Supplementary Information

Exploring Histidine Conformations in the M2 Channel Lumen of the Influenza A Virus at Neutral pH via Molecular Simulations

Hao Dong,[†¶] Giacomo Fiorin,[†¶] William F. DeGrado,[‡] Michael L. Klein[†*]

[†] Institute for Computational Molecular Science, Temple University, 1900 North 12th Street, Philadelphia, Pennsylvania 19122-6078

[‡] Department of Pharmaceutical Chemistry, 555 Mission Bay Blvd South, San Francisco, CA 94158-9001

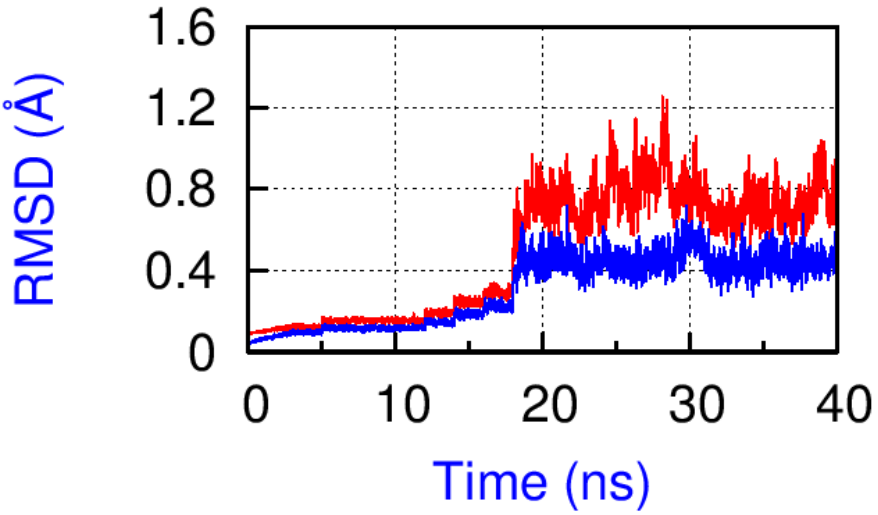

Figure S1. The root mean square deviations (RMSDs) of the backbone Cα atoms from the starting structure in two systems in the classic MD simulations. The "dimer-of-dimers" configuration is in red and the "His-box" configuration is in blue. The RMSDs saturate to a value less than 1 Å, indicating that the protein backbone have been fully relaxed.

