# Multiscale Reactive Molecular Dynamics for Absolute pK_a Predictions and Amino Acid Deprotonation

J. Gard Nelson, Yuxing Peng, Daniel W. Silverstein, and Jessica M. J. Swanson*

Department of Chemistry, Institute for Biophysical Dynamics, and Computation Institute, University of Chicago, 5735 S. Ellis Ave., Chicago, Illinois 60637, United States

S Supporting Information

ABSTRACT: Accurately calculating a weak acid's pK_a from simulations remains a challenging task. We report a multiscale theoretical approach to calculate the free energy profile for acid ionization, resulting in accurate absolute pK_a values in addition to insights into the underlying mechanism. Importantly, our approach minimizes empiricism by mapping electronic structure data (QM/MM forces) into a reactive molecular dynamics model capable of extensive sampling. Consequently, the bulk property of interest (the absolute pK_a) is the natural consequence of the model, not a parameter used to fit it. This approach is applied to create reactive models of aspartic and glutamic acids. We show that these models predict the correct pK_a values and provide ample statistics to probe the molecular

mechanism of dissociation. This analysis shows changes in the solvation structure and Zundel-dominated transitions between the protonated acid, contact ion pair, and bulk solvated excess proton.
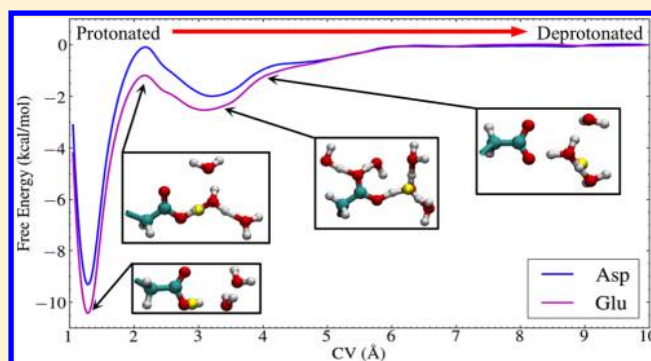
INTRODUCTION

Weak acid ionization plays a central role in many chemical, biological, and industrial processes. Measuring and predicting the free energy of this process (i.e., pK_a values) have become important experimental and theoretical tools, enabling control over and insight into acid/base chemistry and pH-dependent biomolecular transformations. Although a large number of theoretical methods have been developed to predict pK_a values (see recent reviews 1 and 2 and references therein), less attention has been given to understanding the molecular-scale mechanism of acid dissociation. This is largely due to the complexity and cost of analyzing the full free energy surface, which involves not only a chemical reaction but also significant solvent and solute reorganization. Instead, most pK_a methods rely on the use of thermodynamic cycles that break the complex process of deprotonation into computationally tractable steps. For small molecules, this cycle typically combines the gas phase dissociation energy, calculated with high-level quantum mechanical (QM) calculations, with de/solvation free energies of the reactant and product species, calculated with continuum solvent approaches.[2−6] While the gas phase energies are quite accurate, the simple continuum description of solvation free energies often requires parameter adjustment to reproduce all of the entropic and enthalpic complexities that are inherent in molecular solvation. They have been improved by hybrid approaches that include some number of explicit water molecules in the solvation free energy analysis.[7,8] In conjunction with methodological refinements, these QM-continuum solvent models have enabled pK_a calculations of

small acids to within a few units of the experimentally determined values.[4−6,8]

In contrast, the suite of methods that have been developed to predict biological pK_a values have typically used a thermodynamic cycle in which the solute remains in a solvated environment. These approaches range from microscopic to continuum to completely empirical. Each has advantages and drawbacks as well, discussed in ref 2, however none has yet demonstrated the ability to predict pK_a values within a few pK_a units of the experimental value for all test cases. It has been suggested that challenging moieties (e.g., buried residues that have large pK_a shifts or those with strong coupling to conformational changes or other titration sites) will require methods that better capture the underlying physics.[1] These conclusions highlight how interesting and nuanced a biomolecular system can be. While it is clear that protonation states influence everything from stability to binding and reactivity, the details of their influence is still sometimes poorly understood.

Few methods to date have simulated the actual process of interest, the exchange of an excess positive charge from an acid to bulk water. To do so involves calculating the free energy surface for proton dissociation in the bulk phase environment, from which one learns about the molecular mechanism and

rate. This requires three things: (1) a method that describes the underlying physics of bond formation and cleavage in the bulk phase environment, (2) a method that is efficient enough that when combined with enhanced sampling approaches can capture the distribution of phase space that defines a free energy in the condensed phase (where entropy and energy are often a similar magnitude), and (3) a reaction coordinate that is flexible enough to track the transferring excess charge defect.

The first of these requirements is most obviously filled by QM approaches, which have been used in a handful of helpful studies to probe acid dissociation. Ivanov et al. have used Carr–Parrinello molecular dynamics (CPMD) to calculate the free energy profile of histidine dissociation.[9,10] Although the chosen reaction coordinate limited their analysis to stopping just beyond the transition state,[11] they were able to calculate the correct relative pK_a value by subtracting the equivalent profile for water autodissociation.[9] Similarly, Park et al. used CPMD in combination with metadynamics to sample the configurational space of acetic acid in bulk water.[12] While their trajectories were too short to sufficiently converge the free energy surface, they were able to observe and characterize multiple protonation and deprotonation events. These reactions occurred along a well-characterized pathway where the excess proton is briefly shared between the acid and solvating water. They also noted the existence of a shallow minimum in the free energy surface corresponding to the existence of a contact ion pair (CIP) that is separated from the protonated state by a small energy barrier. In an alternative approach that still describes the bond dissociation at the ab initio level, Uddin et al. recently used QM/MM to calculate the free energy surfaces for several small molecules.[13] Although they obtained quite accurate pK_a values, their method cannot describe the mechanism of dissociation since they include only one water molecule and the acid in the QM region. Hence, this method changes the physical process from the dissociation of a delocalized charge defect to proton transfer followed by separation of a localized hydronium cation and acid anion in a solvating MM environment. Just as for thermodynamic cycles, their results show that sampling an alternative pathway between reactant and product states can yield accurate pK_a values for small molecules.

