Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Electronic Supplementary Information for " pK_a cycling of the general acid/base in glycoside hydrolase families 33 and 34"

Haibo Yu

1 Summary of the simulated systems

The explicit solvent simulations with MM and QM/MM are summarised in Table 1. The equilibrium (Eq.) simulations were carried out for 40 ns for MM and 20 ns for DFTB3/MM. Three independent FEP simulations were carried out for MM and QM/MM by taking the initial structures at 20 ns (A), 30 ns (B), and 40 ns (C) Eq. simulations for MM and 10 ns (A), 15 ns (B) and 20 ns (C) Eq. simulations for DFTB3/MM.

Table 1: Summary of the equilibrium molecular dynamics simulations (Eq.) and free energy perturbation (FEP) simulations.

Simulations	PDB id	Resolution Å	Substrate	Number of DFTB3 atoms[e]	Eq. (in ns) MM	Eq. (in ns) DFTB3/MM	FEP per λ (in ns) MM	FEP per λ (in ns) DFTB3/MM
Trypanosoma cruzi trans-sialidase								
1MS3-apo	1MS3 [1]	1.65	-	8	40	20	3	≥ 2
1S0I-apo[a]	1S0I [2]	1.60	-	8	40	20	3	3
1S0I-holo	1S0I [2]	1.60	SLT[b]	8	40	20	≥ 3	≥ 3
influenza neuraminidase								
2QWB-apo[a]	2QWB [3]	2.00	-	8	40	20	3	≥ 2
2QWB-holo	2QWB [3]	2.00	SIA[c]	8	40	20	3	3
3W09-int	3W09 [4]	2.00	covalent int.[d]	65	-	20	-	2

[a] The cocrystalized substrate was deleted in the 1S0I-apo and 2QWB-apo simulations; [b] SLT: lactose sialic acid; [c] SIA: sialic acid; [d] covalent int.: This is a glycosyl-enzyme intermediate state. [e] In all the simulations except 3W09-int, the QM/MM partition boundary is placed between the Cα and Cβ atoms of Asp59 in TcTS and Asp151 in NA; In 3W09-int, since there are no published CHARMM force field available for the glycosyl-enzyme linkage, all the relevant parts are treated with DFTB3 including the side chains of Glu277, Tyr406, Asp151 and the substrate. Two additional link atoms were placed between the Cβ and Cγ atoms in Glu277 and Tyr406.

2 pK_a values of the proposed catalytic acid in the GH33 and GH34 with PROPKA

The pK_a values of the proposed catalytic acid for all the structures available for GH33 (Table 2) and GH34 (Table 3) were predicted with the very efficient package PROPKA 3.1 [5]. The most recent version of PROPKA has shown to be able to provide an improved treatment of ligand-binding effects on the pK_a values of the protein residues. Based on our calculations, it is conclusive that in the apo state, the proposed catalytic aspartic acid (Asp59 in TcTS and Asp151 in NA) is deprotonated with a pK_a well-below 7. However, according to PROPKA, contradictory to the proposed role as a general acid in the holo state, Asp remains deprotonated.

Table 2: pK_a values of the proposed catalytic residues in the GH33.

Protein names	Organism	PDB id	pK_a values	
			apo	holo
Bacteria				
sialidase	Bacteroides thetaiotaomicron VPI-5482	4BBW[a]	3.34	
exo-α-sialidase	Clostridium perfringens str. 13	2BF6[b]	2.45	1.99
sialidase	Micromonospora viridifaciens	2BER[b]	3.89	4.01
sialidase	Salmonella typhimurium TA262	2SIM[c]	3.47	3.18
sialidase	Streptococcus pneumoniae R6	3H72[b]	4.60	2.97
(trans-)sialidase	Streptococcus pneumoniae TIGR4	2VW1[c]	4.23	3.41
neuraminidase	Streptococcus pneumoniae TIGR4	2YA5[b]	4.52	2.39
sialidase/neuraminidase	Vibrio cholerae 569B 395	1W0O[c]	4.24	2.39
Eukaryota				
sialidase	Homo sapiens	1VCU[b]	3.71	3.65
sialidase	Macrobdella decora	1SLI[c]	3.18	3.62
trans-sialidase	Trypanosoma cruzi	1MS3[a]	2.25	
		1MS1[c]	2.31	2.53
		1S0I[b]	2.13	2.56
sialidase	Trypanosoma rangeli	1N1Y[b]	2.27	1.70

[a] denotes that the apo structure was used for PROPKA calculations; [b] denotes that the holo structure in complex with substrate was used for PROPKA calculations; [c] denotes that the holo structure in complex with a transition-state like inhibitor was used for PROPKA calculations.

Table 3: pK_a values of the proposed catalytic residues in the GH34.

