Supporting Information

Predicting catalytic proton donors and nucleophiles in enzymes: how adding dynamics helps elucidate the structure-function relationships

Yandong Huang,[†,‡] Zhi Yue,[†,¶] Cheng-Chieh Tsai,[†] Jack Henderson,[†] and Jana Shen[*,†]

† *Department of Pharmaceutical Sciences, University of Maryland School of Pharmacy, Baltimore, MD 21201.*

‡ *Current address: College of Computer Engineering, Jimei University, Xiamen, Fujian 361021, China.*

¶ *Joint first author.*

E-mail: jana.shen@rx.umaryland.edu

Computational details and protocols

Structure preparation

Five enzymes were studied in this work, β-secretase 1 (BACE1, pdb code 1SGZ), β-secretase 2 (BACE2, pdb code 3ZKQ), cathepsin D (CatD, pdb code 1LYA), hen egg white lysozyme (HEWL, pdb code 2LZT), and a hyperstable variant of staphylococcal nuclease ($\Delta+$PHS SNase, pdb code 3BDC). Hydrogens were added to the proteins using the HBUILD facility[1] in CHARMM.[2] Dummy hydrogen atoms were added to Asp and Glu sidechains and placed in the *syn* position initially.[3] The protein was then placed in a truncated octahedral water box with a minimum distance of 10 Å between the solute and edges of the water box. Water molecules within 2.6 Å of any protein heavy atoms were deleted. The unit-cell lattice parameter for HEWL, SNase, BACE1, CatD and BACE2 were 68, 69, 90, 84 and 84 Å, respectively. While both hybrid-solvent[4] and all-atom CpHMD simulations with particle mesh Ewald electrostatics[5] were performed for HEWL, SNase and BACE1, only hybrid-solvent CpHMD simulations performed for CatD and BACE2, which are close homologs of BACE1.

In the hybrid-solvent CpHMD, no explicit ions were added, as the electrostatic calculations were performed using the generalized Born model in the propagation of titration coordinates.[4] As such, the presence of ions have negligible effects on the pK_{a}'s, as demonstrated in our previous work.[4] Salt effects were taken into account via the ionic strength setting (to the experimental conditions if available) in the Debye-Hückel approximation.[6]

In the all-atom CpHMD simulations, titratable water molecules were added to achieve charge neutrality. Each titratable site was coupled with a titratable water, which can convert to a hydronium ion for the acidic site, and a titratable water, which can convert to a hydroxide ion for the basic site.[7,8] The titratable waters were initially randomly placed in the simulation box and free to move during the simulation. Explicit Na^+ and Cl^- ions were added to neutralize the excess charge and achieve the ionic strength used in the NMR

titration experiments: 50 mM for HEWL,[9] 100 mM for SNase,[10] and 50 mM for BACE1.[11] The composition of the simulation systems are summarized in Table S1.

Each system was first energy minimized by the steepest descent method followed by the adopted basis Newton-Raphson method with protein heavy atoms fixed and then harmonically restrained with a force constant of 5 kcal/mol·Å^2 to relax the hydrogen positions. The system was then heated from 100 to 300 K within 80 ps with the protein heavy atoms harmonically restrained with a force constant of 5 kcal/mol·Å^2. Following heating, the system was equilibrated for 120 ps under harmonic restraints, where the force constant was gradually reduced from 5 (40 ps), to 1 (40 ps), and 0.1 kcal/mol·Å^2 (40 ps). Finally, the system was equilibrated for 300 ps without restraints. Heating and equilibration were performed at the crystallization pH condition or pH 7.

Table S1: Overview of the simulation systems

	Number of residues/molecules[a]							
	Hybrid-solvent					All-atom		
	HEWL	SNase	BACE1	CatD	BACE2	HEWL	SNase	BACE1
Protein	129	129	389	338	385	129	129	389
H_2O	7198	7448	16238	12651	16874	7333	7496	15709
H_3O^+	-	-	-	-	-	1	2	7
OH^-	-	-	-	-	-	9	17	43
Na^+	-	-	-	-	31	2	15	21
Cl^-	-	-	-	-	32	12	24	17
No. atoms	23581	24484	54831	43203	56612	24028	24725	53439

[a] HEWL contains 7 Asp, 2 Glu, 1 His, 6 Lys and 11 Arg residues. SNase contains 6 Asp, 11 Glu, 2 His, 18 Lys and 5 Arg residues. BACE1 contains 21 Asp, 22 Glu, 7 His, 15 Lys and 17 Arg residues. CatD contains 20 Asp, 14 Glu, 5 His, 20 Lys and 8 Arg residues BACE2 contains 17 Asp, 21 Glu, 1 His, 13 Lys and 12 Arg residues.

Simulation protocols

All CpHMD simulations were carried out using the CHARMM package (version c36a6, modified for all-atom CpHMD),[2] where the hybrid-solvent[4] and all-atom CpHMD methods[5,7,8] were implemented (in PHMD module). The pH-based replica-exchange proto-

col[4] was used to enhance sampling in both protonation and conformational space (in REPDSTR module). The CHARMM c22/CMAP force field[12,13] was used to represent proteins. The CHARMM style TIP3P water model was used to represent water.[2] The parameters for titratable water that can convert to hydronium or hydroxide ion were taken from Chen *et al.*[8] In all simulations, the SHAKE algorithm[14] was applied to bonds involving hydrogen atoms to allow a 2-fs time step. Simulations were conducted with periodic boundary conditions at ambient temperature (300 K) and pressure (1 atm) maintained by the Nóse-Hoover thermostat[15] and Langevin piston pressure-coupling algorithm,[16] respectively. Switch function was applied to the van der Waals potential calculation from 10 to 12 Å. Electrostatic potential was calculated using the smooth particle-mesh Ewald method[17,18] with a real-space cutoff of 12 Å and a sixth-order interpolation with approximately 1-Å grid spacing. The neighbor and image lists were updated when necessary. All Asp, Glu and His residues were allowed to titrate; Lys and Arg residues were fixed in the charged state, as their model p$K_{\rm a}$ are at least two units above the titration pH range and none of them is buried such as a p$K_{\rm a}$ downshift would be expected. The titration coordinates (λ) were propagated using the Langevin algorithm with a collision frequency of 5 ps^{-1}. The mass of the fictitious λ particles was set to 10 atomic mass units.

Hybrid-solvent CpHMD. In the hybrid-solvent CpHMD simulations, the titration dynamics is propagated using the generalized-born (GB) model GBSW[19] with the GB input radii by Chen *et al.*[20] The ionic strength in the GB calculations was set to the experimental condition: 50 mM for HEWL,[9] 100 mM for SNase,[10] 50 mM for BACE1,[11] 50 mM for CatD,[21] and 100 mM for BACE2.[22] The default setting in the GBSW module of CHARMM[19] was used for all other GB input options.

