THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB

Article

Accurate and Transferable Reactive Molecular Dynamics Models from Constrained Density Functional Theory

Published as part of The Journal of Physical Chemistry virtual special issue "125 Years of The Journal of Physical Chemistry".

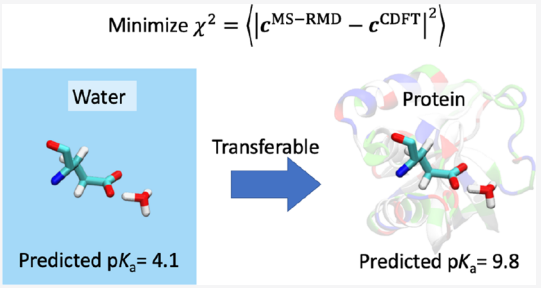
Chenghan Li and Gregory A. Voth*

Cite This: *J. Phys. Chem. B* 2021, 125, 10471−10480

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Chemical reactions constitute the central feature of many liquid, material, and biomolecular processes. Conventional molecular dynamics (MD) is inadequate for simulating chemical reactions given the fixed bonding topology of most force fields, while modeling chemical reactions using *ab initio* molecular dynamics is limited to shorter time and length scales given its high computational cost. As such, the multiscale reactive molecular dynamics method provides one promising alternative for simulating complex chemical systems at atomistic detail on a reactive potential energy surface. However, the parametrization of such models is a key barrier to their applicability and success. In this work, we present reactive MD models derived from constrained density functional theory that are both accurate and transferable. We illustrate the features of these models for proton dissociation reactions of amino acids in both aqueous and protein environments. Specifically, we present models for ionizable glutamate and lysine that predict accurate absolute pK_a values in water as well as their significantly shifted pK_a in staphylococcal nuclease (SNase) without any modification of the models. As one outcome of the new methodology, the simulations show that the deprotonation of ionizable residues in SNase can be closely coupled with side chain rotations, which is a concept likely generalizable to many other proteins. Furthermore, the present approach is not limited to only pK_a prediction but can enable the fully atomistic simulation of many other reactive systems along with a determination of the key aspects of the reaction mechanisms.



INTRODUCTION

Molecular dynamics (MD) can be a powerful computational tool for studying the kinetics and thermodynamics of liquid, material, and biomolecular systems and other associated processes. Conventional MD (referred to herein as classical MD) usually requires a predefined bonding topology of the systems being simulated. Since this topology remains fixed during the course of the simulation, conventional MD is, strictly speaking, suited only for simulating physical processes without chemical reactions. In contrast, the *ab initio* molecular dynamics (AIMD) approach[1] represents a natural choice for simulating reactive processes because it solves the electronic Schrödinger equation on the fly and does not rely on a fixed bonding structure of the system. We also note, on the other hand, that chemical reactions can be intrinsically multiscale, i.e., they can be coupled with multiple collective and corporative motions that span spatial and temporal scales of several orders of magnitude. Given the multiscale nature of reactions, an efficient MD simulation model is needed to sample the required space and time scales in order to obtain statistically meaningful results. However, AIMD calculations can be prohibitively expensive and therefore are often limited to subnanosecond time scales for a system consisting of hundreds of electrons. In addition, the accuracy of AIMD is highly correlated with the adopted electronic structure method underlying the implementation of the AIMD algorithm. A popular approach is the generalized gradient approximation (GGA) within the framework of electronic density functional theory (DFT), which often overestimates charge transfer and partial covalency.[2] Using significantly more computational expensive hybrid functionals has been found to be necessary for achieving quantitative accuracy, e.g., in the description of somewhat simple and nonreactive systems such as liquid water.[3]

An alternative approach is the multiscale reactive molecular dynamics (MS-RMD) method,[4−7] which can be up to 3 orders of magnitude more computationally efficient compared to

Received: July 5, 2021
Revised: August 23, 2021
Published: September 14, 2021

https://doi.org/10.1021/acs.jpcb.1c05992
J. Phys. Chem. B 2021, 125, 10471−10480

AIMD. MS-RMD enables microsecond sampling of complex systems, making it effective for accurately modeling condensed-phase chemical reactions for which the varying bonding topologies can be defined, with sufficient sampling of the coupled, but relatively slow, processes. MS-RMD was previously formulated as a force-matching (FM) algorithm[5−7] in which the models were parametrized to best reproduce the atomic forces of *ab initio* calculations in a least-squares sense. The FM-based MS-RMD model was shown to successfully reproduce the benchmark results of the reference *ab initio* model in both aqueous[5] and biomolecular environments[7] as well as to explain and reproduce experimental thermodynamic and kinetic data.[8−12] These models were parametrized in a case-by-case basis, whereby training *ab initio* calculations were conducted each time the MS-RMD model was applied to a new chemical environment.

In the condensed phase, the MS-RMD approach describes the system as a linear combination of as many as 20−30 diabatic states, each of which corresponds to a resonance form of the system with a different bonding topology. Conventional DFT was adopted as the reference *ab initio* method; however, because of the lack of diabatic state information, only the ground electronic state was fit in the MS-RMD FM scheme. To better address this issue, *ab initio* diabatic methods have been employed in this work to facilitate a more systematic and new parametrization approach for the RMD models at the diabatic state level. Furthermore, this newer diabatic matching (DM) approach can be regarded as a generalization of our original FM method that focuses only on the ground electronic state. The quantum diabatic method chosen here is constrained DFT (CDFT),[13] which is flexible in dealing with various atomic charge prescriptions such as Mulliken,[14] Hirshfeld,[15] and Becke,[16] but we note that other diabatic methods such as the multistate DFT (MS-DFT)[17,18] could also be employed.

As an important category of chemical reactions, proton transport (PT) via Grotthuss proton shuttling[19] along confined water and the titration of protonable residues in biomolecular systems, represents a key step in many protein functional cycles.[20−30] Employing an efficient and accurate computational approach that explicitly models this PT process is crucial for understanding the full working mechanisms associated with these enzymes, channels, and transporters.

As detailed in this work, the deprotonation events of ionizable amino acids in both water and the protein staphylococcal nuclease (SNase) were examined using the new DM method. Importantly, the computational efficiency of MS-RMD allows for the computation of the full free energy profile for the coupled proton dissociation process with local conformational changes in SNase. We show that our model can predict the pK_a of the very different glutamate (Glu; E) and lysine (Lys; K) amino acids in both aqueous solution and solvated SNase, in good agreement with experimental results and without any need to reparametrize the model. Because of this system transferability, we anticipate that the new DM MS-RMD models will be more widely applied to other biomolecular systems to study PT and its coupled phenomenon, notably, the hydration change, global protein conformational change, and ligand transport. Moreover, since neither MS-RMD nor CDFT is limited to PT reactions, our approach could significantly advance the systematic development of accurate and transferrable reactive MD models to simulate other chemical reactions with much greater computational efficiency.

