Supporting Information

pH-dependent population shift regulates BACE1 activity and inhibition

Christopher R. Ellis and Jana Shen*

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

*Corresponding author. Phone: (410) 706-4187; Fax: (410) 706-5017; E-mail: jshen@rx.umaryland.edu.

1 Methods and Protocols

The CpHMD method

Continuous constant pH molecular dynamics (CpHMD)[1,2,3,4,5,6] is a method based on the λ-dynamics approach for free energy simulations[7]. The uniqueness of CpHMD is that it simultaneously samples both conformational and titration degrees of freedom at a given pH with very little additional computational cost[1,2]. In hybrid-solvent CpHMD, conformational dynamics is propagated in explicit solvent, while titration coordinates are propagated based on conformations sampled in explicit solvent and solvation free energies calculated with the generalized-Born (GB) implicit-solvent model. The particular GB model we use is GBSW[8]. To accelerate convergence of the coupled conformational and protonation-state sampling, a replica-exchange protocol in the pH space is used[4]. More discussion of CpHMD and similar methods can be found in the recent reviews[9,10].

Structure preparation and equilibration

All simulations were performed using the CHARMM simulation package (version C37b[11]). The crystal structures with PDB ID 1SGZ and 1FKN were used as the starting configurations for the *apo* and *holo* BACE1 simulations, respectively. All water molecules in the crystal structures were kept, and hydrogen atoms were added using the HBUILD facility in CHARMM. Each system was solvated in an octahedral box with a distance of at least 10 Å between each side of the protein and the edge of the water box. The total number of atoms in the system is about 55,000. Both BACE1 and the inhibitor were represented by the CHARMM22/CMAP all-atom force field[12,13] For the inhibitor, the parameters for Leu*Ala were adapted from analogous amino acids in CHARMM22 (Table S2). The modified TIP3P water model was used to represent water[11].

Each system was first energy minimized using the steepest descent followed by the Newton-Raphson algorithm. The system was then gradually heated over the course of 120 ps from 100 K to 300 K with the protein heavy atoms restrained using a harmonic potential with a force constant of 5 kcal/mol/Å^2 . Following heating, the system was equilibrated for 180 p 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 (100 ps) . Lastly, the system was equilibrated for 100 ps without restraints.

Production simulations

Production simulations were performed with the hybrid-solvent CpHMD coupled with the pH replica-exchange protocol.[4] Each pH replica was simulated in the NPT ensemble at a temperature of 300 K and pressure of 1 atm. Temperature was maintained using a modified Hoover thermostat method[14], while pressure was maintained using the Langevin piston coupling method[15]. The van der Waals interactions were smoothly switched to zero between 10 and 12 Å. The particle mesh Ewald method[16] was used to calculate long-range electrostatics, with a real-space cutoff of 12 Å and a sixth-order interpolation with 1.6 $Å^{-1}$ grid spacing. The SHAKE algorithm was used to constrain bonds involving hydrogen to enable a 2-fs timestep. A GB calculation was invoked every 10 MD steps to update the titration coordinates. In the GB calculation, the default settings were used, consistent with our previous work[4]. The GB input radii were taken from Chen et al.[17]

The *apo* and *holo* proteins were simulated for 21 and 11 ns per pH replica, respectively. In each simulation, all Asp, Glu, and His sidechains were allowed to titrate, while all Arg and Lys sidechains were modeled in the charged state because their model pK_{a} values (12.5 and 10.4 respectively) are significantly higher than the simulation pH range (1—8 or -0.5–9.0). The first nanosecond from each simulation was discarded. The remaining frames were recorded following each exchange attempt (every 500 MD steps or 1 ps) and used for analysis. The *apo* protein was simulated using 24 pH replicas in the pH range 1–8, while the *holo* form was simulated using 32 pH replicas in the pH range -0.5–9.0. The *holo* simulations required additional replicas to capture the full titration of the active site. The pH conditions of the 24 replicas of the *apo* simulations are shown in Figure S1, and the additional replicas used in the *holo* simulations are listed in the caption. The average exchange ratio was 45%.

