

SCIENTIFIC REPORTS



OPEN

Comparative studies of catalytic pathways for *Streptococcus pneumoniae* sialidases NanA, NanB and NanC

Kela Xiao^{1,2}, Xingyong Wang^{1,2}  & Haibo Yu^{1,2,3} 

Streptococcus pneumoniae (*S. pneumoniae*) is a leading human pathogen, which takes large responsibility for severe otitis media, acute meningitis and septicaemia. It encodes up to three distinct sialidases: NanA, NanB and NanC, which are promising drug targets. Recent experimental studies have shown that these three sialidases might work together up to the ultimate step, where NanA and NanB produce *N*-acetylneuraminic acid (Neu5Ac) and 2,7-anhydro-Neu5Ac following the functions of sialidase and intramolecular *trans*-sialidase, whilst NanC carries on a ping-pong mechanism that produces or removes 2-deoxy-2,3-didehydro-Neu5Ac. It is intriguing that these sialidases have similar active sites but operate via three distinct reaction pathways. To clarify this issue, herein we present the first systematic computational investigation on the catalytic pathways for *S. pneumoniae* NanA, NanB and NanC based on combined quantum mechanics/molecular mechanics simulations, and propose the most preferred routes for the three *S. pneumoniae* sialidases. Our findings support the mechanisms of NanA and NanC that were proposed by previous experimental studies, whereas the role of water in NanB was found to differ slightly from our current understandings. The mechanistic insights obtained from this work are expected to assist in the design of potent inhibitors targeting these key enzymes for therapeutic applications.

S. pneumoniae (pneumococcus) is a Gram-positive, α -haemolytic, facultative anaerobic organism that resides in the nasopharynx. *S. pneumoniae* is largely responsible for upper/lower-respiratory infections, meningitis, and septicaemia, and operates in a number of fairly major diseases with a considerably high mortality and incidence¹. The occurrence rate of life-threatening pneumococcal disease in human, especially in children under 5 years of age is higher than HIV, malaria, and measles combined². It has been shown that *S. pneumoniae* encodes up to three distinct neuraminidases (NAs, also called sialidases): NanA, NanB and NanC^{3–5}. Clinical research of pneumococcal isolates of *S. pneumoniae* points out that these three sialidases are present in 100%, 96% and 51% of strains, respectively³. NanA and NanB have been shown to play a crucial role in respiratory tract infection and sepsis¹. Moreover, considerable evidences indicate that NanA is essential to middle ear invasion, nasopharynx and the incidence of otitis media effusion⁶. *S. pneumoniae* sialidases have been shown to have great potential as drug targets or vaccine candidates⁷. Recent studies have shown that the human upper respiratory epitheliums express α 2,6-sialylated glycans while many human bronchial epithelial cells possess α 2,3-linked sialic acid substrates^{8,9}. NanA would mainly be responsible for the sialic acid removal⁴, while NanB would be essential for the survival of the colonised bacteria^{4,9}, and NanC was suggested to be a regulator of NanA^{4,10}. All these findings indicate that the three *S. pneumoniae* sialidases might play distinct roles in pneumococcal virulence.

Sequence identity between NanB and NanC is the greatest (over 50%), while it is limited between NanA and NanB (up to 24%). The crystal structures of NanA, NanB and NanC have been solved and demonstrate conservation of their catalytic domains^{10–13}. Previous studies¹² have suggested NanA to be a typical hydrolytic sialidase that cleaves α 2,3-, α 2,6- and α 2,8-linked sialic acids, and produces *N*-acetylneuraminic acid (α -Neu5Ac). NanB acts as an intramolecular (IT) *trans*-sialidase that preferentially cleaves α 2,3-linked sialic acid substrates

¹School of Chemistry and Molecular Bioscience, University of Wollongong, Wollongong, NSW, 2500, Australia.

²Molecular Horizons, University of Wollongong, Wollongong, NSW, 2500, Australia. ³Illawarra Health and Medical Research Institute, Wollongong, NSW, 2500, Australia. Correspondence and requests for materials should be addressed to H.Y. (email: hyu@uow.edu.au)

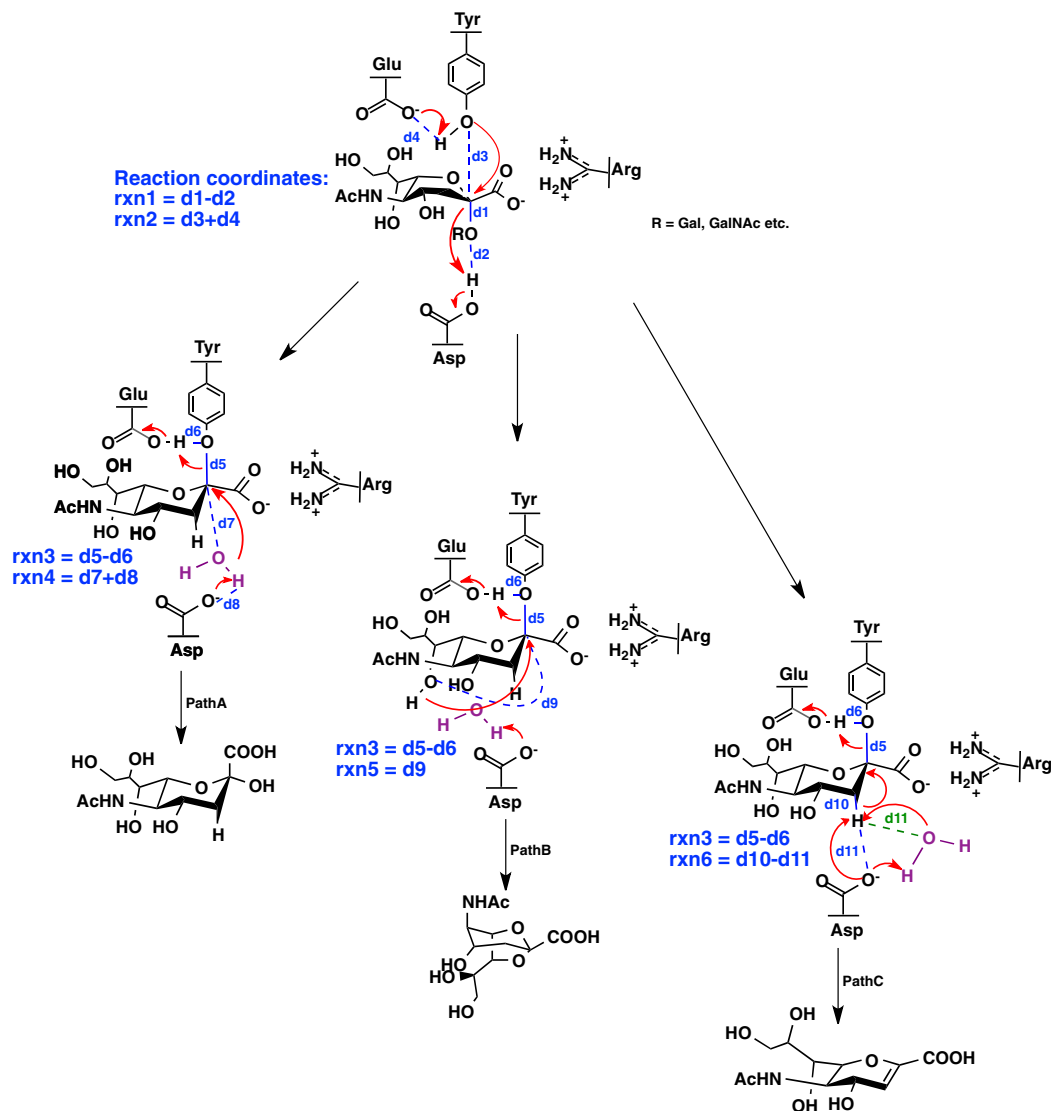


Figure 1. The three predefined pathways for NanA, NanB and NanC. For the first step, d1 refers to the distance between C2 position of the sugar ring and the glycosidic O of methyl group, d2 is the distance between the glycosidic O of methyl group and the aspartic acid H δ 2, d3 describes the distance between the tyrosine O η and C2 position of the sugar ring, and d4 is the one between the tyrosine H η and the glutamic acid O ϵ 2. For the second steps, d5 is the distance between the tyrosine O η and C2 position of the substrate, d6 corresponds to the one between the tyrosine O η and H η , d7 is the one between C2 position of the sugar ring and the catalytic water oxygen, d8 is the one between aspartic acid residue O ϵ 2 and catalytic water hydrogen, d9 defines the one between C2 position and O7 of the sugar ring, and d10 refers to the one between C3 and H32 in substrate. For NanB and NanC, d11 (labelled in blue) accounts for that between H32 of the sugar ring and the aspartic residue O ϵ 2. However, for NanA, due to the reaction is proceeded with an activated water molecule nearby because the more accessible the catalytic cavity in NanA allows more water molecules to reside, d11 (labelled in green) is defined by the distance between H32 of the sugar ring and catalytic water oxygen. The reaction coordinates rxn1-rxn6 are also illustrated in the figure. Tyr, Glu, Asp: Tyr752, Glu647, Asp372 in NanA; Tyr653, Glu541, Asp270 in NanB; Tyr695, Glu584, Asp315 in NanC.