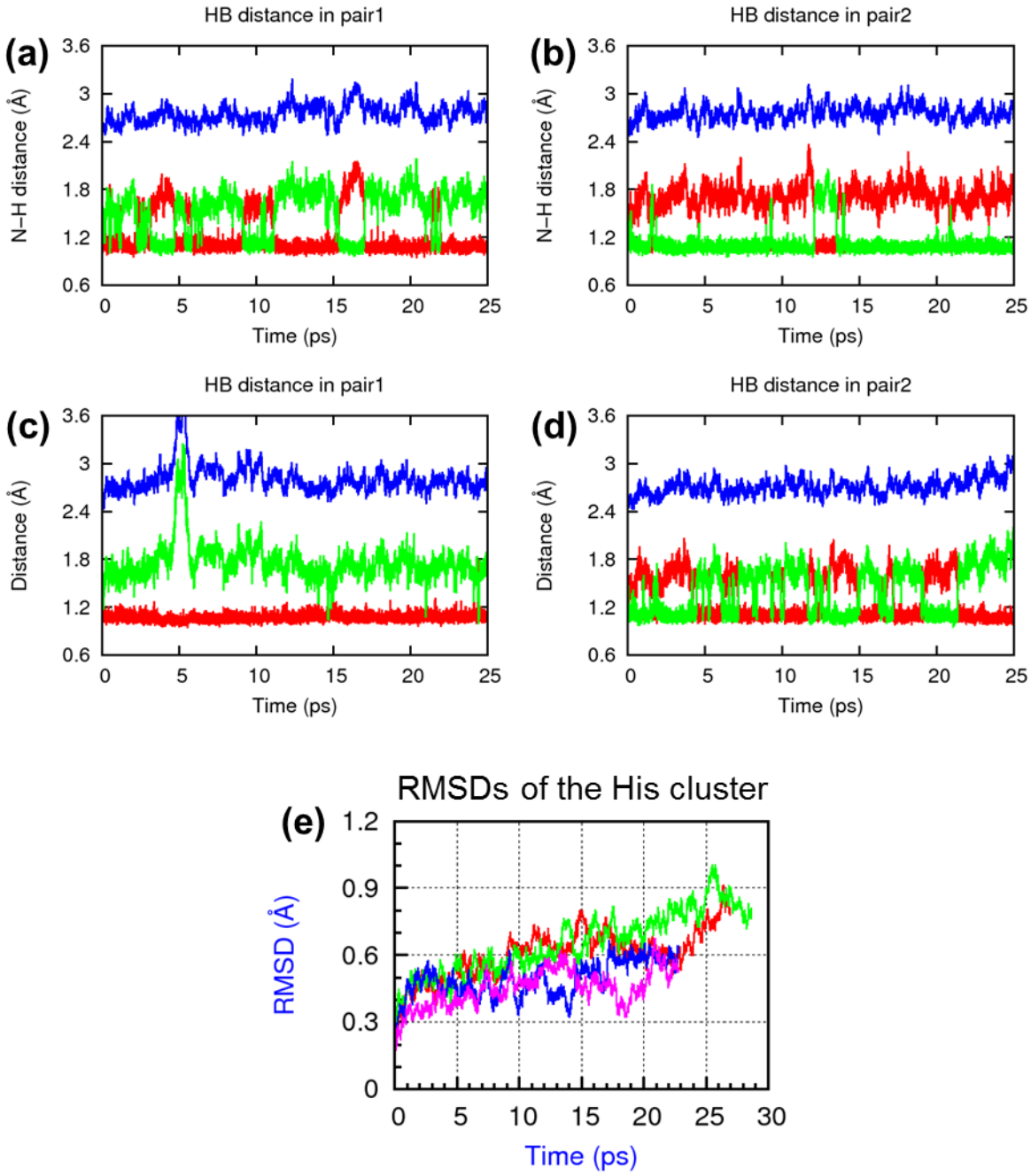

Figure S2. Geometrical parameters of the His cluster in the two parallel QM/MM MD simulations of the "dimer-of-dimers" model. (a) and (b) show the hydrogen bond distance from either histidine pair throughout one simulation run, (c) and (d) show the hydrogen bond distance from a second, uncorrelated simulation run. Distances between the Nδ-H, Nε-H and Nδ…Nε atoms are shown in red, green, and blue, respectively. (e) shows the His37 cluster RMSDs. Data from the "dimer-of-dimers" model is colored in red and green, and the "His-box" model are in blue and pink.