Another approach that describes the bond dissociation process is reactive MD (RMD).[14−17] The efficiency of these models allows them to sample reactive dynamics on much longer length and time scales than ab initio MD. Since they are parametrized to work with MM force fields, they also avoid the boundary issues suffered by QM/MM calculations. Both of these factors make RMD models ideal for simulating reactions in large, complex systems with slow dynamics. Proteins and proton exchange membranes are two important examples of such systems.[18−25] Without sufficient sampling in these types of systems, the final result can be strongly influenced by the initial conditions. A common criticism of RMD models is that they are only approximations to the true quantum mechanical behavior of the system. While this is certainly true, if the functional form is both flexible enough and parametrized correctly, then the reactive model will retain the ability to accurately reproduce the potential and free energy surfaces from which it was parametrized.[26,27]

The multistate empirical valence bond (MS-EVB) method is a RMD force field that has been used to characterize proton transport in many condensed phase and biological environments.[11,28] It is especially well suited to the challenge of acid ionization since it inherently defines a center of excess charge

that can be used to track reactions involving proton transport, thereby avoiding issues with geometric reaction coordinates.[9,11] Although MS-EVB has been used to study weak acids[29,30] and amino acid dissociation,[19,31] these efforts involved a larger degree of empiricism since the pK_a value was needed to fit the MSEVB potential. It has been a long sought goal to find a procedure that would minimize the empiricism in RMD models, such that the bulk property of interest was derived entirely from quantum data in combination with extensive sampling.

This work reports the achievement of that goal, a multiscale approach to calculate the free energy profile for acid ionization that is based entirely on QM data and yields an accurate absolute pK_a value. An iterative procedure, in the spirit of adaptive force matching,[32,33] is used to sample the reactive phase space, obtain reference ab initio data (QM/MM forces), and fit a RMD potential. Care has been taken to modify the functional form such that it is sufficiently flexible to reproduce the QM/MM reference data and, in principle, capture the essential physics. Our approach is applied to aspartic acid (Asp) and to glutamic acid (Glu). Both of these amino acids play crucial roles in proton and ion transport in biological channels and pumps and are also commonly involved in salt bridges that influence protein dynamics, structure, and binding. We show that accurate estimates of the pK_a values of both Asp and Glu are obtained from the resulting reactive models, which is the first time a reactive model has been able to accurately predict the correct pK_a without being explicitly parametrized to do so. Finally, we discuss the mechanistic description our models provide of acid deprotonation in water.

To emphasize that these models were fit entirely to QM/MM data and without adjustments to match the experimental pK_a, we will hereafter refer to them as multiscale reactive molecular dynamics (MS-RMD) models.[26] It should be noted, however, that the MS-RMD Asp and Glu models are parametrized to work with a refined version of the MS-EVB3 hydronium model that accurately describes a proton solvation and transport in water.[28]

■ METHODOLOGY

MS-RMD Description. MS-RMD is similar to the MS-EVB framework that has been successfully used to model proton solvation and transport in many aqueous systems.[11,15,28,31,34,35] The power in this approach lies in its ability to describe electronic delocalization as a linear combination of topologically distinct states. For a given configuration, states $|i\rangle$ and $|j\rangle$ differ only in their bonding topology. Each diabatic state describes a localized protonated species. For example, state $|i\rangle$ could be a protonated amino acid in a box of water (Figure 1a). State $|j\rangle$ would have the same coordinates but have a topology describing a CIP between a solvated hydronium ion and a deprotonated amino acid (Figure 1b). At each step in the simulation, a state search algorithm first determines all the possible states based on geometric criteria and then calculates the energy of each state. The lowest energy state is termed the pivot state and is the start of the state search algorithm at the next step. Since the number of waters, and therefore the number of possible states, grows exponentially with distance from the pivot state, only states in the first three solvation shells are included in the Hamiltonian. This has been shown to provide sufficient accuracy and energy conservation.

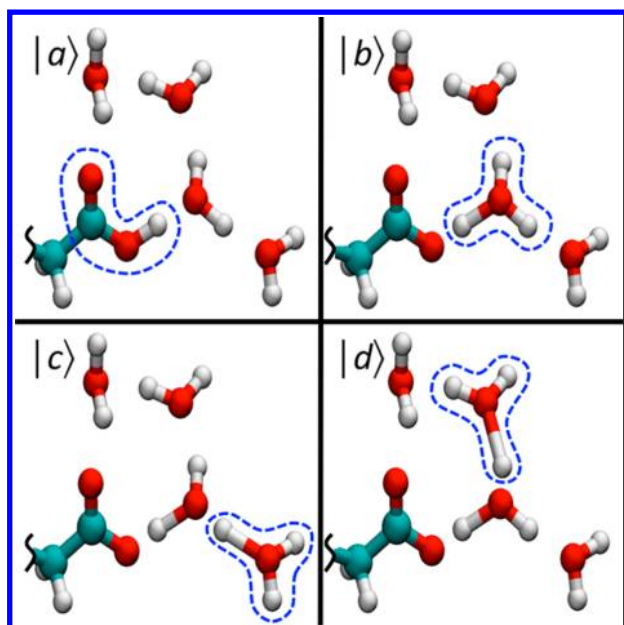The reactive system is described by the following Hamiltonian:

Figure 1. Several different bonding topologies are shown for a given set of nuclear coordinates, **r**. Dashed lines surround the protonated molecule, as defined by the bonding topology. (a) A protonated Asp solvated by water. (b−d) Different possible locations for the hydronium molecule.

$$\mathbf{H}^{\text{EVB}} = \sum_{ij} |i(\mathbf{r})\rangle h_{ij}(\mathbf{r})\langle j(\mathbf{r})| \tag{1}$$

where **r** is the vector describing the nuclear coordinates, h_{ij} is the sum of MM potential terms for state i, and h_{ij} is the coupling between states $|i\rangle$ and $|j\rangle$. The details of these terms will be provided in the Force Field section. With the Hamiltonian defined this way, the relative weight of each state is described by the components of the eigenvector, c_i:

$$\mathbf{H}c = E_0 c \tag{2}$$

Finally, forces are calculated using the Hellmann−Feynman theorem. These forces are then passed to a standard MD integrator such as velocity Verlet.

Free energy calculations require a continuously varying reaction coordinate (RC) to describe the process of interest. The EVB formalism provides a convenient reference for the RC; the center of the excess charge (CEC):

$$\mathbf{r}_{\text{CEC}} = \sum_{i}^{N_{\text{EVB}}} c_i^2(\mathbf{r})\mathbf{r}_i^{\text{COC}} \tag{3}$$

In this equation, $\mathbf{r}_i^{\text{COC}}$ is the geometric center-of-charge for the protonated molecule in state i, and c_i is the relative weight of that state. We then use the distance between the CEC and the center of mass of the Asp or Glu carboxyl group to describe the progress of the deprotonation reaction.

Force Field. The diagonal elements of the MS-RMD Hamiltonian matrix are derived from the standard CHARMM MM force field.[36] The protonated form uses all of the standard parameters with the exception of the acidic O−H bond. This bond is replaced with a Morse potential to allow dissociation:

$$U^{\text{Morse}} = D[1 - \exp(-\alpha(r - r_0))]^2 \tag{4}$$

where r is the bond length and D, α, and r_0 are parameters fit to reproduce the bond length and frequency of the harmonic

potential it replaces while still allowing for a reasonable bond dissociation energy. The deprotonated form of amino acid uses all of the standard CHARMM bonded and nonbonded terms and parameters. Since the CHARMM force field was not originally intended for use in reactive simulations, additional repulsive terms are used between the carboxyl oxygens of the deprotonated model and the hydronium oxygen and hydrogens. These terms help to correct for the overattraction at close distances due to the use of point charges.[37] The functional forms for the repulsions are the same as those used in the MS-EVB3 hydronium model:[28]

$$V_{\text{OO}_k}^{\text{rep}} = B \exp[-b(R_{\text{OO}_k} - d_{\text{OO}}^0)]\cdot \sum_{j}^{3} \exp[-b'(R_{\text{H}_j\text{O}_k})^2] \tag{5}$$

$$V_{\text{HO}_k}^{\text{rep}} = C \exp[-c(R_{\text{HO}_k} - d_{\text{OH}}^0)] \tag{6}$$

where B, b, b', C, and c are fitted parameters. Following the logic in previous work, and to avoid parameter redundancy, d_{OO}^0 and d_{OH}^0 were fixed the same values used in MS-EVB3. The sum of Gaussians term in eq 5 is included to help the fitted potential to reproduce the correct asymmetric solvation structure around the hydronium, and the same switching function is used as given in eq 9 of ref 28.

The Asp and Glu models are designed to work with a refined version of the MS-EVB3 force field, herein referred to as MSEVB 3.1. This refinement uses the same functional forms for the Hamiltonian but revised parameters. Using a genetic algorithm, the fitting parameters were optimized using the original ab initio data sets as well as potential energy surface scans for the Eigen ion at the same level of the original ab initio calculations. Furthermore, the oxygen partial charge of the hydronium is an adjustable parameter (the charge for the hydrogens is automatically determined to maintain the hydronium +1 total charge). MS-EVB 3.1 better describes contact ion pairs. Parameters for this force field are presented in the Supporting Information.