The production simulations were performed using the pH-based replica-exchange (pH-REX) protocol[4] with an exchange between the neighboring pH replicas attempted every 500 MD steps or 1 ps. Data was collected after every exchange attempt (1 ps). For

HEWL, 16 replicas were placed in the pH range 0–9 with an interval of 0.5 or 1 pH unit. Each replica was run for 5 ns per replica, resulting in an aggregate sampling time of 80 ns. For SNase, 16 replicas were placed in the pH range 0–7 with an interval of 0.25, 0.5 or 1 pH unit. Each replica was run for 5 ns, resulting in an aggregate sampling time of 80 ns. For BACE1, 24 replicas were placed in the pH range 1–8 with an interval of 0.25 or 0.5 pH units. Each replica was run for 20 ns, resulting in an aggregate sampling time of 480 ns. For BACE2, 20 replicas were placed in the pH range 1.3–8 with an interval of 0.3 and 0.5 pH units. Each replica was run for 20 ns, resulting in an aggregate sampling time of 400 ns. For CatD, 24 replicas were placed in the pH range 1–8 with an interval of 0.25 or 0.5 pH units. Each replica was run for 30 ns, resulting in an aggregate sampling time of 720 ns. The last 3 ns of each replica were extracted for HEWL and SNase analysis, and the last 10 ns of each replica were extracted for BACE1, BACE2 and CatD analysis. The trajectories for BACE1 and CatD were taken from our previous work.[23,24]

All-atom CpHMD. In the all-atom CpHMD simulations, each protein was simulated for 30 ns per replica. For HEWL, 24 replicas were placed in the pH range -1–10.5 with an interval of 0.5 pH units. The aggregate sampling time was 700 ns. For SNase, 22 replicas were placed in the pH range -1–9.5 with an interval of 0.5 pH units. The aggregate sampling time was 660 ns. For BACE1, 24 replicas were placed in the pH range -2–8 with an interval of 0.25 and 0.5 pH units. The aggregate sampling time was 720 ns. The last 10 ns of each replica were extracted for analysis. The initial trajectories (10 ns per replica) for HEWL and SNase were taken from our previous work.[5]

Following our previous work,[5] a periodic boundary related correction was made to the calculated pK_{a}'s (Table S2).

$$\Delta pK_a^{\mathrm{corr}} = \pm \frac{2\pi\kappa\gamma^{\mathrm{solv}}Q}{3ln(10)RT}\left(\frac{N}{V} - \rho^{\mathrm{pure}}\right), \tag{1}$$

where ρ^{pure} is the number density of pure solvent, which is 0.0333679 Å^3 for water at

ambient temperature and pressure. κ is the electrostatic constant, N and V are the number of solvent molecules and volume of the periodic box, respectively. Q is the charge of the titratable site, i.e., -1 for Asp/Glu or +1 for His/Lys, and γ^{solv} is the quadrupole moment trace of the solvent model relative to a van der Waals interaction site (0.764 e·Å^2 for TIP3P water). The negative sign is for acid groups and positive sign is for basic groups.

Table S2: Finite-size corrections for the PME-based all-atom CpHMD simulations[a].

System	Volume (Å^3)	N^{water}	ΔpK$_a^{\mathrm{corr}}$	
			Asp	Glu
HEWL	2.3561$\times$10^5 $\pm$ 404	7343	-0.9	-0.9
SNase	2.4186$\times$10^5 $\pm$ 447	7515	-0.9	-0.9
BACE1	5.2280$\times$10^5 $\pm$ 844	15759	-1.2	-1.2

[a] Volume was averaged over all pH replicas based on the last 1 ns data. The volume for HEWL and SNase is the same as in our previous work.[5]

Supplementary Figures

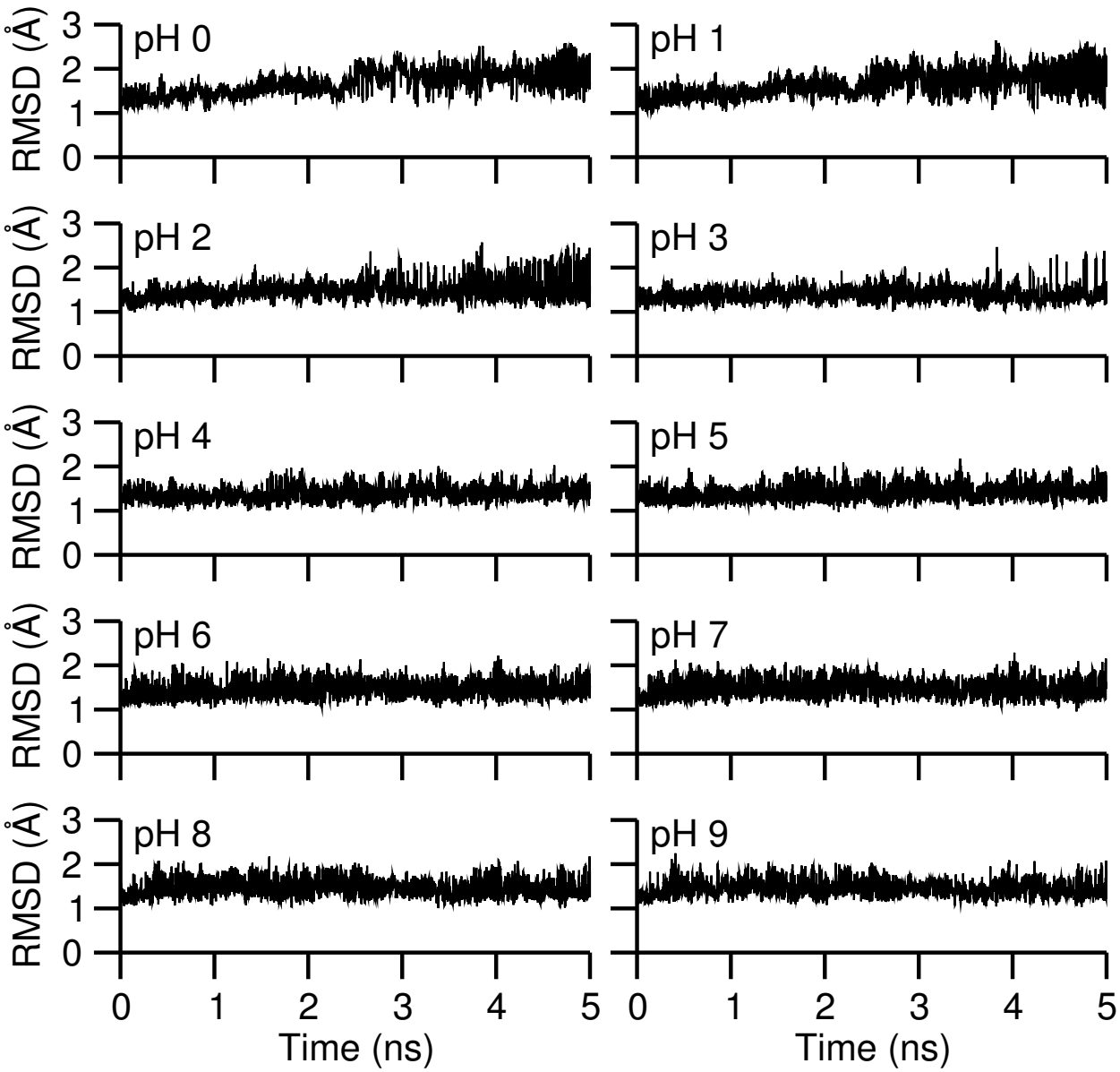

Figure S1: **Stability of HEWL in the hybrid-solvent CpHMD.** Time series of the backbone RMSD of HEWL at different pH conditions.