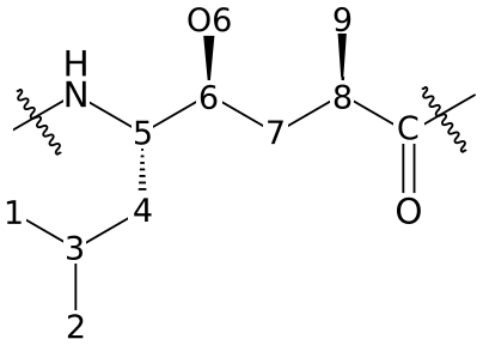
2 Supplementary Tables

Table S1: Crystal structures of BACE1 and its complex with inhibitor OM99-2 resolved at different pH conditions

PDB ID	pH	R	φ
1SGZ[18]	6.5	11.3	10.8
1FKN (OM99-2)[19]	7.4	6.3	-0.8
2ZHV[20]	7.0	7.6	31.2
2ZHU[20]	5.0	7.6	42.5
2ZHT[20]	4.5	7.6	37.6
2ZHS[20]	4.0	7.7	30.3
2ZHR (OM99-2)[20]	5.0	6.6	-0.2

R and φ are the two order parameters used in our study to characterize the conformation of the active site . 1SGZ and 1FKN are the starting structures used in our simulations. 2ZHV, 2ZHU, 2ZHT, 2ZHS and 2ZHR are the crystal structures discussed in the work of Shimizu et al.[20] These crystal structures adopt similar configurations, i.e., RMS deviations between them are lower than 1 Å, R and φ values are similar. Thus, they do not provide structural evidence for the conformation switch that regulates the activity and inhibition of BACE1.

Table S2: Atom types and partial charges for the group Leu*Ala in OM99-2



Name	Type	Charge	Name	Type	Charge	Name	Type	Charge
N	NH1	-0.47	C4	CT2	-0.18	C6	CT1	0.14
HN	H	0.31	H41	HA	0.09	H61	HA	0.09
C5	CT1	0.07	H42	HA	0.09	O6	OH1	-0.66
H51	HB	0.09				HO6	H	0.43
C	C	0.51	C3	CT1	-0.09	C7	CT2	-0.18
O	O	-0.51	H31	HA	0.09	H71	HA	0.09
						H71	HA	0.09
			C2	CT3	-0.27	C8	CT1	-0.09
			H21	HA	0.09	H81	HB	0.09
			H22	HA	0.09			
			H23	HA	0.09			
			C1	CT3	-0.27	C9	CT3	-0.27
			H11	HA	0.09	H91	HA	0.09
			H12	HA	0.09	H92	HA	0.09
			H13	HA	0.09	H93	HA	0.09

The atom types and partial charges are adapted from the sidechains of Leu, Thr (hydroxyl) and Ala in the CHARMM22 protein force field.[12]