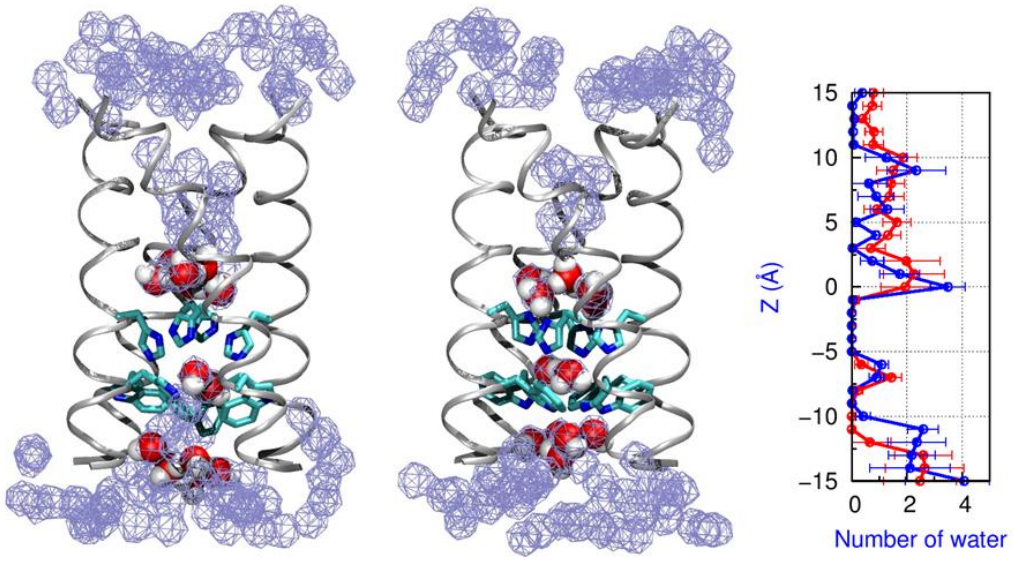

Figure S3. Averaged water density in the QM/MM MD simulations. The "dimer-of-dimers" (left panel) and the "His-box" (middle panel) models are shown, respectively. The backbone is shown in silver ribbon. His37 and Trp41 side chains are shown in the stick mode, and the three water clusters are shown in the space-filling mode. Water density profiles in the two systems are shown at the right panel. The membrane normal was defined as the z-direction, and the center of lipid bilayers was defined as the origin, Data from the "dimer-of-dimers" model is shown in red, and from the "His-box" model is in blue. The errorbar represents the standard deviation of the number of water at each point along the pore.

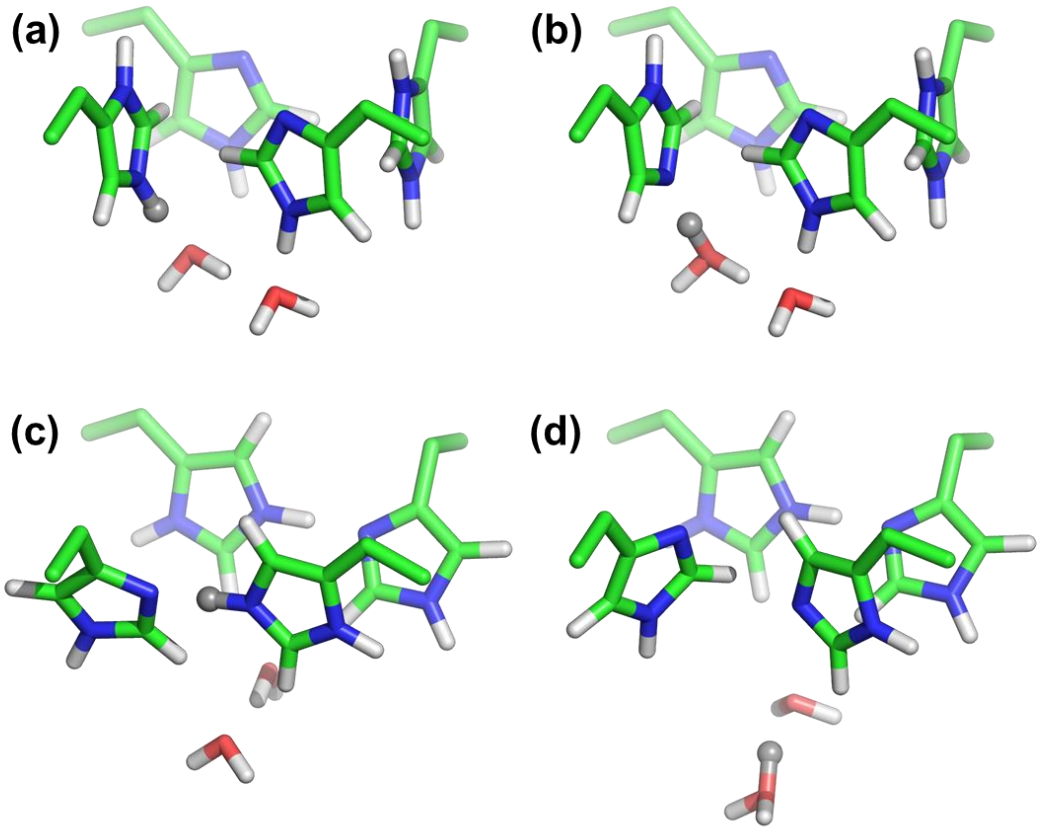

Figure S4. QM/MM optimized His37 tetrad structures with the two water molecules underneath during the deprotonation. One proton was released from a charged His37 to water, forming a hydronium group at the cage of His37 and Trp41, which is energetic unfavorable in either model. (a), (b) Optimized structures before and after deprotonation in the "His-box" model. (c), (d) Optimized structures in the "dimer-of-dimers" model.