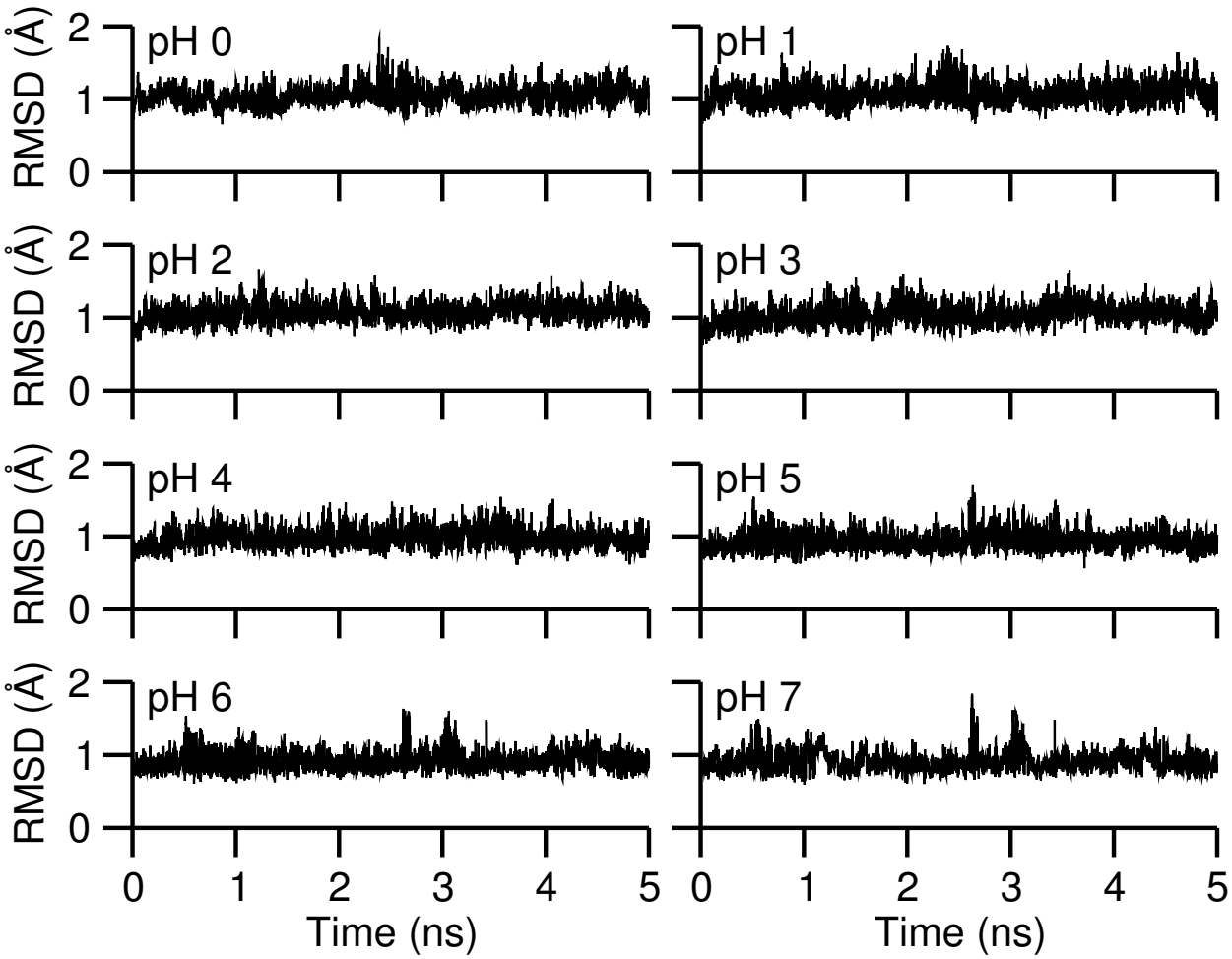

Figure S2: **Stability of SNase in the hybrid-solvent CpHMD.** Time series of the backbone RMSD of SNase at different pH conditions.

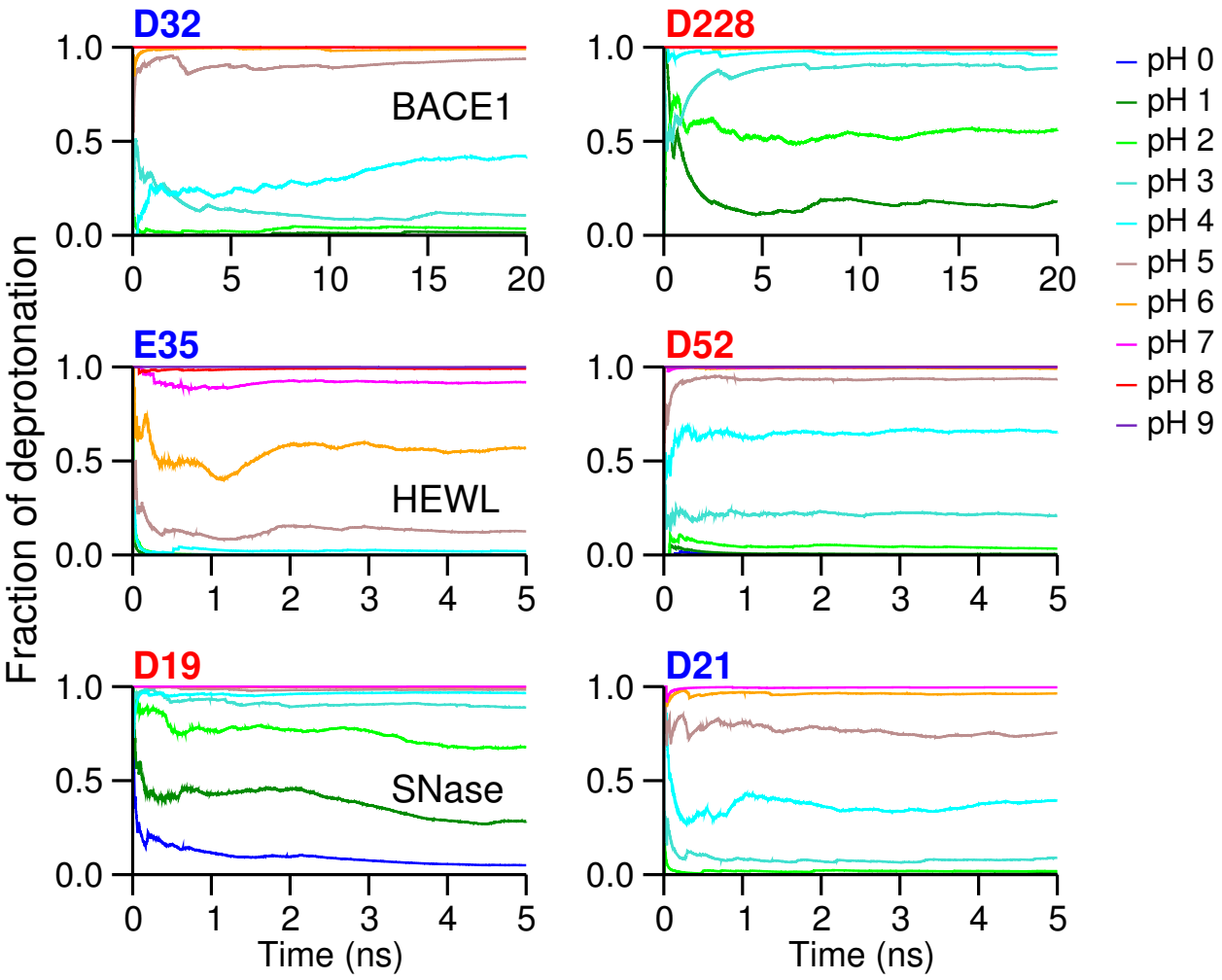

Figure S3: **Convergence of the unprotonated fractions in the hybrid-solvent CpHMD.** Unprotonated fractions of the dyad residues in HEWL, BACE1 and SNase as a function of simulation time in the hybrid-solvent CpHMD. For clarity, only integer pH conditions are shown. The unprotonated fraction is calculated cumulatively starting from the beginning of the simulation. The proton donor and nucleophile are labeled in blue and red, respectively.