In this study, the limiting species are the de/protonated Asp or Glu models and the MS-EVB 3.1 solvated proton in SPC-F$_\text{W}$ water. The MS-RMD force field is parametrized to smoothly switch from one protonated species to another. The asymmetry of the reaction necessitates several modifications to the original MS-EVB equations, which were designed with water in mind. The off diagonal coupling term, h_{ij}, is designed to reproduce the correct transition state geometry. Previous MS-EVB models used a complex off-diagonal term to describe a symmetric reaction pathway between two waters. However, the asymmetry of the Asp-water system was better fit by a simple Gaussian potential:

$$h_{ij}^{\text{Asp}}(r_{\text{OH}}) = c_1 \exp[-c_2(r_{\text{OH}} - c_3)^2] \tag{7}$$

where r_{OH} is the distance between the deprotonated Asp oxygen and the excess proton. The other parameters determine the shape and size of the Gaussian and are fit to the QM/MM data. In practice, protonation reactions occurring between different species require a constant energy term (V_{ii}) to be added to the deprotonated state. This term accounts for the difference in energy resulting from different MM force fields. Without it, the protonated and deprotonated Asp potential energy surfaces would be offset by hundreds of kilocalories per mole, and no reaction would ever take place. This difference is due primarily to electrostatics and can be estimated by

subtracting the coulomb energy of the most favorable hydronium state (containing the favorable electrostatic interactions between the hydronium cation and amino acid anion) from the coulomb energy of the protonated (and neutral) amino acid state. Figure 2 shows a plot of the average
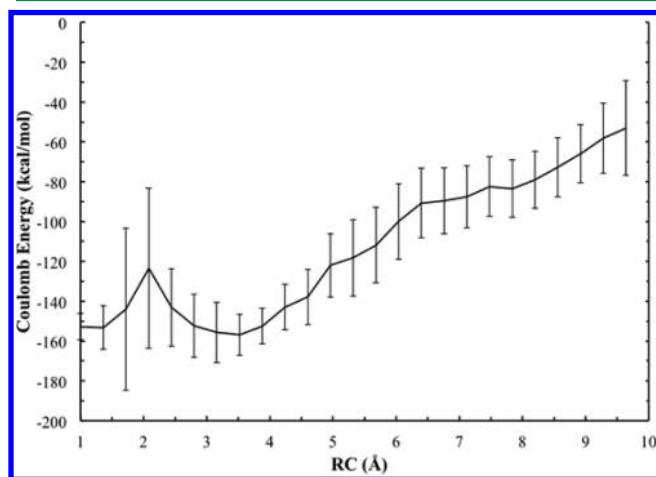

Figure 2. Plot of the difference in coulomb energy between the protonated Asp state and the lowest energy non-Asp state. At short distances, this quantity is a good initial estimate of V_{ii}.

value of this term as a function of RC. Since V_{ii} controls the relative stability of the protonated and deprotonated states, it was used in past work to tune the behavior of the model to reproduce experimental quantities such as pK_a.[31] In the current work, an iterative fitting approach is used to rigorously determine the value of V_{ii} along with all other MS-RMD parameters.

Fitting Procedure. Our fitting procedure is described graphically in Figure 3. It is similar in some ways to that used in previous work in which MS-RMD models were force matched from AIMD data.[26,38] There are, however, a few important differences, including the use of QM/MM instead of AIMD, the use of empirical functional forms instead of tabulated
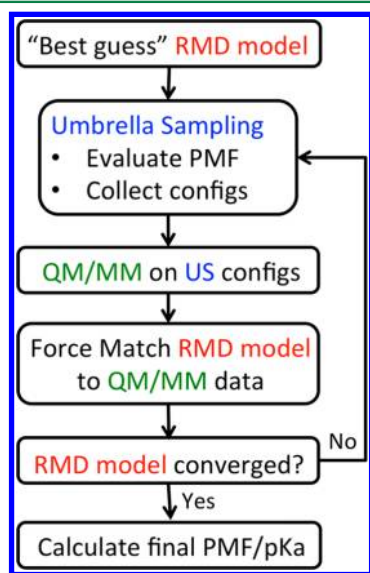

Figure 3. Schematic representation of the iterative fitting procedure used in this paper. Details of each step are discussed in the text.

potentials, and the use of an iterative scheme for sampling configuration space starting from a guessed Hamiltonian. These choices were partially motivated by the success that has been demonstrated with the adaptive force matching method by Wang and co-workers.[32,39] The general fitting procedure is conceptually simple. First, umbrella sampling simulations are run with a "best guess" reactive Hamiltonian in order to generate a set of configurations along the reactive pathway. Next, high-level QM/MM calculations are performed to collect the atomic forces for those configurations. The new reactive Hamiltonian is then used to generate new configurations via umbrella sampling, and the process is repeated until the parameters reach the desired convergence.

Due to the nonlinearity of several of the reactive force field parameters, the reactive model is fit to the QM/MM reference data using a genetic algorithm[40] to minimize the residual:

$$\chi^2 = \frac{1}{3N_C N_A} \sum_{j=1}^{N_C} \sum_{i=1}^{N_A} w(\mathbf{r}_{ij}) |\mathbf{F}_{ij} - \mathbf{F}_{ij}^{QM/MM}|^2$$

(8)

Here, N_C and N_A are the number of configurations and the number of atoms in each configuration, respectively. $w(\mathbf{r}_{ij})$ is the weight for each atom that, unless specified otherwise, is set to 1. $\mathbf{F}_{ij}$ is the atomic force from the current MS-RMD parameter set and $\mathbf{F}_{ij}^{QM/MM}$ is the reference force from the QM/MM calculation.

The fitting calculations were run in parallel and used 4000 genomes (parameter sets) per generation. To lessen the danger of overfitting, a two-step minimization scheme was used. The first step followed the uniformly distributed mutation scheme outlined in section 3.1 of ref 40. The range of parameters was chosen to be as broad as possible without allowing unreasonable values. For example, c_3 in the off-diagonal term would be unreasonable if it were greater than the AspO−O$_W$ distance. Once a converged parameter set was reached with the uniform distribution, the mutation scheme was switched to the normally distributed scheme described in section 3.2 of ref 40. This scheme is very good at quickly finding a local minimum and was therefore used to find the best genome in the vicinity of the result from the first step. Since the second scheme is easily trapped in local minima, it was not ideal for use on its own. The combination described here was found to quickly and reliably converge on the optimum solution. The discrete recombination scheme described in section 2.2 of ref 40 is used with both mutation schemes.