3 Supplementary Figures

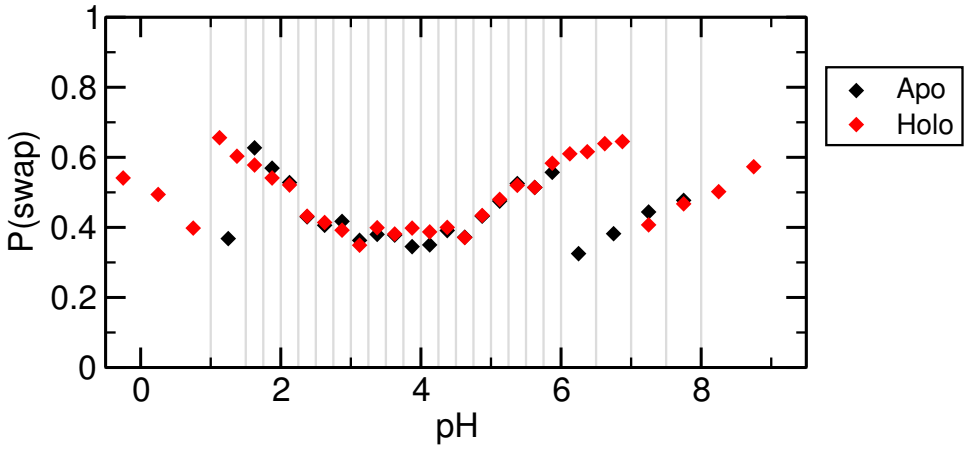


Figure S1: **Acceptance rate of replica-exchange attempts.** The black and red diamonds present the probabilities of exchange between adjacent replicas from the *apo* and *holo* BACE1 simulations, respectively. The vertical grey lines represent the 24 pH replicas from the *apo* simulation. The *holo* simulation has eight additional replicas at pH -0.5, 0.0, 0.5, 1.25, 6.25, 6.75, 8.50, 9.00. The average exchange rates across replicas in the *apo* and *holo* simulations are 44$\pm$8 and 49$\pm$9 %, respectively.

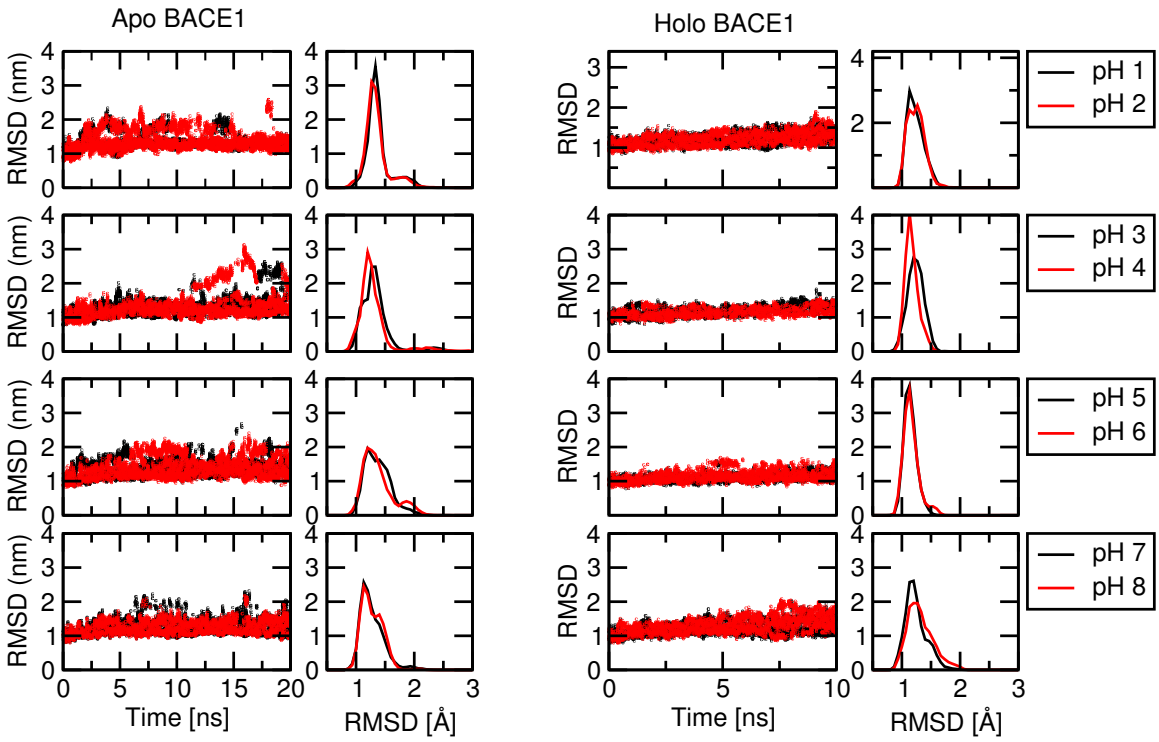


Figure S2: **BACE1 stability**. The time series of backbone heavy atom RMSD and corresponding probability distributions of *apo* and *holo* BACE1 are displayed in the two left and two right columns, respectively. Plots on the same row are from two different pH replicas as indicated in the legend box. Although only 8 replicas are shown, the RMSD remained below 2.5 Å for all pH conditions during the entire simulation time.