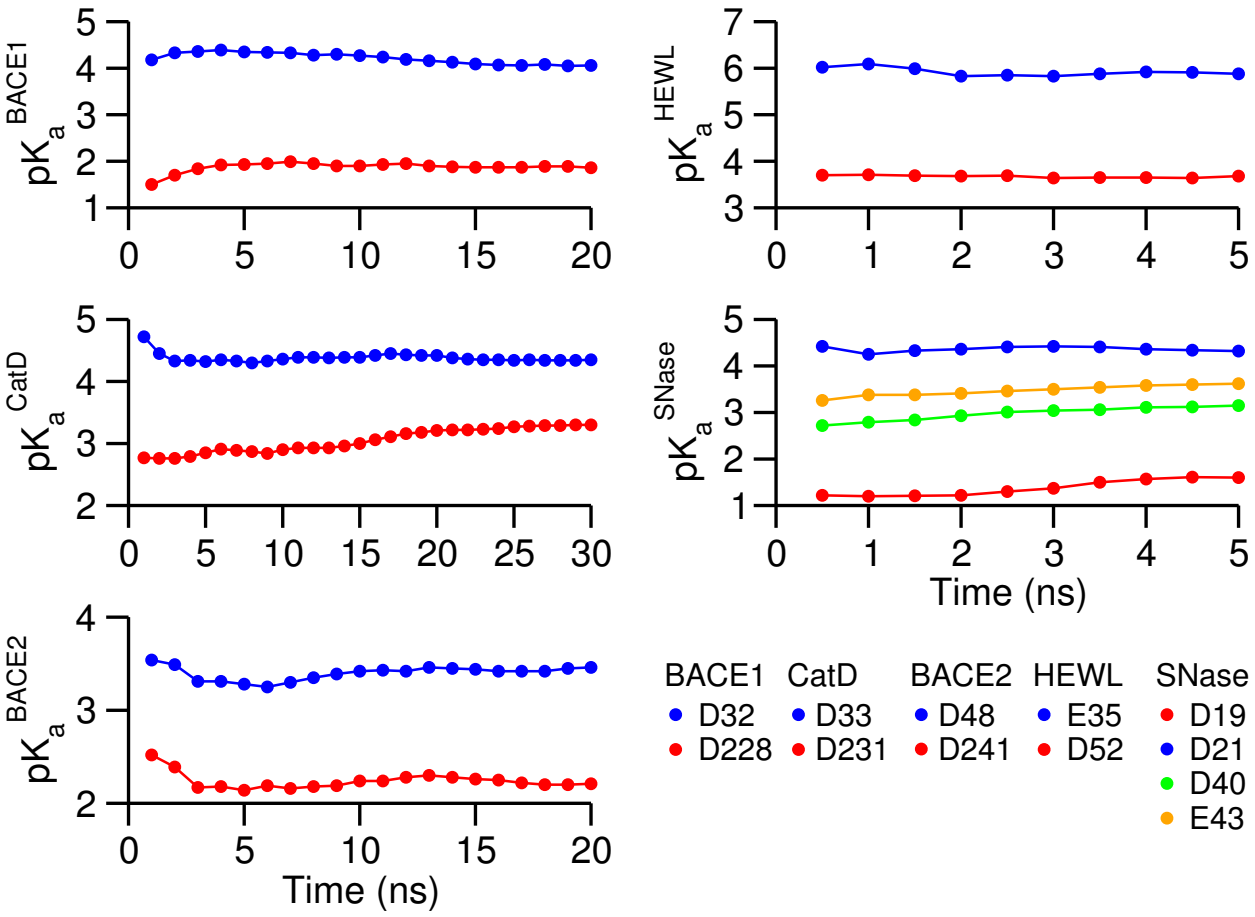

Figure S4: **Calculated pK_a's for the catalytic residues in HEWL, SNase, BACE1, CatD and BACE2 as a function of simulation time per replica in the hybrid-solvent CpHMD.**

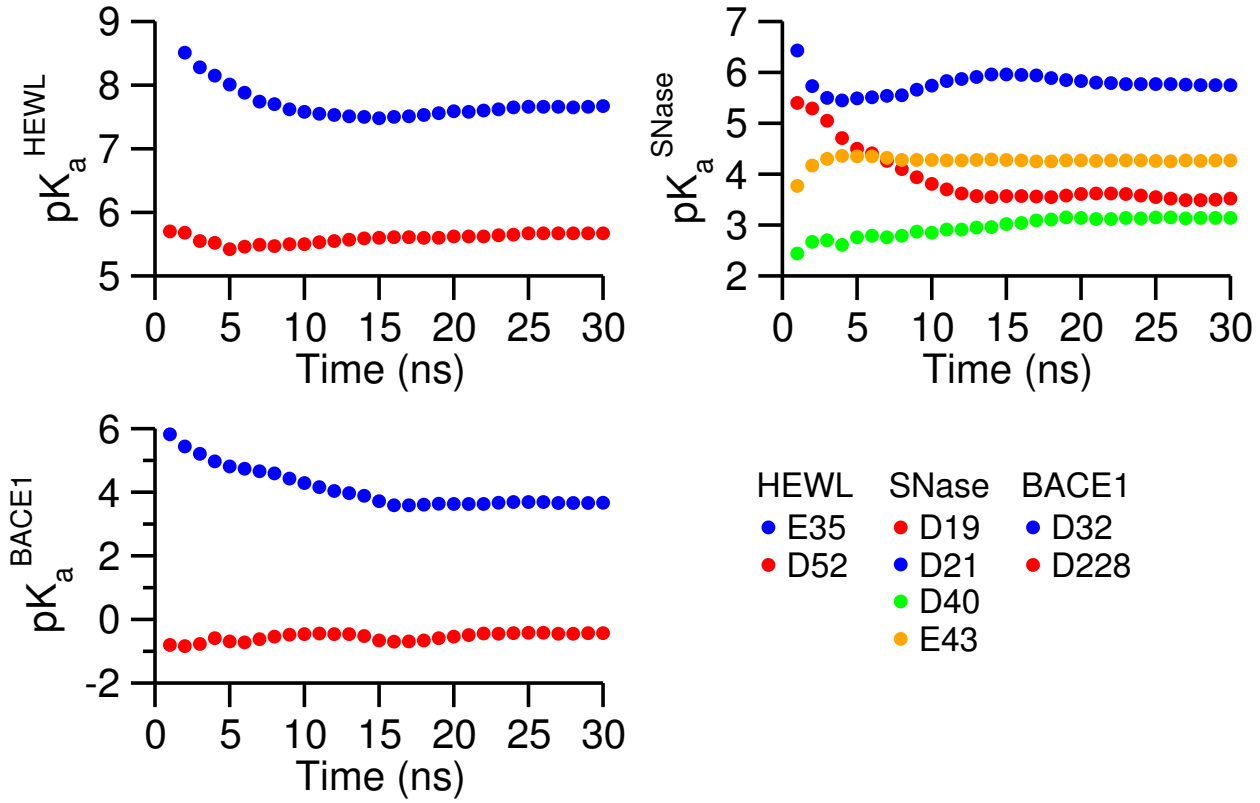

Figure S5: **Calculated p*K*$_a$'s for HEWL, SNase and BACE1 as a function of simulation time per replica in the all-atom CpHMD.**