Protein names	Organism	PDB id	pK_a values	
			apo	holo
Viruses				
neuraminidase	A/Tanzania/205/2010(H3N2)	4GZQ[b]	5.09	4.67
neuraminidase	A/Brevig Mission/1/1918(H1N1)	3B7E[c]	4.92	4.92
neuraminidase	A/California/04/2009(H1N1)	3TI5[c]	4.88	3.58
neuraminidase	A/duck/Alberta/60/1976(H12N5)	3SAN[c]	4.75	3.07
neuraminidase	A/duck/England/1/1956(H11N6)	1W1X[c]	4.70	3.47
neuraminidase	A/duck/Ukraine/1/1963(H3N8)	2HTR[c]	5.17	4.84
neuraminidase	A/Memphis/31/98(H3N2)	2AEQ[a]	4.45	
neuraminidase	A/Mexico/4108/2009(H1N1)	3NSS[a]	3.93	
neuraminidase	A/mink/Sweden/E12665/84(H10N4)	2HTW[c]	6.21	5.92
neuraminidase	A/RI/5+/1957(H2N2)	4H53[b]	4.41	4.50
neuraminidase	A/tern/Australia/G70C/1975(H11N9)	2QWB[b]	4.87	3.60
neuraminidase	A/Tokyo/3/1967(H2N2)	2BAT[b]	5.32	3.25
neuraminidase	A/Viet Nam/1203/2004(H5N1)	3CL2[c]	6.79	4.01
neuraminidase	B virus	1NSC[b]	4.68	4.05
neuraminidase	B/BEIJING/1/87	1A4G[c]	4.63	3.38
neuraminidase	B/Perth/211/2001	3K37[c]	5.49	2.92

[a] denotes that the apo structure was used for PROPKA calculations; [b] denotes that the holo structure in complex with substrate was used for PROPKA calculations; [c] denotes that the holo structure in complex with a transition-state like inhibitor was used for PROPKA calculations.

3 Benchmark of DFTB3/MM for proton affinities with the acidic oxygen atom in aspartic acid

Three different link atom based frontier treatments in DFTB3/MM simulations were evaluated by comparing the proton affinities of the acidic oxygen in Asp. The proton affinities from the calculations, where the whole system were treated with either G3B3 or DFTB3, were taken as a reference. The divided frontier charge (DIV), in which the partial charge associated with the frontier MM atom is evenly distributed to the other MM atoms within the same group, provides the best description. This has been noted in previous study when partitioning amino acids across the Cα-Cβ bond in amino acids as in our case here, SLA and EXGR cause significant errors [6]. The calculation with the DFTB3/DIV scheme predicted a slightly over-estimated proton affinity for both water and Asp (1.8 kcal/mol for water and 1.7 kcal/mol for Asp). Thus in all DFTB3/MM simulations, the DIV scheme was adopted.

Table 4: Proton affinities with different methods (in kcal/mol).

	G3B3	DFTB3[a]	DFTB3/MM[b]		
			SLA	EXGR	DIV
Water	398.4	400.2	-	-	-
NH_3-Asp-OCH_3	333.9	339.8	355.8	355.2	335.6

[a] The full system was treated with DFTB3; [b] DFTB3/MM was used in which three different link-atom schemes were adopted: SLA, EXGR and DIV. See König et al. [6] for more details.

4 pK_a calculations with explicit solvent simulations

To overcome the known limitations of pK_a calculations based on continuum electrostatics or empirical models, we carried out additional pK_a shift calculations with explicit solvent simulations. The pK_a shift is defined as Δ pK_a = $\Delta\Delta G/(2.303k_{\mathrm{B}}T)$, with $\Delta\Delta G = \Delta G_{\mathrm{protein}} - \Delta G_{\mathrm{model}}$, where $\Delta G_{\mathrm{model}}$ refers to the free energy of deprotonation of an isolated residue in bulk solution. This value is estimated to be -47.3 kcal/mol with MM [7] and 121.2 kcal/mol with DFTB3/MM. $\Delta G_{\mathrm{protein}}$ refers to the free energy of deprotonation of the corresponding resiude in the protein. For both terms, a direct unbiased alchemical perturbation of a protonated Asp into a deprotonated Asp were carried out. 11 evenly spaced windows between 0 and 1 were adopted in the current simulations. All the simulations were triplicated by taking different snapshots from the equilibrium simulations as the starting structure for the FEP simulations (see Table 1). The differences between the three sets were estimated to be about 2.0 kcal/mol (1.4 pK_a units).

Table 5: $\Delta G_{\text{E-RCOO(D/H)}}$ of Asp59 in TcTS or Asp151 in NA based explicit solvent MM simulations (in kcal/mol)[a].

Structure	MM FEP/MD (in kcal/mol)[b]		
	A	B	C
Asp59 in *Trypanosoma cruzi* trans-sialidase			
1MS3-apo	-47.9 (1.0/2.0)	-47.6 (1.0/2.0)	-48.8 (2.0/2.0)
1S0I-apo	-51.0 (1.0/2.0)	-49.5 (1.0/2.0)	-50.1 (1.0/2.0)
1S0I-holo	-39.1 (1.0/2.0)	-38.5 (1.0/3.0)	-40.3 (1.0/3.0)
Asp151 in influenza neuraminidase			
2QWB-apo	-45.2 (1.0/2.0)	-44.9 (1.0/2.0)	-44.3 (1.0/2.0)
2QWB-holo	-41.0 (1.0/2.0)	-40.2 (1.0/2.0)	-39.4 (1.0/2.0)

[a] $\Delta G(\text{Asp})$ is estimated to be -47.3 kcal/mol with MM. The three different sets were carried out by taking the three different snapshots in the Eq. MM simulations at 20 ns, 30 ns and 40 ns respectively. [b] The numbers in parentheses are the equilibration time and the production time, respectively.

