Relative Binding Free Energy Predictions for Inhibitors of Tetrameric Influenza Virus Neuraminidase

Billy J. Williams-Noonan, Elizabeth Yuriev and David K. Chalmers**

Medicinal Chemistry, Monash Institute of Pharmaceutical Sciences, Monash University, 381 Royal Pde Parkville, Victoria 3052, Australia.

ABSTRACT

Accurate methods to predict the free energies of protein-ligand interactions have great potential to assist rational drug design. In this work, we used molecular dynamics simulations with alchemical perturbation to predict the binding of carbohydrate-based ligands to influenza virus neuraminidase (N2). This specific drug target is a challenging test system for alchemical free energy methods because it has flexible binding site motifs. We use a molecular dynamics protocol that works for longer time scales than are often reported in previous molecular dynamics studies of N2. We demonstrated that N2-ligand complex stability and that accurate N2 150-loop dynamics, on a 350 ns time scale, are both force field-dependent (AMBER99SB-ILDN, GAFF and TIP4P water are an optimal combination). Further, we showed that crystallographic waters must be retained to maintain stability of N2-ligand complexes over 350 ns. Using our modelling protocol, we were able to determine relative binding free energy values between neuraminidase ligands which correlated strongly with experimental differences in pIC₅₀ values ($R = -0.96$, $\rho = 0.86$, $N = 13$, $sig < 0.0001$). It is anticipated that the molecular dynamics protocol and the relative binding free energy methods reported here, will both be useful in expediting the discovery of novel therapeutics for N2 and other homologous glycohydrolases.

INTRODUCTION

Rational drug design is underpinned by an understanding of the interactions between a potential drug and its biological target. In some projects, the nature of the interactions may be only poorly understood while in others, the interactions can be well characterised through crystallography or other methods that provide a detailed, atomic scale picture of the ligand in the binding site. However, even when good structural information is available, it is frequently difficult to use this information to design more potent compounds. With the development of computational power and improvements in algorithms, *in silico* free energy methods have expanded potential to guide the development of biologically active

compounds, by affording estimates of protein-ligand interaction strength. Here, we conduct a retrospective study, to investigate the ability of alchemical free energy methods to reproduce experimental measures of binding strength for the glycohydrolase enzyme neuraminidase, which is a key virulence factor in the influenza virus. We provide here an optimal simulation protocol for neuraminidase-ligand complexes and complement existing evidence that molecular modelling methods have scope in assisting the design of novel anti-influenza drugs, as reviewed by Patel *et al.*¹

It is common to rank compounds by binding affinity, in terms of the free energy difference between the ‘bound’ and ‘free’ states (ΔG_{bind}). Binding affinity is often expressed terms of the dissociation constant, K_D (Equation 1) where $[P]$, $[L]$ and $[PL]$ are the concentrations of the protein, ligand and complex. The free energy of binding is related to the K_D as shown in Equation 2, R is the gas constant and T is the absolute temperature.²

$$K_D = \frac{[P][L]}{[PL]} \quad 1$$

$$\Delta G_{bind} = RT \ln K_D \quad 2$$

Since many physical events (e.g. protein-ligand binding) occur outside the timescales that are easily accessible by MD, Zwanzig proposed that a system could be perturbed between two different states, but with respect to the variable λ , instead of time.³ Methods that do this are referred to as ‘alchemical’. The original alchemical perturbation method is ‘Free Energy Perturbation’ (FEP),³ where the phase space between two states is sampled with respect to the scaling variable (λ). The system is in state A when $\lambda = 0$, the, and in state B when $\lambda = 1$. As shown in Figure 1, this methodology can be used to calculate the difference in free energies of binding ($\Delta\Delta G$) of two ligands, A and B. This approach is commonly known as a relative free energy calculation. In this instance, state A includes ligand A, and state B uses ligand B. Two series of simulations are performed where ligand A is non-physically transformed into ligand B by addition/deletion of atoms, transformation of atoms from one element to another or modification of bonds (e.g. transformation of single to double bonds). One set of transformation calculations is performed with the ligand in solution and the other with the ligand in the binding site. By subtraction of the free energy changes in the binding site and in solution, one can calculate the difference in binding free energies of ligands A and B (Figure 1). In a recent study,⁴ we compared a number of methods for predicting the binding affinity of ligands. In that work, we found the relative alchemical perturbation method to be most accurate method and of sufficient accuracy to be useful in drug development. Relative alchemical perturbation methods are increasingly being applied in industrial and academic research settings.⁵⁻⁷

Figure 1. Prediction of binding free energies using the relative method. Arrows depicting the perturbation reaction being simulated over a series of λ -states are in grey. Arrows for long time-scale reactions are in yellow.

Influenza is a debilitating disease with a high mortality rate and there have been multiple influenza virus pandemics in the past century.^{8,9} Influenza has been successfully targeted by a number of anti-influenza therapeutics including Relenza (zanamivir)^{10,11} and Tamiflu (oseltamivir)¹²⁻¹⁵. Relenza and Tamiflu are ester prodrugs that are cleaved to the active forms, Relenza carboxylate and Tamiflu carboxylate (Figure 2), following absorption. The free acids act by inhibiting the influenza virus enzyme neuraminidase, a glycohydrolase that cleaves 5-*N*-acetyl neuraminic acid (Neu5Ac) residues from cell surface glycosides. Other inhibitors of neuraminidase include Neu5Ac2en and glucuronic acid (GlcA) derivatives combine the isopentyl ether group of Tamiflu with the unsaturated pyranose scaffold of Relenza (Figure 2).

Neuraminidase plays a crucial role in viral replication in allowing newly synthesised viral progeny to escape the host cell. Inhibition of viral neuraminidase is a validated mechanism for targeting influenza virus. Blocking the active site of influenza virus neuraminidase means that the viral progeny remain attached to the host cell, and the host immune system is then able to remove the trapped virions. There are 9 subtypes of neuraminidase (N1 to N9).^{16,17} All subtypes of viral neuraminidase exist as tetramers,¹⁸ as shown in Figure 3A. An illustration of the influenza virus neuraminidase active site with Tamiflu bound is shown in Figure 3B. The overlay of crystal structures of all neuraminidase subtypes reveals very similar binding site architecture among each variant of neuraminidase. The active site contains two flexible protein motifs. Conformational change in the 150-loop occurs when inhibitors of neuraminidase bind to the active site. Specifically, the volume underneath the 150-loop (the 150-cavity, Figure 3B) expands to allow for the binding of bulkier substituents. Besides the global SARS-CoV-2 pandemic that continues into 2021, other recent pandemics have been caused by the influenza virus containing the following hemagglutinin and neuraminidase subtypes: H1N1 (in 1918), H2N2 (1957), H3N2 (1968)

and H5N1 (in 2006).^{8,9} Therefore, the N1 and N2 subtypes of neuraminidase are the most relevant with respect to human disease.