to release 2,7-anhydro-Neu5Ac^{12,13}. NanC is shown to be specific for α 2,3-linked sialic acids, producing the primary product 2-deoxy-2,3-didehydro-N-acetyl- neuraminic acid (Neu5Ac2en) that can be hydrolysed to Neu5Ac by attacking C2 of the oxocarbenium ion¹⁰. The reaction mechanisms for the three sialidases NanA, NanB and NanC in *S. pneumoniae* have been proposed by previous experimental studies based on X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, kinetic analysis and site-directed mutagenesis (Fig. 1)^{4,10-15}. For the conserved catalytic residues shared by these three sialidases, the carboxylate group of sialic acids interacts with the tri-arginine clusters, and the tyrosine residue close to the glutamic acid acts as a nucleophile, while an aspartic residue acts as an acid/base (Fig. 1). The first step shows a common reaction pathway, whereas the second step exhibits intriguing differences. The C2 position of the oxocarbenium ion is attacked by a water molecule, the sialyl cation O7 atom, and the carbohydrate acceptor for NanA, NanB and NanC, respectively. Subsequent

products (Neu5Ac, 2,7-anhydro-Neu5Ac and Neu5Ac2en) are then formed. The general picture is consistent with other retaining glycosidases¹⁶. It is further supported by the captured sialyl-enzyme covalent intermediate on the conserved tyrosine residue in trans-sialidase TcTS and the bacterial sialidase *C. perfringens* NanI^{17–19}. The experimental findings^{4,10} indicated that these three sialidases might work together up to the ultimate step, where NanA and NanB produce Neu5Ac and 2,7-anhydro-Neu5Ac following the functions of sialidase and IT *trans*-sialidase, whilst NanC could carry on a ping-pong mechanism that produces or removes Neu5Ac2en.

It is intriguing that these three different sialidases have similar active sites but operate via three distinct reaction pathways⁴. To fully understand the molecular factors dictating their respective catalyses, we explored the hydrolysis mechanisms of NanA, NanB and NanC by systematically comparing the energy profiles of all possible reaction pathways on the basis of combined quantum mechanics/molecular mechanics (QM/MM) simulations. Based on their structural and sequence homology to other well-studied (trans)sialidases, it is believed that their first steps share a common reaction pathway, while the second steps present intriguing differences which produce three different products, supported by NMR and thiobarbituric assay studies⁴. Therefore, we focus on the reaction barriers for the second step. The most preferred routes have been determined for all three NAs. In addition, based on the combined QM/MM simulations of the intermediate state of NanA, a detailed analysis on the active site structures and dynamics including water has been carried out. These computational results, together with previous experimental findings, thus provide a better understanding of the details on the catalytic mechanisms of these three *S. pneumoniae* sialidases.

Results and Discussions

Equilibrated structures from MM and QM/MM simulations of RC. The equilibrated structures of the reactant complexes (RC) from SCC-DFTB/MM simulations are shown in Fig. 2. The aspartic acid residues (NanA, Asp372; NanB, Asp270; NanC, Asp315) are protonated, and the glutamic acids (NanA, Glu647; NanB, Glu541; NanC, Glu584) are deprotonated in RC. The sugar rings are in a boat conformation (²B₅). For the key residues and substrates, the heavy atom positional root-mean-square-deviations (RMSD) from the crystal structures were approximately 1.2 Å over the simulation time. Also, classical MM simulations were carried out for RC to probe the key interactions. Encouragingly, both MM and SCC-DFTB/MM simulations provide a consistent structure that is suitable for the study of the hydrolysis mechanism. The final snapshots from these QM/MM simulations (Supporting Information Table 1) were taken as the initial structures for minimal energy pathway (MEP) calculations.

Potential energy profiles for NanA, NanB and NanC along three different reaction pathways. We assumed that NanA, NanB and NanC could produce three promiscuous products by following three distinct reaction pathways. A comparative minimal energy pathway (MEP) simulation was performed along the hypothetical reaction pathways presented in Fig. 1. The reaction coordinates used to drive the reactions are also shown in Fig. 1.

The first step in NanA, NanB and NanC. The MEP calculations were carried out along those predefined reaction coordinates. Iterative cycles scanning from the reactant state to the product state forward and backward were performed to achieve convergence. For the first step of NanA, NanB and NanC, rxn1 was harmonically restrained from -0.12 to 2.44 Å, -0.13 to 2.51 Å and -0.85 to 2.5 Å respectively, and rxn2 was scanned from 4.86 to 2.30 Å, 4.34 to 2.26 Å and 5.29 to 2.33 Å, respectively. These two reaction coordinates drive the nucleophilic attack to the anomeric carbons of the substrates and the proton transfer from tyrosines to glutamic acids (NanA, Tyr752 and Glu647; NanB, Tyr653 and Glu541; NanC, Tyr695 and Glu584). The scans were constructed with a harmonic force constant of $2,000.0$ kcal/mol/Å² in steps of 0.08 Å. A total number of 257, 209 and 347 energy points have been obtained in the 2D MEP calculations for the first step in NanA, NanB and NanC, respectively. The contour plots of the energy profiles are shown in Fig. 3, describing the reaction pathways from RC to the intermediate complex (IC) in NanA, NanB and NanC. The white dashed lines highlighted the minimum energy paths that estimate the most likely pathways. Regarding the energetics, the reaction barriers for the first steps are 13.4, 12.3 and 12.5 kcal/mol for NanA, NanB and NanC, respectively, and the path is downhill to the IC once the transition complex of the first reaction step (TS1) is reached, indicating a nucleophilic attack to the anomeric carbon of the substrate and a proton transfer from tyrosines to glutamic acids (NanA, Tyr752 and Glu647; NanB, Tyr653 and Glu541; NanC, Tyr695 and Glu584).

The second step in NanA, NanB and NanC. Subsequently equilibrium SCC-DFTB/MM simulations were performed on the IC structure obtained from the MEPs of the first step (Supporting Information Table 1). Similarly, iterative cycles MEP scanning were performed based on the equilibrated IC structures until convergence was achieved. We investigated the second steps of all the three pathways (PathA, PathB and PathC) for NanA, NanB and NanC, respectively. The setup for reaction coordinates rxn3 to rxn6, as well as the total number of energy points scanned, are listed in Supporting Information Table 2. The scans were conducted with a 0.08 Å step using a harmonic force constant of $2,000.0$ kcal/mol/Å².

The potential energy profiles for the second steps along PathA (PathB/PathC) of NanA (NanB/NanC) are shown in Fig. 4 and Supporting Information Fig. 3. The white dashed lines display the minimum energy paths from IC to the transition complexes involved in the second steps (TS2) and then to the product complex (PC). The related reaction barriers for each path of NanA, NanB and NanC are summarised in Fig. 5. It is evident that PathA for NanA, PathB for NanB and PathC for NanC are most favourable, which is consistent with previous proposals. Interestingly, the reaction barriers difference between PathA and PathC for NanA is not large (~ 0.7 kcal/mol), implying PathC might be a possible pathway for NanA while PathA is more thermodynamically favourable (18.5 kcal/mol vs 24.4 kcal/mol, Fig. 5a).

