Extensive Free Energy Simulations Identify Water as the Base in Nucleotide Addition by DNA Polymerase

Daniel Roston,[1]* Darren Demapan,[2,3] and Qiang Cui[2]*

[1]Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093

[2]Departments of Chemistry, Physics and Biomedical Engineering, Boston University, 590 Commonwealth Avenue, Boston, MA 02215

[3]Department of Chemistry, University of Wisconsin, Madison, 1101 University Avenue, Madison, WI 53706

*droston@ucsd.edu, qiangcui@bu.edu

Supporting Information

www.pnas.org/cgi/doi/10.1073/pnas.1914613116

Detailed Methods

All simulations were done using CHARMM(1, 2) and followed the general methods we have used in other recent studies of enzymatic phosphate chemistry.(3-10) Initial models of both the reactant and product states were taken from a single crystal structure of DNA Polymerase η generated through time-resolved crystallography (PDB: 4ECV).(11) To generate models for the reactant, product, and pre- and post-translocated states, various pieces of the 4ECV structures were deleted *in silico*; additionally, protons were added or deleted in various states in order to generate the setup necessary for different mechanisms or pK_a calculations. These various setups are summarized in Figure 2 and Tables 1-2 of the main text.

Simulations used the generalized solvation boundary potential (GSBP),(12, 13) which treats an inner region of a system as flexible during molecular dynamics simulations while outside of this region the atoms are frozen and treated with an implicit solvation scheme. Thus, the initial structures for each state were overlayed with a spherical droplet of water with a 25 Å radius centered at one of the active site Mg^{2+} atoms. This flexible 25 Å spherical region contained ca. 7060 atoms (depending on specific state) while the frozen outer region contained ca. 3200 atoms. Active site atoms were treated with quantum mechanics at the DFTB3 level(14-18) using two different QM regions to test the effect of QM region size on various properties. The larger of the two QM regions contained ca. 210 atoms, while the smaller contained ca. 100 atoms, depending on what state of the reaction cycle and what particular mechanism was under investigation (Figure 2 of the main text). Atoms outside of the QM region used the Charmm36 force field(19-21) with a modified TIP3P water.(22)

To test the reliability of some aspects of the computations, we used similar methods for systems where more experimental data are available. Simulations of human carbonic anhydrase II (CAII, PDB: 3KS3, ref. (23)) and *E*. *coli* Alkaline phosphatase (AP, PDB: 1ED8, ref. (24)) were treated in analogous manners to DNAP. The QM region for AP was identical to that used in refs. (5, 6) (although S102 was protonated) and that for CAII included the active site Zn^{2+}, the sidechains of its 3 protein ligands, and 5 active site water molecules. The H94D mutation of CAII was generated *in silico*. Simulations of ions in aqueous solution were also treated with the DFTB3/MM/GSBP method, where the QM region included the first solvation sphere and the spherical inner region had a 20 Å radius.

Each system underwent a brief initial geometry optimization, which was followed by heating and equilibration at 298 K for 150 ps using 1 fs timesteps. SHAKE constrained all bonds to hydrogen during heating and equilibration; during production simulations, we released SHAKE on reactive hydrogens and reduced the integration timestep to 0.5 fs. To calculate free energy surfaces we used multi-walker metadynamics(25) with the program PLUMED(26) interfaced with CHARMM. Metadynamics sampling varied somewhat for different sampling goals (see below), but in general we deposited Gaussian biasing potentials every 100 fs (200 time steps); the potentials had a height and width of 0.1 kcal/mol and 0.1 Å, respectively. Sampling of 1D surfaces typically used 5-10 walkers, while 2D surfaces used as many as 50-100 walkers. Various walkers were spawned either from equilibration simulations or from prior metadynamics simulations. Simulation of each walker continued for ca. 500 ps, depending on sampling goals, apparent convergence, and how many parallel walkers the system contained (Figure S1). Thus, each of the 2D free energy surfaces below include data from a minimum of 25 ns of total MD sampling and are the sum of at least

250,000 Gaussian potentials. See Table S1 for statistics on the extent of sampling for each mechanism.