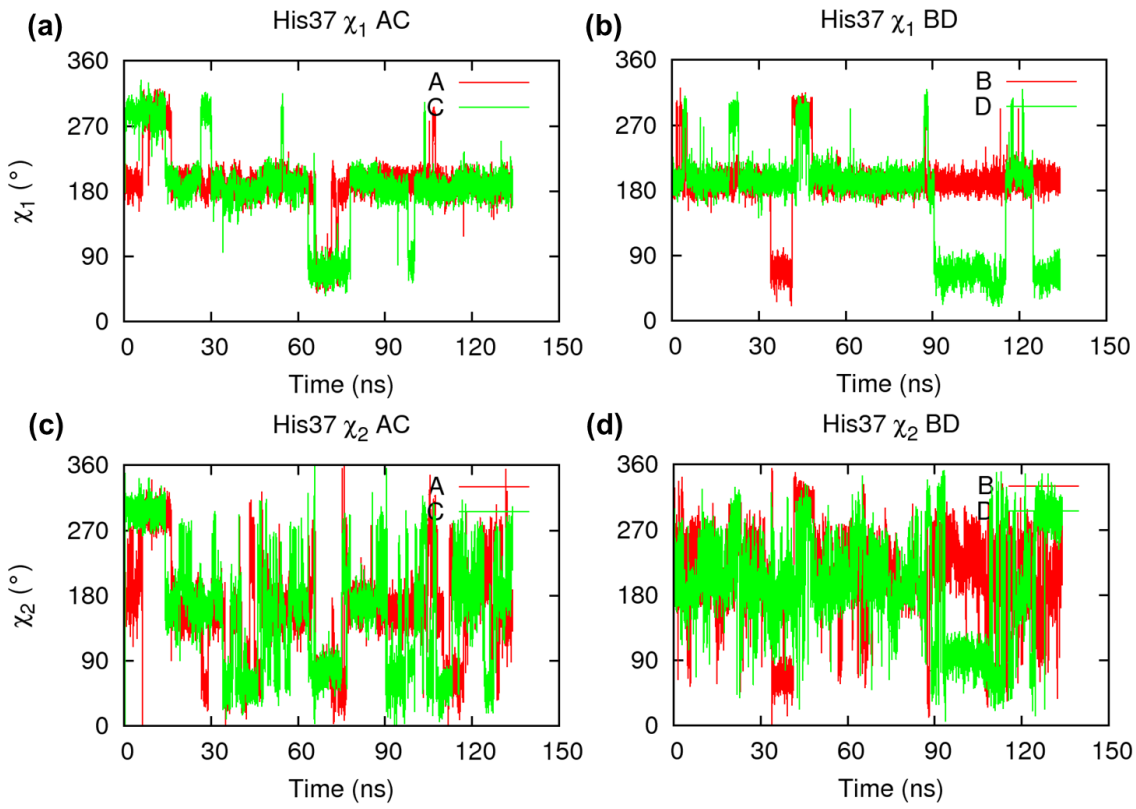

Figure S5. His37 sidechain χ_1 and χ_2 angles in the enhanced free energy sampling.