Figure 2. Structures of 5-*N*-acetyl neuraminic acid and (Neu5Ac) and neuraminidase inhibitors.

Figure 3. (A) The N2 neuraminidase tetramer from PDB-ID 4D8S.¹⁹ Protein chains are shown as ribbons with transparent surfaces. Calcium ions in the active site of each monomer are shown as green spheres. Bound inhibitors are shown as sticks with grey carbon, white hydrogen, red oxygen and blue nitrogen. **(B)** Overlay of the binding sites of two neuraminidase subtypes: N2 (PDB-ID: 4K1K, white)²⁰ and N9 (PDB-ID: 6HEB, light brown). The ligand Tamiflu carboxylate (**1**) is shown in black. A ribbon representation of residues within 12 Å of the 150- and 430-loops framing the active site are shown as ribbon representations coloured red and black. The cavities underneath these loops are referred to as the 150- and 430-cavities

The inhibitors shown in Figure 2 make specific interactions with the active site of influenza virus neuraminidase. Nine conserved residues in the neuraminidase active site make contact with Neu5Ac2en-based inhibitors.^{10, 11, 19, 20} A cluster of three arginine residues in the active site (Arg-118, Arg-272 and Arg-371) interact with the carboxylate group of Neu5Ac²¹ and related ligands. When R₁ (red, Figure 2) is a glycerol side chain (e.g. in Relenza carboxylate), the hydroxyl groups of this moiety make hydrogen bonding interactions with Glu-276 and active site water molecules. Tamiflu-like ligands contain an isopentyl ether group as R₁, which binds to neuraminidase by an induced-fit effect, where Glu-276 in the receptor flips to interact with Arg-224, and active site waters are displaced so that the isopentyl ether can make hydrophobic interactions with the receptor.^{19, 20} The R₂ group (blue, Figure 2) projects into a pocket covered by the active site 150-loop (Figure 3B). This protein loop is flexible, and shifts to accommodate the bulkier 4-guanidino group of Relenza-like ligands. The 4-hydroxy group of Neu5Ac2en interacts favourably with Glu-119. Likewise, the 4-amino group of Tamiflu carboxylate forms a salt bridge between Glu-119,^{19, 20} while the guanidinium of Relenza carboxylate interacts with Glu-119 and Glu-227 (Figure 3B).^{10, 11}

In 2012, Bhatt *et al.* demonstrated that unsaturated glucuronic acid (GlcA) analogues inhibit neuraminidase with low to mid μM IC₅₀ values (Table 1).¹⁹ It was shown by NMR that the modest affinity of these inhibitors is explained by differences between conformations of the GlcA ligand in solution and the bound GlcA ligand.¹⁹ The energetic penalty associated with adopting the bound conformation explains the weaker binding affinity for ligands **2-6** compared with Tamiflu carboxylate or Relenza carboxylate, which are sub-nanomolar binders to neuraminidase.

Table 1. IC50 values for neuraminidase inhibition by inhibitors with Tamiflu and glucuronic acid (GlcA) scaffolds.

Compound	X	Y	Neuraminidase IC50 (nM)	pIC50 ^d
1	H	-	0.37 (0.09) ^a	9.43 (9.26 – 9.72)
2	-	OH	320 (40) ^b	6.49 (6.40 – 6.62)
3	-	CH ₂ CH ₃	28000 (3600) ^b	4.55 (4.45 – 4.68)
4	-	CH ₃	1450 (190) ^b	5.84 (5.74 – 5.97)
5	-	CH ₂ CH=CH ₂	212000 (28000) ^b	3.67 (3.57 – 3.81)
6	-	CH ₂ C(=O)CH ₃	5390 (700) ^b	5.27 (5.17 – 5.40)
7	-C(=O)H	-	1290 (110) ^c	5.89 (5.82 – 5.97)
8	-C(=O)CH ₃	-	240 (10) ^c	6.62 (6.59 – 6.66)
9	C(=O)CH=CH ₂	-	2980 (210) ^c	5.53 (5.47 – 5.59)

^aFrom Gubareva *et al*²² measured against N2 neuraminidase. Number in brackets is a standard deviation (SD) from triplicate experiments. ^bFrom Bhatt *et al*¹⁹ measured against N2 neuraminidase. The SD values in brackets is approximated from information given by Bhatt *et al*. ^cFrom Alexander *et al*²³ measured against N1 neuraminidase. Standard deviation in brackets is from the paper by Alexander *et al*. ^dThe pIC50 values are determined from the molar IC50. Numbers in brackets are a range of pIC50 values, spanning ± 2 SD in IC50.

One of the first applications of molecular dynamics to predict binding affinities to neuraminidase was reported in a study by Taylor and von Itzstein in 1994.²⁴ Neu5Ac, Neu5Ac2en, 4-amino Neu5Ac2en and Relenza carboxylate were docked into the binding site of N9 neuraminidase and a series of energy minimisations and short MD simulations were then run to produce low energy structures for the complexes. By analysing these structures and measuring the distances between interacting residues in the binding site and the ligand, the binding enthalpies were predicted. These binding enthalpies were reported to correlate with experimental measurements of binding strength.

More recently, a variety of free energy methods have been applied to predict binding of neuraminidase ligands. Using the thermodynamic integration method, Lawrenz *et al*.²⁵ predicted the binding free energy of the inhibitor peramivir binding to the N1 neuraminidase tetramer to within 1.1 kJ mol⁻¹ of the experimental value. They sampled bound conformations using replicate simulations with the GROMOS-96 force field, totalling to 800 ns of simulation time. In a separate study,²⁶ they compared the ability of the GROMOS-96 and AMBER-99SB force fields to predict binding free energies for two neuraminidase ligands, in the presence and absence of binding site calcium ions. They found that the GROMOS-96 and AMBER-99SB force fields performed comparably well in predicting binding strength, and the presence of the binding site calcium ion did not significantly affect inhibitor binding strength.