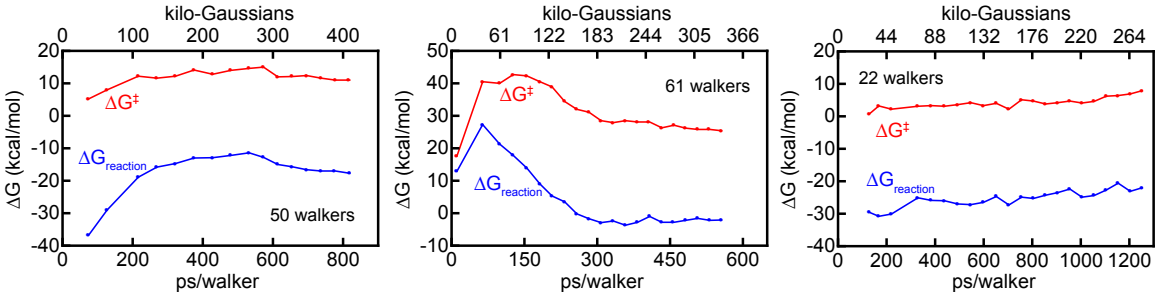

Figure S1: Examples of computed thermodynamic parameters for three different mechanisms as a function of MD simulation time. Each mechanism used a different number of walkers in the multi-walker metadynamics, but each walker deposited Gaussians at the same rate. The three different mechanisms are as follows: Left, water-as-base with 3 Mg^{2+} and protonated pyrophosphate; Middle, Asp115-as-base with 3 Mg^{2+} and deprotonated leaving group; Right, self-activated mechanism with 3 Mg^{2+}.

To calculate pK_a values, we employed the free energy perturbation method developed in ref. (27). This method uses thermodynamic integration to determine the free energy of deprotonation based on simulations at λ ranging from 0 to 1, representing varying amounts of charge on the labile proton (charge = 1-λ). The average proton affinity vs. λ tends to be linear, so we integrated least-squares fits of the data to straight lines using *Mathematica*.(28) In some notable cases, data near λ =1 deviated conspicuously from linearity (Figure S2). This is an artifact of conducting MD with a dummy proton, with charge ~0, such that little or no repulsion exists between the proton and cations, such as metals. Proton affinities in some conformations, therefore, are highly unfavorable and the values can fluctuate wildly during a simulation. Previous studies alleviated this type of artifact by using restraints on the dummy protons.(18) Rather than restraining the system, we chose to accept the data "as-is", but use the standard deviation of the data for each λ as a weight in the linear regression. Due to the large fluctuations in data near λ = 1, then, any deviation from linearity does not contribute significantly to the fitted line.

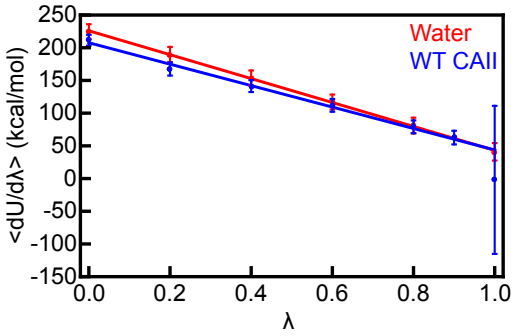

Figure S2: An example of results from pK_a calculations. MD sampling occurs at a range of values of λ, which parameterizes the charge on the labile proton (charge = 1-λ). The average proton affinity (dU/dλ) as a function of λ is then integrated to obtain the free energy of protonation. Species that interact closely with cations, such as the Zn^{2+} in the active site of CAII, tend to exhibit deviations from linearity near λ = 1, but using additional values of λ and weighting the linear regression according to the fluctuations in each window alleviate this problem. The difference in the integrals of the two curves here is equal to the stabilization of the deprotonated form of water (i.e., hydroxide) in the active site of CAII and corresponds to the shift in its pK_a.