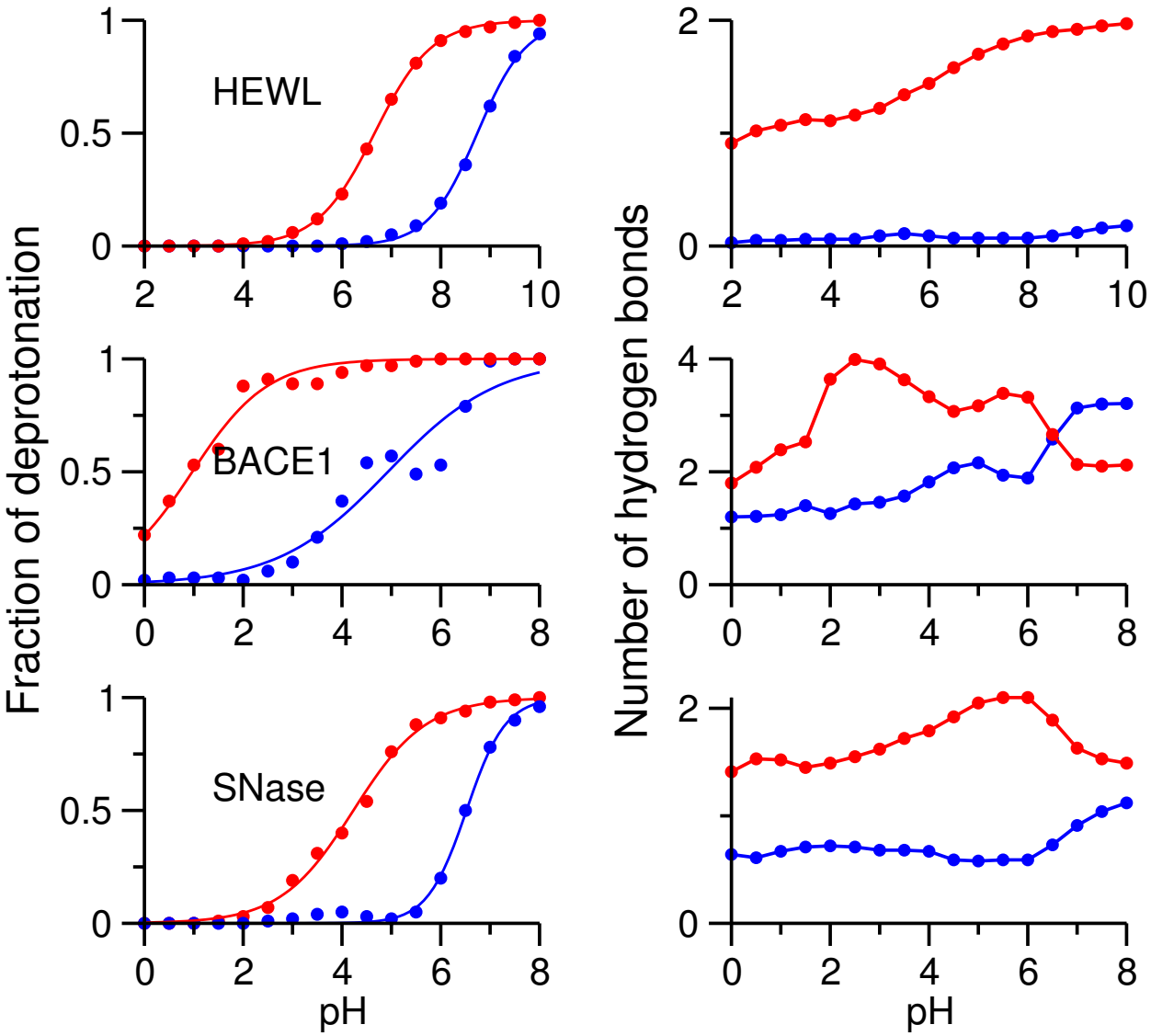

Figure S6: **Number of hydrogen bonds for the catalytic residues in HEWL, BACE1, and SNase from the all-atom CpHMD simulations.** A hydrogen bond is considered present if the donor-acceptor heavy-atom distance is within 3.5 Å and the donor-H-acceptor angle is greater than 150°. The proton donor and nucleophile are colored blue and red, respectively.