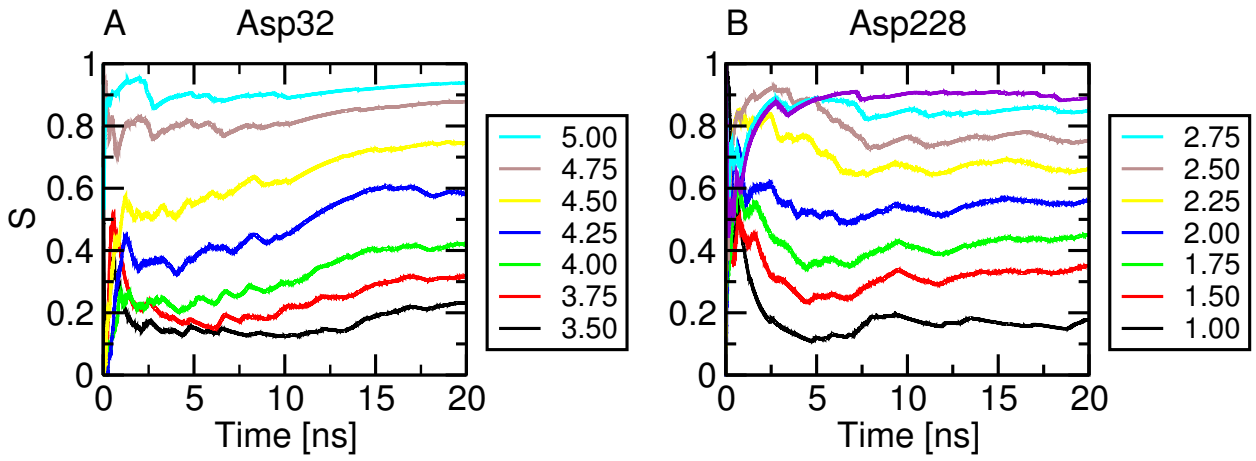


Figure S3: **Convergence of unprotonated fractions of the catalytic dyad in *apo* BACE1**. The cumulatively calculated fraction of the unprotonated states (S) of Asp32 and Asp228 is shown in the left and right panels, respectively.

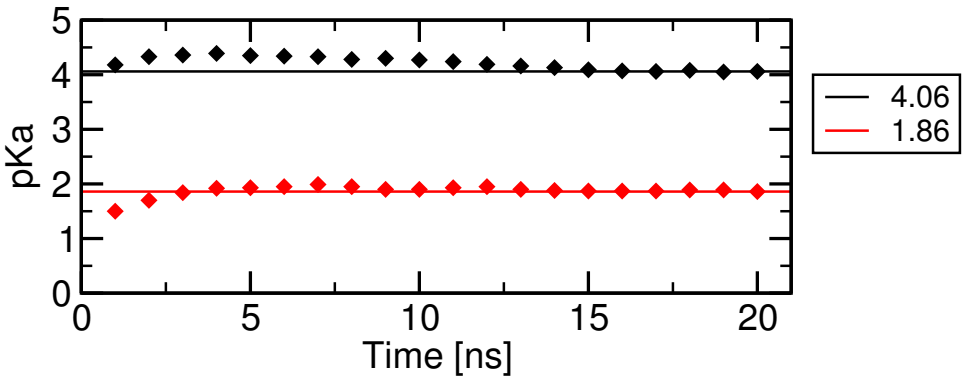


Figure S4: **Convergence of the calculated pK_a's of the catalytic dyad in *apo* BACE1 .** The cumulatively calculated pK_a's of Asp32 (black) and Asp228 (red) as a function of the simulation time. The legend reports the final calculated pK_a's. The calculated pK_a's for the two residues remain nearly constant between 15 ns and 20 ns.