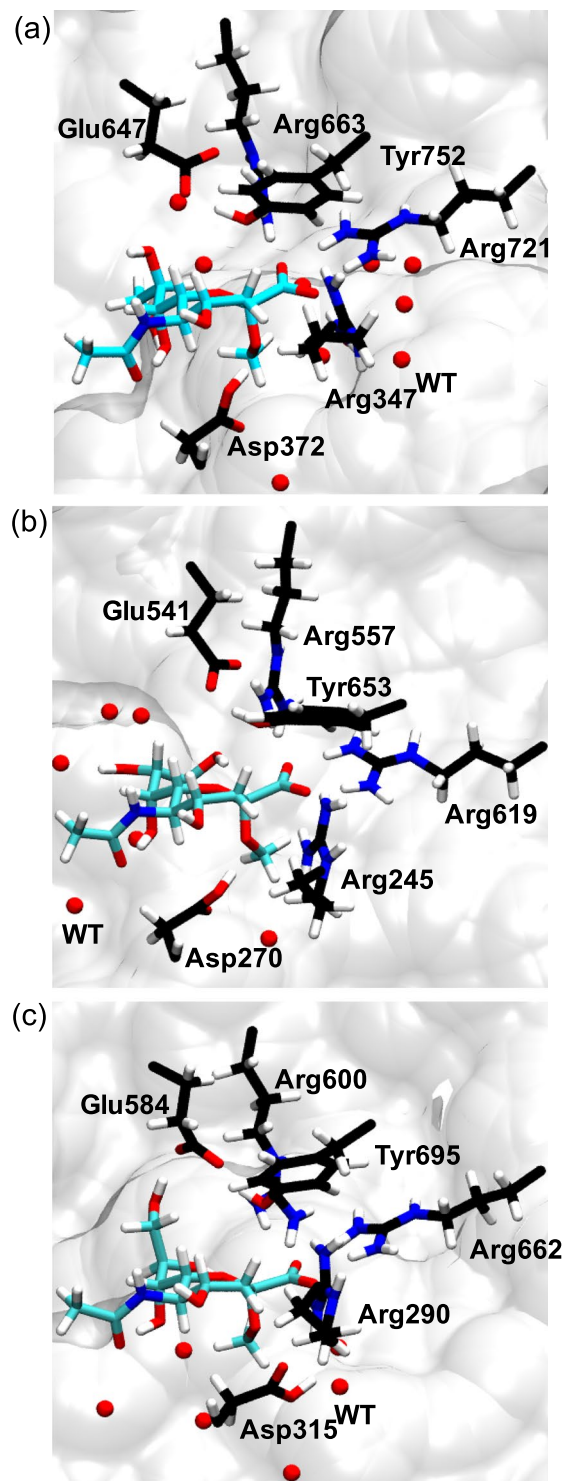


Figure 2. The final snapshots from the explicit solvent SCC-DFTB/MM simulations. The key residues and substrate are shown explicitly. The aliphatic hydrogen atoms are omitted for clarity. The water molecules within approximately 5 Å of O ϵ 2 of Glu647, Glu541 and Glu695 are shown in red, respectively. (a) NanA; (b) NanB; (c) NanC.

It is worth noting that PathB for NanB slightly differs from that proposed in the literature^{4,10}, in which a proton relay through a catalytic water is involved. However, our simulations indicate an alternative pathway (Fig. 5b) without a water is more favourable than that with a catalytic water (for comparison see Supporting Information Fig. 9). This can be rationalised by the relatively hydrophobic environment of its active site (Fig. 2).

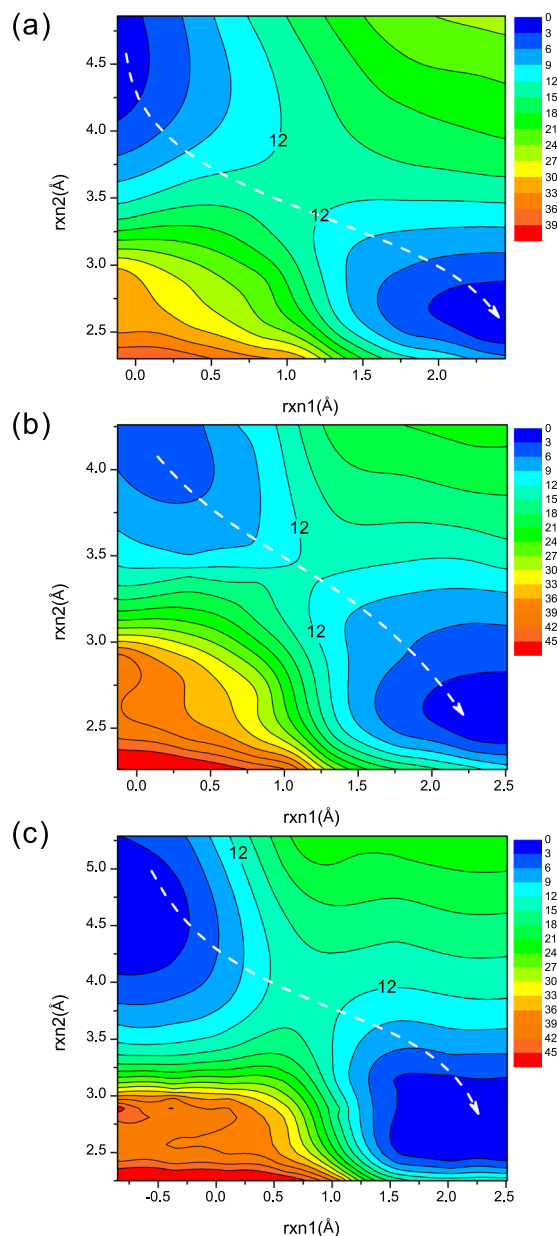


Figure 3. Potential energy contour plot for the first reaction step of (a) NanA, (b) NanB, and (c) NanC. Energies are in kcal/mol. The white dashed line illustrates the minimum potential energy path.

Description of the mechanisms for NanA, NanB and NanC from the reactant state to the product state.

The MEPs for the three different reaction pathways, PathA, PathB and PathC of NanA, NanB and NanC all proceed diagonally, indicating an $A_N D_N$ type mechanism. This represents a concerted bimolecular “ S_N2 ” displacement²⁰, where the reaction coordinates rxn1 and rxn2, rxn3 and rxn4, rxn3 and rxn5, rxn3 and rxn6 are highly correlated. The $A_N D_N$ reaction can be classified into either an associative or a dissociative mechanism according to the relationship between attacking and departing groups, such as the nucleophilic attack by tyrosine (Tyr752 in NanA, Tyr653 in NanB and Tyr695 in NanC) and the departure of the methyl group in the first step. The 2D More O’Ferrall-Jencks style diagram²¹ was depicted, aiming to provide relevant information regarding the mechanistic scenario for the IC and TS formations which has been proved to be a controversial subject of the catalytic mechanism of glycosidases²². In addition, from the 2D More O’Ferrall-Jencks style diagram, we can gain useful information about their hydrolysis mechanisms related to their different reaction pathways (e.g. associative or dissociative $A_N D_N$ mechanism, and how they achieve their different PCs through different TS structures). The Pauling bond order²³ was used to represent the sialic acid bond making and breaking, which was determined as follows