The nature of the reactive model introduces additional complexity into this otherwise simple scheme. V_{ii}, which directly controls the depth of the protonated well relative to the deprotonated CIP, is a constant that is only included for states in which the amino acid is protonated. Hence, its influence is negligible beyond the transition state. The off-diagonal term acts over a broader range spanning the protonated amino acid and CIP configurations but also has a negligible effect once the CIP dissociates. In contrast, the repulsive terms, which are only applied to the *deprotonated* amino acid, are significant at longer distances and negligible at distances below the transition state. Furthermore, since both V_{ii} and the repulsive terms are added to the diagonal matrix elements, they tend to have correlated behavior if fit simultaneously. Thus, the value of V_{ii} for the first iteration after generating new QM/MM data is chosen to be the average value described in Figure 2. This value has been found to be within a few kilocalories per mole of the final value of V_{ii} and is therefore a good initial approximation. Holding the

value of V_{ii} fixed, the off-diagonal and repulsive terms can be fit to the full set of QM/MM reference forces. Once the off-diagonal and repulsive terms reach their optimum value for the current value of V_{ii}, umbrella sampling is used to generate a PMF with the model. Since V_{ii}'s effect is most important at and around the transition state, a range of RC values immediately surrounding the transition state is chosen from the PMF, and configurations from that range are then used to fit a new value of V_{ii}. Using this new value of V_{ii}, the off-diagonal and repulsive terms are refit while V_{ii} is again held constant. This procedure fits an empirical model from solely QM/MM data and results in models that correctly predict the bulk phase pK_a values for both Asp and Glu.

Computational Details. Solution simulations of Asp and Glu include one solute molecule solvated by 486 and 467 waters, respectively. To reduce interactions with the charged backbone protonatable sites, the N-termini were capped with an acetyl group and the C-termini were capped with methylamine. The systems were equilibrated in the NPT ensemble at 1 atm and 310 K for 200 ps before changing to the NVT ensemble with a cubic box of 24.55486 Å on a side for Asp and 24.24938 Å for Glu. All reactive simulations were run with a modified version of the LAMMPS simulation package.[41,42] The RC was chosen to be the distance between the CEC and the center-of-mass of the amino acid's side chain carboxylic acid. Umbrella sampling was used to evenly sample the RC. For the Asp system, 25 windows were spaced every 0.36 Å starting at a RC value of 1.0 Å. Umbrella restraints were chosen to be 20.0 kcal/mol. The Glu system used 36 windows spaced every 0.25 Å with 20.0 kcal/mol restraints. Each production window was further equilibrated for 100 ps, and then statistics were collected for 1.7 ns. The weighted histogram analysis method was used to integrate all PMFs.[43,44] The pK_a is calculated from the resulting PMF using the following equation:

$$\mathrm{p}K_a = \log\left[C^0\left(\int_0^\dagger 4\pi r^2 \exp[-\beta w(r)] \, \mathrm{d}r\right)^{-1}\right] \tag{9}$$

where $w(r)$ is the free energy from the PMF and β is the product of the simulation temperature and the Boltzmann constant. The integral is calculated from zero to the transition state, as denoted by $\dagger$. C^0 is the standard state concentration whose value is 1660 Å³/molecule and results from the entropic freedom that is gained by the proton when it dissociates from the acid.[45,46]

Atomistic configurations for the QM/MM calculations were selected from the umbrella sampling trajectories. In order to ensure a uniform distribution along the RC, individual configurations were sorted by their RC value into 60 0.067 Å wide bins. Five configurations were then selected at random from each bin, giving a total of 300 configurations for each iteration. QM/MM was performed in CP2K[47] using B3LYP[48,49] with the double-zeta basis set. Unrestricted DFT was used to allow the most accurate calculation of forces near the dissociation barrier. B3LYP was chosen over MP2 because recent studies have shown comparable accuracy with a reduced computational cost for water and the solvated excess proton.[39,50,51] Calculations were run on 256 Intel Sandy Bridge cores and on average completed within 10 min. The QM region included the entire amino acid residue and all of the waters found by the MS-RMD state search. Because proximity to classical point charges reduces the accuracy of forces calculated on QM atoms, an additional 3 Å buffer layer of QM water was added around the QM atoms. All together, this results in a QM region containing an average of 208 atoms, or one Asp/Glu and 61 waters. In total, four iterations were needed for Asp parameters to converge, while Glu parameters converged in three iterations.

RESULTS AND DISCUSSION

MS-RMD Model Parameters. Parameters for the new Asp and Glu models are presented in Table 1. These parameters

Table 1. Final Parameters for the Asp and Glu Models

	Asp	Glu		Asp	Glu
B	0.579 671	1.013 770	V_{ii}	−150.505 417	−150.291 915
b	1.506 881	1.418 695	c_1	−25.099 920	−25.013 330
b'	0.000 108	1.084 593	c_2	2.799 262	3.018 957
d_{OO}^0	2.4	2.4	c_3	1.299 994	1.280 649
C	0.579 841	0.987 714	D	143.003	143.003
c	0.931 593	1.146 188	α	1.8	1.8
d_{OH}^0	1.0	1.0	r_0	0.975	0.975
r_s	3.5	3.5			
r_s	4.0	4.0			

were fit to QM/MM reference forces using the previously described iterative scheme to ensure convergence. It must be noted that the Asp and Glu models were parametrized using the same method, but entirely independent data sets. The fact that the final parameters for each model are so similar is not surprising, however, given the molecular similarity and small difference in experimental pK_a values. As the next section discusses, the pK_a predicted by each model is a very good estimate of the experimental value. The fact that the models are able to capture such subtle differences is evidence that the methodology works well.