METHODS

MS-RMD Methodology. The MS-RMD framework can be regarded as a linear combination of force fields of conventional molecular mechanics (MM). The latter approach assumes a fixed bonding topology that cannot be changed during the MD simulation. As such, MM lacks the ability to describe chemical reactions. However, the MS-RMD approach considers multiple resonance states, including the reactant state and the product state of the reaction. In the context of proton solvation and transport, each diabatic state is associated with a different localized protonated species, e.g., a solvated neutral Glu as state $|1\rangle$ in Figure 1A or the contact ion pair formed by a charged Glu and hydronium as state $|2\rangle$. The diabatic states are resolved by proposing hypothetical PT reactions. For instance, state $|2\rangle$ is resolved as the product state of $|1\rangle$ after a proton transfer from the protonated Glu, and state $|3\rangle$ originates from a proton transfer from state $|2\rangle$. The aforementioned search process is essentially a depth-first search that follows the hydrogen bond networks, and a maximum search depth of 3, i.e., up to three solvation shells of the excess proton, is shown to be sufficient when modeling most PT reactions while limiting the number of states in a computationally feasible range. The Hamiltonian of the system is then expressed in the following diabatic state representation:

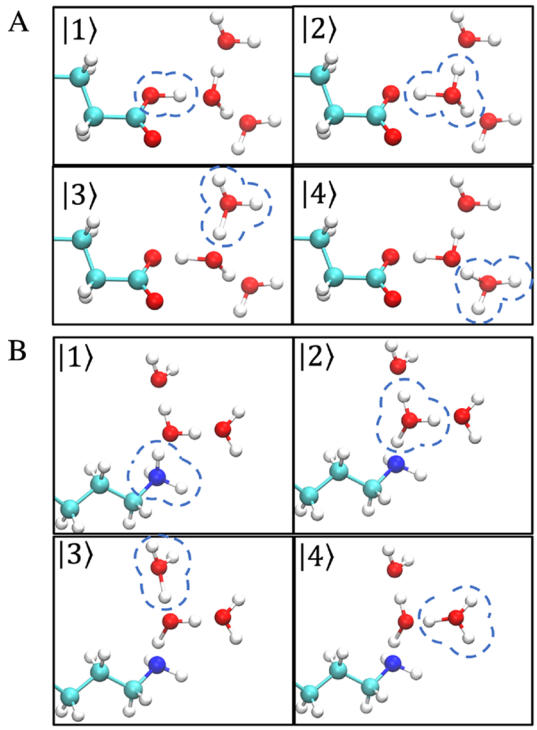


Figure 1. Illustration of MS-RMD states for (A) Glu in water and (B) Lys in water. The protonated moiety in each diabatic state is circled in blue. Note that typically 20−30 coupled bonding topology states are included in the condensed-phase calculations, but only four states are shown here for clarity.

$$\mathbf{H}^{\mathrm{RMD}} = \sum_{ij} |i\rangle h_{ij} \langle j| \tag{1}$$

https://doi.org/10.1021/acs.jpcb.1c05992
J. Phys. Chem. B 2021, 125, 10471−10480

The diagonal term h_{ii} is usually taken to be the energy function described by a suitable MM force field, based on the assumption that the classical force field accurately depicts a diabatic state given a well-defined bonding topology. The off-diagonal element h_{ij} describes the coupling between two states $|i\rangle$ and $|j\rangle$ and is modeled by a physically inspired ansatz in a MM form. (These terms are more fully defined in the Force Field section.) The ground state of the reactive system can be obtained through solving the following eigenvalue problem on the fly as a function of nuclear configuration such that

$$\mathbf{H}^{\mathrm{RMD}}\mathbf{c} = E\mathbf{c} \tag{2}$$

The eigenvector $\mathbf{c} = \{c_i\}$ with the lowest eigenenergy is the diabatic representation of the ground state, which depends on the instantaneous configuration of the nuclear coordinates and also corresponds to the relative weights of the diabatic state contributions to the adiabatic ground state description of the system. The latter relation may be better understood by noting that the ground state energy can also be expressed as the c_i weighted diabatic energies

$$E = \sum_{ij} c_i c_j h_{ij} \tag{3}$$

The atomic forces, as the energy gradient, are c_i weighted diabatic forces according to the Hellmann−Feynman theorem

$$\mathbf{F} = \sum_{ij} c_i c_j \mathbf{F}_{ij} \tag{4}$$

where $\mathbf{F}_{ij}$ is the diabatic force defined as $-\nabla h_{ij}$. Furthermore, the $\mathbf{c}$ vector provides a convenient way to track the effective position of the excess proton, which is the so-called center of excess charge (CEC),[31]

$$\mathbf{r}_{\mathrm{CEC}} = \sum_{i} {c_i}^2 \mathbf{r}_i^{\mathrm{COC}} \tag{5}$$

$\mathbf{r}_i^{\mathrm{COC}}$ is the center of charge of the protonated moiety in state $|i\rangle$. The CEC defined in such a way can be regarded as a ${c_i}^2$ weighted average of the excess protonic charge defect position. (Because of bond breaking and forming during the PT process, the identity of H^+ is not associated with any specific proton in the system.) Thus, the CEC is needed to indicate the position of H^+ and to further define continuous reaction coordinates (RCs) for enhanced free energy sampling simulations of PT reactions, which is discussed in greater detail in the Simulation Details section.

Force Field (FF). The FF functional forms used in MS-RMD are provided in this section, while the detailed training method and procedure of the models will be provided in a later discussion of the parametrization of MS-RMD using CDFT.

The diagonal h_{ii} terms are described by the standard CHARMM36 FF[32] with several adaptations clarified below. The original harmonic O−H bond in a protonated Glu carboxylic group and the N−H bond in the Lys ammine group are replaced by a Morse bond to better describe the proton disassociation curve

$$U_{\mathrm{Morse}}^{\mathrm{bond}} = D(1 - \mathrm{e}^{-\alpha(r-r_0)})^2 \tag{6}$$

where r represents the bond length and D, α, and r_0 are bond parameters taken from ref 6 for Glu. The bond parameters for Lys were optimized to match an energy scan of the bond disassociation, in which the proton was scanned with respect to nitrogen within a distance range of 0.9 to 1.9 Å in 0.1 Å intervals. The reference energies were computed with the functional/basis-set combination ωB97X[33]/TZV2P in CP2K[34] coupled with Libxc.[35] In any diabatic state wherein a protonated Glu or Lys was found, such as state $|1\rangle$ shown in Figure 1, a trainable energy offset V_{ii}^{const} was added to the total energy of the system to bridge zero points of the classical MM potential energy surfaces (PESs), which were originally parametrized for nonreactive protonated and deprotonated forms of the amino acids.

The hydronium FF parameters including bonds, angles, and Lennard-Jones (LJ) terms were obtained from ref 36. Additionally, additive repulsions between hydronium and Glu carboxyl/Lys amine were employed to correct the over-attraction between point charges present in the MM FFs, as given in the following equations:

$$U_{\mathrm{OX}}^{\mathrm{rep}} = B\exp(-b(r_{\mathrm{OX}} - d_{\mathrm{OX}}^0))\sum_{j=1}^{3}\exp(-b'\mathbf{q}_{\mathrm{H}_j\mathrm{X}}^2) \tag{7}$$

and

$$U_{\mathrm{HX}}^{\mathrm{rep}} = C\exp(-c(r_{\mathrm{HX}} - d_{\mathrm{HX}}^0)) \tag{8}$$

r_{OX} represents the distance between the hydronium oxygen and the Glu carboxylic oxygen/Lys amine nitrogen, and r_{HX} represents the associated distance with the hydronium hydrogen. The vector $\mathbf{q}_{\mathrm{H}_j\mathrm{X}}$ is the proton-sharing vector defined as follows

$$\mathbf{q}_{\mathrm{H}_j\mathrm{X}} = \mathbf{r}_{\mathrm{OH}_j} - \frac{1}{2}\mathbf{r}_{\mathrm{OX}} \tag{9}$$

where the O and X atoms follow the same definition as above and H_j represents one of the three hydrogens of the hydronium structure. B, b, b', C, and c are tunable parameters, while d_{OX}^0 = 2.4 Å and d_{HX}^0 = 1.0 Å were fixed and do not need fitting since they can be absorbed into the B and C prefactors.

As part of the MM FF, some of the nonbonded LJ interactions were also tuned to better describe proton-transfer reactions. While the standard 12-6 LJ functional form was used as shown here

$$U_{\mathrm{LJ}} = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{10}$$

LJ parameters σ and ϵ for interactions between the carboxyl/ammine and water/hydronium were adjusted. To be specific, the modified interactions were among eight atom pairs: (1) protonated Glu carboxyl oxygen (OEP) and water oxygen (Ow), (2) Glu carboxyl proton (HEP) and Ow, (3) deprotonated Glu carboxyl oxygen (OE) and hydronium oxygen (OH), (4) OE and hydronium proton (HH), (5) protonated Lys ammine nitrogen (NKP) and Ow, (6) protonated Lys ammine proton (HKP) and Ow, (7) deprotonated Lys ammine nitrogen (NK) and OH, and (8) NK and HH.