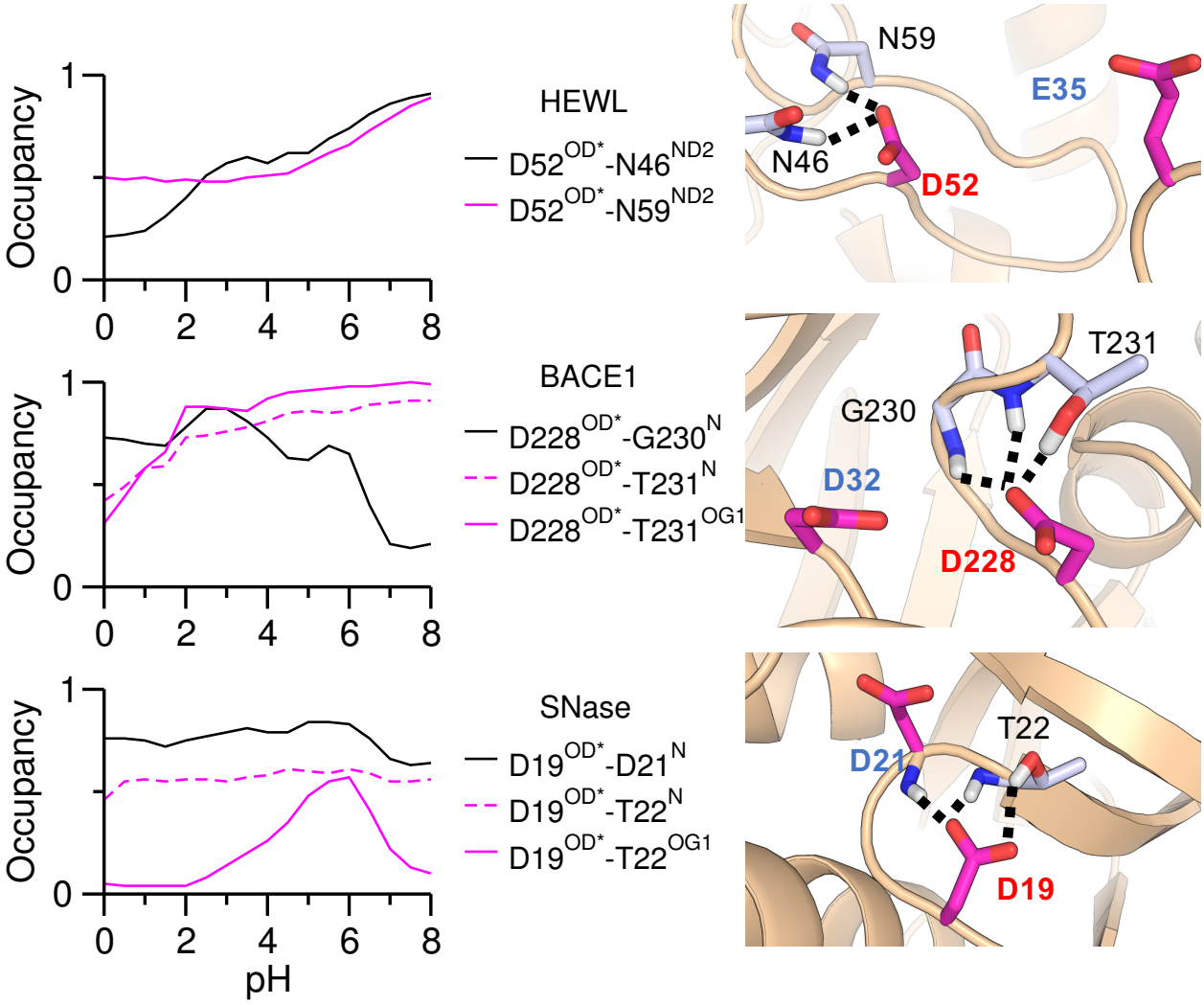

Figure S7: **Occupancy of the individual hydrogen bonds formed by the the catalytic nucleophile (base) in HEWL, SNase and BACE1 from the all-atom CpHMD simulations.** On the right panel, the proton donor and nucleophile are labeled in blue and red, respectively.

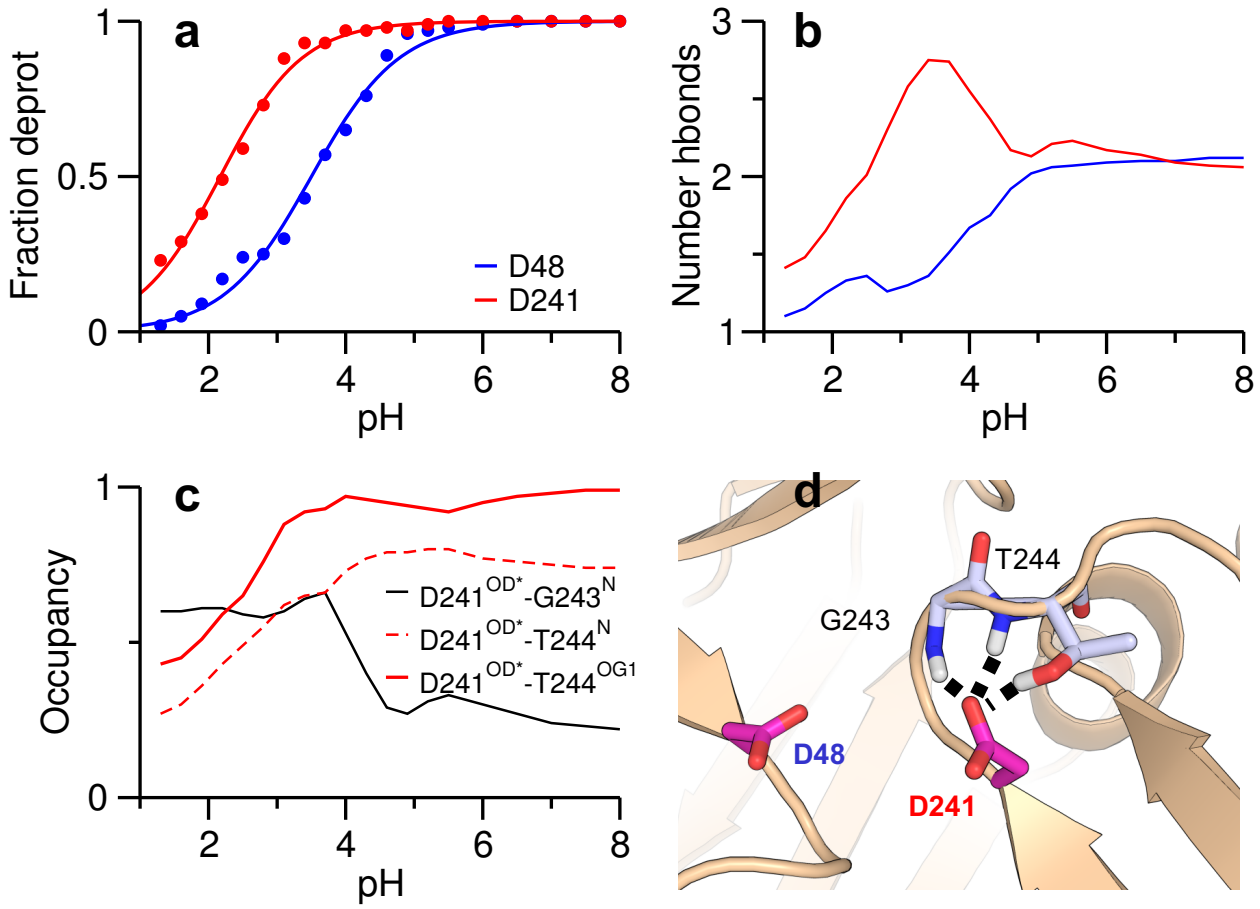

Figure S8: **Titration and hydrogen bond environment of the catalytic carboxylates in BACE2 from the hybrid-solvent CpHMD simulation.** a) Fraction of deprotonation as a function of pH for Asp48 (blue) and Asp241 (red). b) Total number of hydrogen bonds formed by Asp48 (red) and Asp241 (blue) as a function of pH. c) Occupancy of the individual hydrogen bonds formed by Asp241 as a function of pH. d) A snapshot showing the hydrogen bonds formed by Asp241. The catalytic proton donor (Asp48) and nucleophile (Asp241) are labeled in blue and red, respectively. A hydrogen bond is present if the donor-acceptor distance is closer than 3.5 Å and the donor-H-acceptor angle is larger than 150°.