Table S1: Sampling statistics for free energy surfaces for DNAP

Figure # of surface	No. of Walkers[a]	Simulation Time per Walker (ns)[a]	Total Simulation Time (ns)[a]	Kilo-Gaussians[a]
1D Surfaces				
Figure S3	7	0.80	5.6	56
Figure S4	10	0.14	1.4	14
2D Surfaces				
Figure 3, Left	139	0.41	57.2	572
Figure 3, Right	22	1.25	27.5	275
Figure 4, Upper Left	90	0.41	36.5	365
Figure 4, Lower Left	43	0.87	37.6	376
Figure 4, Upper Right	61	0.56	33.9	339
Figure 4, Lower Right	43	1.02	44.0	440
Figure 5, Upper Left	62	0.69	43.0	430
Figure 5, Lower Left	116	0.35	41.0	410
Figure 5, Upper Right	37	1.70	63.0	630
Figure 5, Lower Right	50	0.82	41.0	410

[a]As a result of the large biasing potentials that accumulate along collective variables, some walkers entered states that were not relevant to the reaction being studied. For example, at late stages of sampling a proton transfer from the 3'-O to Asp115, the biasing potential leaves the proton unbound to either one of the two heavy atoms. This enables—and favors—spontaneous proton transfers to other heavy atoms. Additional restraints prevent most of this behavior, but during long simulations with large biasing potentials, such processes were sometimes unavoidable. When this happened we only used data that accumulated prior to the walker entering the artifactual state. Thus, the number of walkers listed for each surface is the number that never had such a problem, but the total number of Gaussians and total simulation time refer to all the walkers. Simulation time per walker is the ratio between the total time given here and the number of walkers given here.

Supporting Results

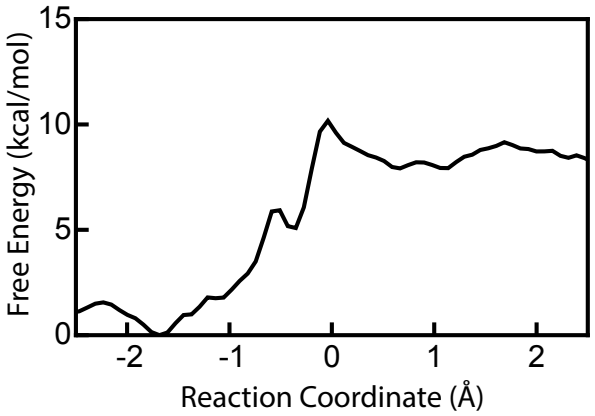

Figure S3: Free energy profile for proton transfer from the 3'-OH to the pyrophosphate in the post-translocated state. The post-translocated state was generated from the crystal structure of the product, containing three Mg^{2+} ions. The terminal nucleotide in that structure was deleted and additional water was added to fill the void. The reaction coordinate is defined as the difference in length of the breaking and forming O-H bonds.