The h_{ij} off-diagonal energy term between amino acid and water proposed in ref 6 is used here as well:

$$h_{ij} = c_1\exp(-c_2(r_{\mathrm{HX}} - c_3)^2) \tag{11}$$

Atom X represents the carboxyl oxygen or the ammine nitrogen, and the H atom represents the shared proton between Glu/Lys and water. c_1, c_2, and c_3 were treated as trainable parameters.

https://doi.org/10.1021/acs.jpcb.1c05992
J. Phys. Chem. B 2021, 125, 10471−10480

CDFT Theory. CDFT has been well documented in the literature.[13] As the name indicates, CDFT is essentially DFT with constraints:

$$E^{\mathrm{CDFT}} = \min_{\rho} \max_{\lambda} \left(E[\rho] + \lambda \left(\int \mathrm{d}\mathbf{r}\, w(\mathbf{r})\, \rho(\mathbf{r}) - N \right) \right) \tag{12}$$

This is a minimization problem with the constraint on the electronic density. The $E[\rho]$ here is the density functional, and λ is the Lagrange multiplier. A constraint is defined by a weighting function $w(\mathbf{r})$ and the target value N:

$$\int \mathrm{d}\mathbf{r}\, w(\mathbf{r})\, \rho(\mathbf{r}) \equiv N \tag{13}$$

In this work, the total Becke atomic charges on molecules were constrained and the associated weighting function adopted the following form:

$$w(\mathbf{r}) = \sum_I w_I^{\mathrm{Becke}}(\mathbf{r}) \tag{14}$$

The Becke population function $w_I^{\mathrm{Becke}}(\mathbf{r})$ is a smooth indicator function which becomes unity when close to atom I and decays to zero when more distant. The constraint target number N was computed from the "promolecule" density $\tilde{\rho}(\mathbf{r})$

$$N = \int \mathrm{d}\mathbf{r}\, w(\mathbf{r})\, \tilde{\rho}(\mathbf{r}) = \int \mathrm{d}\mathbf{r}\, w(\mathbf{r})(\rho_{\mathrm{A}}(\mathbf{r}) + \rho_{\mathrm{B}}(\mathbf{r})) \tag{15}$$

where ρ_{A} and ρ_{B} were the ground state electronic densities of fragments A (identified as the protonated species) and B (identified as the remainder). Accordingly, the summation provided in eq 14 runs over all atoms in fragment A. As a result, eq 12 with the constraint in eq 13 seeks the diabatic state whose electronic structure mimics a direct superposition of a pure fragment A plus a pure fragment B.

Considering a system featuring one Glu/Lys sharing a proton with one water, two diabatic states can be defined: a protonated Glu/Lys plus a neutral water (denoted as $|1\rangle$) and a deprotonated Glu/Lys plus a hydronium (denoted as $|2\rangle$). To compute state $|1\rangle$ in CDFT, fragment A was defined as the Glu/Lys and fragment B was the water. In state $|2\rangle$, fragment A was the hydronium and fragment B was the negatively charged Glu/Lys. After the diabatic energies and wave functions are obtained, the ground state of the system can be expressed in the diabatic representation from a configuration interaction (CI) calculation

$$\mathbf{H}^{\mathrm{CDFT}}\mathbf{c} = E\mathbf{S}\mathbf{c} \tag{16}$$

where the CDFT Hamiltonian is composed of the CDFT diabatic energies and the couplings between diabatic states

$$\mathbf{H}^{\mathrm{CDFT}} = \begin{pmatrix} E_1 & \langle \Psi_1 | \hat{H} | \Psi_2 \rangle \\ \langle \Psi_2 | \hat{H} | \Psi_1 \rangle & E_2 \end{pmatrix} \tag{17}$$

and the matrix $\mathbf{S}$ describes the overlap between diabatic states

$$\mathbf{S} = \begin{pmatrix} 1 & \langle \Psi_1 | \Psi_2 \rangle \\ \langle \Psi_2 | \Psi_1 \rangle & 1 \end{pmatrix} \tag{18}$$

Parametrization of MS-RMD Using CDFT. The original FM-based MS-RMD parametrization was designed to minimize the force residual between ground state MS-RMD forces and the ground state *ab initio* forces such that

$$\chi^2 = \langle |\mathbf{F}^{\mathrm{RMD}} - \mathbf{F}^{\mathrm{QM}}|^2 \rangle \tag{19}$$

where the bracket indicates some sort of ensemble average. The MS-RMD atomic forces $\mathbf{F}^{\mathrm{RMD}}$ are computed from the Hellmann–Feynman theorem (eq 4) and depend on the FF parameters noted in the Force Field section. We note from eq 4 that the ground state forces depend on the contributions from both the ground state wave function $\mathbf{c}$ and the diabatic forces $\mathbf{F}_{ij}$. The parametrization based solely on the ground state force residual (eq 19) is likely to overfit due to the error cancellation in $\mathbf{c}$ and $\mathbf{F}_{ij}$. The rationale of our new approach is to separate the ground state wave function $\mathbf{c}$ and the diabatic forces in the training procedure. Accordingly, two residuals, defined by utilizing the diabatic information computed by CDFT, should be minimized with respect to MS-RMD parameters

$$\chi_c^2 = \langle |\mathbf{c}^{\mathrm{RMD}} - \mathbf{c}^{\mathrm{CDFT}}|^2 \rangle \tag{20}$$

and

$$\chi_F^2 = \left\langle \sum_{ij} |\mathbf{F}_{ij}^{\mathrm{RMD}} - \mathbf{F}_{ij}^{\mathrm{CDFT}}|^2 \right\rangle \tag{21}$$

where the CDFT diabatic forces are defined as $\mathbf{F}_{ij}^{\mathrm{CDFT}} = -\nabla H_{ij}^{\mathrm{CDFT}}$. Residual eq 21 differs from eq 19 since the diabatic forces ($\mathbf{F}_{ij}$) are compared individually instead of using the comparison of the total ground state force, which is c_i^2-weighted $\mathbf{F}_{ij}$ according to the Hellman–Feynman theorem. We note that, as an eigenvector, the ground state wave functions $\mathbf{c}^{\mathrm{RMD}}$ and $\mathbf{c}^{\mathrm{CDFT}}$ are determined by the diabatic energies in $\mathbf{H}^{\mathrm{RMD}}$ and $\mathbf{H}^{\mathrm{CDFT}}$, respectively. Thus, minimizing χ_c^2 can be regarded as an implicit energy-matching process. In this sense, the DM approach systematically matches the energy and force through the minimization of χ_c^2 and χ_F^2 sequentially. This technique can be beneficial since it is reported that including both energy and forces in the loss function results in a more reliable machine-learned FF.[37]

The multistate empirical valence bond (MS-EVB) version 3.2 model[36] has been successful in describing the solvation and transport of an excess proton in pure water. Therefore, our parametrization of MS-RMD kept the hydronium-water PT model unchanged, focusing instead on diagonal and off-diagonal terms that correspond to the proton transfer between Glu/Lys and its first solvation shell proton-acceptor water. The training set was generated by placing one water at r_{OO} (the distance between the water oxygen and carboxyl oxygen of Glu) or r_{ON} (the distance between the water oxygen and amine nitrogen of Lys), ranging from 2.2 to 2.8 Å with 0.1 Å spacing, and for each r_{OO} or r_{ON}, the shared proton was placed at nine equally spaced positions with r_{OH} ranging from 1.0 Å to r_{OO} − 1.0 Å for Glu and r_{NH} ranging from 1.0 Å to r_{ON} − 1.0 Å for Lys. The CDFT calculations were based on these configurations at a DFT/basis-set level of BLYP/TZV2P or ωB97X/TZV2P, followed by a CDFT-CI calculation performed using the DFT Kohn–Sham surrogates for the diabatic state wave functions. To reduce the computational cost and achieve better convergence, the ωB97X-level CDFT was carried out with Glu/Lys side chains only. (We note that calculations that included backbones at the BLYP level showed no difference in $\mathbf{c}$ from the side-chain-only ones.) All CDFT calculations were performed using the CP2K package.