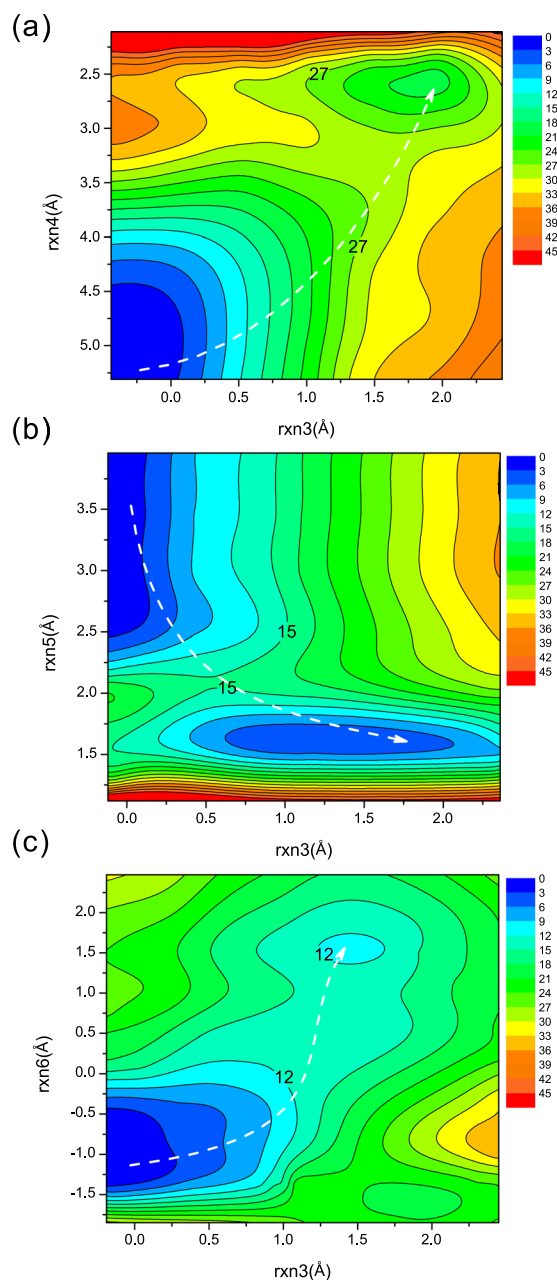


Figure 4. Potential energy contour plot for the second step along (a) PathA in NanA, (b) PathB in NanB, and (c) PathC in NanC. Energies are in kcal/mol. The white dashed line illustrates the minimum potential energy path.

$$n_x = n_0 \exp \left[\frac{(R_0 - R_x)}{0.6} \right] \quad (1)$$

where R_0 is the reference bond length (R_0 equals to 1.4 Å and 1.09 Å for C–O and C–H, respectively), whose bond order is defined as n_0 (=1), and R_x denotes the instantaneous distance between two atoms of interest. Fig. 6a and Supporting Information Figs 18–19 present the More O’Ferrall–Jencks plots for the critical points for the formation of IC during the sialidase activity of NanA, NanB and NanC. The axes represent the Pauling bond orders for the tyrosine (Tyr752 in NanA, Tyr653 in NanB and Tyr695 in NanC) O η –sialyl bond making and O (leaving group)–sialyl bond breaking. It is clear that the leaving methyl group departs substantially when the TS1 is reached, indicating that the first steps of the three systems follow a dissociative A_ND_N mechanism.

The axes in the More O’Ferrall–Jencks plots for PC formation (Fig. 6b and Supporting Information Figs 18–19) denote the Pauling bond orders from C2 at the axial position to the catalytic water or the tyrosine residues (Tyr752 in NanA, Tyr653 in NanB and Tyr695 in NanC). For all the three cases along PathA (6b and Supporting Information Figs 18–19), the tyrosine O η –sialyl bond has already been cleaved while the O (water)–sialyl bond

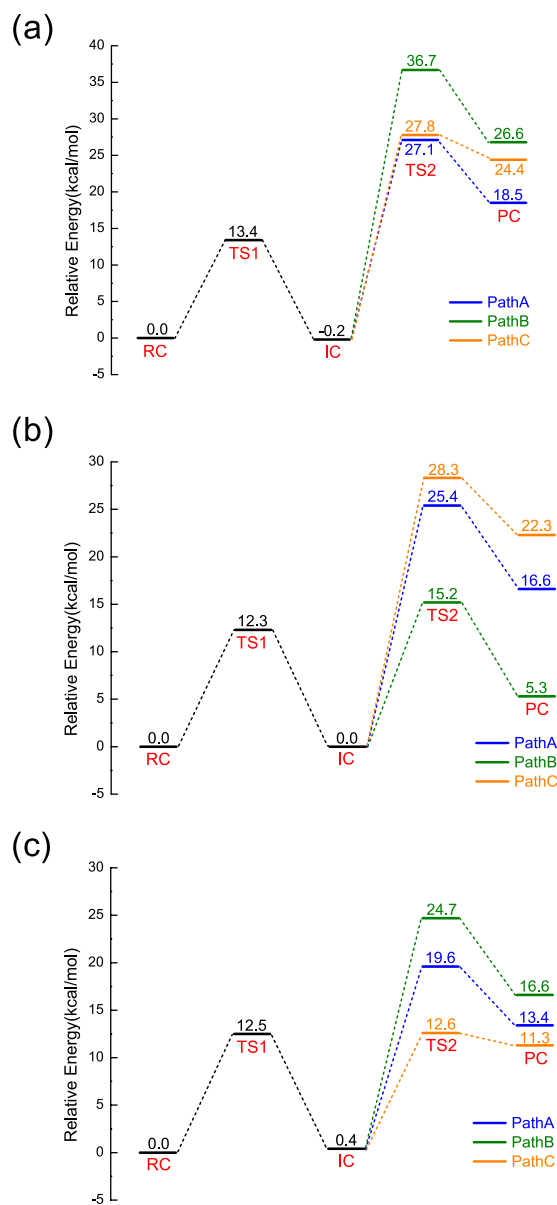


Figure 5. Potential energy profiles (in kcal/mol) for (a) NanA, (b) NanB, and (c) NanC along different reaction pathways.

formation is beginning, indicating that the hydrolysis reaction of the IC along PathA involves a dissociative $A_N D_N$ mechanism. Similarly, for the three systems in PathB, the tyrosine $O\eta$ -sialyl bond has been cleaved before the formation of the $O7$ (sialyl)-C2 (sialyl) bond (Supporting Information Fig. 19). However, unlike the reaction along PathB of NanA, where a proton has transferred from the catalytic water to the Asp372 while the water molecule is protonated by HO7 of the sialyl cation (Supporting Information Figure 18), no catalytic water molecules were observed to be involved in the reactions of NanB and NanC. Interestingly, the HO7 of sialyl cation is transferring or has transferred to the aspartic acid residues (Asp270 in NanB and Asp315 in NanC). Supporting Information Figs 18–19 reveal that in PathC, the tyrosine $O\eta$ -sialyl bond is broken before TS2 is reached. At this point, the proton transfer between the catalytic water molecule and the Asp372 of NanA is about to occur, and the proton transfer between the aspartic acid residue (Asp270 in NanB and Asp315 in NanC) and H32 of the sialyl cation has happened. Taking into account all of the above results, it can be concluded that the $A_N D_N$ mechanisms involved in all pathways are dissociative, indicating that the loss of bonding in the TS has progressed further than bond making. This dissociative mechanism has been observed in the computational study of the catalytic mechanisms of TcTS and TrSA^{24,25}. It is worth noting that the proton transfers between the glutamic acids and the tyrosines are not fully completed at the PC states along all the pathways. The subsequent reactions are provided in Supporting Information Figs 18–19. We will focus on the energy profiles shown above in the following context as we are particularly interested in the different nucleophilic attacks at the C2 position of the sialic acid along the three reaction pathways and the underlying molecular factors.