Simulations where biasing forces are applied to pull a ligand from the binding site are referred to as steered simulations. The maximum force required to pull a ligand from a binding site is commonly referred to as F_{max} , or the rupture force. Mai *et al.* used the AMBER-99SB force field and showed that F_{max} correlated with experimental binding free energies for monomeric N1 neuraminidase, with $R = 0.70$.²⁷ In a separate study,²⁸ the same authors found a correlation between F_{max} and ΔG_{bind} for 24 protein-ligand complexes involving variants of neuraminidase, with $R = 0.97$. Two novel anti-viral agents were included in the latter study, one of which (CS-8958) has shown efficacy both *in vitro* and during clinical trials.^{27, 29, 30} In these two studies, Mai *et al.*^{27, 28} compared steered MD to the MM-PBSA method and found that steered MD was more accurate for influenza virus neuraminidase inhibitors.

Nguyen *et al.*³¹ also performed MM-PBSA endpoint calculations on neuraminidase ligands trialling a series of different protein force fields and water models. All simulations were performed with tetrameric neuraminidase. The authors found that the predicted binding free energy data correlated strongly with the density of water molecules in the active site ($R = 0.90$), using the AMBER-99SB force field and TIP4P water molecules. The use of the OPLS force field also gave a good correlation between active site water density and the ligand binding free energy ($R = 0.67$) while the CHARMM and GROMOS (where only two ligands in the set were predicted to bind) did not model protein-ligand complex stability well. This suggests that force field choice and the presence of water molecules in the active during a simulation has an effect on the predicted binding free energy.

Amaro *et al.*³² investigated the dynamics of the active site 150- and 430-loops with respect to the N1 and N2 variants of neuraminidase. They ran 100 ns molecular dynamics simulations using an AMBER force field and a TIP4P water model with tetrameric neuraminidases. The 150-loop was maintained in position by a salt bridge between Asp-147 and His-150. Importantly, this ionic lock was observed to break intermittently over the short 100 ns timescale, resulting in changes to the 150-cavity volume. Simulations with neuraminidase inhibitors bound in the active site were not reported.

Recently, there is a trend toward using simple endpoint methods, such as MM-PBSA, to determine the absolute binding free energy values of ligands binding to wild-type and mutant neuraminidase variants.³³⁻³⁷ These simulations have potential in assisting the design of novel anti-influenza drugs specific to wild-type and mutant neuraminidase, with scope in combating anti-viral resistance in influenza virus. Since 2018, authors who have reported the use of endpoint methods to study neuraminidase-ligand binding³³⁻³⁷ used different MD protocols and system constructs in each individual study. Notably, Ren *et al.*³⁴ report the use of systems comprising monomeric neuraminidase-ligand complexes for 20 ns MD simulations, where the trajectories were subject to analysis by MM-PBSA. A total of 16 different forcefield combinations for the protein and water models were compared, in terms of the accuracy of the resulting MM-PBSA calculation. Only 3-point water models were trialled. These authors found that variants of the AMBER forcefield afforded the most accurate measures of protein-ligand binding free energy with respect to experiment.

Together, these studies show that modelling ligands binding to neuraminidase is challenging. In particular, the protein dynamics of the 150-loop must be accurately reproduced by the chosen force in both the free protein and when the protein is in complex with inhibitor. Additionally, the binding site arginine residues make important interactions with the carboxylate of most neuraminidase ligands, and the three arginine side chains are shown to be within 3 Å of each other. Arginine residues are protonated at physiological pH, so the chosen force field must be parameterised such that these positively charged side chains do not repel each other on long timescales.

The aim of this work was to establish a modelling approach that could usefully predict the binding affinities of inhibitors of neuraminidase and potentially other glycohydrolases. In a recent study,⁵ we investigated the ability of a number of free energy methods to predict the binding of ligands to HIV integrase. Of the methods investigated, we found that the relative Bennet's acceptance ration (BAR (relative)) gave the most accurate predictions. We therefore chose to apply this free energy prediction method to neuraminidase. The previous work on neuraminidase modelling, discussed above, has shown that to be successful, we would need to use an appropriate force field, water model, and protein model. It is not clear from previous work what molecular dynamics protocol is best for modelling neuraminidase. In previous studies, modelling has often been performed with tetrameric neuraminidase,^{31,32} but no comparison between monomeric and tetrameric simulations has been reported. Since conformational changes in the active site are important for ligand binding, force field choice will affect binding free energy prediction accuracy if some force fields are better at modelling the binding site flexibility than others. Accordingly, we have used an initial set of unbiased long scale MD simulations to establish a suitable computational model for integrase. Using this model, we have performed BAR (relative) calculations on a series of Tamiflu and GlcA analogues to study the binding of neuraminidase ligands, which we compare to experimental pIC50 values derived from the literature.

METHODS

Neuraminidase tetramer model: A crystal structure of GlcA (**2**) (see Table 1) bound to N2 neuraminidase (PDB-ID: 4D8S), with 2.40 Å resolution,¹⁹ and also a 1.60 Å resolution structure of N2 neuraminidase with bound Tamiflu carboxylate **1** (PDB-ID: 4K1K)²⁰ were used for modelling. Structure PDB-ID: 4K1K was used to model the protein and binding site water molecules: the glycosylating carbohydrate residues were deleted, missing protein loops modelled with PrimeX,³⁸ and a tetrameric model constructed using the symmetry mates. Binding site waters within 10 Å of bound Tamiflu carboxylate were retained. A model of compound **2** bound to neuraminidase was generated by using the Protein Structure Alignment Tool in Maestro 2018.1 to align the proteins from PDB-IDs 4D8S and 4K1K and transferring the GlcA **2** coordinates. The binding site calcium ion was retained. Bound models for ligands **3-6** were manually constructed based on the ligand coordinates of the GlcA **2** model.

Structures for **7-9** were built from the coordinates of **1** bound to N2 from PDB-ID 4K1K. The Protein Preparation Wizard tool in Maestro was used to protonate structures at pH 7.0 and remove clashes. To remove steric strain, these manually built ligands were minimized in Maestro in the presence of the binding site waters and protein to generate predicted bound poses.