Optimizing MS-RMD parameters was carried out using an in-house Python script to minimize the residual in eq 20 via

https://doi.org/10.1021/acs.jpcb.1c05992
J. Phys. Chem. B 2021, 125, 10471−10480

the Nelder–Mead method[38] initiated from the previous MS-RMD model parameters[6] followed by a refinement by the Broyden–Fletcher–Goldfarb–Shanno (BFGS) method; all parameters thus obtained are summarized in Table 1. Because

Table 1. Optimized MS-RMD Model Parameters Using ωB97X CDFT Data[a]

	Glu	Lys		Glu	Lys
B	3.94793	1.01096	V_{ii}	−153.284	−67.0979
b	1.41638	1.41969	$\epsilon^{LJ}_{OE/NK-HH}$	0.23117	0.0175896
b'	1.08444	1.07948	$\sigma^{LJ}_{OE/NK-HH}$	1.39561	1.68765
C	3.8605	0.989666	$\epsilon^{LJ}_{Ow-HEP/HKP}$	0.72595	0.0107436
c	1.14669	1.14233	$\sigma^{LJ}_{Ow-HEP/HKP}$	1.29196	1.77364
c_1	−25.0422	−25.015	$\epsilon^{LJ}_{OE/NK-OH}$	0.125249	0.115627
c_2	2.99968	3.02538	$\sigma^{LJ}_{OE/NK-OH}$	3.00742	3.20591
c_3	1.40533	1.37087	$\epsilon^{LJ}_{OEP/NKP-Ow}$	0.162054	0.162548
D	143.003	157.014	$\sigma^{LJ}_{OEP/NKP-Ow}$	3.07772	3.23758
α	1.8	1.70825			
r_0	0.975	1.02389			

[a]The units of the listed parameters use kcal/mol as the energy unit and Å as the length unit. The Glu Morse bond parameters were taken from ref 6.

of the nonorthogonality of CDFT diabatic states, the $\mathbf{c}^{CDFT}$ vector does not normalize to unity. However, since only the relatively stability of $|1\rangle$ and $|2\rangle$ matters in the fitting, we were able to achieve good agreement between $\mathbf{c}^{RMD}$ and normalized $\mathbf{c}^{CDFT}$ (Figure S1) via directly minimizing eq 20. Note that the models obtained from minimizing eq 20 exhibited good accuracy and transferability even prior to applying further improvements using force residual eq 21, which is detailed in the Results and Discussion section.

Simulation Details. The Glu/Lys-in-water simulations were conducted with one protonated solute solvated in 241 water molecules for Glu and in 237 water molecules for Lys in a cubic box with a side length of 20 Å. The temperature was controlled by a Nose–Hoover chain[39] at 300 K, and a time step of 1 fs was used to integrate the system. To enhance the sampling of proton disassociation from Glu/Lys, well-tempered metadynamics (WT-MTD)[40] was performed. The RC for Glu driven in WT-MTD was defined as the distance between the excess proton CEC and the closer Glu carboxyl oxygen, implemented as

$$\xi_{CEC}^{Glu} = -\frac{1}{\kappa}\log[\exp(-\kappa(r_1 - \bar{r})) + \exp(-\kappa(r_2 - \bar{r}))] + \bar{r} \quad (22)$$

where κ = 40 Å^{-1}, r_1 and r_2 denote the CEC separation from the two carboxyl oxygen atoms, and $\bar{r} = (r_1 + r_2)/2$. The RC, ξ_{CEC}^{Lys} for Lys, was the distance between CEC and the amine nitrogen. The initial Gaussian height for WT-MTD was 0.8 kcal/mol, and the bias factor was set to 12. The Gaussians were deposited every 1 ps with a fixed width of 0.1 Å. To restrain the sampling in regions of interest, a harmonic potential soft wall with a 25 kcal/mol/Å^{-2} force constant was added to the RC ξ_{CEC} if its value exceeded 8 Å. The WT-MTD was run for ∼20 ns for both Glu and Lys, and the potential of mean force (PMF) was obtained by summing the deposited Gaussians. The error bar of the PMF was estimated by partitioning the full trajectory into eight blocks and calculating the standard deviation of PMFs computed from the last five blocks.

The pK_a of Glu/Lys was computed from their PMFs via[41]

$$pK_a = \log\left(c_0 \int_0^{\dagger} 4\pi\xi_{CEC}^2 e^{-\beta(F(\xi_{CEC})-F(+\infty))}\, d\xi_{CEC}\right) \quad (23)$$

where c_0 = 1/1660 Å^{-3} is the standard state concentration (1 M) expressed in number density, $F(\xi_{CEC})$ is the PMF, and $F(\infty)$ is the value when ξ_{CEC} is sufficiently large and $F(\xi_{CEC})$ reaches a plateau. The integral was evaluated up to the dividing surface (denoted by †) between protonated and deprotonated Glu/Ly. However, the integral value did not change if the integral went further because the free energy passing the dividing surface was sufficiently high for weak acids and the exponential integrand at further distances vanishes. Note that $F(\xi_{CEC})$ used in the pK_a calculation (shown in Figure 2) differs

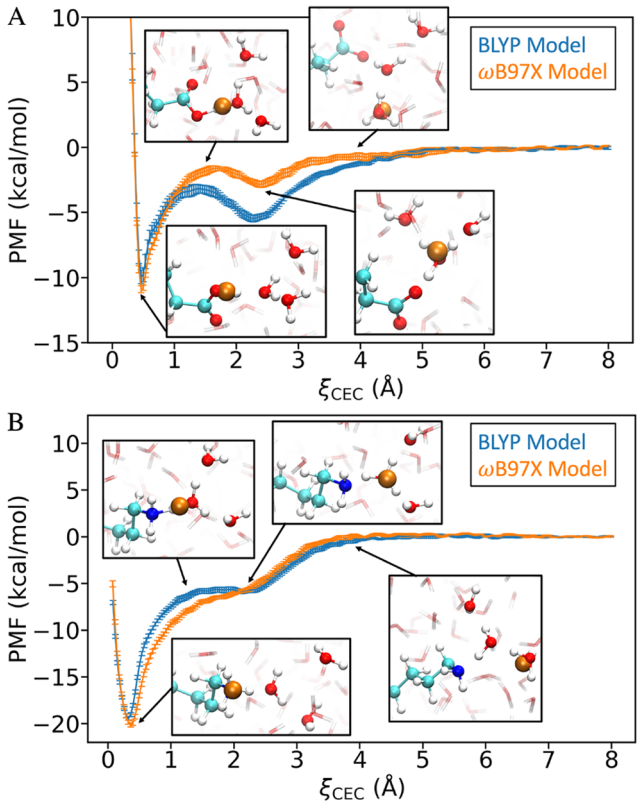


Figure 2. Proton dissociation potential of mean force (PMF) of (A) Glu in water and (B) Lys in water, computed from BLYP or ωB97X DFT level-parametrized MS-RMD models. The insets show representative molecular configurations at the corresponding positions on the PMFs. The center of excess charge (CEC, i.e., the effective position of H^+) is rendered as an orange sphere.

from the PMF directly summed from Gaussians (denoted as $F'(\xi_{CEC})$) via the relation

$$4\pi\xi_{CEC}^2 e^{-\beta F(\xi_{CEC})} = e^{-\beta F'(\xi_{CEC})} \quad (24)$$

In essence, eq 24 converts the probability density for the system to visit $[\xi_{CEC}, \xi_{CEC} + d\xi_{CEC}]$ into the probability density of visiting the infinitesimal spherical shell $4\pi\xi_{CEC}^2 d\xi_{CEC}$. The resulting $F(\xi_{CEC})$ thus becomes a plateau with a sufficiently large ξ_{CEC}, in contrast to the behavior of $F'(\xi_{CEC})$ that reflects the increasing probability density in $[\xi_{CEC}, \xi_{CEC} + d\xi_{CEC}]$ due to the larger accessible volume of the $4\pi\xi_{CEC}^2 d\xi_{CEC}$ spherical shell.