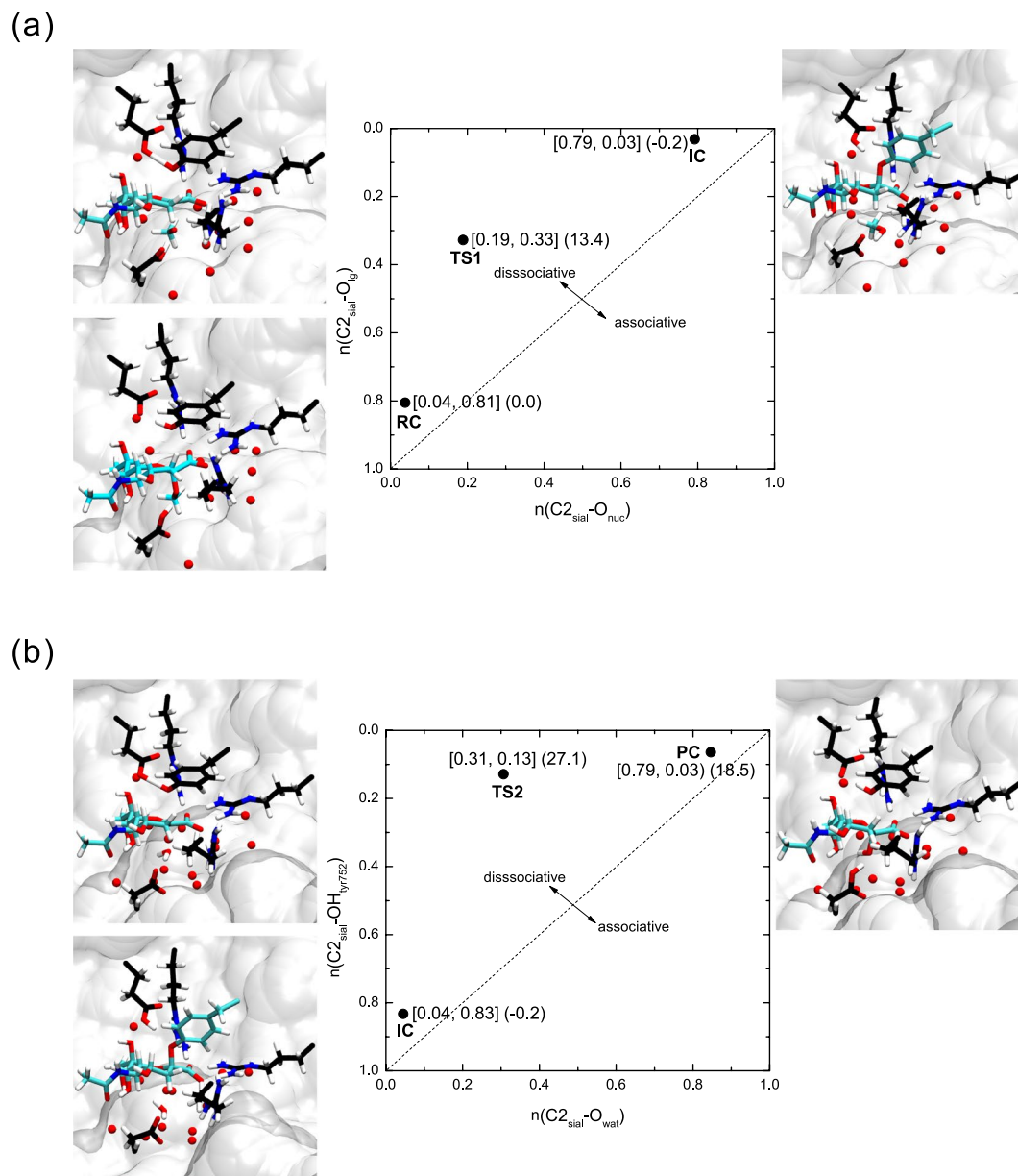


Figure 6. The More O'Ferrall-Jencks plot for the (a) IC and (b) PC formation of NanA. The axes are the Pauling bond orders from the anomeric C to the nucleophile ($n(\text{C2}_{\text{sial}}-\text{O}_{\text{nuc}})$) and the leaving group ($n(\text{C2}_{\text{sial}}-\text{O}_{\text{lg}})$). The RC, TS1, and IC configurations are shown in the bottom left, top left, and top right, respectively. See Fig. 2 for more information. The water molecules are shown in red with CPK.

The values for the distances at the critical points are presented in (Supporting Information Tables 3–8). At TS1, for all the three systems, the proton transfer between the tyrosine and the glutamic acid residue is almost completed ($d_4 \approx 1.18 \text{ \AA}$, 1.14 \AA and 1.21 \AA , respectively), and the tyrosine hydroxyl oxygen is moving closer to the anomeric carbon of sialic acid. The tyrosines in the three cases are more negatively charged, which enhances their nucleophilicity. Unlike the traditional carboxylate nucleophile, it has been argued that Tyr752 here offers an easier electrostatic approach to the anomeric carbon of sialic acid due to the more favourable electrostatic interaction with the negatively charged substrate^{19,24,26}. The distance from the hydroxyl oxygen of the nucleophilic tyrosine to the C2 position of the substrate in NanB is shorter than those in NanA and NanC by 0.16 and 0.40 \AA , respectively, making Tyr653 of NanB a better nucleophile (see details in Supporting Information Table 6). Additionally, the tyrosine $\text{O}\eta$ -sialyl has already been cleaved ($d_1 \approx 2.07 \text{ \AA}$, 2.24 \AA and 2.10 \AA , respectively) and the aspartic acid residue proton has transferred to the methyl group ($d_2 \approx 1.22 \text{ \AA}$, 1.05 \AA and 1.02 \AA , respectively). Significant conformational changes of the sugar ring have been observed during the first reaction step. The sialic acid adopts a 2B_5 conformation in RC and changes to 4H_5 half-chair conformation at the oxocarbenium ion-like TS1, and finally reaches a 2C_5 chair form at IC. This itinerary is consistent with the proposed mechanisms based on theoretical studies of other retaining (trans)sialidase^{17,27,28}.

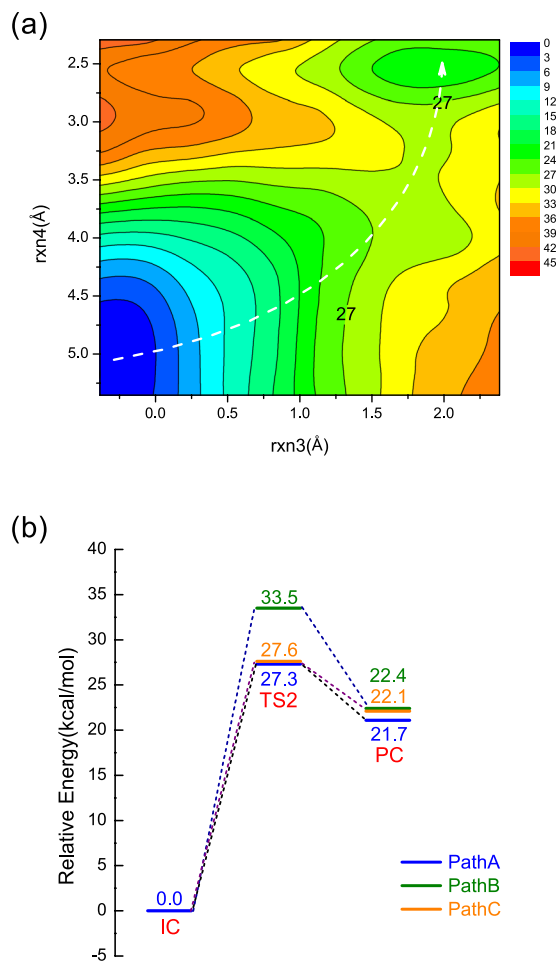


Figure 7. (a) Free energy contour plot for PathA of NanA. Energies are in kcal/mol. The white dashed line illustrates the minimum free energy path; (b) Free energy profiles for PathA (blue), PathB (green) and PathC (orange) of NanA.

In the IC state for NanA, a stable hydrogen bond is formed between the catalytic hydrogen and Asp372 O ϵ 2. The average distance between the anomeric carbon and the oxygen atom of the catalytic water molecule is less than 4 Å ($d7 \approx 3.26$ Å) and the average angle for hydroxyl oxygen (Tyr752)–C2(sialic acid)–O(water) is $\approx 178^\circ$, providing an ideal position for a nucleophilic attack by a water molecule and the subsequent proton transfer. In contrast, in the IC state for NanB and NanC, the distances between C2 position of the sugar ring and the catalytic water oxygen are 4.59 Å and 5.08 Å, respectively. This is due to the more hydrophobic pockets surrounding the catalytic clefts of NanB and NanC. At TS2 in PathA, the tyrosine–sialic acid bonds have been cleaved ($d5 \approx 2.63$ Å, 2.83 Å and 2.47 Å, respectively), while the proton transfers between the glutamic acids and the tyrosines are about to happen ($d6 \approx 1.15$ Å, 1.21 Å and 1.22 Å, respectively). This proton transfer is also observed at TS2 in PathC when the O7 (sialyl)–C2 (sialyl) bond formation is beginning in PathB.

Free energy profiles for NanA along three different paths. To better understand the chemical reaction for the second step of NanA, the potential of mean force (PMF) simulations have been carried out using the same reaction coordinates based on the structures obtained from the above MEP simulations. Figure 7a and Supporting Information Fig. 20 show the free energy contour plots for the three different reaction pathways from IC to PC and Fig. 7b summarises the reaction energetics from IC to PCs based on the PMF calculations. The minimum free energy paths highlighted with white dashed lines show similar patterns with those obtained from the MEP calculations, which validates the catalytic mechanisms of NanA obtained from MEP calculations. Besides, the MEP and PMF reaction barriers are close to each other, with the largest difference being less than 3.5 kcal/mol. Based on the MEP results, it was concluded that the second step is the rate-limiting step (Fig. 5). Our PMF simulations predicted an activation barrier of 27.3 kcal/mol along PathA for NanA (Fig. 7).