MD Simulations: MD simulations were performed using GROMACS 2016.3.^{39, 40} The AMBER99SB-ILDN force field⁴¹ was used for the protein and ions. The Generalised Amber Force field (GAFF)⁴² was used to parameterize all ligands. Periodic boundary conditions⁴³ were set up so that the protein-ligand complex was centred in an $6.0 \times 10.0 \times 10.0$ nm periodic cell. Simulations used a 2 fs time step, and neighbour searching was performed every 80 ps. The default Verlet integrator was used.⁴⁴ The LINCS algorithm⁴⁵ was used to maintain rigid bond-lengths. In each case, LINCS-iter was set to 4 and LINCS-order was set to 6. Coulombic and Lennard-Jones potential cut-offs of 1.4 nm were used for short-range interactions and the Particle Mesh Ewald (PME)⁴⁶ algorithm was used to account for electrostatic interactions beyond this range. The protein-ligand system was solvated to a density of approximately 1.00 g mL^{-1} with explicit TIP4P-Ew water.⁴⁷ The overall charge of the system was made neutral through the addition of sodium or chloride ions. Additional sodium and chloride ions were added to make a total salt concentration of $\sim 150 \text{ mM}$.

Each system was equilibrated using the following procedure. The solvated protein-ligand complex was first subjected to a restrained, 10,000 step, steepest descent energy minimization, with harmonic position restraints ($k = 25,000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$) on all protein and ligand atoms. This restrained minimisation was followed by an unrestrained 10,000 step steepest descent energy minimization, to relieve any residual steric strain on the system. It was found that the neuraminidase 150-loop and binding site arginine residues were prone to move, and care had to be taken to prevent this during the equilibration phase. The system was therefore equilibrated using series of simulations as shown in Table 2. Throughout the equilibration phases, complexes were monitored for stability using Visual Molecular Dynamics (VMD). By prior alignment of the trajectory protein coordinates, movement of the ligand by more than $4 \text{ \AA}$ RMSD from the initial pose was typically associated with unbinding events. Complexes that remained stable for the duration of equilibration phase 8 were then subjected to unbiased MD and alchemical perturbation, using the same ensemble.

Table 2. Equilibration protocol for neuraminidase complexes.

Phase	Time (ps)	Thermostat	Temp. (K)	Barostat	Restraint force constant on protein heavy atoms (kJ mol ⁻¹ nm ⁻²)	Restraint force constant on ligand atoms (kJ mol ⁻¹ nm ⁻²)
1	300	Berendsen	50	-	25,000	25,000
2	50	Berendsen	100	-	2500	2500
3	50	Berendsen	150	-	2500	2500
4	50	Berendsen	250	-	2500	2500
5	50	Berendsen	300	-	250	250
6	1000	V-rescale	300	Berendsen	25	25
7	300	V-rescale	300	Parrinello-Rahman	1 ^a	1
8	10,000	V-rescale	300	Parrinello-Rahman	-	-

^a Restraints on C α atoms only.

Models for the simulations of the free ligand were constructed in cubic periodic cells with dimensions, $5.0 \times 5.0 \times 5.0$ nm. The ligand was solvated to a density of 1.00 g mL⁻¹, and water molecules were replaced with Na⁺ or Cl⁻ ions to bring the salt concentration to 150 mM and to neutralize the electric charge of the system. Equilibration of the free ligand in solution began with an unrestrained 10,000 step steepest descent energy minimization. Phases 5, 6 and 8 (Table 2) were used to bring the solvated ligand to a temperature of 300 K and a pressure of 1.0 bar, except no harmonic position restraints were used, and the protein was not present. Free ligand runs used phase 8 conditions.

BAR (relative): Dual topologies for GROMACS with the AMBER99SB-ILDN force field were constructed with FESetup.⁴⁸ Monomer A of the tetramer had a bound ligand, while the three remaining monomeric units had ligand coordinates removed. A total of 14 relative perturbations were performed between Tamiflu carboxylate (**1**) and ligands **2-9**, whereby one ligand was transformed into another. Each perturbation used 20 λ -states; 10 for the Coulombic interaction perturbation, and then separately, 10 for the vdW radii transformations. Perturbations for the Coulombic and then vdW parameters were performed linearly. Similarly, the bond, dihedral and mass terms were perturbed linearly across the 20 λ -states. Prior to each simulation, a steepest descent unrestrained minimization was performed at each λ -state. Each λ -state simulation was run for 4 ns for the complex simulations, and 10 ns for the simulations of the free ligand. The first half of each simulation was used as an additional equilibration phase and analysis was performed on the second half of the simulation. The $\Delta\Delta G$ values were calculated as shown in Figure 1. Hysteresis correction was performed on the closed cycles in the manner described by Wang *et al.*²⁰ with an in-house Python script.

RESULTS AND DISCUSSION

In this study, we aimed to predict the binding strengths of complexes of N2 neuraminidase and the Tamiflu-like ligands using alchemical perturbation methods. As a first step, we trialled multiple simulations protocols for their ability to produce stable neuraminidase complexes with ligands **1** and **2**. Table 3 shows the conditions investigated, which varied the force field, water model, and the protein oligomerization state. From these simulations, we found that modelling the native tetramer was necessary for neuraminidase-ligand complex stability. In models containing only a single protomer, we saw conformational changes in the protein face that would have been stacked against an adjacent monomer, increased 150-loop flexibility and distortion of the active site residues. Specifically, the added flexibility to the 150-loop resulted in an increase in the 150-cavity volume and eventual displacement of the arginine residues in the binding site. The increased volume in the 150-cavity resulted in the displacement of R₂ (blue, Figure 2) from the 150-cavity within 50 ns. This was true for every force field trialled, and was independent of the solvation model chosen. This observation is consistent with studies by Amaro *et al.*³² who proposed that the 150-loop is stabilised by two salt bridges between Asp147 and His150, and between Asp147 and Arg-107 of the adjacent protein monomer. These salt bridges are observed by crystallography²⁰ and were shown by the MD simulations of Amaro *et al.* to intermittently break and reform on a 100 ns time scale. The instability of the 150-loop in the monomer simulations can be rationalized by the loss of the stabilizing interaction between Asp-147 (Chain A) and Arg-107 (Chain D).