Our simulations of SNase consisted of one SNase protein (PBD id 1U9R[42]) solvated in a cubic water box of 70 Å on a

https://doi.org/10.1021/acs.jpcb.1c05992
J. Phys. Chem. B 2021, 125, 10471−10480

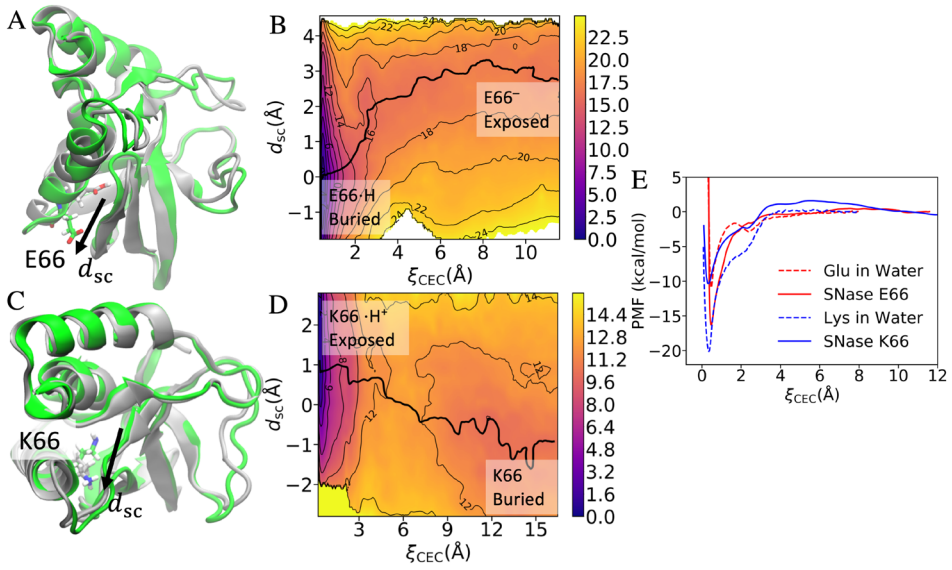


Figure 3. Coupling between proton dissociation and side chain orientation in SNase. (A) Classical equilibrated configurations of V66E mutants when E66 is deprotonated (green) and protonated (gray). (B) Potential of mean force in kcal/mol of E66 ionization and its side chain rotation. The minimum free energy path is shown as a black curve. The PMF statistical error can be found in Figure S3A of the Supporting Information. (C) Classical equilibrated configurations of V66K mutants when K66 is deprotonated (green) and protonated (gray). (D) PMF of K66 ionization and its side chain rotation. The minimum free energy path is shown in black. The PMF error can be found in Figure S3B of the Supporting Information. (E) Comparison of ionization PMFs of Glu (red) and Lys (blue) in water (dashed) and in SNase (solid; computed via eq 29).

side with 0.15 M sodium chloride added. Buried residue V66 was mutated into Glu/Lys to form the V66E and V66K mutants. For both the protonated and deprotonated E66/V66 mutants, classical MD equilibrations were conducted in the constant NPT ensemble at 298 K under 1 atm for 200 ns. The temperature and pressure were controlled by a Nose−Hoover chain and by the Parrinello−Rahman barostat,[43] respectively. All of the bonds involving hydrogens were constrained using the LINCS algorithm,[44] and a time step of 2 fs was used to propagate the system. All classical equilibrations were carried out in the GROMACS MD package.[45]

The SNase MS-RMD simulations, initiated from classical equilibrations, were conducted in the NVT ensemble at 298 K using a time step of 1 fs. The RC for proton disassociation from E66/K66 was defined to be the same as the ξ_{CEC} for Glu/Lys in water. A second RC d_{SC} was defined as follows

$$d_{SC} = \mathbf{v}_{SC}\cdot\mathbf{n}_{prot} \tag{25}$$

to characterize the side chain rotation of E66/K66, given that protonated/deprotonated forms of E66/K66 were found to display distinct side chain orientations during classical equilibrations (Figure 3). In eq 25, $\mathbf{v}_{SC}$ represents the vector pointing from the E66/K66 α-carbon to the geometric center of the carboxylic group (COO) of E66 or the amine nitrogen of K66 and $\mathbf{n}_{prot}$ is a vector defined to reflect the protein's overall orientation

$$\mathbf{n}_{prot} = \frac{\mathbf{r}_{C2} - \mathbf{r}_{C1}}{|\mathbf{r}_{C2} - \mathbf{r}_{C1}|} \tag{26}$$

where $\mathbf{r}_{C1}$ is the center of the backbone atoms of residues 15−19 and 61−65 and $\mathbf{r}_{C2}$ is the center of the backbone of residues 90−94 (illustrated in Figure S2 of the Supporting Information). To restrain the lateral diffusion of the hydrated proton when completely dissociated from E66/K66, a harmonic potential soft wall u_{res} was added to a collective variable $r_\perp$ defined as the length of the orthogonal part of $\mathbf{v}_{CEC}$ with respect to $\mathbf{n}_{prot}$

$$r_\perp = |\mathbf{v}_{CEC} - (\mathbf{v}_{CEC}\cdot\mathbf{n}_{prot})\mathbf{n}_{prot}| \tag{27}$$

where $\mathbf{v}_{CEC}$ is the vector pointing from the E66 COO center or K66 nitrogen to the CEC. The force constant of the wall was 10 kcal/mol/Å^2, and the restraining potential was switched on for $r_\perp \geq 7$ Å. To enhance the sampling of both ξ_{CEC} and d_{SC}, two-dimensional umbrella sampling (2D-US) on the two RCs was performed. Each US window was run for ∼1 ns, and the total simulation time was ∼1 μs for both V66E and V66K. The weighted histogram analysis method (WHAM)[46] was used to combine the 2D-US data and compute the PMF, $F(\xi_{CEC}, d_{SC})$, from which the one-dimensional (1D) PMF (Figure 3E) for proton dissociation was obtained by integrating out the d_{SC} degree of freedom (DOF). Specifically,

$$F(\xi_{CEC}) = -\beta^{-1}\ln\left(\int e^{-\beta F(\xi_{CEC}, d_{SC})}\, dd_{SC}\right) \tag{28}$$

Then, following the derivation in ref 47, the pK_a of E66/K66 is given by

$$pK_a = \log\left(c_0 S_u \int_0^{\dagger} e^{-\beta(F(\xi_{CEC}) - F(+\infty))}\, d\xi_{CEC}\right) \tag{29}$$

where the meaning of c_0 and the integral range are the same as in eq 23 and $S_u = \int_0^\infty 2\pi r_\perp e^{-\beta u_{res}(r_\perp)}\, dr_\perp$ corrects for the introduced radial restraint on CEC. The errors reported for PMFs and pK_a were obtained from partitioning the trajectories of all US windows into six equally sized blocks and calculating the standard deviation using the final four blocks. All MS-RMD simulations were performed using the LAMMPS MD package[48] coupled with RAPTOR[4] for reactions and PLUMED 2[49] for the free energy sampling.

https://doi.org/10.1021/acs.jpcb.1c05992
J. Phys. Chem. B 2021, 125, 10471−10480

RESULTS AND DISCUSSION

Benchmark of Reactive Models in Water. The MS-RMD parameters obtained using the DM approach with ωB97X CDFT data are provided in Table 1, and model parameters determined from the BLYP data are given in Table S1. The resulting MS-RMD Glu/Lys models were benchmarked in water by computing the PMF of their ionization (Figure 2).