We note that the barrier of 27.3 kcal/mol is higher for a typical enzymatic system²⁹. However, this result is similar to those reported for other retain GHs including *S. marcescens* ChiB (25.8 kcal/mol)³⁰, TcTS (26.8 kcal/mol)²⁵, 1,3-1,4- β -glucanase (32.0 kcal/mol)³¹ and Golgi α -mannosidase II (GMII, 23.0 kcal/mol)³². Previous experimental studies showed that the catalytic efficiency of NanA is higher than NanB and NanC^{4,12,33}. This is inconsistent with our results that NanA has a higher reaction energy barrier (Fig. 5). It has been suggested in the literature that the carbohydrate-binding module (CBM) can greatly affect the catalytic efficiency of large sialidases³⁴. However,

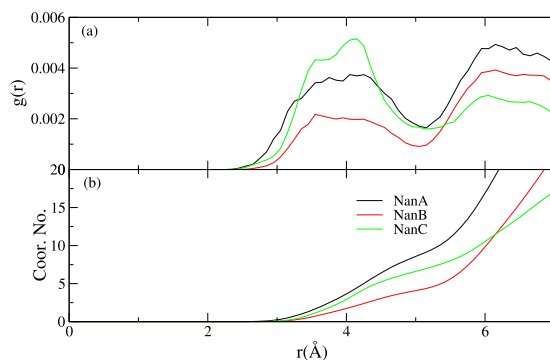


Figure 8. (a) Radial distribution function of water oxygens around C2 in the substrate in the IC states of NanA, NanB and NanC; (b) The coordination number of water around C2 in the substrate in the IC states of NanA, NanB and NanC.

we only focus on the catalytic domains of NanA, NanB and NanC with the GSBP setup in our simulations, bringing considerable approximations during the calculations. No structural studies have been performed on the interactions between the CBM40 domains and the substrates of NanA, thus it is envisaged that CBM40 domains of NanA may have the biggest effect on its catalytic efficiency. Additionally, previous work has shown that different buffers, pH optimum and substrates can also affect the enzymatic activity of sialidases^{33,34}. On the other hand, it has been noted that SCC-DFTB/MM overestimates the reaction barriers for the classical reaction pathways in glycoside hydrolases^{35,36}, likely due to the intrinsic limitation of the SCC-DFTB method. Future work with high level DFT or *ab initio* method is required to reproduce quantitatively the observed kinetic data. Nevertheless, our current study qualitatively reproduced the observed mechanistic difference for three sialidases.

What are the major determinants for their distinct catalytic pathways? We carried out perturbative analyses on the effects of residues of interest on the activation barrier (see more details in Supporting Information Section 11). We focused on the different residues among NanA, NanB and NanC, which are thought to be the main determinants for their distinct catalytic pathways. In the perturbative analysis, their partial charges were set to zero and the activation barrier was recalculated based on the corresponding optimised structures of the wild type^{26,37}. In all the cases, the resultant activation barriers are relatively similar to that of PathA in the wild type (Supporting Information Table 9), which indicated that these residues have relatively minor effects on the reaction energetics. However, it is worth noting that these residues are relatively far away from the reactive centre and their effects might be largely screened by the environment.

Instead, we further hypothesised that the local water distribution around the active site plays a more significant role. We characterised this by looking at the water distribution in their respective IC simulations (Fig. 8). For NanA, the overall values of the coordination number for the water molecules around C2 position of the substrate are much higher than those for NanB and NanC. This means that in the IC, the C2 of the sugar in NanA is more exposed to water, and this is consistent with the hypothesis that the water molecule can easily attack C2 position in NanA. Previous analyses of their structural properties in the intermediate states further support this hypothesis (see Description of the Mechanisms for NanA, NanB and NanC). In summary, the differences in their sequence/structure at the active site will likely introduce substantial differences in their local solvation environment, which then contributes to their preferred catalytic pathways. This observation will have significant implications for designing TS analogues as potent inhibitors. For instance, TS analogues with different substitutions around the C7 position of the sialyl cation of NanB (hydrophobic vs hydrophilic) will likely affect their binding affinities. In return, this will help to (dis)approve the hypothesis regarding the local environment in the active sites.

Conclusions

In this work, the detailed catalytic mechanisms for three closely related *S. pneumoniae* sialidases were studied. Based on the comparative MEP simulations, the complete pathways from the reactant state to the product state were characterised. It was confirmed that NanA and NanC operate via the mechanisms consistent with those proposed in the literature. In contrast, the mechanism for NanB is slightly different from the previously proposed one, which involves a catalytic water in the second step. Our simulations strongly suggested that the local hydration environment in the active site strongly favours an alternative mechanism without a catalytic water molecule in the second step. Combining all the data together, we suggest the lowest reaction barriers for the most preferred PathA in NanA, PathB in NanB and PathC in NanC, respectively. It was noted that very different solvation environments in the intermediate states might play a crucial role in the determination of their preferred reaction pathways. Additionally, analyses of the transition states and intermediate states indicate that all three enzymes operate through a dissociative A_ND_N mechanism. The detailed characterisation of the TS complex will provide insights into inhibitor design.

Methods

Simulation setup. The crystal structures of NanA (PDB code: 2VVZ¹¹, resolution 2.50 Å), NanB (PDB code: 2VW1¹², resolution 2.39 Å) and NanC (PDB code: 4YW3¹⁰, resolution 2.05 Å) were used as the starting structures for all the simulations. The protonation states of all the titratable residues, particularly the catalytic residues of NanA (Glu647, Tyr752 and Asp372), NanB (Glu541, Tyr653 and Asp270) and NanC (Glu584, Tyr695 and Asp315), were predicted based on the empirical method PROPKA³⁸ which has been proved to be capable of providing satisfactory pK_a predictions and assigning proper protonation states in our previous work³⁷. The Glu647, Glu541 and Glu584 were predicted to be deprotonated, while Asp372, Asp270 and Asp315 were shown to be protonated in the Michaelis complex (reactant complex RC). This is consistent with their catalytic roles either as a general base (Glu) or a general acid (Asp) in the first step from RC to the Intermediate complex (IC). All QM/MM simulations were carried out with CHARMM (version c38a2)³⁹. CHARMM force field c36⁴⁰ was used to describe the protein and sugars. The self-consistent charge density functional tight binding (SCC-DFTB)^{28,41,42} was adopted to describe the quantum region. The choice of SCC-DFTB as the QM method in our QM/MM simulations is based on the following considerations. Firstly, SCC-DFTB/MM simulations have been proved to be one of the most popular tools to study enzymatic reactions thanks to their efficiencies⁴². Secondly, SCC-DFTB/MM or DFTB3/MM simulations have been successfully applied to study other sialidases^{24,25}. The link atom was placed between C β and C γ , C β and C γ , and C β and C α for the glutamic acid, tyrosine and aspartic residue (Glu647, Tyr752 and Asp372 in NanA; Glu541, Tyr653 and Asp270 in NanB; Glu584, Tyr695 and Asp315 in NanC), respectively. Notably, the R group in the substrate was modified to a methyl group considering the computational cost.