We also found that only the combination of the AMBER99SB-ILDN protein force field, the GAFF force field for the ligand and the four-point water TIP4P-ew water model produced stable protein/ligand models. Amaro *et al.*³² previously used the AMBER99SB force field to model the neuraminidase protein binding site. Masukawa *et al.*⁴⁹ and Nguyen *et al.*³¹ have previously observed that retaining active site waters for an MD simulation is important for neuraminidase-ligand complex stability and for accurate binding free energy calculations. In this work, we found that the 3-point water models trialled did not produce stably bound ligands. However, the 4-point TIP4P-Ew model, which is optimized for the PME algorithm,⁴⁷ did produce stable protein/ligand complexes. It is possible that any 4-point water model would afford stably bound active site water molecules by long-timescale MD simulations, however we report the use of TIP4P-Ew here (Table 3).

Table 3. Summary of simulation protocols evaluated to establish a method that gives stable neuraminidase-ligand complexes.

Protein oligomerization state	Protein force field	Ligand force field	Water model	Protein-ligand complex stability
Monomer	GROMOS54a7	ATB	SPC/e	Low ^a
Monomer	OPLS-AA/L	LigParGen	TIP3P	Low
Monomer	CHARMM27	CGenFF	TIP3P	Low
Monomer	AMBER99SB-ILDN ⁴¹	GAFF	TIP3P	Low
Tetramer	GROMOS54a7	ATB	SPC/e	Low
Tetramer	OPLS-AA/L	LigParGen	TIP3P	Low
Tetramer	CHARMM27	CGenFF	TIP3P	Low
Tetramer	AMBER99SB-ILDN	GAFF	TIP3P	Moderate ^b
Tetramer	OPLS-AA/L	LigParGen	TIP4P-Ew	Low
Tetramer	CHARMM27	CGenFF	TIP4P-Ew	Low
Tetramer	AMBER99SB-ILDN	GAFF	TIP4P-Ew	High ^c

^aKey interactions between the protein and ligand break, and ligand dissociates within 50 ns. ^bLigand dissociation occurs between 50 to 350 ns. ^cComplex remains stable for 350 ns.

Figure 4 shows the RMSD of the protein backbone, binding site arginine residues and ligand during the 350 ns simulations of Tamiflu carboxylate (**1**) and GlcA analogue (**2**) bound to N2 neuraminidase. The movement of **1** is measured relative to the crystal structure and the movement of GlcA **2** is measured relative to the synthetic starting model. In both simulations, the protein heavy atom RMSD converged to 1.4 Å RMSD from the original crystal structure (Figure 4A) and the binding site arginine residues were stable, converging to ~1.0 Å RMSD (Figure 4B). In both cases, both the active site arginine cluster (Arg-118, Arg-272 and Arg-371) as well as Arg-224 did not move significantly from the crystal structure. The ligand atoms converged to 0.6 Å and 2.0 Å RMSD from crystal structures, for complexes with **1** and **2**, respectively (Figure 4C).

These results presented in Table 3 and Figure 4 improve on the findings of Ren *et al.*,³⁴ who investigated 16 different protein and water forcefield combinations to determine the best MD parameters for 20 ns simulations subject to MM-PBSA analysis. These authors also reported that the AMBER forcefield variants result in the most accurate binding free energy calculations by MM-PBSA, consistent with data in Table 3 and Figure 4. Ren *et al.* detail the use of monomeric neuraminidase simulation systems with 3-point water models. From the data presented in Table 3, we suggest that Ren *et al.* would observe dissociation of the protein-ligand complexes in their simulation systems on longer timescales than 20 ns. We show here an optimal MD protocol for simulating N2 neuraminidase-ligand complexes.

Figure 4. RMSD relative to crystal pose over simulation time for neuraminidase in complex with Tamiflu carboxylate (blue) and **2** (red). **(A)** Plots of protein heavy atom RMSD for tetrameric neuraminidase. **(B)** RMSD of Arg residues within 5 Å of the bound ligand. **(C)** Ligand RMSD (right).

To determine how well relative alchemical perturbation could predict differences in binding free energy ($\Delta\Delta G$) for neuraminidase ligands, we performed a set of 18 alchemical perturbations for ligands **1-9** (detailed in supplementary information Table S1). Four transformations produced unrealistically large $\Delta\Delta G$ values, which were discarded, leaving set of 14 perturbations connecting 9 compounds. The resulting perturbation map is shown in Figure 5. In performing the calculations, we found that transformations involving the charged ammonium group of compound **1a** were generally problematic. It was, however, possible to transform **1a** into other compounds through intermediate **1b**, which contains a neutral amino group. The transformations formed a number of closed cycles. An in-house Python script identified a total of 15 closed cycles in the perturbation map. Hysteresis correction was performed by cycle closure across the perturbation map using the method described by Wang *et al.*²⁰ and raw values of $\Delta\Delta G$ were corrected to make all closed cycles sum to 0 kJ mol⁻¹ in the statistically most likely way, given the assumption that the error in every perturbation is the same. The cycle with the largest hysteresis was: **1b** → **2** → **4** → **6** → **9** → **1b** where, instead of affording a $\Delta\Delta G$ value of 0 kJ mol⁻¹, the closed cycle had a hysteresis of 84.4 kJ mol⁻¹. This cycle is large and contains transformations between ligands with the chemically different Tamiflu and GlcA-based scaffolds. The hysteresis correction algorithm reduced the hysteresis in each of the 15 closed cycles to 0 kJ mol⁻¹.

Figure 5. Perturbation map for relative alchemical transformations between ligands based on Tamiflu carboxylate (1) and the GlcA cores. Raw $\Delta\Delta G$ values are shown in red text (top value). $\Delta\Delta G$ values corrected by the hysteresis correction algorithm are shown in green (bottom value). Structures with a Tamiflu scaffold are in black and GlcA analogues are in blue. Structures 1a and 1b differ in the protonation state of the amino group.

Figure 6. (A) Raw predicted $\Delta\Delta G$ against experimental $\Delta pIC50$. (B) Predicted $\Delta\Delta G$ values against experimental $\Delta pIC50$ after correcting for hysteresis in the predicted $\Delta\Delta G$ values using the method by Wang *et al.*²⁰ Errors bars representing ± 2 SD in the experimental $IC50$ are shown. Lines of best fit from linear regression are shown as a solid line. The 95% CI limits of the two lines of best fit are shown as dashed lines. The RMSE values represent the average residual in predicted $\Delta pIC50$ from the line of best fit.