The potential of mean force (PMF) contains arguably the most valuable information about the system because it provides the full free energy profile as a function of the reaction progress (monitored by the RC value). At $\xi_{CEC} \approx 0.5$ Å, a narrow and deep free energy well was observed in both the Glu and Lys PMFs. The molecular configurations corresponding to the well are protonated Glu/Lys (Figure 2 insets), and this well in the PMF reveals the weak acid nature of both Glu and Lys. The steep wall at $\xi_{CEC} < 0.5$ Å is caused by the large energy penalty of compressing the O−H/N−H bond of Glu/Lys. The smooth uphill increase in the region of 0.5 Å < ξ_{CEC} < 1.4 Å arises from the energy barrier of proton disassociation from Glu/Lys to its first solvation shell water. The free energy was observed to reach a peak at around 1.4 Å, the point at which the proton is shared by Glu/Lys and the water, forming a Glu/Lys-replaced Zundel ($H_5O_2^+$) configuration (Figure 2 insets). The second free energy well was noted at $\xi_{CEC} \approx 2.3$ Å when the excess proton is found within the first solvation shell water. The resulting H_3O^+ and negatively charged Glu closely interact with each other to form a contact ion pair (CIP) (Figure 2A inset). The favorable Coulombic interaction yields a deeper potential well for Glu compared to that for Lys. In the latter case, a CIP cannot be formed between a neutral deprotonated Lys and H_3O^+, but the system is stabilized by forming an Eigen cation (H_3O^+ $(H_2O)_3$) with one water substituted by Lys (Figure 2B inset).

It is seen that a second free energy potential well at around 2.3 Å is shallower within the ωB97X model in comparison with that of the BLYP model, which is consistent with the expectation that the GGA functional BLYP overstabilizes the CIP due to its overestimation of charge transfer compared with the hybrid functional ωB97X. After moving beyond that well, the value of the free energy increases within the range of 2.3 Å < ξ_{CEC} < 4 Å, which corresponds to proton dissociation from the first shell water. Although the excess proton mostly resides on a second shell water when $\xi_{CEC} \approx 4$ Å (Figure 2 insets), this layer of water is sufficiently diffusive to become indistinguishable from bulk water, given that both PMFs reach a plateau after this point.

For a direct comparison with the experimental results, we calculated the pK_a of Glu and Lys in water from the PMF using eq 23. The resulting data are summarized in Table 2. We found that the ωB97X models accurately predicted the pK_a values for both Glu and Lys. By contrast, our results based on the BLYP models were rather less accurate but still within ∼1 pH unit of error. It is important to note that these MS-RMD models were not fit in the aqueous phase; rather, we fit the parameters from gas-phase CDFT calculations and transferred that model to the bulk water environment.

This success of the DM-based MS-RMD models encouraged us to apply them to the more complicated aqueous protein system, SNase, without reparameterization and in a directly transferable manner. Given that the ωB97X models showed better agreement with the experimental pK_a in water than the BLYP models, the latter models were not implemented in the subsequent SNase simulations.

Table 2. MS-RMD Predicted and Experimental pK_a Values of Glu and Lys

		in water	in SNase V66 mutants
Glu	simulation (ωB97X)	4.1 ± 0.2	9.8 ± 0.3[a]
	simulation (BLYP)	3.4 ± 0.4	
	experiment	4.15[b]	9.00−9.10[c] (8.73−9.28[d])
Lys	simulation (ωB97X)	10.7 ± 0.2	5.8 ± 0.3[a]
	simulation (BLYP)	10.0 ± 0.1	
	experiment	10.67[b]	5.61−6.05[e] (6.25−6.45[f])

[a]The PDB structure (1U9R) used in the simulation was the PHS form of SNase (engineered with three substitutions: P117G, H124A, and S128L). Note that the missing 45−50 loop in the solved structure was not modeled in our simulations, and thus the simulated system more resembles the Δ + PHS form (additional G50F, V51N, and 44−49 deletion from PHS). [b]Taken from ref 50. [c]pK_a in Δ + PHS by potentiometry taken from ref 51. [d]pK_a in PHS from chemical denaturation taken from ref 51. [e]pK_a in Δ + PHS from chemical denaturation taken from ref 52. [f]pK_a in PHS by potentiometry take from ref 53.

Benchmark of Reactive Models in SNase. SNase is a well-known protein model system for studying titration pK_a values of internal ionizable groups, including many different mutants.[51−53] Several SNase mutants have been generated by focusing on one of the buried hydrophobic residues, V66, and substituting it with Asp, Glu, and Lys. The pK_a of the mutated residue has been measured experimentally, and large pK_a shifts were observed in favor of the neutral forms of the residues.[51−53] Therefore, the SNase mutants provide an ideal and well-characterized protein environment for benchmarking our reactive MD models against experimental observations.

Classical equilibration simulations of the V66E and V66K mutants were performed for both protonated and deprotonated forms of E66 and K66. For the neutral form of E66 (deprotonated) and K66 (protonated), the residue adopted a buried conformation, while a charged E66/K66 was found to be exposed to water through the rotation of its side chain to stabilize its charge in the more highly polarizable medium (Figure 3A,C). This observation is consistent with the experimental pK_a shifts for Glu and Lys toward stabilizing a neutral form.

Considering the distinct conformations that are dependent on the protonation state, an RC, d_{SC}, reflecting the side chain orientation was defined as noted earlier and explicitly sampled along with the proton dissociation RC, ξ_{CEC}, which describes proton disassociation in the reactive 2D-US (details in Simulation Details). The next section provides a detailed discussion of these informative 2D PMF findings (Figure 3B,D) calculated by 2D-US, but we first discuss the pK_a values of E66 and K66 as calculated from the PMFs (Figure 3E). According to Table 2, the calculated pK_a values were found to be in very good agreement with those of their experimentally determined counterparts, again with no reparameterization of the DM MS-RMD model beyond that done in the gas phase. We found that the agreement with the experimental results was closer for the Lys model, which shows an excellent match. The Lys and Glu models were parametrized following the same procedure, except that the Morse bond of Lys was fit to a gas-phase energy scan while Glu was taken from our previous work,[6] where the bond parameters were chosen to fit the original CHARMM harmonic bond potential. The better

https://doi.org/10.1021/acs.jpcb.1c05992
J. Phys. Chem. B 2021, 125, 10471−10480

performance of our Lys model suggests that the current Glu model may be readily improved by a bond energy scan followed by a reparameterization. Notably, we used exactly the same MS-RMD parameters of Glu/Lys in water when simulating SNase, and even then the MS-RMD models demonstrated high accuracy and transferability in predicting the acidic and basic residue pK_a values in these distinct environments.