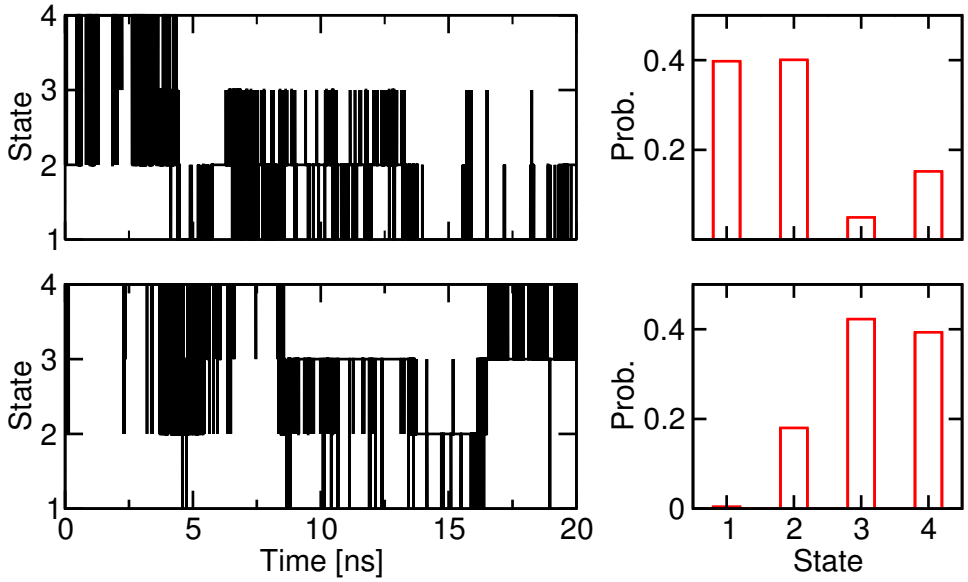


Figure S5: **Free energy surface (FES) sampling.** The left and right panels present time traces and populations of two replicas through the various regions of the FES. States 1, 2, 3, and 4 represent the Tyr-inhibited, binding-competent, Gln-inhibited, and diffuse states, respectively.

