. We then used these values to compare STD NMR signals with the predicted epitope map based on MD.

Generally, the results of the STD NMR and MD show partial, but not complete, agreement. In the case of the NEU3–4 complex, we find that both methods identify the C4-methyl and H8 of the aglycone to have close contact to the enzyme. MD simulations predicted that the Neu5Ac residue of 4 should have its closest contacts to the enzyme at H3(eq), H6′ and H8′. Both STD NMR and MD simulation agree on a close contact for H8′, as well as H4′, H6′ and H7′. We also note that trends between the relative signals of H3′(ax)/H3′(eq) and H7′/H8′/H9′ agree between both theory and experiment. The only major disagreement is found for H5′, which shows moderate STD NMR signal, but is predicted to have significantly reduced contact with the protein by MD.

For the NEU3–2 complex, we find that the aglycone has close contact to the protein in both methods. From MD simulations, the Glc and Gal residues were predicted to have their closest contacts at H2′, H4′, H5′ and H2′. Although all of these residues show significant STD NMR signal, the relative signal strength did not track directly with these predicted distances. Several specific sites are consistent between both methods, such as the increased contact found for H2″ and H4″/H5″. Additional comparison of the Glc and Gal residues, when treated by A- or B-faces, shows agreement between the methods (see Supplementary data, Table S13). MD simulations give significant agreement with experiment. The strongest contact from STD NMR was H8″, which also shows close contact with the protein in simulations. Note that the other residues of the glycerol side chain are also in agreement, with moderate contacts at H6″, H7″ and H9″. Among contacts to the pyranose ring, the order of the Neu5Ac H3″ (ax) vs. H3″(eq) signals is in agreement for both methods. The H4″ site is predicted to have moderate contact with the enzyme, consistent with the experimental data. However, H5″ shows significant disagreement between the two methods as this site is predicted to have reduced contact, which is not seen in the STD NMR data.

Discussion

STD NMR is a valuable technique for the study of carbohydrate–protein interactions (Maggioni et al. 2012; Plum et al. 2011). Here, we have mapped the interactions of NEU3 with two known substrates by STD NMR: 4MU-NANA (4) and octyl-GM3 (2). Currently, the only reported crystal structures for the hNEU family are for NEU2 (Chavas et al. 2005, 2010). The hydrophobic nature of the membrane-associated NEU3 likely increases the difficulty in obtaining a crystal structure of the protein. In the absence of crystallographic data, homology models of the enzyme, based on the crystal structure of NEU2, have been developed (Albohy et al. 2010). However, experimental validation of these models has, until now, been restricted to site-directed mutagenesis of the enzyme (Albohy et al. 2010). In this work, we have used the results of STD NMR to interrogate the NEU3 homology model. We find that while certain essential features of the model are well supported, some aspects of ligand binding are not clearly predicted by the model. We discuss potential reasons for this disagreement below.

We considered several factors that could interfere with the reliability of the STD NMR results. We employed an inactive mutant of NEU3, NEU3(Y370F), to avoid degradation of the substrate during acquisition of the STD NMR data. Additionally, we used an analog of GM3 (2) that significantly reduced the formation of micelle artifacts in the STD NMR data, allowing us to make observations using the aliphatic region of the spectrum. It is interesting to note that STD NMR may be useful for the observation of lipid–lipid interactions within a micellar structure based on our findings with GM3 (Supplementary data, Figure S12d). Indeed, STD NMR has been used previously to study lipid–peptide interactions (Chatterjee et al. 2004; Chatterjee and Mukhopadhyay 2005). Competition experiments with DANA and ligands 2 and 4 demonstrate that these three ligands bind at the same site in the protein (Supplementary data, Figure S110). The lack of STD NMR signal for DANA in our experiments is likely the result of a slow off-rate for the inhibitor. Another potential source of variability in the epitope maps determined here is the lack of stability of the enzyme. The protein does lose activity over time, and as a result, the sample may contain some portion of unfolded protein. In previous experiments, we have found that the enzyme can maintain activity for ~3 h at room temperature, and our experiments were conducted within this window. However, we cannot rule out the possibility that a portion of unfolded protein may also interact with the substrate, which could result in a biased epitope map.