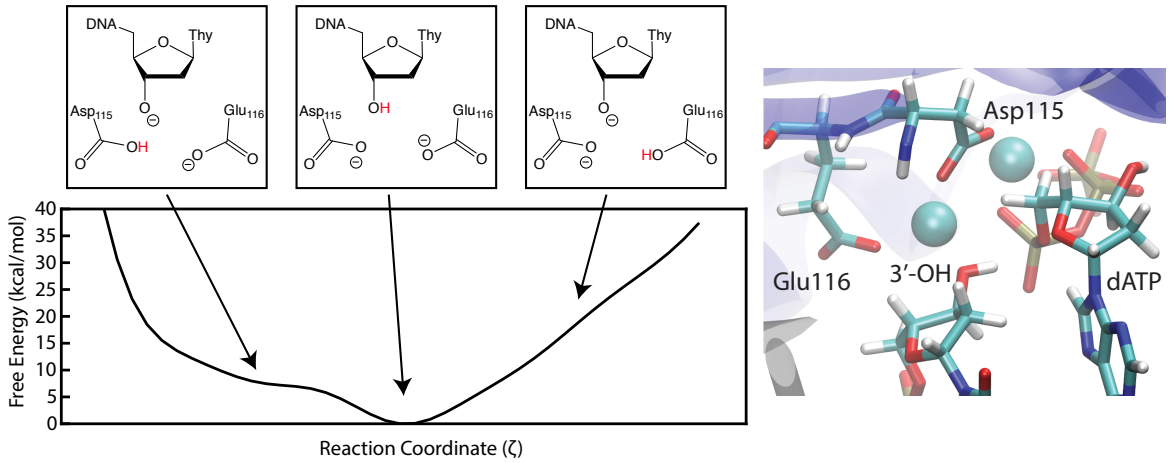

Figure S4: Test of the ability of either Asp115 or Glu116 to deprotonate the 3'-OH. The structure at right shows the proximity of these groups in the crystal structure. The reaction coordinate is defined in terms of the center of excess charge (i.e., the position of the proton highlighted red)(29) along a path from Asp to 3'-O to Glu. The most stable state is when the 3'-O is protonated, but Asp115 may be a suitable base to transiently deprotonate the 3'-OH. The steep increase in free energy when Glu116 abstracts the proton disqualifies it as a catalytic base.