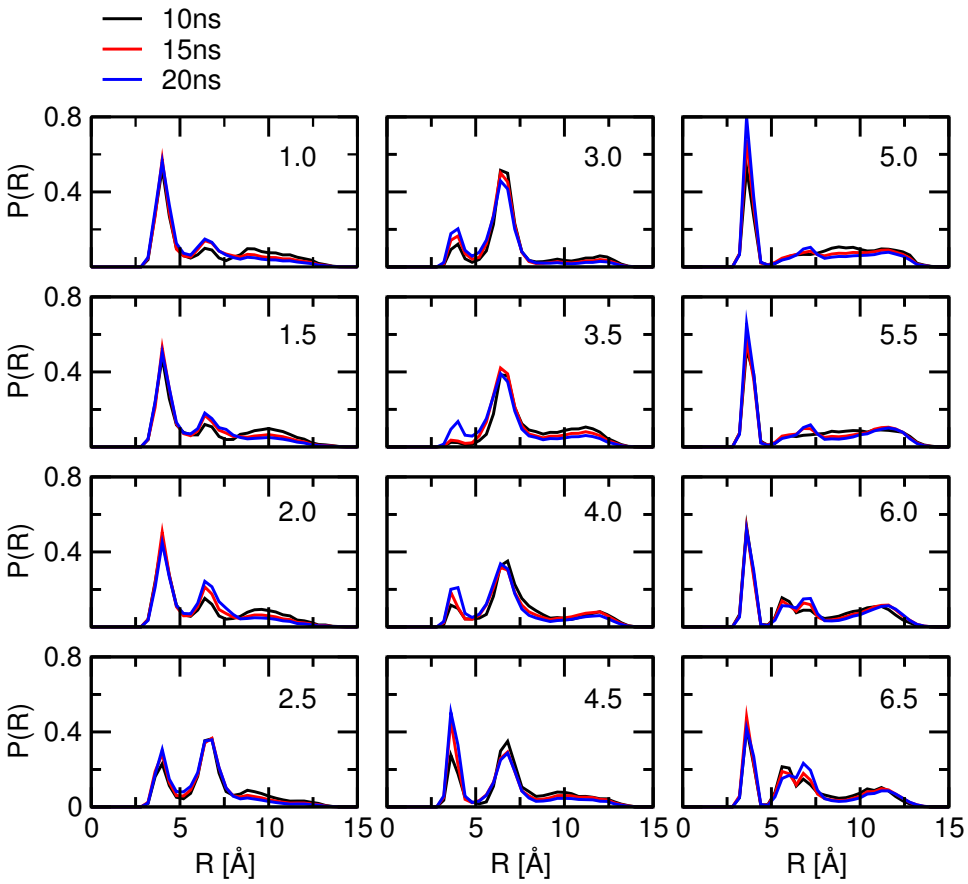


Figure S6: **Convergence of** R. Histograms of R after 10, 15, and 20 ns of simulation at various pH conditions.

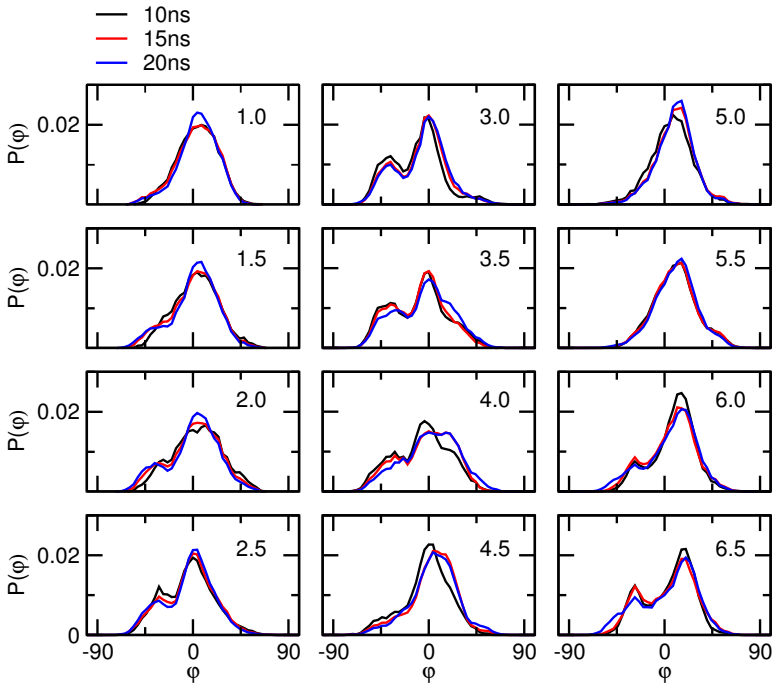


Figure S7: **Convergence of** φ. Histograms of φ after 10, 15, and 20 ns of simulation at various pH conditions.

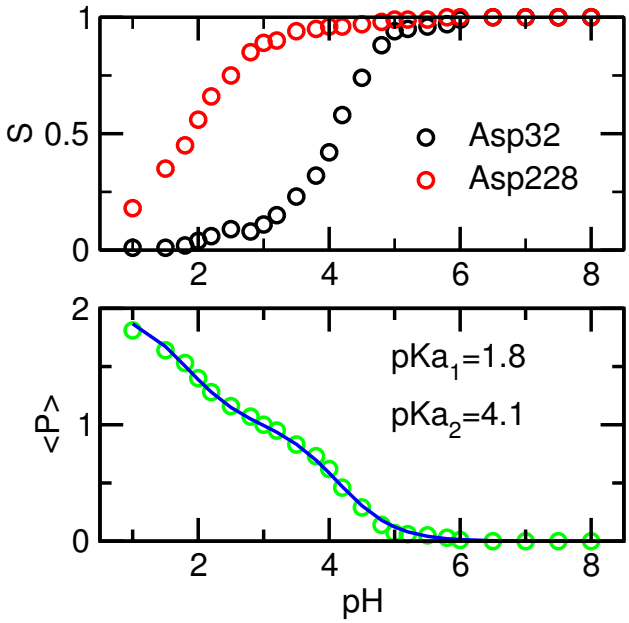


Figure S8: **Macroscopic pK_{a}'s**. Top. Unprotonated fractions of Asp32 and Asp228 obtained from the *apo* simulations. Bottom. The average number of total bound protons <P> at different pH. Solid curve is the best fit to Eq. 24 from Ref.[5] The two corresponding macroscopic pK_{a} values, pK_{a1} and pK_{a2}, are indicated.