Both epitope maps of 4 and 2 support the involvement of the substrate aglycone in binding to NEU3. We previously proposed a model for recognition of glycolipids by NEU3 based on relative hydrolysis rates of GM3 analogs (Sandbhor et al. 2011). The previously proposed model suggested that at least two sites were required for recognition by the enzyme, including the Neu5Ac residue and the hydrophobic aglycone. Our results here provide the first direct evidence that the aglycone interacts with the protein. For example, the
maximum STD effect was seen with the hydrophobic aglycone in both 4MU and octyl cases. It is notable that the aglycone of 4MU-NANA (4) shows STD signals when the intervening Gal residue of 2 has relatively moderate STD NMR response. Additionally, the epitope maps show significant binding for the Neu5Ac residue of each substrate. The most notable contacts appear to be at H5, H8 and H9 of the Neu5Ac residues. The interaction of the Neu5Ac glycerol side chain has previously been proposed to be important for human NEU recognition based on inhibitor studies (Albohy et al. 2011). The epitope map of octyl-GM3 (2) also suggests that Glc provides an additional point of contact between the ligand and the enzyme. In particular, the reducing end of the sugar (H1' and H2' of Glc) shows some of the strongest STD NMR signals within the entire substrate.

Molecular modeling can be used to interpret STD NMR mapping results. Previous studies have analyzed STD NMR using a range of methods (Krishna and Jayalakshmi 2008; Balazs et al. 2013; Guzzi et al. 2013). Here we compared an averaged set of distances for each H atom obtained from MD simulations of the relevant complex. The close contacts between the aglycone of octyl-GM3 with two hydrophobic patches on the enzyme are consistent with the strong STD NMR effect seen for this portion of the ligand. We identified two hydrophobic patches on NEU3 within the model which may be advantageous for complexation of diacyl sphingolipids in the native substrates. The MD results for both substrates indicate consistent interactions between the protein and the Neu5Ac residue, including the C1-carboxylate interactions with the arginine triad and contacts between the glycerol side chain (H6, H7 and H8 positions of Neu5Ac).

In comparing the STD NMR and MD results for the NEU3–2 complex, we were struck by the apparent disagreement found for the Glc and Gal residues. The STD results give little differentiation between sites on these two residues, and they show moderate signals indicating contact with the protein. Most of the sites found within Glc and Gal have a much smaller range of STD NMR signals (39–50%) when compared with Neu5Ac (10–73%), with the notable exception of H1' and H2'. Additionally, we observed evidence of facial selectivity in the Glc and Gal residues based on the predicted binding mode of the ligand (Figure 5) (Albohy et al. 2010). Close examination of the data for the A- vs. B-faces of Glc and Gal shows agreement between MD and STD NMR (Supplementary data, Table SI3; Figures 3 and 4). However, the degree of difference between individual signals is still small when compared with that of Neu5Ac. We speculate that this finding may indicate a dynamic mode of binding for the lactoside portion of the substrate which tempers the expected differences in this portion of the epitope as observed by STD NMR.

Although we concluded that our binding model was in general agreement with the STD NMR data, we also considered that potential shortcomings of the homology model may contribute to our observations. Previous structural studies of NEU2 have found that substrate recognition by the enzyme is dynamic and depends on rearrangement of a flexible loop in the protein (the D46 loop in NEU2) (Chavas et al. 2005). The homologous loop in NEU3 contains D50, a site predicted to be in contact with the substrate, and mutation of which greatly reduces enzyme activity (Albohy et al. 2010). The D50 loop of NEU3 shows high homology to the sequence of NEU2; therefore, we expect that the position of the loop in our model is reliable. However, we also note that several loops on the same face of the model have much lower homology to NEU2, and some of these may be close enough to influence substrate binding directly. We identified three candidate loops with low homology to NEU2: Q16–122, Y182–H199 and C388–C394. Due to the low homology of these regions of the protein with the NEU2 sequence, their conformations may be suspect (see Supplementary data, Figure SI9). The loops of NEU3 have been noted to be particularly variable and longer than in other sialidase enzymes (Giacopuzzi et al. 2011, 2012). Additional contacts to NEU3 through one of these loops could result in a different predicted epitope map for the ligand in MD simulations. Future experimental work is planned to examine the role of these loops through site-directed mutagenesis of the protein.