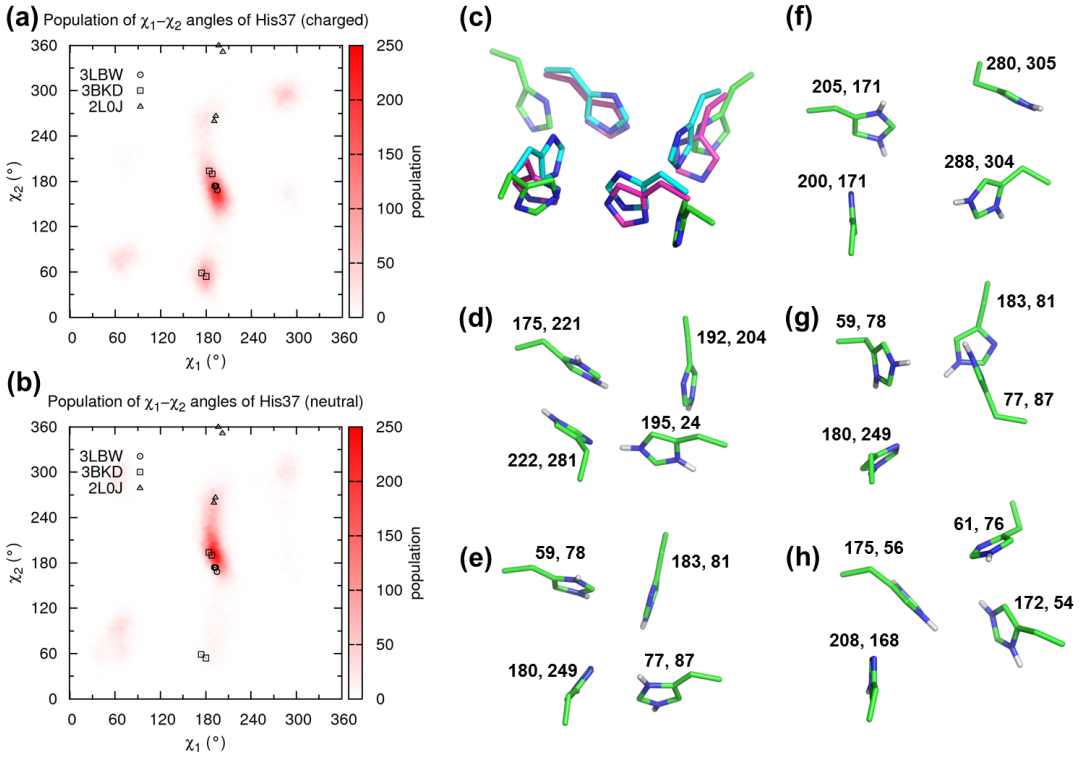

Figure S6. His37 sidechain χ_1, χ_2 dihedral angles in the sampling. (a~b) Distribution and population of the dihedral angles in charged and neutral His37. (c) Superposition of the H37 tetrad in three structure, 3LBW (in cyan, representing the "His-box" configuration), 3KBD (in green, representing the "dimer-of-dimers" configuration) and 2L0J (in magenta, representing the His37 tetrad configuration at low pH condition). (d~h) Representative snapshots obtained from the sampling. χ1 and χ2 angles of the snapshots were shown.

Supplementary computational details

(a) ONIOM protocol to optimize the His37…Trp41 quartet

As the starting structure, the high resolution X-ray crystallographic structure (PDB entry: 3LBW, 1.65 Å) [1] was adopted for the all the calculations in present work. However, as the hydrogen information is missing in the X-ray structure, we firstly used ONIOM method [2] implemented in Gaussian 03 package [3] to optimize both configurations, the "dimer-of-dimers" and the "histidine-box", with the presence of water clusters nearby.

Each system contains the His37, Leu38, Leu40, and Trp41, plus 3 neighboring water clusters (13 water molecules). The backbone heavy atoms were fixed to the positions in the X-ray structure, and the broken bonds were saturated with hydrogen atoms. In the "histidine-box" configuration, the histidine side chains on chain B/D were assigned to be in the Im+ tautomeric form, and those of chains A/C are in the τ tautomeric form, to mimic the +2 state at neutral pH condition.

To optimize the "dimer-of-dimers" configuration, the χ1 and χ2 angles of His37 and Trp41 were manually rotated to approximate the SSNMR reported values.[4] Histidine side chains on chain B/D were assigned to be in the τ tautomeric form, and those of chains A/C are in the Im+ tautomeric form.

In the ONIOM protocol, the system was treated with two layers. The inner layer including His37, the amino group of Leu38, Trp41, and the carboxyl group of Leu40 in each chain, as well as the 13 water molecules, was calculated with B3LYP/6-31G(d,p). The outer layer, containing the remaining part of each system, was treated by the AM1 semi-empirical method. The systems were optimized with all the backbone heavy atoms fixed.

(b) Classical molecular dynamics simulations to pre-equilibrate the system

The optimized His37…Trp41 quartet and water clusters in either configuration were then put into the 3LBW structure embedded in the fully hydrated palmitoyl oleoyl phosphatidyl choline (POPC) bilayers, which had been fully equilibrated in previous simulations. The box size is ~75×75×95 $Å^3$, containing ~54 600 atoms.