The predicted $\Delta\Delta G$ values before and after hysteresis correction, were correlated with experimental $\Delta pIC50$ values (Figure 6). The raw predicted $\Delta\Delta G$ values (before hysteresis correction) have modest predictive value and correlate with $\Delta pIC50$ with Pearson's $R = -0.75$ and Spearman's $\rho = 0.41$ (Figure 6A). By correcting for hysteresis (Figure 6B) the correlation between predicted $\Delta\Delta G$ and $\Delta pIC50$ values improves to $R = -0.96$ and $\rho = 0.86$. Thus, hysteresis correction substantially improved correlation to experiment. The corrected ρ^2 value of 0.74 means that approximately 74 percent of the variance in the rank order of differences in experimental $pIC50$ can be explained by the modelling. Further, the 95% CI in the line of best fit between raw $\Delta\Delta G$ predictions and experimental $\Delta pIC50$ values was considerably broader than when the hysteresis correction was performed (Figure 6A and 6B). A total of six datapoints are outside this 95% CI of the line of best fit in Figure 6A, while two are outside the narrower 95% CI in the line of best fit in Figure 6B. This decrease in both the number of predictions outside of the 95% CI in the line of best fit, and in the range in each line of best fit, indicates the improved predictive power of our model after the hysteresis correction is applied to the raw data. The RMSE between experimental and predicted $\Delta pIC50$ values from our linear regression model was 1.56 and 0.62 when using the raw and corrected data for the line of best fit (Figures 6A and 6B). These RMSE values represent the average residual in $\Delta pIC50$ from the line of best fit. By comparison with experimental errors, the range in the experimental $pIC50$ value for Tamiflu carboxylate (**1**) was 0.46, representing ± 2 SD in the $IC50$ value. Our binding free energy calculations are therefore reliable in predicting which compounds more strongly inhibit neuraminidase in hit-to-lead and lead-optimisation projects, approaching the accuracy of experiment. This level of accuracy is similar to that previously reported for binding free energy calculations with neuraminidase ligands. For example, Mai *et al.*²⁸ used steered MD to estimate ligand binding strength for a similar sample size of neuraminidase ligands and obtained a correlation coefficient of $R = 0.97$ when comparing F_{max} with experimental ΔG .

CONCLUSION

We have established a set of computational conditions that are suitable for modelling ligand binding to influenza virus neuraminidase. It has been shown that neuraminidase must be simulated as the native tetramer for stable complexes. It is also shown that AMBER99SB-ILDN and GAFF force fields, with the TIP4P-Ew water model, afforded the most stable N2 neuraminidase complexes by MD simulation over a 350 ns timescale used here. We used alchemical perturbation with BAR to predict the relative binding affinities ($\Delta\Delta G$) of a set of 9 neuraminidase inhibitors. The raw perturbation data was corrected for hysteresis, which greatly improved the correlation with experimental data. After hysteresis correction, we found a correlation between $\Delta\Delta G$ and $\Delta pIC50$ values, with $R = -0.96$ for a total of 14 transformations. The RMSE in $\Delta pIC50$ values from the linear regression model on the corrected $\Delta\Delta G$ values was 0.62, approaching the range in the $pIC50$ values from repeated experimental measurements, which was 0.46 for Tamiflu carboxylate (given by ± 2 SD in the $IC50$ value). Therefore, the level of error in our models approaches the experimental error of the original assays. Additionally, there are

differences in how previous modelling studies have treated neuraminidase, with force field choices varying as either CHARMM27, OPLS, GROMOS96 or AMBER99SB, and either 3-point or 4-point water models used to model solvent. We report here an optimal MD protocol. These MD methods also have potential applications for homologous glycohydrolases. We anticipate that these methods for molecular dynamics simulations and binding free energy calculations can be applied in the future to assist with the discovery of new drugs targeting neuraminidase-like proteins. Examples where our MD protocol and perturbation methods could potentially be applied are for ligands binding to: the homologous haemagglutinin-neuraminidase (HN) from human parainfluenza virus,⁵⁰⁻⁵² bacterial neuraminidases⁵³ or the putative pseudaminidase reported in 2008 by Xu *et al.*,⁵⁴ which is thought to cleave pseudaminic acid⁵⁵ from Gram negative bacteria cell-surfaces and to produce biofilm.

ACKNOWLEDGMENTS

We acknowledge the MASSIVE and NCI supercomputing facilities for computational resources and thank them for technical support. BJW-N gratefully thanks the Australian government for his research training program (RTP) grant. Discussions with Jeff Dyason and Assoc. Prof. Milton Kiefel of the Griffith University Institute for Glycomics have been very helpful in this work, and we offer them our thanks.

REFERENCES

1. Patel, H.; Kukol, A., Integrating molecular modelling methods to advance influenza A virus drug discovery. *Drug Discov.* **2020**.
2. Copeland, R. A.; Pompliano, D. L.; Meek, T. D., Drug–target residence time and its implications for lead optimization. *Nat. Rev. Drug Discov.* **2006**, *5* (9), 730.
3. Zwanzig, R. W., High - temperature equation of state by a perturbation method. I. nonpolar gases. *J. Chem. Phys.* **1954**, *22* (8), 1420-1426.
4. Williams-Noonan, B. J.; Yuriev, E.; Chalmers, D. K., Free Energy Prediction for Binding of Allosteric Inhibitors of Human Immunodeficiency Virus-1 Integrase: a Comparative Study. *Submitted* **2020**.
5. Williams-Noonan, B. J.; Yuriev, E.; Chalmers, D. K., Free Energy Methods in Drug Design: The Prospects of ‘Alchemical Perturbation’ in Medicinal Chemistry. *J. Med. Chem.* **2017**.
6. Perez-Benito, L.; Casajuana-Martin, N.; Jimenez-Roses, M.; van Vlijmen, H.; Tresadern, G., Predicting Activity Cliffs with Free-Energy Perturbation. *J. Chem. Theory Comput.* **2019**, *15* (3), 1884-1895.
7. Roos, K.; Wu, C.; Damm, W.; Reboul, M.; Stevenson, J. M.; Lu, C.; Dahlgren, M. K.; Mondal, S.; Chen, W.; Wang, L., OPLS3e: Extending force field coverage for drug-like small molecules. *J. Chem. Theory Comput.* **2019**, *15* (3), 1863-1874.
8. Taubenberger, J. K.; Morens, D. M., 1918 Influenza: the mother of all pandemics. *Revista Biomed.* **2006**, *17* (1), 69-79.
9. Alexander, D., Avian influenza viruses and human health. *Dev. Biol.* **2006**, *124*, 77-84.
10. von Itzstein, M.; Wu, W.-Y.; Kok, G. B.; Pegg, M. S.; Dyason, J. C.; Jin, B.; Van Phan, T.; Smythe, M. L.; White, H. F.; Oliver, S. W., Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* **1993**, *363* (6428), 418.