Table 6: $\Delta G_{\text{E-RCOO(D/H)}}$ of Asp59 in TcTS or Asp151 in NA based explicit solvent DFTB3/MM simulations (in kcal/mol)[a].

Structure	DFTB3/MM FEP/MD (in kcal/mol)		
	A	B	C
Asp59 in *Trypanosoma cruzi* trans-sialidase			
1MS3-apo	117.5 (1.0/2.0)	116.6 (1.0/1.0)	117.2 (1.0/1.0)
1S0I-apo	116.7 (1.0/2.0)	117.3 (1.0/2.0)	117.9 (1.0/2.0)
1S0I-holo	126.7 (1.0/2.0)	127.5 (1.0/2.0)	128.6 (1.0/3.0)
Asp151 in influenza neuraminidase			
2QWB-apo	121.8 (1.0/1.0)	121.5 (1.0/2.0)	121.8 (1.0/2.0)
2QWB-holo	128.1 (1.0/2.0)	127.3 (1.0/2.0)	128.3 (1.0/2.0)
3W09-int	125.0 (1.0/1.0)	124.8 (1.0/1.0)	124.1 (1.0/1.0)

[a] $\Delta G(\text{Asp})$ is estimated to be 121.2 kcal/mol with DFTB3/MM. The three different sets were carried out by taking the three different snapshots in the Eq. DFTB3/MM simulations at 10 ns, 15 ns and 20 ns respectively. [b] The numbers in parentheses are the equilibration time and the production time, respectively.

5 The reorganisation energy during deprotonation from the DFTB3/MM perturbative analyses

The reorganisation energy from water, protein and substrate is related to the variation of their contribution to the free enegy derivatives as a function of λ. They can be characterised by comparing their contributions at λ = 0 and λ = 1. In all cases, the reorganisation from water is substantially larger than that from the protein. Additionally in the substrate-bound state, the substrate also undergoes noticeable reorganisation. This is consistent with the conclusion that the substrate binding and the local solvation environment are the dominating factors modulating the protonation state of the catalytic acid.

Table 7: The free energy derivatives from the perturbative analyses on explicit solvent DFTB3/MM simulations (in kcal/mol)[a].

Systems	λ	Ref.[b]	ΔWat[c]	ΔProt[c]	ΔLig[c]
Asp59 in *Trypanosoma cruzi* trans-sialidase					
1S0I-apo	0.0	199.8	132.8 (67.0)	275.9 (-76.1)	-
	1.0	25.4	92.3 (-66.9)	124.0 (-98.6)	-
1S0I-holo	0.0	199.6	211.2 (-11.6)	287.0 (-87.4)	127.4 (72.2)
	1.0	37.3	146.4 (-109.1)	133.6 (-96.3)	5.8 (31.5)
Asp151 in influenza neuraminidase					
2QWB-apo	0.0	204.2	126.6 (77.6)	290.0 (-85.8)	-
	1.0	23.7	90.2 (-66.5)	128.8 (-105.1)	-
2QWB-holo	0.0	209.6	193.3 (16.3)	293.0 (-83.4)	161.5 (48.1)
	1.0	32.8	145.5(-112.7)	131.0 (-98.2)	2.1 (30.7)

[a] The numbers in parenthese are the differences between the reference and the perturbed systems. [b] Ref. refers to the perturbative analyses on the originally simulated systems while keeping the partial charges as it is. [c] refers to the perturbative analyses on the originally simulated systems while zeroing out the charges on the part of interests (ΔWat: water; ΔProt: protein; ΔLig: substrate).

References

[1] Buschiazzo, A., Amaya, M. F., Cremona, M. L., Frasch, A. C. & Alzari, P. M. The crystal structure and mode of action of trans-sialidase, a key enzyme in Trypanosoma cruzi pathogenesis. *Mol. Cell.* **10**, 757–768 (2002).

[2] Amaya, M. F. *et al.* Structural Insights into the Catalytic Mechanism of Trypanosoma cruzi trans-Sialidase. *Structure* **12**, 775–784.

[3] Varghese, J. N. *et al.* Drug design against a shifting target: a structural basis for resistance to inhibitors in a variant of influenza virus neuraminidase. *Structure* **6**, 735–746 (1998).

[4] Kim, J.-H. *et al.* Mechanism-based covalent neuraminidase inhibitors with broad-spectrum influenza antiviral activity. *Science* **340**, 71–75 (2013).

[5] Søndergaard, C. R. & Olsson, M. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of p K a values. *J. Chem. Theo. Comput.* **7**, 2284–2295 (2011).

[6] König, P. H., 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).

[7] Yu, H., Ratheal, I. M., Artigas, P. & Roux, B. Protonation of key acidic residues is critical for the K+-selectivity of the Na/K pump. *Nat. Struct. Mol. Biol.* **18**, 1159–1163 (2011).