The systems were equilibrated under harmonic position restraints applied to the heavy atoms of the protein backbone. The systems were heated to 310K in 10 ns, with the constraint force constant of 20 kcal/mol. Then starting at 10 kcal/mol, the force constant gradually decreased to 0 in 8 ns. The systems were then accumulated for another 22 ns under constant *NPT*. The

His37…Trp41 quartet structure obtained from ONIOM optimization was fixed throughout the simulations.

Periodic boundary conditions were applied, with the particle mesh Ewald method used for treating long-range electrostatic interactions.[5] Constant temperature was achieved by running Langevin dynamics,[6] and constant pressure and surface tension were achieved by the Nose-Hoover Langevin piston method.[7-8] The SHAKE algorithm [9] was applied to constrain all bonds involving hydrogen atoms, which allowed for an integration time step of 2 fs. The snapshots at 18 and 22 ns in the production phase were then minimized with protein heavy atoms fixed, and were used as the initial structures for the subsequent two parallel QM/MM MD simulations.

NAMD version 2.8 [10] was used to run the simulations.

(c) QM/MM MD simulations to study the stability of the configurations

Starting with the aforementioned minimized structure, hybrid quantum mechanical/molecular mechanical (QM/MM) method was employed with the CP2K software package.[11] The QM region is identical to the inner layer defined in the above ONIOM calculations. The rest part was treated with MM method.

The QM region was treated at the DFT level of theory in the Gaussian plane wave (GPW) approximation.[12] The dispersion-corrected-DFT at the D2 level,[13-15] the molecular optimized triple-ζ basis sets with two polarization functions,[16] and the Goedecker, Teter, and Hutter type pseudo-potentials [17] were used. An energy cutoff of 320 Ry was employed in the plane wave representation of the density. A wavelet-based Poisson solver [18] was used to remove the spurious interactions of the QM region with its periodic images. The QM system box size is 32×32×32 $Å^3$ that the buffer between QM atoms and the box edge is ~8 Å. The broken bonds were saturated with hydrogens, the forces on which were treated via the IMOMM scheme,[19] where the scaling factors of 1.355, 1.384 and 1.416 were used to relate the QM C-H bond distance to the MM Cα-C, Cα-N and Cα-Cβ distances, respectively. The CHARMM force field [20] was used for the protein and lipid atoms, and the TIP3P model [21] was used for water in the MM region. The electrostatic interactions of the total QM/MM system were treated with the particle mesh Ewald method.[5]

In the two parallel QM/MM MD simulations in either configuration, the system was run with a time step of 0.4 fs for approximately 25 ps (~100 ps in total) at 310 K using the Nose´-Hoover

chains thermostat [22] with a time constant of 1 ps.

(d) QM/MM optimization to study the deprotonation

The final snapshots from the QM/MM MD simulations were taken to study the deprotonation in either configuration. For each configuration, we used QM/MM calculation to optimize the geometries of two snapshots: one with all excess proton bound to the histidines, and one with a water molecule transformed into a hydronium (H_3O^+) species.

The QM region is identical to those defined in the ONIOM calculations, and was treated with B3LYP. The rest part was treated with CHARMM force field.[20] The calculations were performed by using NWChem version 5.1.[23]

(e) Enhanced sampling calculations to sample the configuration space

Metadynamics simulations [24-25] using collective variables (CVs) [26] implemented in NAMD version 2.8 [10] were carried out to run the free energy calculations. The RMSD of the His37 and Trp41 were used as the two CVs. A HILLS height of 0.1 kcal/mol was used in both CVs. Upper boundary limit for both CVs was set to 15 Å. To avoiding from being trapped by the significant change in the protein backbone structure, the backbone Cα atoms were restrained to have a RMSD less than 1.5 Å to the initial structure as the reference. But this CV was not included in the metadynamics simulations.

1. Acharya, R.; Carnevale, V.; Fiorin, G.; Levine, B. G.; Polishchuk, A. L.; Balannik, V.; Samish, I.; Lamb, R. A.; Pinto, L. H.; DeGrado, W. F.; Klein, M. L. Structure and mechanism of proton transport through the transmembrane tetrameric M2 protein bundle of the influenza A virus. *Proc. Natl. Acad. Sci. U. S. A.* **2010,** *107* (34), 15075-15080.

2. Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. ONIOM: A multilayered integrated MO+MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactions and Pt(P(t-Bu)(3))(2)+H-2 oxidative addition. *J. Phys. Chem.* **1996,** *100* (50), 19357-19363.

3. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; J. A. Montgomery, J.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.;

Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, Revision E.01*, Gaussian, Inc.: Wallingford, CT, 2004.

4. Sharma, M.; Yi, M. G.; Dong, H.; Qin, H. J.; Peterson, E.; Busath, D. D.; Zhou, H. X.; Cross, T. A. Insight into the Mechanism of the Influenza A Proton Channel from a Structure in a Lipid Bilayer. *Science* **2010,** *330* (6003), 509-512.

5. 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* (12), 10089-10092.

6. Adelman, S. A.; Doll, J. D. Generalized Langevin Equation Approach for Atom-Solid-Surface Scattering - General Formulation for Classical Scattering Off Harmonic Solids. *J. Chem. Phys.* **1976,** *64* (6), 2375-2388.

7. Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant-Pressure Molecular-Dynamics Algorithms. *J. Chem. Phys.* **1994,** *101* (5), 4177-4189.

8. Feller, S. E.; Zhang, Y. H.; Pastor, R. W.; Brooks, B. R. Constant-Pressure Molecular-Dynamics Simulation - the Langevin Piston Method. *J. Chem. Phys.* **1995,** *103* (11), 4613-4621.

9. 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* (3), 14.

10. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005,** *26* (16), 1781-802.

11. VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. QUICKSTEP: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.* **2005,** *167* (2), 103-128.

12. Lippert, G.; Hutter, J.; Parrinello, M. A hybrid Gaussian and plane wave density functional scheme. *Mol. Phys.* **1997,** *92* (3), 477-487.

13. Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic-Behavior. *Phys. Rev. A* **1988,** *38* (6), 3098-3100.

14. Lee, C. T.; Yang, W. T.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron-Density. *Physical Review B* **1988,** *37* (2), 785-789.

15. Grimme, S. Semiempirical hybrid density functional with perturbative second-order correlation. *J. Chem. Phys.* **2006,** *124* (3).

16. VandeVondele, J.; Hutter, J. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *J. Chem. Phys.* **2007,** *127* (11).

17. Goedecker, S.; Teter, M.; Hutter, J. Separable dual-space Gaussian pseudopotentials. *Physical Review B* **1996,** *54* (3), 1703-1710.

18. Genovese, L.; Deutsch, T.; Goedecker, S. Efficient and accurate three-dimensional Poisson solver for surface problems. *J. Chem. Phys.* **2007,** *127* (5).

19. Maseras, F.; Morokuma, K. Imomm - a New Integrated Ab-Initio Plus Molecular Mechanics Geometry Optimization Scheme of Equilibrium Structures and Transition-States. *J. Comput. Chem.* **1995,** *16* (9), 1170-1179.

20. MacKerell, A. D., Jr.; Banavali, N.; Foloppe, N. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers* **2000,** *56* (4), 257-65.

21. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; M.L., K. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983,** *79*, 10.

22. Martyna, G. J.; Klein, M. L.; Tuckerman, M. Nose-Hoover Chains - the Canonical Ensemble Via Continuous Dynamics. *J. Chem. Phys.* **1992,** *97* (4), 2635-2643.

23. Valiev, M.; Bylaska, E. J.; Govind, N.; Kowalski, K.; Straatsma, T. P.; Van Dam, H. J. J.; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T. L.; de Jong, W. NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Comput. Phys. Commun.* **2010,** *181* (9), 1477-1489.

24. Iannuzzi, M.; Laio, A.; Parrinello, M. Efficient exploration of reactive potential energy surfaces using Car-Parrinello molecular dynamics. *Phys. Rev. Lett.* **2003,** *90* (23).

25. Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99* (20), 12562-12566.

26. Henin, J.; Fiorin, G.; Chipot, C.; Klein, M. L. Exploring Multidimensional Free Energy Landscapes Using Time-Dependent Biases on Collective Variables. *J. Chem. Theory Comput.* **2010,** *6* (1), 35-47.