Proton Transport and Conformation Coupling in SNase. As discussed in the prior section, we compared the predicted pK_a values with the experimental values. It is worth noting that a reactive MD model is not only useful for calculating reaction equilibrium constants but also can predict reaction rates when combined with appropriate kinetic theory and modeling, such as activated rate theory[54] or a more complex Markov state model.[55] More importantly, atomistic details obtained from reactive MD simulations enable a detailed exploration of the reaction mechanisms as well as the identification of crucial factors and interactions affecting the mechanism. In terms of SNase, an important question arises regarding how the proton transport is coupled with the protein conformations. As discussed earlier, the classical equilibrium data indicate that the side chain orientation of E66/K66 depends on its protonation state. However, in the absence of knowledge of the explicit reaction pathways, classical simulations can determine only the metastable basin states. In contrast, reactive MD can reveal whether the side chain rotation facilitates the protonation/deprotonation process or is instead a result of the amino acid conformation responding to a protonation state change. Interestingly, the answer is likely neither alone according to our 2D PMF results (Figure 3B,D). The most probable reaction pathway (or, equivalently, the minimum free energy path, MFEP) shows that the protonation dissociation and the side chain rotation are not two independent processes that happen in a stepwise manner. Instead, these two processes are highly and reciprocally coupled, which can be verified by the ramped slope of the MFEP connecting the two end points of the charged, exposed E66/K66 state and the neutral buried state. It is worth noting that the coupling between ξ_{CEC} and d_{SC} in the V66E mutant is the highest in the ξ_{CEC} < 4 Å range, while the MFEP passing ξ_{CEC} = 4 Å has a flatter shape. The strong coupling in the proton dissociation/association region is attributed to the solvation penalty of both a buried charged Glu and the hydrated excess proton. In contrast, the V66K mutant exhibits weaker coupling in the proton dissociation phase because the charged Lys adopts a water-exposed position that provides the solvation environment for the dissociated excess proton. Because of the ramped slope of the MFEP, if the PT degree of freedom ξ_{CEC} is driven by some external force such as a proton gradient and if this driving is a quasistatic process, then the movement of ξ_{CEC} will result in a response of the movement in d_{SC} along the curvy MFEP (and vice versa). Importantly, in addition to the side chain orientation observed in this study, we found in our previous work that this type of coupling typically occurs between PT and water hydration in confined hydrophobic spaces of proteins and other materials.[8,10,12,30,56] This outcome may also be the case for PT and ligand transport in some proton-driven transporters; indeed, if this is the case, then it is understandable why the excess proton and the ligand can exhibit a coupled driving force for the transport.

CONCLUSIONS

As a generalization to the previous FM MS-RMD approach, we have presented here a DM framework to systematically parametrize reactive MD models from CDFT. We used the proton dissociation reactions of Glu and Lys as our case studies and found that the DM-based models can reproduce the experimental absolute pK_a of the acid and the basic amino acid in water. We also confirmed that the same reactive MD models were able to accurately capture the large pK_a shifts introduced by the apolar SNase protein environment. In addition, the efficiency of the reactive MD enabled us to compute the joint free energy surface of the excess proton CEC coordinate, along with the side chain orientation of E66/K66, from which we discovered that PT is coupled to this local conformational change. As such, the two processes can be considered to be mutually cooperative, in contrast to being a sequential process whereby the protonation state change occurs either before or after a conformational change. Since the results indicate that our models are transferable, we believe that they can likely be readily applied to other biomolecular and biomaterial systems. Moreover, the noted efficiency of these MS-RMD models can facilitate the discovery of other protein motions and degrees of freedom coupled to the PT process, such as hydration, global conformational changes, and ligand transport. Finally, we anticipate that the DM MS-RMD approach can provide a potentially powerful tool for developing accurate and transferrable reactive MD models for other chemical reactions in the future.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.1c05992.

Additional table of the optimized MS-RMD model parameters using BLYP CDFT data; additional figures showing the fitted MS-RMD model wave function **c**, illustrating the reactive coordinate definitions in SNase, and showing the statistical errors in PMFs of SNase (PDF)

AUTHOR INFORMATION

Corresponding Author

Gregory A. Voth − *Department of Chemistry, Chicago Center for Theoretical Chemistry, James Franck Institute, and Institute for Biophysical Dynamics, University of Chicago, Chicago, Illinois 60637, United States;* orcid.org/0000-0002-3267-6748; Email: gavoth@uchicago.edu

Author

Chenghan Li − *Department of Chemistry, Chicago Center for Theoretical Chemistry, James Franck Institute, and Institute for Biophysical Dynamics, University of Chicago, Chicago, Illinois 60637, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.1c05992

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research was supported in part by the National Institute of General Medical Sciences (NIGMS) of the U.S. National

Institutes of Health (NIH) through grant R01 GM053148 and in part by the U.S. Department of Energy (DOE), Office of Basic Energy Sciences, Separation Science Program of the Division of Chemical Sciences, Geosciences, and Biosciences under award number DE-SC0018648. Computational resources were provided by the Research Computing Center (RCC) at the University of Chicago. The authors appreciate insightful discussions with Dr. Xinyou Ma.