The system was solvated with the TIP3P water model⁴³. The General Solvent Boundary Potential (GSBP)⁴⁴ setup was used in conjunction with explicit water molecules in the inner sphere for the QM/MM simulation that includes: (i) a spherical region of 20 Å radius centred on the glutamic acid (Glu647 in NanA, Glu541 in NanB and Glu584 in NanC) O ϵ 2 was treated as the inner region and the remaining protein is treated as the outer region; (ii) a spherical region of radius 18 Å was defined as the reaction region where Newtonian dynamics was solved; (iii) the region between 18 Å and 20 Å was treated as buffer region where Langevin dynamics is solved. A non-polar cavity potential was adopted to confine the inner region water molecules and prevent them leaving this region. The dielectric constants in the protein and solvent were assumed to be 1.0 and 80.0, respectively. The simulated temperature was set up to 298.15 K. The bonds involving hydrogen were constrained via the SHAKE algorithm⁴⁵. To treat the bonded interactions at the QM/MM boundary, the bonds between the QM and MM regions were cut by a link atom with the divided frontier charge model (DIV)⁴⁶. Such a treatment has been shown to provide a balanced description of proton affinities for molecules of interest²⁶. Further more, the puckering of the sugar was converted to a boat conformation (2B_5) based on the Cremer-Pople puckering coordinates⁴⁷ during the QM/MM minimisation for consistency with the conformation in the proposed catalytic mechanism²⁸. This was achieved by applying a harmonic restraint on the reaction coordinates with the predefined reference values. The whole substrate was also included in the QM region. Full QM/MM minimisation was performed prior to MD simulations. Supporting Information Table 1 summaries the relevant information for the starting structures and equilibrium molecular dynamics simulations. We refer to Supporting Information Section 1 and 2 for more information.

The minimum energy pathway (MEP) calculations. A comparative MEP simulation was performed along the hypothetical reaction pathways presented in Fig. 1. The reaction coordinates used to drive the reactions are also shown in Fig. 1. In the MEP calculations, the adapted basis, Newton-Raphson approach, is used for energy minimisation, with a gradient threshold of 0.05 kcal/mol/Å for the last several (\approx 4) adiabatic mapping cycles⁴⁸. The transition states were then identified by locating the potential energy maxima along these approximate reaction paths.

From RC to IC. From RC to IC, the three reaction pathways share the same reaction coordinates. rxn1 (d1 – d2) describes the departure of the leaving group. rxn2 (d3 + d4) represents the proton transfer from tyrosine to glutamic acid and the nucleophilic attack on the C2 position of the sugar ring by the deprotonated tyrosine. d1 is the distance between the C2 position of the sugar ring and the glycosidic O of the methyl group. d2 describes the distance between the glycosidic O of the methyl group and the aspartic acid residue (NanA, Asp372; NanB, Asp270; NanC, Asp315) H δ 2. d3 refers to the distance between the tyrosine (NanA, Tyr752; NanB, Tyr653; NanC, Tyr695) O η and C2 position of the sugar ring, and d4 represents the distance between the tyrosine H η and the glutamic acid O ϵ 2 (NanA, Tyr752 and Glu641; NanB, Tyr653 and Glu541; NanC, Tyr695 and Glu584).

From IC to PC. During the second step (from IC to the Product complex PC), the reaction is driven by 2D reaction coordinates. rxn3 defines the cleavage of the tyrosine–sialic acid bond and the protonation of tyrosine by a proton of the glutamic acid. d5 refers to the distance between the tyrosine O η and C2 position of the sugar ring, and d6 corresponds to the one between the tyrosine O η and H η (NanA, Tyr752; NanB, Tyr653; NanC, Tyr695). rxn4, rxn5 and rxn6 define the nucleophilic attack on C2 position of the sugar ring with the deprotonated catalytic water molecule by the aspartic acid residue (PathA), or O7 of the sialyl cation and the carbohydrate acceptor (PathB and PathC), respectively (Glu647, Tyr752 and Asp372 in NanA; Glu541, Tyr653 and Asp270 in NanB; Glu584, Tyr695 and Asp315 in NanC).

PathA: rxn3 = d5 – d6 and rxn4 = d7 + d8 were used to define the reaction pathway from IC to PC Neu5Ac (PathA). d7 is the distance between the C2 position of the sugar ring and the catalytic water oxygen, and d8 is the one between the aspartic acid residue (NanA, Asp372; NanB, Asp270; NanC, Asp315) O ϵ 2 and the catalytic water hydrogen.

PathB: rxn3 = d5 – d6 and rxn5 = d9 were used to describe the pathway from IC to PC 2,7-anhydro-Neu5Ac (PathB). d9 defines the distance between C2 position and O7 of the sugar ring.

PathC: rxn3 = d5 – d6 and rxn6 = d10 – d11 were used to describe the one from IC to PC Neu5Ac2en (PathC). d10 refers to the distance between C3 and H32 in substrate. d11 is the distance between H32 of the sugar ring and Oε2 of the aspartic acid residue for NanB and NanC (NanB, Asp270; NanC, Asp315) but between H32 of the sugar ring and catalytic water oxygen for NanA.

When simulating PathC in NanA, unlike the catalytic mechanism for PathC in NanC (Supporting Information Fig. 4) proposed by previous studies, in which the C3 proton of the substrate is attacked directly by the catalytic base Asp315^{4,10}, the reaction in our setup proceeds with an activated water molecule nearby because the more accessible catalytic cavity in NanA allows more water molecules to reside in (Fig. 1). This mechanism is more accordant with the actual environment in NanA. In contrast, due to the similarity between NanB and NanC in the catalytic mechanisms, in which the distance between the anomeric carbon C2 and aspartic acid residue is rather short and the pocket around the substrate is hydrophobic^{4,10}, the reaction involved in the second steps of PathC for NanB were undertaken without activated water molecules nearby (Fig. 1). The water distributions around C2, HO7 and H32 of the substrate were rationalised during the simulations. In comparison, we performed MEP calculations on the reaction path for the second step of NanA based on the mechanism of NanC (See details in Supporting Information Section 5). We also carried out MEP calculations on the reaction paths for the second steps of NanB and NanC based on the mechanisms that involve a catalytic water molecule (See details in Supporting Information Section 7).

The potential of mean force (PMF) calculations. In this work, we performed PMF simulations on the second step of NanA and compared the difference between the MEP and PMF reaction barriers to probe the effect of thermal fluctuations and flexibilities. The PMF simulations were carried out with a harmonic force constant 50.0 kcal/mol/Å² for each reaction coordinate except 350 kcal/mol/Å² for rxn5. The total number of windows was around 266, 980 and 931 for PathA, PathB and PathC, respectively. For each window, 200 ps SCC-DFTB/MM simulations following the same GSBP protocol were performed. The 2D Weighted Histogram Analysis Method (WHAM) method was used to unbias the systems and to evaluate the free energy profiles along the predefined reaction coordinates^{49,50}.