11. Varghese, J. N.; Epa, V. C.; Colman, P. M., Three-dimensional structure of the complex of 4-guanidino-Neu5Ac2en and influenza virus neuraminidase. *Protein science : a publication of the Protein Society* **1995**, *4* (6), 1081-7.
12. Rohloff, J. C.; Kent, K. M.; Postich, M. J.; Becker, M. W.; Chapman, H. H.; Kelly, D. E.; Lew, W.; Louie, M. S.; McGee, L. R.; Prisbe, E. J., Practical total synthesis of the anti-influenza drug GS-4104. *J. Org. Chem.* **1998**, *63* (13), 4545-4550.
13. Sidwell, R. W.; Huffman, J. H.; Barnard, D. L.; Bailey, K. W.; Wong, M.-H.; Morrison, A.; Syndergaard, T.; Kim, C. U., Inhibition of influenza virus infections in mice by GS4104, an orally effective influenza virus neuraminidase inhibitor. *Antivir. Res.* **1998**, *37* (2), 107-120.
14. Li, W.; Escarpe, P. A.; Eisenberg, E. J.; Cundy, K. C.; Sweet, C.; Jakeman, K. J.; Merson, J.; Lew, W.; Williams, M.; Zhang, L., Identification of GS 4104 as an orally bioavailable prodrug of the influenza virus neuraminidase inhibitor GS 4071. *Antimicrob. Agents Chemother.* **1998**, *42* (3), 647-653.
15. Lew, W.; Chen, X.; Kim, C. U., Discovery and development of GS4104 (oseltamivir) an orally active influenza neuraminidase inhibitor. *Curr. Med. Chem.* **2000**, *7* (6), 663-672.
16. Webster, R. G.; Bean, W. J.; Gorman, O. T.; Chambers, T. M.; Kawaoka, Y., Evolution and ecology of influenza A viruses. *Microbiol. Mol. Biol. Rev.* **1992**, *56* (1), 152-179.
17. Yoon, S.-W.; Webby, R. J.; Webster, R. G., Evolution and ecology of influenza A viruses. In *Influenza Pathogenesis and Control-Volume I*, Springer: 2014; pp 359-375.
18. Colman, P. M.; Varghese, J.; Laver, W., Structure of the catalytic and antigenic sites in influenza virus neuraminidase. *Nature* **1983**, *303* (5912), 41.
19. Bhatt, B.; Böhm, R.; Kerry, P. S.; Dyason, J. C.; Russell, R. J.; Thomson, R. J.; von Itzstein, M., Exploring the interactions of unsaturated glucuronides with influenza virus sialidase. *J. Med. Chem.* **2012**, *55* (20), 8963-8968.
20. Wang, L.; Deng, Y.; Knight, J. L.; Wu, Y.; Kim, B.; Sherman, W.; Shelley, J. C.; Lin, T.; Abel, R., Modeling local structural rearrangements using FEP/REST: application to relative binding affinity predictions of CDK2 inhibitors. *J. Chem. Theory Comput.* **2013**, *9* (2), 1282-93.
21. Varghese, J. N.; McKimm - Breschkin, J. L.; Caldwell, J. B.; Kortt, A. A.; Colman, P. M., The structure of the complex between influenza virus neuraminidase and sialic acid, the viral receptor. *Proteins* **1992**, *14* (3), 327-332.
22. Gubareva, L. V.; Webster, R. G.; Hayden, F. G., Comparison of the activities of zanamivir, oseltamivir, and RWJ-270201 against clinical isolates of influenza virus and neuraminidase inhibitor-resistant variants. *Antimicrob. Agents Chemother.* **2001**, *45* (12), 3403-3408.
23. Wang, K.; Yang, F.; Wang, L.; Liu, K.; Sun, L.; Lin, B.; Hu, Y.; Wang, B.; Cheng, M.; Tian, Y., Synthesis and biological evaluation of NH₂-acyl oseltamivir analogues as potent neuraminidase inhibitors. *Eur. J. Med. Chem.* **2017**, *141*, 648-656.
24. Taylor, N. R.; von Itzstein, M., Molecular Modeling Studies on Ligand Binding to Sialidase from Influenza Virus and the Mechanism of Catalysis. *J. Med. Chem.* **1994**, *37* (5), 616-624.
25. Lawrenz, M.; Baron, R.; McCammon, J. A., Independent-trajectories thermodynamic-integration free-energy changes for biomolecular systems: determinants of H5N1 avian influenza virus neuraminidase inhibition by peramivir. *J. Chem. Theory Comput.* **2009**, *5* (4), 1106-1116.
26. Lawrenz, M.; Wereszczynski, J.; Amaro, R.; Walker, R.; Roitberg, A.; McCammon, J. A., Impact of calcium on N1 influenza neuraminidase dynamics and binding free energy. *Proteins* **2010**, *78* (11), 2523-2532.
27. Kiso, M.; Kubo, S.; Ozawa, M.; Le, Q. M.; Nidom, C. A.; Yamashita, M.; Kawaoka, Y., Efficacy of the new neuraminidase inhibitor CS-8958 against H5N1 influenza viruses. *PLOS Pathog.* **2010**, *6* (2), e1000786.