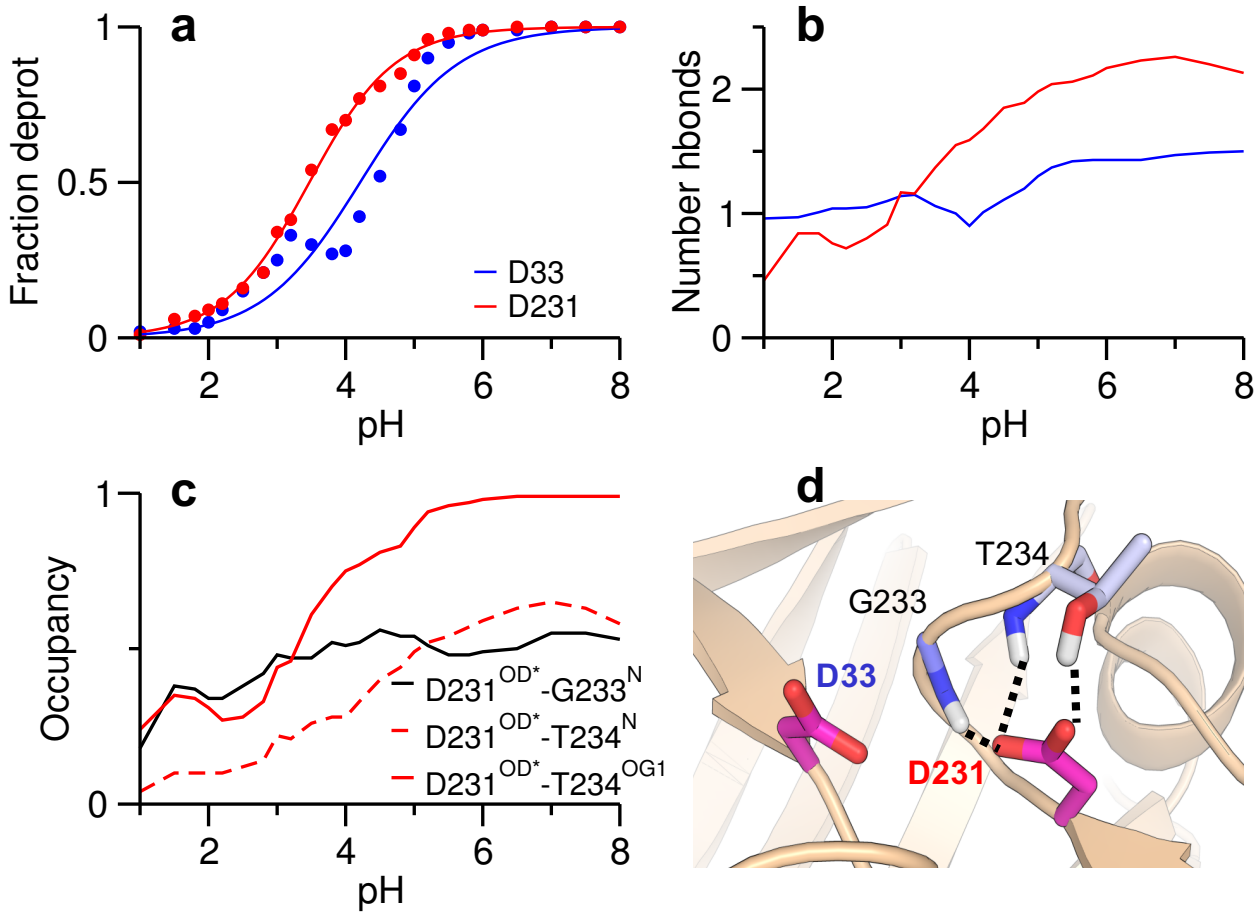

Figure S9: **Titration and hydrogen bond environment of the catalytic carboxylates in CatD from the hybrid-solvent CpHMD simulation.** a) Deprotonation fractions as a function of pH for Asp33 (blue) and Asp231 (red). b) Total number of hydrogen bonds formed by Asp33 (blue) and Asp231 (red) as a function of pH. c) Occupancy of the individual hydrogen bonds formed by Asp231 as a function of pH. d) A snapshot showing the hydrogen bonds formed by Asp231. The catalytic proton donor (Asp33) and nucleophile (Asp231) are labeled in blue and red, respectively. A hydrogen bond is present if the donor-acceptor distance is closer than 3.5 Å and the donor-H-acceptor angle is larger than 150°.

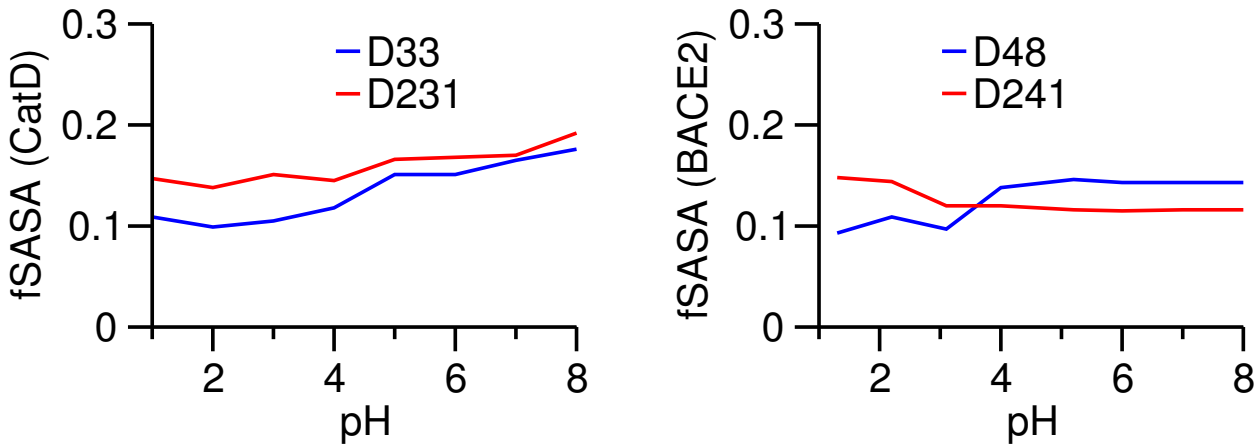

Figure S10: **Fractional solvent accessible surface area (SASA) of the catalytic carboxylates in CatD and BACE2 from the hybrid-solvent CpHMD simulations.** Fractional SASA was calculated as the ratio between the SASA of the carboxylate oxygens in the protein environment and that in solution. The data for the proton donor and nucleophile are colored blue and red, respectively.