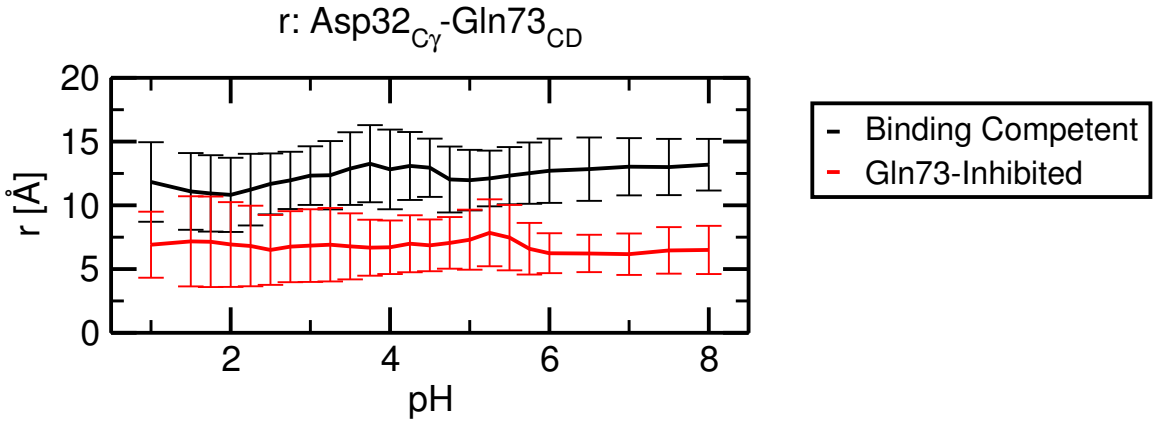


Figure S9: **Interaction between Asp32 and Gln73**. The average (solid lines) and RMSF (error bars) of the distance between Gln73:CD and Asp32:Cγ in the binding-competent (black) and Gln-inhibited (red) states at different pH.

References

1. Lee, M. S.; Salsbury, Jr., F. R.; Brooks III, C. L. *Proteins* **2004**, *56*, 738–752.
2. Khandogin, J.; Brooks III, C. L. *Biophys. J.* **2005**, *89*, 141–157.
3. Khandogin, J.; Brooks III, C. L. *Biochemistry* **2006**, *45*, 9363–9373.
4. Wallace, J. A.; Shen, J. K. *J. Chem. Theory Comput.* **2011**, *7*, 2617–2629.
5. Wallace, J. A.; Shen, J. K. *J. Chem. Phys.* **2012**, *137*, 184105.
6. Chen, W.; Wallace, J.; Yue, Z.; Shen, J. *Biophys. J.* **2013**, *105*, L15–L17.
7. Kong, X.; Brooks, III, C. L. *J. Chem. Phys.* **1996**, *105*, 2414–2423.
8. Im, W.; Lee, M. S.; Brooks III, C. L. *J. Comput. Chem.* **2003**, *24*, 1691–1702.
9. Wallace, J. A.; Shen, J. K. *Methods Enzymol.* **2009**, *466*, 455–475.
10. Chen, W.; Morrow, B. H.; Shi, C.; Shen, J. K. *Mol. Simulat.* **2014**, *40*, 830–838.
11. Brooks, B. R. et al. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
12. MacKerell Jr., A. D. et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
13. MacKerell, Jr., A. D.; Feig, M.; Brooks, III, C. L. *J. Am. Chem. Soc.* **2004**, *126*, 698–699.
14. Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.
15. Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.* **1995**, *103*, 4613–4621.
16. Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
17. Chen, J.; Im, W.; Brooks III, C. L. *J. Am. Chem. Soc.* **2006**, *128*, 3728–3736.
18. Hong, L.; Tang, J. *Biochemistry* **2004**, *43*, 4689–4695.
19. Hong, L.; Koelsch, G.; Lin, X.; Wu, S.; Terzyan, S.; Ghosh, A. K.; Zhang, X. C.; Tang, J. *Science* **2000**, *290*, 150–153.
20. Shimizu, H.; Tosaki, A.; Kaneko, K.; Hisano, T.; Sakurai, T.; Nukina, N. *Mol. Cell. Biol.* **2008**, *28*, 3663–3671.