Despite the potential shortcomings of the NEU3 homology model, it is worthwhile to mention that the model is still predictive for aspects of enzyme specificity (Smuova et al. 2014). For example, O4' in the NEU3–2 model forms a hydrogen bond with R48, which may help to order the D50 loop for catalysis (Albohy et al. 2010). Additionally, we note that the O4' position would be the site of GlcNAc branching in glycolipid substrates such as GM1 and GM2. These branched glycolipids are known to be poor substrates for the enzyme, and this finding is consistent with blocking of the optimal conformation of the D50 loop of NEU3 and reducing the substrate activity of GM1 and GM2 (Figure 5).

Conclusions
We have used STD NMR to map the epitopes of two substrates bound to the NEU3 enzyme. Additionally, we used MD to test the predictions of a homology model of the protein when compared with the epitope maps. Our results confirm the importance of the Neu5Ac residue for recognition by NEU3, as well as provide direct evidence of recognition of the hydrophobic aglycone by the enzyme. There is general agreement between our homology model and the STD NMR data. Our models provide general predictions that agree with the STD NMR data including the facial interactions of both Glc and Gal residues, as well as confirming interactions of the Neu5Ac glycerol side chain and the lipid aglycone with the protein. We conclude that the homology model of NEU3 may require some refinement of the conformation of protein loops that lack homology to the known structure of NEU2.

Materials and methods
Synthetic ligands
GM3(1) was purchased from Avanti Polar Lipids, Inc. Compounds 2–4 were synthesized as reported (Sandbhor et al. 2011; Zhang et al. 2013).

Protein expression
Protein was isolated by culturing Escherichia coli TB1 cells containing plasmids pMBP-Neu3-Y370F (for expression of the NEU3 inactive mutant) or pMal (for expression of the MBP protein alone) in LB medium containing ampicillin (100 μg mL⁻¹) to OD₆₀₀ of 0.3 at 37°C. Production of the fusion protein was induced by addition of IPTG to a final concentration of 0.3 mM at 20°C. Cells were harvested by centrifugation after growing overnight from induction. The pellet was resuspended (50 mL L⁻¹ of medium) in buffer (20 mM morpholinopropane sulfonic acid (MOPS), pH 7.2, 200 mM NaCl, 1 mM EDTA and 0.1% Triton X-100) supplemented with one complete protease inhibitor tablet (Roche). The lysate was passed through a cell disruptor once at 20,000 psi and then immediately pelleted by centrifugation at 105,000 × g for 60 min at 4°C. The supernatant was loaded onto an amylose column (New England Biolabs) equilibrated with 20 mM MOPS buffer (200 mM NaCl, pH 7.2). MBP-fusion protein was eluted with running buffer containing 10%
glycerol (v/v) and 10 mM maltose. The chromatogram showed one single peak for the pure protein and it has one major band on SDS–PAGE at molecular weight of ~93 kDa. The final protein concentration was determined by measuring UV absorbance at 280 nm. The NEU3(Y370F) mutant was previously found to show reduced activity when compared with wild-type protein (Albohy et al. 2010).

NMR sample preparation
All samples were prepared in the NMR buffer (40 mM sodium phosphate buffer, pH 5.65, in D2O). The protein sample was dialyzed to the sample buffer and concentrated using an Amicon spin column (MWCO 30 kDa). The final protein concentration was adjusted to 25 µM and the substrate (4 or 2) was added to a final concentration of 2.5 mM (100 times the protein concentration). The sample was added to either 5 or 3 mm NMR tubes. All the components (buffer, enzyme and substrate) were subjected to STD experiments separately to check for interference. To confirm that the substrates bind to NEU3 and not MBP, 2 was tested with the MBP protein. No STD signal was observed under these conditions, ruling out MBP–substrate interactions interfering with the measurement. In addition, 3 was tested with the NEU3(Y370F) protein and showed STD signal only at the octyl side chain, which we attribute to weak, but specific binding as the protein alone and the substrate alone show no signal (Supplementary data, Figure SI2).