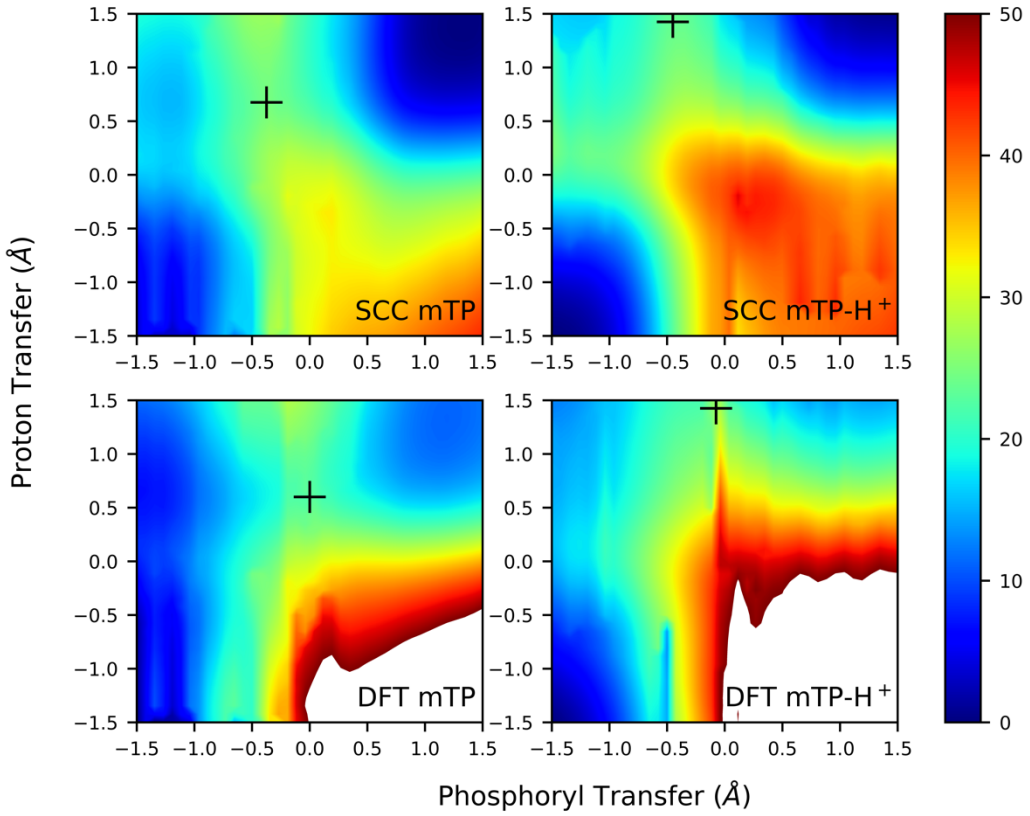

Figure S5: Potential Energy surfaces of a model reaction comparing SCC-DFTB3 with density functional theory (B3LYP/6-311G**). The reaction uses methyl triphosphate (mTP) as the nucleotide and methanol as the nucleophile in a water droplet, but the mechanism is otherwise the same as the water-catalyzed mechanism in the main text: a Mg^{2+}-coordinated hydroxide deprotonates the methanol (the proton transfer coordinate) and the methanol substitutes at P_α. The γ-phosphate is protonated with (mTP-H^+) and fully deprotonated with mTP. In addition to the methyl-triphosphate, each model also includes 2 magnesium ions coordinated with waters, one hydroxyl and one methanol (model for the 3'-OH group) in the QM region; the relative position and orientation of different QM groups are based on the corresponding groups in the crystal structure of DNAP used in the main text. The QM region is solvated in a 10 Å TIP3P water droplet. DFTB3 potential energy surfaces (top) were obtained by adiabatically scanning using a 0.05 Å grid spacing from the reactant to product and vice versa until convergence. The DFT surfaces were then generated via single point energy calculations of the DFTB3 adiabatic surface structures. The transition state is the highest energy point along the minimum energy path from the reactant to product. In clockwise order starting from SCC mTP, the energy barriers are 24.7, 23.6, 22.7 and 26.8 kcal/mol. Similarly, the energy difference between product and reactant are -2.9, 0.6, 11.0, and 14.2 kcal/mol. Therefore, qualitatively similar to our previous comparison of DFTB3/3OB and B3LYP for the case of ATP hydrolysis in myosin,(3) the DFTB3/3OB model is semi-quantitatively consistent with B3LYP calculations, although the exothermicity of the phosphoryl transfer is overestimated.