Deprotonation PMF and pK_a. The potential of mean force (PMF) of deprotonation describes the free energy of the process as a function of the RC. To obtain a free energy change between any two points on the PMF, one simply integrates the PMF. Figure 4a shows the PMF for deprotonation of Asp, and Figure 4b shows the PMF for Glu (black lines). As depicted in Figure 5, the RC gives the distance between the center-of-mass of the carboxylic moiety and the CEC. RC values around 1.3 Å correspond to the protonated amino acid, while values between 2.1 and 5 Å correspond to the stable CIP. Values above 5.5 Å represent an excess proton separated from the anionic acid by bulk solvent.

The protonated forms of Asp and Glu are much more stable than the deprotonated forms, which is consistent with their status as weak acids. Evaluating the integral in eq 9 estimates the pK_a values for Asp and Glu to be 3.82 ± 0.17 and 4.56 ± 0.18, respectively. These are both in very good agreement with the experimental ranges of 3.71 to 3.90 for Asp and 4.15 to 4.31 for Glu.[52,53]

The CIP is a significant feature of weakly acidic systems. It is stabilized by the electrostatic attraction between oppositely charged ions. The barrier that prevents protonation is due to the cost of forming a Zundel-like arrangement between the acid and hydronium. The depth and extent of the CIP varies with the system; however its existence has been predicted in a variety of weak acid systems simulated with different methods.[12,31,54,55] When preparing a simulation to calculate a PMF of deprotonation and a pK_a, care must be taken to ensure that the RC is sampled far enough beyond the CIP to ensure
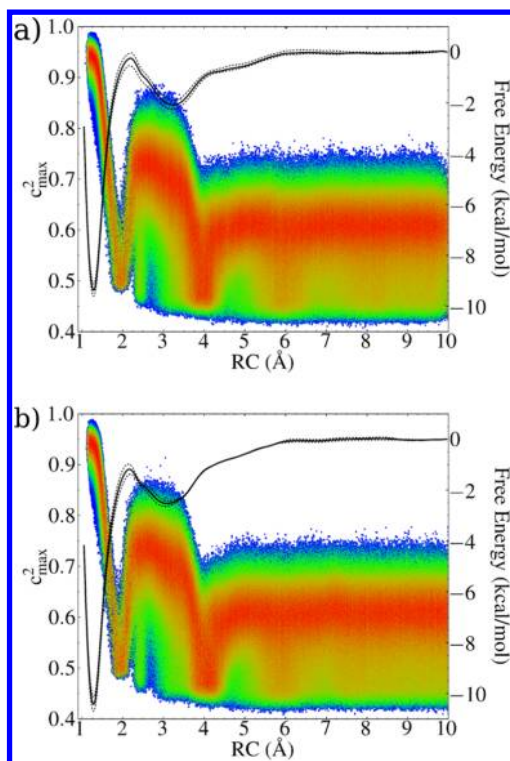
Figure 4. PMFs and c_{max}^2 distributions from the final (a) Asp and (b) Glu models. The PMF describes the free energy as a function of the RC. The pK_a is calculated from the PMF. The colormap in the background gives the c_{max}^2 probability distribution normalized for each RC value. This distribution describes the extent of charge delocalization and, by extension, the solvation structure.

bulk-like properties. If this is not done, then the PMF from which the pK_a is calculated will be shifted and the resulting "pK_a" will neither be accurate nor reflect the quality of the model. Fortunately, sampling well beyond the CIP is not a problem with MS-RMD models due to their efficiency. However, care needs to be taken when running QM/MM simulations to ensure that a large enough QM box and enough solvating waters are used to not only sample beyond the CIP but also to avoid boundary issues.[13]