28. Mai, B. K.; Li, M. S., Neuraminidase inhibitor R-125489—a promising drug for treating influenza virus: steered molecular dynamics approach. *Biochem. Biophys. Res. Commun.* **2011**, *410* (3), 688-691.
29. Yamashita, M.; Tomozawa, T.; Kakuta, M.; Tokumitsu, A.; Nasu, H.; Kubo, S., CS-8958, a prodrug of the new neuraminidase inhibitor R-125489, shows long-acting anti-influenza virus activity. *Antimicrob. Agents Chemother.* **2009**, *53* (1), 186-192.
30. Watanabe, A.; Chang, S.-C.; Kim, M. J.; Chu, D. W.-s.; Ohashi, Y.; Group, M. S., Long-acting neuraminidase inhibitor laninamivir octanoate versus oseltamivir for treatment of influenza: a double-blind, randomized, noninferiority clinical trial. *Clin. Infect. Dis.* **2010**, *51* (10), 1167-1175.
31. Nguyen, T. T.; Viet, M. H.; Li, M. S., Effects of water models on binding affinity: evidence from all-atom simulation of binding of tamiflu to A/H5N1 neuraminidase. *Sci. World J.* **2014**, *2014*.
32. Amaro, R. E.; Swift, R. V.; Votapka, L.; Li, W. W.; Walker, R. C.; Bush, R. M., Mechanism of 150-cavity formation in influenza neuraminidase. *Nat. Commun.* **2011**, *2*, 388.
33. Jusoh, N.; Zainal, H.; Hamid, A. A. A.; Bunnori, N. M.; Abd Halim, K. B.; Abd Hamid, S., In silico study of carvone derivatives as potential neuraminidase inhibitors. *J. Mol. Model.* **2018**, *24* (4), 1-13.
34. Ren, J.; Yuan, X.; Li, J.; Lin, S.; Yang, B.; Chen, C.; Zhao, J.; Zheng, W.; Liao, H.; Yang, Z., Assessing the performance of the g_mmpbsa tools to simulate the inhibition of oseltamivir to influenza virus neuraminidase by molecular mechanics Poisson–Boltzmann surface area methods. *J. Chin. Chem. Soc.* **2020**, *67* (1), 46-53.
35. Hanpaibool, C.; Leelawiwat, M.; Takahashi, K.; Rungrotmongkol, T., Source of oseltamivir resistance due to single E119D and double E119D/H274Y mutations in pdm09H1N1 influenza neuraminidase. *J. Comput. Aided Mol. Des.* **2020**, *34* (1), 27-37.
36. Mahalapbutr, P.; Sangkhawasi, M.; Kammarabutr, J.; Chamni, S.; Rungrotmongkol, T., Rosmarinic acid as a potent influenza neuraminidase inhibitor: In vitro and in silico study. *Curr. Top. Med. Chem.* **2020**.
37. Kato, Y.; Takahashi, K.; Ito, F.; Suzuki, S.; Fukui, K.; Mimaki, M.; Suzuki, K., Novel oseltamivir-resistant mutations distant from the active site of influenza B neuraminidase. *J. Biomol. Struct. Dyn.* **2020**, 1-10.
38. Bell, J.; Cao, Y.; Gunn, J.; Day, T.; Gallicchio, E.; Zhou, Z.; Levy, R.; Farid, R., PrimeX and the Schrödinger computational chemistry suite of programs. 2006; pp 534-538.
39. Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G., PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185* (2), 604-613.
40. Abraham, M.; van der Spoel, D.; Lindahl, E.; Hess, B., the GROMACS development team, GROMACS User Manual version 5.1. 2, 2016.
41. Lindorff - Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E., Improved side - chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78* (8), 1950-1958.
42. Sprenger, K.; Jaeger, V. W.; Pfaendtner, J., The general AMBER force field (GAFF) can accurately predict thermodynamic and transport properties of many ionic liquids. *J. Phys. Chem. B* **2015**, *119* (18), 5882-5895.
43. de Leeuw, S. W.; Perram, J. W.; Smith, E. R. In *Simulation of electrostatic systems in periodic boundary conditions. I. Lattice sums and dielectric constants*, Proc. Royal Soc. Lond. A, The Royal Society: 1980; pp 27-56.
44. Verlet, L., Computer “experiments” on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.* **1967**, *159* (1), 98.

45. Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G., LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18* (12), 1463-1472.
46. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98* (12), 10089-10092.
47. Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T., Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120* (20), 9665-9678.
48. Loeffler, H. H.; Michel, J.; Woods, C., FESetup: automating setup for alchemical free energy simulations. *J. Chem. Inf. Model.* **2015**.
49. Masukawa, K. M.; Kollman, P. A.; Kuntz, I. D., Investigation of Neuraminidase-Substrate Recognition Using Molecular Dynamics and Free Energy Calculations. *J. Med. Chem.* **2003**, *46* (26), 5628-5637.
50. Spriggs, M.; Murphy, B.; Prince, G.; Olmsted, R.; Collins, P., Expression of the F and HN glycoproteins of human parainfluenza virus type 3 by recombinant vaccinia viruses: contributions of the individual proteins to host immunity. *J. Virol.* **1987**, *61* (11), 3416-3423.
51. Winger, M.; von Itzstein, M., Exposing the flexibility of human parainfluenza virus hemagglutinin-neuraminidase. *J. Am. Chem. Soc.* **2012**, *134* (44), 18447-18452.
52. Guillon, P.; Dirr, L.; El-Deeb, I. M.; Winger, M.; Bailly, B.; Haselhorst, T.; Dyason, J. C.; Von Itzstein, M., Structure-guided discovery of potent and dual-acting human parainfluenza virus haemagglutinin–neuraminidase inhibitors. *Nature Commun.* **2014**, *5*, 5268.
53. Gaskell, A.; Crennell, S.; Taylor, G., The three domains of a bacterial sialidase: a β -propeller, an immunoglobulin module and a galactose-binding jelly-roll. *Structure* **1995**, *3* (11), 1197-1205.
54. Xu, G.; Ryan, C.; Kiefel, M. J.; Wilson, J. C.; Taylor, G. L., Structural studies on the *Pseudomonas aeruginosa* sialidase-like enzyme PA2794 suggest substrate and mechanistic variations. *J. Mol. Biol.* **2009**, *386* (3), 828-840.
55. Zunk, M.; Williams, J.; Carter, J.; Kiefel, M. J., A new approach towards the synthesis of pseudaminic acid analogues. *Org. Biomol. Chem.* **2014**, *12* (18), 2918-2925.