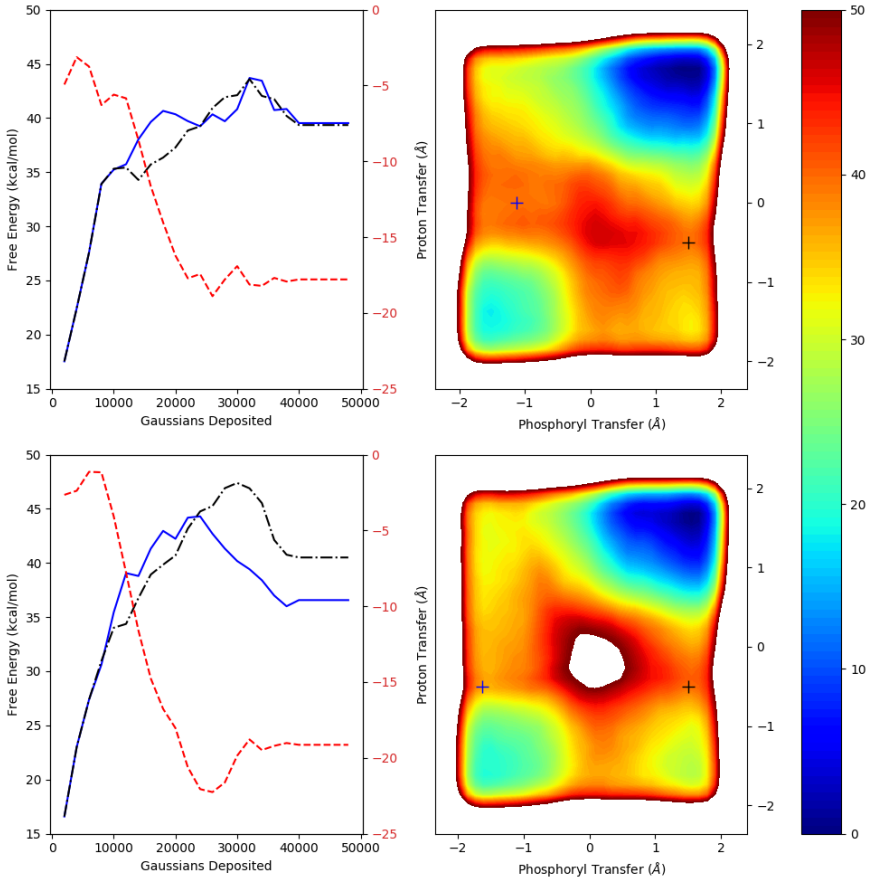

Figure S6: Free energy surfaces of the model reaction from Figure S5 in a water droplet using metadynamics simulations. The top and bottom represent results for mTp and mTp-H$^+$, respectively. 20 walkers were employed for each model and gaussians with a height of 0.25 kcal/mol were deposited every 0.1 ps. After the first 400 ps, we conducted an additional 800 ps of well-tempered metadynamics with the same initial height and deposition rate of gaussians. The panels on the left test the convergences of the simulations, similarly to Figure S1. The "gaussians deposited" in the panels on the left reflect the total number of non-negligible gaussians (height $> 10^{-6}$ kcal/mol); the results illustrate the convergence of the barrier heights (blue and black, left axis) and exothermicity (red, right axis) for both pathways visible in the surfaces. The proton-transfer then phosphoryl transfer mechanism is the blue trace and phosphoryl-transfer then proton-transfer mechanism is the black trace. Plus signs on the free energy surfaces indicate the locations of the transition states for both pathways, determined as the maximum of the minimum free energy pathway from the reactant to product.

Movie S1: An animation of the water-as-base mechanism with three Mg^{2+} and a protonated leaving group. The trajectory for the animation was generated using steered molecular dynamics on the biased potential energy surface generated from the metadynamics simulations.