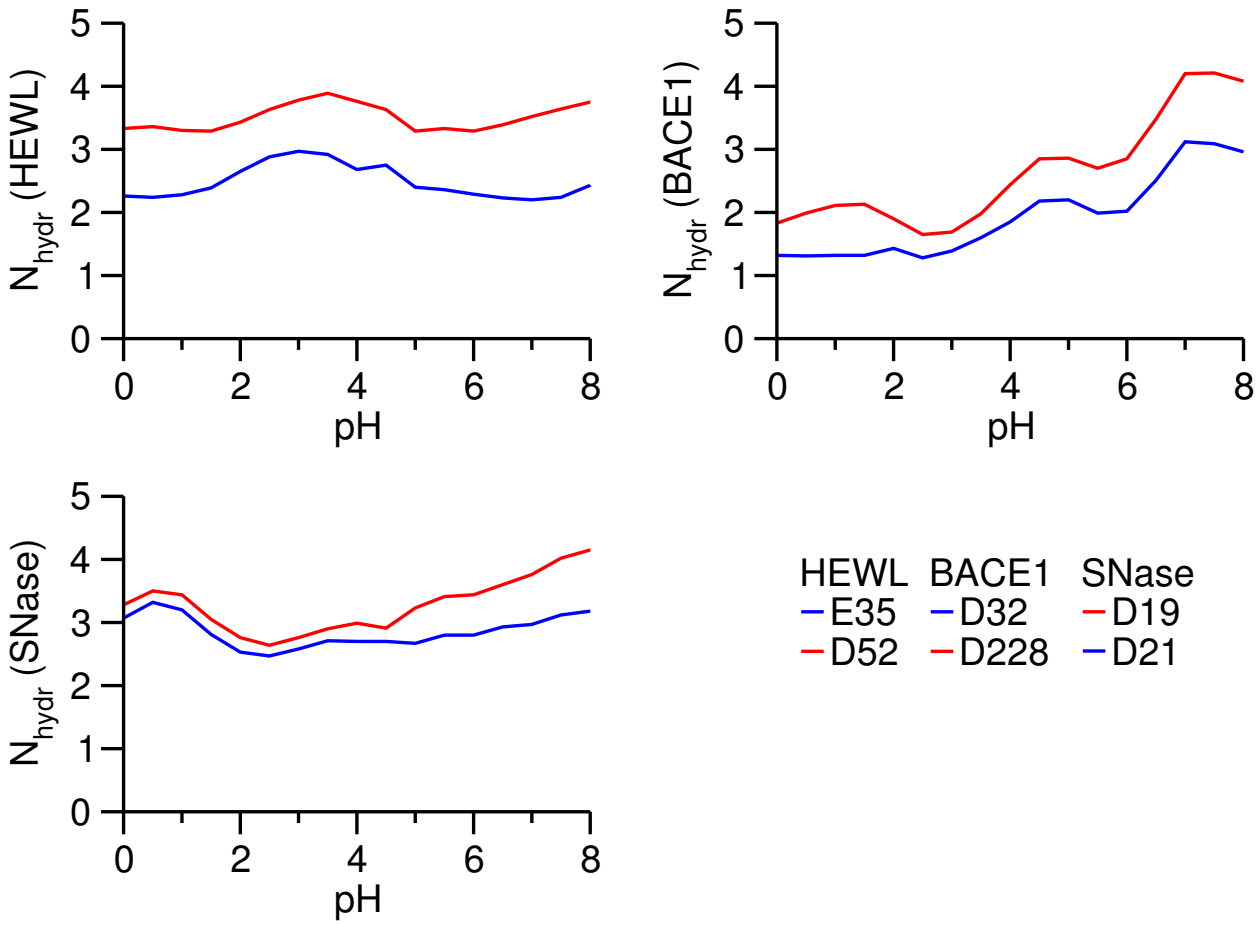

Figure S11: **Hydration number of the carboxylate dyad in HEWL, BACE1 and SNase as a function of pH from the all-atom CpHMD simulations.** Hydration number was calculated as the number of water (oxygen atom) within 3.5 Å of the carboxylate oxygen atoms. The data for the proton donor and nucleophile are colored blue and red, respectively.

References

(1) Brünger, A. T.; Karplus, M. Polar hydrogen positions in proteins: Empirical energy placement and neutron diffraction comparison. *Proteins* **1988**, *4*, 148–156.

(2) Brooks, B. R.; Brooks, III, C. L.; Mackerell, Jr., A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S. et al. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.

(3) Khandogin, J.; Brooks, III, C. L. Constant pH molecular dynamics with proton tautomerism. *Biophys. J.* **2005**, *89*, 141–157.

(4) Wallace, J. A.; Shen, J. K. Continuous constant pH molecular dynamics in explicit solvent with pH-based replica exchange. *J. Chem. Theory Comput.* **2011**, *7*, 2617–2629.

(5) Huang, Y.; Chen, W.; Wallace, J. A.; Shen, J. All-Atom Continuous Constant pH Molecular Dynamics With Particle Mesh Ewald and Titratable Water. *J. Chem. Theory Comput.* **2016**, *12*, 5411–5421.

(6) Khandogin, J.; Brooks, III, C. L. Toward the accurate first-principles prediction of ionization equilibria in proteins. *Biochemistry* **2006**, *45*, 9363–9373.

(7) Wallace, J. A.; Shen, J. K. Charge-leveling and proper treatment of long-range electrostatics in all-atom molecular dynamics at constant pH. *J. Chem. Phys.* **2012**, *137*, 184105.

(8) Chen, W.; Wallace, J.; Yue, Z.; Shen, J. Introducing titratable water to all-atom molecular dynamics at constant pH. *Biophys. J.* **2013**, *105*, L15–L17.

(9) Webb, H.; Tynan-Connolly, B. M.; Lee, G. M.; Farrell, D.; O'Meara, F.; Søndergaard, C. R.; Teilum, K.; Hewage, C.; McIntosh, L. P.; Nielsen, J. E. Remeasuring

HEWL pK_a values by NMR spectroscopy: methods, analysis, accuracy, and implications for theoretical pK_a calculations. *Proteins* **2011**, *79*, 685–702.

(10) Castañeda, C. A.; Fitch, C. A.; Majumdar, A.; Khangulov, V.; Schlessman, J. L.; García-Moreno E., B. Molecular determinants of the pK_a values of Asp and Glu residues in staphylococcal nuclease. *Proteins* **2009**, *77*, 570–588.

(11) Toulokhonova, L.; Metzler, W. J.; Witmer, M. R.; Copeland, R. A.; Marcinkeviciene, J. Kinetic studies on β-site amyloid precursor protein-cleaving enzyme (BACE). *J. Biol. Chem.* **2003**, *278*, 4582–4589.

(12) MacKerell, Jr., A. D.; Bashford, D.; Bellott, M.; Dunbrack, Jr., R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S. et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(13) MacKerell, Jr., A. D.; Feig, M.; Brooks, III, C. L. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25*, 1400–1415.

(14) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of *n*-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

(15) Hoover, W. G. Canonical dynamics: Equilibration phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695–1697.

(16) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **1995**, *103*, 4613–4621.

(17) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(18) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(19) Im, W.; Lee, M. S.; Brooks, III, C. L. Generalized Born model with a simple smoothing function. *J. Comput. Chem.* **2003**, *24*, 1691–1702.

(20) Chen, J.; Im, W.; Brooks, III, C. L. Balancing solvation and intramolecular interactions: toward a consistent generalized Born force field. *J. Am. Chem. Soc.* **2006**, *128*, 3728–3736.

(21) Baldwin, E. T.; Bhat, T. N.; Gulnik, S.; Hosur, M. V.; II, R. C. S.; Cachau, R. E.; Collins, J.; Silva, A. M.; Erickson, J. W. Crystal structures of native and inhibited forms of human cathepsin D: implications for lysosomal targeting and drug design. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 6796–6800.

(22) Banner, D. W.; Gsell, B.; Benz, J.; Bertschinger, J.; Burger, D.; Brack, S.; Cuppuleri, S.; Debulpaep, M.; Gast, A.; Grabulovski, D. et al. Mapping the conformational space accessible to BACE2 using surface mutants and cocrystals with Fab fragments, Fynomers and Xaperones. *Acta Crystallogr. D* **2013**, *69*, 1124–1137.

(23) Ellis, C. R.; Shen, J. pH-Dependent Population Shift Regulates BACE1 Activity and Inhibition. *J. Am. Chem. Soc.* **2015**, *137*, 9543–9546.

(24) Ellis, C. R.; Tsai, C.-C.; Lin, F.-Y.; Shen, J. Conformational dynamics of cathepsin D and binding to a small-molecule BACE1 inhibitor. *J. Comput. Chem.* **2017**, *38*, 1260–1269.