Solvation Structure. With the PMF in hand, one can additionally probe the molecular mechanism governing weak acid dissociation. The MS-RMD methodology conveniently provides a measure of the structure of the solvated complex. It has been shown that the square of the largest component of the ground state eigenvector (c_{max}^2, from c in eq 3) describes the solvated structure.[28] A value of 1 indicates the limiting case of an entirely localized excess charge, while values less than that quantify the degree of charge delocalization. Eigen structures, where the excess charge is shared among a central hydronium and three H-bound water molecules, have a c_{max}^2 value around 0.65. Zundel structures, which share the excess charge more equally between two water molecules, have a c_{max}^2 value closer to 0.5. It should be emphasized that there is always additional delocalization to other surrounding water molecules, even for these limiting Eigen- and Zundel-like cations. By plotting the c_{max}^2 probability density as a function of the RC, we gain insight into the structural changes that accompany deprotonation. This probability density is plotted as a colormap in Figure 4. At low RC values, both Asp and Glu have an average c_{max}^2 greater than 0.9, which indicates that the structure and dynamics mostly resemble the limiting case of a protonated acid in bulk water with little charge transfer to the surrounding water molecules. This is notably different than the protonated water molecule, which always involves a significant amount of charge delocalization.[56] As the proton begins to dissociate, the c_{max}^2 decreases as the proton goes through a clear Zundel-like transition between the acid and solvating water. Beyond the transition state, the system begins to resemble an Eigen cation where one of the solvating waters is replaced with the acid anion. However, the Eigen structure seen in the CIP is distinctly different from that seen in bulk water, as evidenced by a c_{max}^2 of ~0.73. The close proximity of the anionic acid helps to stabilize the hydronium, leading to a more localized charge. Further evidence of the uniqueness of the CIP is that there is a well-defined Zundel transition around 4 Å linking the CIP to the bulk-like structure at longer distances. While proton transfers always occur via Eigen−Zundel−Eigen transitions in bulk, the water around the anionic amino acid is more structured, as evident in the distinct Zundel-like transition at a RC value around 4 Å. Figure 5 shows snapshots of the typical structures found at different RC values and characteristic of
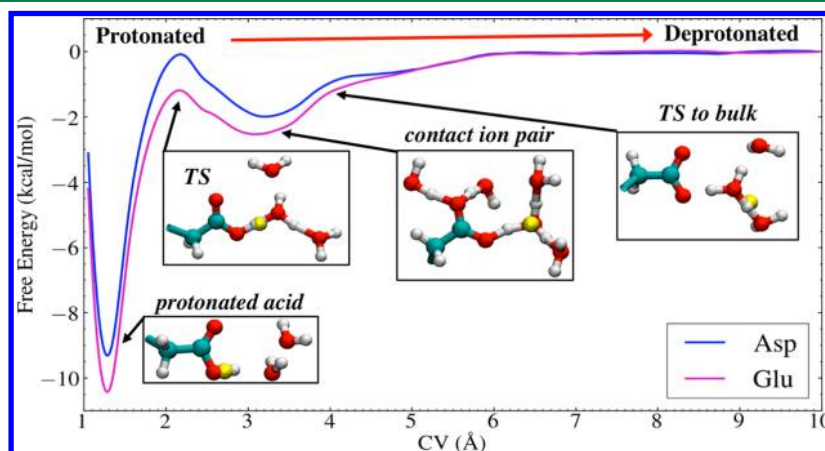

Figure 5. Snapshots from umbrella sampling trajectories illustrating the types of structures suggested by the c_{max}^2 distributions and the 2D-RDFs. Arrows indicate the region of the PMF where the windows were run, while the distances in the box are the center of the harmonic restraint. Atoms are colored according to their type. The position of the CEC is shown as a yellow sphere. In Eigen structures, the CEC is nearly aligned with the central water's oxygen. In Zundel structures, the CEC resides close to the midpoint between neighboring oxygens.

significant regions in the PMF. A yellow sphere indicates the location of the CEC. Its position relative to the atoms reinforces the analysis presented herein. The bonding topology reflects O−H distances less than or equal to 1.6 Å, which is commonly considered to be the length of a hydrogen bond.

It is also interesting to look at the structural changes that occur directly around the CEC. To do so, the two-dimensional radial distribution functions (2D-RDFs) between the CEC and surrounding water oxygens (O_W) and water hydrogens (H_W) are shown in Figure 6. These results corroborate the c_{max}^2
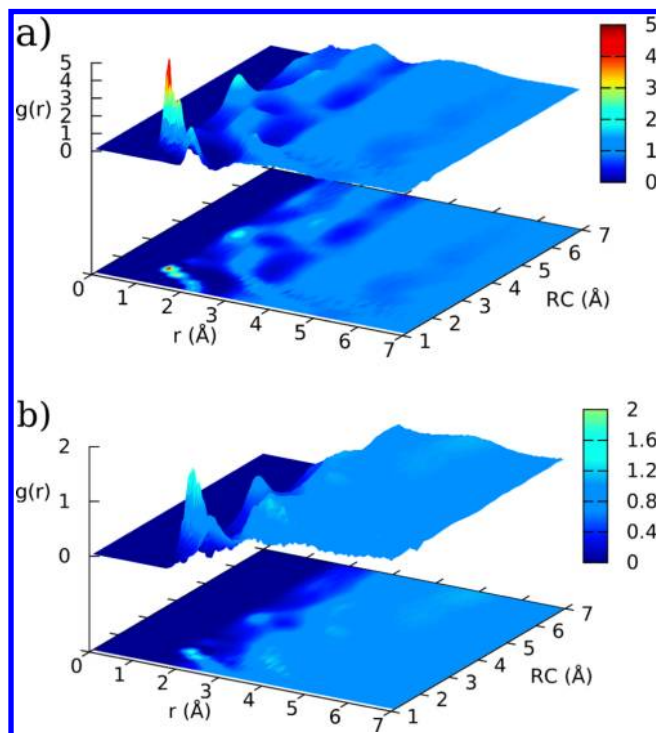

Figure 6. 2D-RDFs for (a) CEC−O_W and (b) CEC−H_W with Asp. These plots show how the water structure changes around the CEC as deprotonation occurs.

analysis. A slice along the first dimension of the 2D-RDFs gives a traditional 1D-RDF, which describes the probability of finding the indicated pair a given distance apart. Scanning the second dimension shows how the solvation structure changes as the reaction progresses according to the RC. In combination with the c_{max}^2 analysis, the 2D-RDFs show that the solvent has well-defined rearrangements as the proton dissociates from Asp and travels to bulk. Figure 6a shows the 2D-RDF for Asp's CEC−O_W distance. Once Asp is deprotonated, the first peak is near 2.5 Å, which is characteristic of the Eigen structure. When the RC reaches 4 Å, the peak distance drops to around 1.5 Å, which is characteristic of a slightly asymmetric Zundel cation where the CEC resides close to the midpoint of the O−O vector. The 2D-RDF showing Asp's CEC−H_W distribution shows similar behavior and is displayed in Figure 6b. Figures 4b and 7 show that Glu behaves similarly to Asp.