References

- Manco, S. *et al.* Pneumococcal neuraminidases A and B both have essential roles during infection of the respiratory tract and sepsis. *Infect. Immun.* **74**, 4014–4020 (2006).
- Adegbola, R. A. Childhood pneumonia as a global health priority and the strategic interest of the Bill & Melinda Gates Foundation. *Clin. Infect. Dis.* **54**, S89–S92 (2012).
- Pettigrew, M. M., Fennie, K. P., York, M. P., Daniels, J. & Ghaffar, F. Variation in the presence of neuraminidase genes among *Streptococcus pneumoniae* isolates with identical sequence types. *Infect. Immun.* **74**, 3360–3365 (2006).
- Xu, G. *et al.* Three *Streptococcus pneumoniae* sialidases: Three different products. *J. Am. Chem. Soc.* **133**, 1718–1721 (2011).
- Walther, E. *et al.* Antipneumococcal activity of neuraminidase inhibiting artocarpin. *Int. J. Med. Microbiol.* **305**, 289–297 (2015).
- Long, J., Tong, H. & DeMaria, T. Immunization with native or recombinant *Streptococcus pneumoniae* neuraminidase affords protection in the chinchilla otitis media model. *Infect. Immun.* **72**, 4309–4313 (2004).
- Jedrzejak, M. J. Pneumococcal virulence factors: structure and function. *Microbiol. Mol. Biol. Rev.* **65**, 187–207 (2001).
- Shinya, K. *et al.* Avian flu: Influenza virus receptors in the human airway. *Nature* **440**, 435–436 (2006).
- Chandrasekaran, A. *et al.* Glycan topology determines human adaptation of avian H5N1 virus hemagglutinin. *Nat. Biotechnol.* **26**, 107–113 (2008).
- Owen, C. D. *et al.* *Streptococcus pneumoniae* NanC structural insights into the specificity and mechanism of a sialidase that produce a sialidase inhibitor. *J. Biol. Chem.* **290**, 27736–27748 (2015).
- Xu, G., Li, X., Andrew, P. W. & Taylor, G. L. Structure of the catalytic domain of *Streptococcus pneumoniae* sialidase nana. *Acta Crystallogr. Sect. F: Struct. Biol. Cryst. Commun.* **64**, 772–775 (2008).
- Xu, G. *et al.* Crystal structure of the NanB sialidase from *Streptococcus pneumoniae*. *J. Mol. Biol.* **384**, 436–449 (2008).
- Gut, H., King, S. J. & Walsh, M. A. Structural and functional studies of *Streptococcus pneumoniae* neuraminidase B: An intramolecular *trans*-sialidase. *FEBS Lett.* **582**, 3348–3352 (2008).
- Walther, E. *et al.* Dual acting neuraminidase inhibitors open new opportunities to disrupt the lethal synergism between streptococcus pneumoniae and influenza virus. *Front. Microbiol.* **7** (2016).
- Xu, Z. *et al.* Sequence diversity of NanA manifests in distinct enzyme kinetics and inhibitor susceptibility. *Sci. Reports* **6**, 357–366 (2016).
- Rye, C. S. & Withers, S. G. Glycosidase mechanisms. *Curr. Opin. Chem. Biol.* **4**, 573–580 (2000).
- Amaya, M. F. *et al.* Structural insights into the catalytic mechanism of *Trypanosoma cruzi* *trans*-sialidase. *Structure* **12**, 775–784 (2004).
- Newstead, S. L. *et al.* The structure of *Clostridium perfringens* NanI sialidase and its catalytic intermediates. *J. Biol. Chem.* **283**, 9080–9088 (2008).
- Watts, A. G. *et al.* *Trypanosoma cruzi* *trans*-sialidase operates through a covalent sialyl-enzyme intermediate: Tyrosine is the catalytic nucleophile. *J. Am. Chem. Soc.* **125**, 7532–7533 (2003).
- Guthrie, R. D. & Jencks, W. P. IUPAC recommendations for the representation of reaction mechanisms. *Accounts Chem. Res.* **22**, 343–349 (1989).
- O’Ferrall, R. M. Relationships between E2 and E1cB mechanisms of β-elimination. *J. Chem. Soc. B: Phys. Org.* 274–277 (1970).
- Zechel, D. L. & Withers, S. G. Dissection of nucleophilic and acid–base catalysis in glycosidases. *Curr. Opin. Chem. Biol.* **5**, 643–649 (2001).
- Pauling, L. Atomic radii and interatomic distances in metals. *J. Am. Chem. Soc.* **69**, 542–553 (1947).
- Pierdominici-Sottile, G., Horenstein, N. A. & Roitberg, A. E. Free energy study of the catalytic mechanism of *trypanosoma cruzi* *trans*-sialidase. from the michaelis complex to the covalent intermediate. *Biochemistry* **50**, 10150–10158 (2011).
- Bueren-Calabuig, J. A., Pierdominici-Sottile, G. & Roitberg, A. E. Unraveling the differences of the hydrolytic activity of *Trypanosoma cruzi* *trans*-sialidase and *Trypanosoma rangeli* sialidase: A quantum mechanics/molecular mechanics modeling study. *J. Phys. Chem. B* **118**, 5807–5816 (2014).
- Yu, H. & Griffiths, T. M. pK_a cycling of the general acid/base in glycoside hydrolase families 33 and 34. *Phys. Chem. Chem. Phys.* **16**, 5785–5792 (2014).
- Davies, G. J., Planas, A. & Rovira, C. Conformational analyses of the reaction coordinate of glycosidases. *Accounts Chem. Res.* **45**, 308–316 (2011).

28. Montgomery, A. P., Xiao, K., Wang, X., Skropeta, D. & Yu, H. Computational glycobiology: Mechanistic studies of carbohydrate-active enzymes and implication for inhibitor design. *Adv. Protein Chem. Struct. Biol.* **109**, 25–76 (2017).
29. Wolfenden, R. Benchmark reaction rates, the stability of biological molecules in water, and the evolution of catalytic power in enzymes. *Annu. Rev. Biochem.* **80**, 645–667 (2011).
30. Jitnonom, J., Limb, M. A. & Mulholland, A. J. QM/MM free-energy simulations of reaction in *Serratia marcescens* chitinase b reveal the protonation state of asp142 and the critical role of tyr214. *J. Phys. Chem. B* **118**, 4771–4783 (2014).
31. Biarnés, X., Ardèvol, A., Iglesias-Fernández, J., Planas, A. & Rovira, C. Catalytic itinerary in 1, 3-1, 4- β -glucanase unraveled by qm/mm metadynamics. charge is not yet fully developed at the oxocarbenium ion-like transition state. *J. Am. Chem. Soc.* **133**, 20301–20309 (2011).
32. Petersen, L., Ardevol, A., Rovira, C. & Reilly, P. J. Molecular mechanism of the glycosylation step catalyzed by Golgi α -mannosidase II: A QM/MM metadynamics investigation. *J. Am. Chem. Soc.* **132**, 8291–8300 (2010).
33. Hayre, J. K. *et al.* Optimization of a direct spectrophotometric method to investigate the kinetics and inhibition of sialidases. *BMC Biochem.* **13**, 19–25 (2012).
34. Thobhani, S., Ember, B., Siriwardena, A. & Boons, G.-J. Multivalency and the mode of action of bacterial sialidases. *J. Am. Chem. Soc.* **125**, 7154–7155 (2003).
35. Liu, J., Wang, X. & Xu, D. QM/MM study on the catalytic mechanism of cellulose hydrolysis catalyzed by cellulase Cel5aa from *Acidothermus cellulolyticus*. *J. Phys. Chem. B* **114**, 1462–1470 (2010).
36. Xiong, J. & Xu, D. Insights into the catalytic mechanism of unsaturated glucuronyl hydrolase of *Bacillus* sp. GL1. *J. Phys. Chem. B* **121**, 931–941 (2017).
37. Xiao, K. & Yu, H. Rationalising pK_a shifts in *Bacillus circulans* xylanase with computational studies. *Phys. Chem. Chem. Phys.* **18**, 30305–30312 (2016).
38. Søndergaard, C. R., Olsson, M. H., Rostkowski, M. & Jensen, J. H. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pK_a values. *J. Chem. Theory Comput.* **7**, 2284–2295 (2011).
39. Brooks, B. R. *et al.* Charmm: the biomolecular simulation program. *J. Computat. Chem.* **30**, 1545–1614 (2009).
40. MacKerell, A. Jr. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
41. Elstner, M. *et al.* Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B* **58**, 7260–7268 (1998).
42. Riccardi, D. *et al.* Development of effective quantum mechanical/molecular mechanical (QM/MM) methods for complex biological processes. *J. Phys. Chem. B* **110**, 6458–6469 (2006).
43. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
44. Schaefer, P., Riccardi, D. & Cui, Q. Reliable treatment of electrostatics in combined QM/MM simulation of macromolecules. *J. Chem. Phys.* **123**, 014905 (2005).
45. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).
46. König, P., Hoffmann, M., Frauenheim, T. & Cui, Q. A critical evaluation of different QM/MM frontier treatments with SCC-DFTB as the QM method. *J. Phys. Chem. B* **109**, 9082–9095 (2005).
47. Cremer, D. & Pople, J. General definition of ring puckering coordinates. *J. Am. Chem. Soc.* **97**, 1354–1358 (1975).
48. Yang, Y., Yu, H. & Cui, Q. Extensive conformational transitions are required to turn on atp hydrolysis in myosin. *J. Mol. Biol.* **381**, 1407–1420 (2008).
49. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *J. Comput. Chem.* **13**, 1011–1021 (1992).
50. Grossfield, A. WHAM: the weighted histogram analysis method, version 2.0.9.1, <http://membrane.urmc.rochester.edu/content/wham> (2018).

Acknowledgements

H.Y. is the recipient of an Australian Research Council Future Fellowship (Project number FT110100034) and X.W. is the recipient of the University of Wollongong Vice-Chancellor's Postdoctoral Research Fellowship. K.X. thanks the Australian Government for an Australian International Postgraduate Award scholarship. We wish to acknowledge that this research was undertaken with the assistance of resources provided at the NCI National Facility systems at the Australian National University through the National Computational Merit Allocation Scheme supported by the Australian Government (v15).

Author Contributions

K.X., X.W. & H.Y. designed research. K.X. conducted experiments and analysed the data. K.X. wrote the main manuscript and prepared figures and tables. All authors have reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-38131-z>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019