REFERENCES

(1) Marx, D.; Hutter, J. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods.*; Cambridge University Press: 2009.
(2) Gillan, M. J.; Alfe, D.; Michaelides, A. Perspective: How good is DFT for water? *J. Chem. Phys.* **2016**, *144*, 130901.
(3) Marsalek, O.; Markland, T. E. Quantum Dynamics and Spectroscopy of Ab Initio Liquid Water: The Interplay of Nuclear and Electronic Quantum Effects. *J. Phys. Chem. Lett.* **2017**, *8*, 1545−1551.
(4) Yamashita, T.; Peng, Y.; Knight, C.; Voth, G. A. Computationally Efficient Multiconfigurational Reactive Molecular Dynamics. *J. Chem. Theory Comput.* **2012**, *8*, 4863−4875.
(5) Knight, C.; Lindberg, G. E.; Voth, G. A. Multiscale reactive molecular dynamics. *J. Chem. Phys.* **2012**, *137*, 22A525.
(6) Nelson, J. G.; Peng, Y.; Silverstein, D. W.; Swanson, J. M. J. Multiscale Reactive Molecular Dynamics for Absolute pK(a) Predictions and Amino Acid Deprotonation. *J. Chem. Theory Comput.* **2014**, *10*, 2729−2737.
(7) Lee, S.; Liang, R. B.; Voth, G. A.; Swanson, J. M. J. Computationally Efficient Multiscale Reactive Molecular Dynamics to Describe Amino Acid Deprotonation in Proteins. *J. Chem. Theory Comput.* **2016**, *12*, 879−891.
(8) Lee, S.; Swanson, J. M. J.; Voth, G. A. Multiscale Simulations Reveal Key Aspects of the Proton Transport Mechanism in the ClC-ec1 Antiporter. *Biophys. J.* **2016**, *110*, 1334−1345.
(9) Lee, S.; Mayes, H. B.; Swanson, J. M. J.; Voth, G. A. The Origin of Coupled Chloride and Proton Transport in a Cl-/H+ Antiporter. *J. Am. Chem. Soc.* **2016**, *138*, 14923−14930.
(10) Liang, R.; Swanson, J. M. J.; Peng, Y.; Wikström, M.; Voth, G. A. Multiscale simulations reveal key features of the proton-pumping mechanism in cytochrome c oxidase. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 7420−7425.
(11) Liang, R. B.; Swanson, J. M. J.; Wikstrom, M.; Voth, G. A. Understanding the essential proton-pumping kinetic gates and decoupling mutations in cytochrome c oxidase. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 5924−5929.
(12) Wang, Z.; Swanson, J. M. J.; Voth, G. A. Modulating the Chemical Transport Properties of a Transmembrane Antiporter via Alternative Anion Flux. *J. Am. Chem. Soc.* **2018**, *140*, 16535−16543.
(13) Kaduk, B.; Kowalczyk, T.; Van Voorhis, T. Constrained Density Functional Theory. *Chem. Rev.* **2012**, *112*, 321−370.
(14) Mulliken, R. S. Electronic Population Analysis on Lcao-Mo Molecular Wave Functions 0.2. Overlap Populations, Bond Orders, and Covalent Bond Energies. *J. Chem. Phys.* **1955**, *23*, 1841−1846.
(15) Hirshfeld, F. L. Bonded-Atom Fragments for Describing Molecular Charge-Densities. *Theor Chim Acta* **1977**, *44*, 129−138.
(16) Becke, A. D. A Multicenter Numerical-Integration Scheme for Polyatomic-Molecules. *J. Chem. Phys.* **1988**, *88*, 2547−2553.
(17) Cembran, A.; Song, L. C.; Mo, Y. R.; Gao, J. L. Block-Localized Density Functional Theory (BLDFT), Diabatic Coupling, and Their Use in Valence Bond Theory for Representing Reactive Potential Energy Surfaces. *J. Chem. Theory Comput.* **2009**, *5*, 2702−2716.
(18) Ren, H. S.; Provorse, M. R.; Bao, P.; Qu, Z. X.; Gao, J. L. Multistate Density Functional Theory for Effective Diabatic Electronic Coupling. *J. Phys. Chem. Lett.* **2016**, *7*, 2286−2293.
(19) Cukierman, S. Et tu, Grotthuss! and other unfinished stories. *Biochim. Biophys. Acta, Bioenerg.* **2006**, *1757*, 876−885.
(20) Bush, D. R. Proton-Coupled Sugar and Amino-Acid Transporters in Plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1993**, *44*, 513−542.
(21) Decoursey, T. E. Voltage-gated proton channels and other proton transfer pathways. *Physiol. Rev.* **2003**, *83*, 475−579.
(22) Weber, J.; Senior, A. E. ATP synthesis driven by proton transport in F1F0-ATP synthase. *FEBS Lett.* **2003**, *545*, 61−70.
(23) Turina, P.; Samoray, D.; Graber, P. H+/ATP ratio of proton transport-coupled ATP synthesis and hydrolysis catalysed by CF0F1-liposomes. *EMBO J.* **2003**, *22*, 418−426.
(24) Pinto, L. H.; Lamb, R. A. The M2 proton channels of influenza A and B viruses. *J. Biol. Chem.* **2006**, *281*, 8997−9000.
(25) Nevo, Y.; Nelson, N. The NRAMP family of metal-ion transporters. *Biochim. Biophys. Acta, Mol. Cell Res.* **2006**, *1763*, 609−620.
(26) Wraight, C. A. Chance and design - Proton transfer in water, channels and bioenergetic proteins. *Biochim. Biophys. Acta, Bioenerg.* **2006**, *1757*, 886−912.
(27) Buch-Pedersen, M. J.; Pedersen, B. P.; Veierskov, B.; Nissen, P.; Palmgren, M. G. Protons and how they are transported by proton pumps. *Pfluegers Arch.* **2009**, *457*, 573−579.
(28) Zhang, Y.; Voth, G. A. The Coupled Proton Transport in the ClC-ec1 Cl-/H+ Antiporter. *Biophys. J.* **2011**, *101*, L47−L49.
(29) Parker, J. L.; Li, C.; Brinth, A.; Wang, Z.; Vogeley, L.; Solcan, N.; Ledderboge-Vucinic, G.; Swanson, J. M. J.; Caffrey, M.; Voth, G. A.; et al. Proton movement and coupling in the POT family of peptide transporters. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 13182−13187.
(30) Li, C.; Yue, Z.; Espinoza-Fonseca, L. M.; Voth, G. A. Multiscale Simulation Reveals Passive Proton Transport Through SERCA on the Microsecond Timescale. *Biophys. J.* **2020**, *119*, 1033−1040.
(31) Cuma, M.; Schmitt, U. W.; Voth, G. A. A multi-state empirical valence bond model for weak acid dissociation in aqueous solution. *J. Phys. Chem. A* **2001**, *105*, 2814−2823.
(32) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. 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.* **2012**, *8*, 3257−3273.
(33) Chai, J. D.; Head-Gordon, M. Systematic optimization of long-range corrected hybrid density functionals. *J. Chem. Phys.* **2008**, *128*, 084106.
(34) Kuhne, T. D.; Iannuzzi, M.; Del Ben, M.; Rybkin, V. V.; Seewald, P.; Stein, F.; Laino, T.; Khaliullin, R. Z.; Schutt, O.; Schiffmann, F.; et al. CP2K: An electronic structure and molecular dynamics software package - Quickstep: Efficient and accurate electronic structure calculations. *J. Chem. Phys.* **2020**, *152*, 194103.
(35) Lehtola, S.; Steigemann, C.; Oliveira, M. J. T.; Marques, M. A. L. Recent developments in LIBXC - A comprehensive library of functionals for density functional theory. *Softwarex* **2018**, *7*, 1−5.
(36) Biswas, R.; Tse, Y. L. S.; Tokmakoff, A.; Voth, G. A. Role of Presolvation and Anharmonicity in Aqueous Phase Hydrated Proton Solvation and Transport. *J. Phys. Chem. B* **2016**, *120*, 1793−1804.
(37) Zhang, L. F.; Han, J. Q.; Wang, H.; Car, R.; Weinan, E. Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics. *Phys. Rev. Lett.* **2018**, *120*, 143001.
(38) Gao, F.; Han, L. Implementing the Nelder-Mead simplex algorithm with adaptive parameters. *Computational Optimization and Applications* **2012**, *51*, 259−277.
(39) Martyna, G. J.; Klein, M. L.; Tuckerman, M. Nose-Hoover Chains - the Canonical Ensemble Via Continuous Dynamics. *J. Chem. Phys.* **1992**, *97*, 2635−2643.
(40) Barducci, A.; Bussi, G.; Parrinello, M. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008**, *100*, 020603.
(41) Maupin, C. M.; Wong, K. F.; Soudackov, A. V.; Kim, S.; Voth, G. A. A multistate empirical valence bond description of protonatable amino acids. *J. Phys. Chem. A* **2006**, *110*, 631−639.

https://doi.org/10.1021/acs.jpcb.1c05992
J. Phys. Chem. B 2021, 125, 10471−10480

(42) Denisov, V. P.; Schlessman, J. L.; Garcia-Moreno, B.; Halle, B. Stabilization of internal charges in a protein: Water penetration or conformational change? *Biophys. J.* **2004**, *87*, 3982−3994.
(43) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(44) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(45) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *Softwarex* **2015**, *1*−*2*, 19−25.
(46) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules 0.1. The Method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(47) Roux, B.; Andersen, O. S.; Allen, T. W. Comment on ″Free energy simulations of single and double ion occupancy in gramicidin A″ [J. Chem. Phys. 126, 105103 (2007)]. *J. Chem. Phys.* **2008**, *128*, 227101.
(48) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular-Dynamics. *J. Comput. Phys.* **1995**, *117*, 1−19.
(49) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604−613.
(50) Haynes, W. M. *CRC Handbook of Chemistry and Physics*; CRC Press: 2014.
(51) Karp, D. A.; Stahley, M. R.; Garcia-Moreno, E. B. Conformational Consequences of Ionization of Lys, Asp, and Glu Buried at Position 66 in Staphylococcal Nuclease. *Biochemistry* **2010**, *49*, 4138−4146.
(52) Fitch, C. A.; Karp, D. A.; Lee, K. K.; Stites, W. E.; Lattman, E. E.; García-Moreno, E. B. Experimental pKa values of buried residues: analysis with continuum methods and role of water penetration. *Biophys. J.* **2002**, *82*, 3289−3304.
(53) García-Moreno, B. E.; Dwyer, J. J.; Gittis, A. G.; Lattman, E. E.; Spencer, D. S.; Stites, W. E. Experimental measurement of the effective dielectric in the hydrophobic core of a protein. *Biophys. Chem.* **1997**, *64*, 211−224.
(54) Chandler, D. *Introduction to Modern Statistical Mechanics*; Oxford University Press: New York, 1987.
(55) Bowman, G. R.; Pande, V. S.; Noé, F. *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*; Springer Science & Business Media: 2013; Vol. 797.
(56) Peng, Y.; Swanson, J. M. J.; Kang, S.-g.; Zhou, R.; Voth, G. A. Hydrated Excess Protons Can Create Their Own Water Wires. *J. Phys. Chem. B* **2015**, *119*, 9212−9218.

https://doi.org/10.1021/acs.jpcb.1c05992
J. Phys. Chem. B 2021, 125, 10471−10480