References

1. B. R. Brooks *et al.*, CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **30**, 1545-1614 (2009).
2. B. R. Brooks *et al.*, CHARMM - A Program For Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **4**, 187-217 (1983).
3. X. Y. Lu, V. Ovchinnikov, D. Demapan, D. Roston, Q. Cui, Regulation and Plasticity of Catalysis in Enzymes: Insights from Analysis of Mechanochemical Coupling in Myosin. *Biochemistry* **56**, 1482-1497 (2017).
4. D. Roston, Q. Cui, QM/MM Analysis of Transition States and Transition State Analogues in Metalloenzymes. *Methods Enzymol.* **577**, 213-250 (2016).
5. D. Roston, Q. Cui, Substrate and Transition State Binding in Alkaline Phosphatase Analyzed by Computation of Oxygen Isotope Effects. *J. Am. Chem. Soc.* **138**, 11946-11957 (2016).
6. D. Roston, D. Demapan, Q. Cui, Leaving Group Ability Observably Affects Transition State Structure in a Single Enzyme Active Site. *J. Am. Chem. Soc.* **138**, 7386-7394 (2016).
7. D. Roston, X. Y. Lu, D. Fang, D. Demapan, Q. Cui, Analysis of Phosphoryl-Transfer Enzymes with QM/MM Free Energy Simulations. *Methods Enzymol.* **607**, 53-90 (2018).
8. Y. Nomura, D. Roston, E. J. Montemayor, Q. Cui, S. E. Butcher, Structural and mechanistic basis for preferential deadenylation of U6 snRNA by Usb1. *Nucleic Acids Res.* **46**, 11488-11501 (2018).
9. G. H. Hou, Q. Cui, QM/MM Analysis Suggests That Alkaline Phosphatase (AP) and Nucleotide Pyrophosphatase/Phosphodiesterase Slightly Tighten the Transition State for Phosphate Diester Hydrolysis Relative to Solution: Implication for Catalytic Promiscuity in the AP Superfamily. *J. Am. Chem. Soc.* **134**, 229-246 (2012).
10. G. H. Hou, Q. Cui, Stabilization of Different Types of Transition States in a Single Enzyme Active Site: QM/MM Analysis of Enzymes in the Alkaline Phosphatase Superfamily. *J. Am. Chem. Soc.* **135**, 10457-10469 (2013).
11. T. Nakamura, Y. Zhao, Y. Yamagata, Y. J. Hua, W. Yang, Watching DNA polymerase eta make a phosphodiester bond. *Nature* **487**, 196-U177 (2012).
12. W. Im, S. Berneche, B. Roux, Generalized solvent boundary potential for computer simulations. *J. Chem. Phys.* **114**, 2924-2937 (2001).
13. P. Schaefer, D. Riccardi, Q. Cui, Reliable treatment of electrostatics in combined QM/MM simulation of macromolecules. *J. Chem. Phys.* **123**, 14 (2005).
14. Q. Cui, M. Elstner, E. Kaxiras, T. Frauenheim, M. Karplus, A QM/MM implementation of the self-consistent charge density functional tight binding (SCC-DFTB) method. *J. Phys. Chem. B* **105**, 569-585 (2001).
15. M. Elstner *et al.*, Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B* **58**, 7260-7268 (1998).
16. M. Gaus, Q. Cui, M. Elstner, DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *J. Chem. Theory Comput.* **7**, 931-948 (2011).

17. M. Gaus, X. Y. Lu, M. Elstner, Q. Cui, Parameterization of DFTB3/3OB for Sulfur and Phosphorus for Chemical and Biological Applications. *J. Chem. Theory Comput.* **10**, 1518-1537 (2014).
18. X. Y. Lu, M. Gaus, M. Elstner, Q. Cui, Parametrization of DFTB3/3OB for Magnesium and Zinc for Chemical and Biological Applications. *J. Phys. Chem. B* **119**, 1062-1082 (2015).
19. R. B. Best *et al.*, Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone phi, psi and Side-Chain chi(1) and chi(2) Dihedral Angles. *J. Chem. Theory Comput.* **8**, 3257-3273 (2012).
20. O. Guvench, E. Hatcher, R. M. Venable, R. W. Pastor, A. D. MacKerell, CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. *J. Chem. Theory Comput.* **5**, 2353-2370 (2009).
21. A. D. MacKerell *et al.*, All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586-3616 (1998).
22. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **79**, 926-935 (1983).
23. B. S. Avvaru *et al.*, A Short, Strong Hydrogen Bond in the Active Site of Human Carbonic Anhydrase II. *Biochemistry* **49**, 249-251 (2010).
24. B. Stec, K. M. Holtz, E. R. Kantrowitz, A revised mechanism for the alkaline phosphatase reaction involving three metal ions. *J. Mol. Biol.* **299**, 1303-1311 (2000).
25. A. Laio, M. Parrinello, Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12562-12566 (2002).
26. M. Bonomi *et al.*, PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **180**, 1961-1972 (2009).
27. D. Riccardi, P. Schaefer, Q. Cui, pKa Calculations in Solution and Proteins with QM/MM Free Energy Perturbation Simulations: A Quantitative Test of QM/MM Protocols. *J. Phys. Chem. B* **109**, 17715-17733 (2005).
28. Wolfram Research, Inc., Mathematica, Version 11.0, Champaign, IL (2019).
29. P. H. Konig *et al.*, Toward theoretical analyis of long-range proton transfer kinetics in biomolecular pumps. *J. Phys. Chem. A* **110**, 548-563 (2006).