4 Full citation list

1a) Hussain, I.; Powell, D.; Howlett, D. R.; Tew, D. G.; Meek, T. D.; Chapman, C.; Gloger, I. S.; Murphy, K. E.; Southan, C. D.; Ryan, D. M.; Smith, T. S.; Simmons, D. L.; Walsh, F. S.; Dingwall, C.; Christie, G. *Mol. Cell. Neurosci.* **1999**, *14*, 419–427

1b) Sinha, S.; Anderson, J. P.; Barbour, R.; Basi, G. S.; Caccavello, R.; Davis, D.; Doan, M.; Dovey, H. F.; Frigon, N.; Hong, J.; Jacobson-Croak, K.; Jewett, N.; Keim, P.; Knops, J.; Lieberburg, I.; Power, M.; Tan, H.; Tatsuno, G.; Tung, J.; Schenk, D.; Seubert, P.; Suomensaari, S. M.; Wang, S.; Walker, D.; Zhao, J.; McConlogue, L.; John, V. *Nature* **1999**, *402*, 537–540

1c) Vassar, R.; Bennett, B. D.; Babu-Khan, S.; Kahn, S.; Mendiaz, E. A.; Denis, P.; Teplow, D. B.; Ross, S.; Amarante, P.; Loeloff, R.; Luo, Y.; Fisher, S.; Fuller, J.; Edenson, S.; Lile, J.; Jarosinski, M. A.; Biere, A. L.; Curran, E.; Burgess, T.; Louis, J.-C.; Collins, F.; Treanor, J.; Rogers, G.; Citron, M. *Science* **1999**, *286*, 735–741

2b) Bateman, R. J.; Xiong, C.; Benzinger, T. L. S.; Fagan, A. M.; Goate, A.; Fox, N. C.; Marcus, D. S.; Cairns, N. J.; Xie, X.; Blazey, T. M.; Holtzman, D. M.; Santacruz, A.; Buckles, V.; Oliver, A.; Moulder, K.; Aisen, P. S.; Ghetti, B.; Klunk, W. E.; McDade, E.; Martins, R. N.; Masters, C. L.; Mayeux, R.; Ringman, J. M.; Rossor, M. N.; Schofield, P. R.; Sperling, R. A.; Salloway, S.; Morris, J. C. *N. Engl. J. Med.* **2012**, *367*, 795–804

4a) Luo, Y.; Bolon, B.; Kahn, S.; Bennett, B. D.; Babu-Khan, S.; Denis, P.; Fan, W.; Kha, H.; Zhang, J.; Gong, Y.; Martin, L.; Louis, J.-C.; Yan, Q.; Richards, W. G.; Citron, M.; Vassar, R. *Nature Neurosci.* **2001**, *4*, 231–232

4b) Nishitomi, K.; Sakaguchi, G.; Horikoshi, Y.; Gray, A. J.; Maeda, M.; Hirata-Fukae, H.-F.; Becker, A. G.; Hosono, M.; Sakaguchi, I.; Minami, S. S.; Nakajima, Y.; Li, H.-F.; Takeyama, C.; Kihara, T.; Ota, A.; Wong, P. C.; Aisen, P. S.; Kato, A.; Kinoshita, N.; Matsuoka, Y. *J. Neurochem.* **2006**, *99*, 1555–1563

12) Brooks, B. R.; Brooks III; C. L.; Mackerell, Jr., A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartles, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545–1614