NMR experiments
Spectra were recorded on a Varian 700 MHz spectrometer equipped with a cryo probe. The STD NMR spectra were collected with 3072 scans. During each scan, the protein was saturated with a series of 20 GAUSSIAN-shaped pulses (50 ms, 1 ms delay between pulses), for a total saturation time of 1.02 s (Mayer and Meyer 2001). Selective saturation of protein resonances was done at −0.3 ppm with 2 and 4 to excite the protein but not the substrate. Control experiments with the complex of 2 and NEU3(Y370) with on-resonance saturation at 7 ppm showed a substantially similar epitope, with a slight increase in the STD NMR signal of H1' (Rademacher et al. 2007). Several on-resonance values were tested for 1 but none were able to excite the protein without the substrate (on resonance at −0.25 ppm is shown in Supplementary data, Figure SI2d). The off-resonance frequency used was 27 ppm. A WATERGATE (Pirotto et al. 1992; Sklenar et al. 1993) pulse sequence was used to suppress the water signal in the STD spectrum, and the FID was collected for 2.0 s afterwards.

MD simulations
The previously reported GM3–NEU3 model was used as the basis for our simulations (Albohy et al. 2010). Since STD experiments were performed with 2, we replaced GM3 with this substrate in the complex. Previously reported partial charges for the octyl side chain were used (Dupradeau et al. 2008; Abel et al. 2011), and the sugar residue names were fixed according to GLYCAM notations (Abel et al. 2011). The LEAP module (Zhang et al. 2012) in AMBER tools was used to generate AMBER topology and coordinate files after the solvation and neutralization of the complex. Water (TIP3P) was added in an octahedral box of 7 Å around the complex with 3 Na+ ions added (Jorgensen et al. 1983). MD simulations were performed using the AMBER 10 package (Case et al. 2005). The AMBER ff99SB force field (Hornak et al. 2006) was used for amino acids, and GLYCAM 06 was used to treat the carbohydrates (Kirschner et al. 2008). The protocol used was guided by an earlier report (Islam et al. 2011).

Before the production run of the MD simulation, the system was minimized and equilibrated. Water molecules were minimized while first keeping the substrate constrained. This was followed by a minimization of the entire system. In both minimization steps, a steepest descent energy minimization was carried out for 50 cycles followed by 4950 cycles of conjugate gradient minimization. A total of 600 ps of annealing similar to earlier reported protocol was used during which the temperature was increased every 50 ps from 5 to 300 K and then cooled back from 300 to 5 K (Islam et al. 2011). This step was followed by a short equilibration run of 200 ps during which the temperature of the systems was gradually increased from 5 to 300 K over 150 ps, and then kept constant for 50 ps. The production simulation (10 ns) was performed under constant pressure and temperature conditions with periodic boundaries. The temperature was kept at 300 K and the pressure at 1 atm to simulate the experimental conditions. The particle mesh Ewald method was used to control long-range electrostatic interactions and the SHAKE (Ryckaert et al. 1977) algorithm was used to constrain bond lengths involving hydrogen to their equilibration values. The models were evaluated using protein structure and model assessment tools under the SWISS-MODEL workspace (Guex and Peitsch 1997; Benkert et al. 2008). For analysis purposes, 50,000 frames of the MD trajectory were collected during the run and clustered into 20 clusters. Close contacts were chosen from the clusters and the distances were monitored during the stable part of the MD run (10 ns for NEU3–2 and the last 7.5 ns for NEU3–4). The three shortest distances in each frame were selected, and the average of these three was used for comparisons of relative distance to the protein.

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

Acknowledgements
The authors thank the Alberta Glycomics Centre and the Department of Chemistry NMR facility (University of Alberta) for NMR access. We thank M. Miskolzie (University of Alberta) for assistance with NMR experiments, Y. Zou, J. Cartmell and M. S. Sandhbor (University of Alberta) for preparation of substrates, and H. Taha (University of Alberta) for helpful discussions. A.A. was supported by an Alberta Innovates Health Solutions graduate fellowship.

Conflict of interest statement
None declared.

Funding
This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC).

Abbreviations
DANA, 2-deoxy-2,3-didehydro-α-D-N-acetylneuraminic acid; Gal, galactose; GH, glycosyl hydrolase; Glc, glucose; hNEU, human neuraminidase; MBP, maltose-binding protein; MD, molecular dynamics; 4MU-NANA, 2′-(4-methylumbelliferyl)-α-D-N-acetylneuraminic acid; MOPS, morpholinopropane sulfonic acid; NEU, neuraminidase; NMR, nuclear magnetic resonance; Sia, sialic acid; STD, saturation transfer difference.