■ CONCLUSION

We have presented a procedure for fitting the parameters of a MS-RMD model to reproduce the forces from QM/MM calculations. Due to the efficiency of MS-RMD, we are able to collect and analyze data from large time and length scales that can reveal intricate behavior and is essential for converged bulk
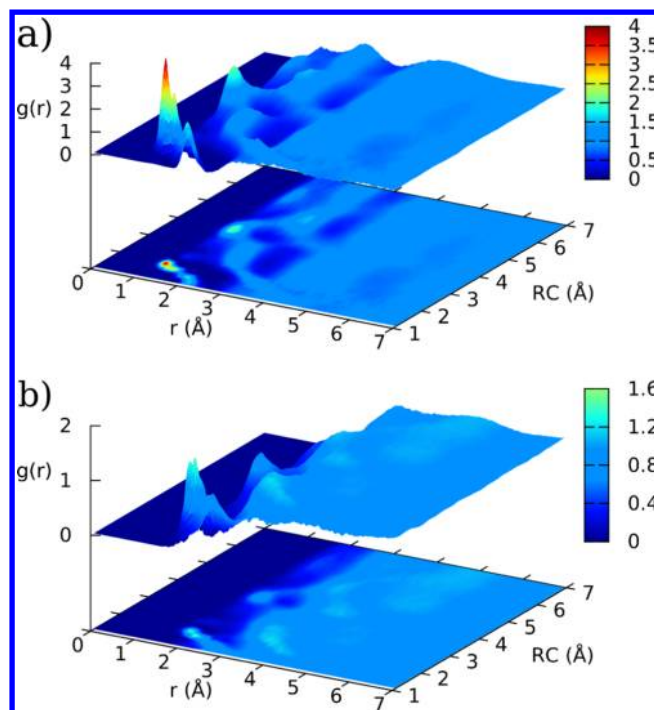

Figure 7. 2D-RDFs for (a) CEC−O_W and (b) CEC−H_W with Glu.

phase thermodynamic properties. This new methodology was used to parametrize new models of both Asp and Glu that correctly predict the correct experimental pK_a values. This is the first time that a MS-RMD model has been able to reproduce the experimental pK_a values without any fitting to experimental data. Multi-nanosecond trajectories were analyzed to extract important information about the mechanism underlying the deprotonation process. It is shown that the deprotonation process follows a well-defined series of Zundel to Eigen transitions as the excess proton travels from Asp to bulk solution.

This work is considered continued progress toward the goal of having flexible and systematic algorithms to reliably make a multiscale connection between accurate electronic structure calculations and efficient empirical models to describe complex reactive processes. As more accurate ab initio methodologies (e.g., those explicitly accounting for electronic correlation) become computationally tractable for QM/MM of condensed phase systems, the procedure presented herein and related force matching algorithms will be invaluable tools for mapping high level data into efficient RMD models that can significantly extend the time and length scales of reactive simulations to explore complex phenomena.

While the results presented herein are important, the true test of this procedure will be to apply it in biological systems where the pK_a of individual residues can be shifted by several pK_a units. Proton transport in proteins is very sensitive to the underlying protonatable residue models, and it is therefore crucial to have a procedure to ensure that those models are physically accurate. Furthermore, biological systems are orders of magnitude larger and have much slower dynamics than the systems presented here. While it is conceivable that QM/MM MD simulations could reach the same sort of time scales as the solution simulations presented here, they would be prohibitively expensive to run the length-and time-scales needed to account for the slow dynamics commonly found in biological

processes. The MS-RMD methodology has already been shown to scale well to hundreds of thousands of atoms and can realistically reach tens of nanosecond simulations with large explicit biomembrane systems. Work is already underway to implement these reactive models in protein environments and to use the presented methodology to fine-tune these and other individual amino acid models to behave properly in their specific protein environments.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Parameters for the revised MS-EVB3 force field are provided. This information is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: jmswanson@uchicago.edu.

Author Contributions
The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Alexov, E.; Mehler, E. L.; Baker, N.; Baptista, A. M.; Huang, Y.; Milletti, F.; Nielsen, J. E.; Farrell, D.; Carstensen, T.; Olsson, M. H.; Shen, J. K.; Warwicker, J.; Williams, S.; Word, J. M. Progress in the prediction of pKa values in proteins. *Proteins* **2011**, *79* (12), 3260–3275.

(2) Ho, J.; Coote, M. L. A universal approach for continuum solvent pK a calculations: are we there yet? *Theor. Chem. Acc.* **2010**, *125* (1–2), 3–21.

(3) Maurer, P.; Thomas, V.; Rivard, U.; Iftimie, R. A computational study of ultrafast acid dissociation and acid-base neutralization reactions. I. The model. *J. Chem. Phys.* **2010**, *133* (4), 044108.

(4) Zhang, S.; Baker, J.; Pulay, P. A Reliable and Efficient First Principles-Based Method for Predicting pKa Values. 2. Organic Acids. *J. Phys. Chem. B* **2010**, *114*, 432–442.

(5) Matsui, T.; Baba, T.; Kamiya, K.; Shigeta, Y. An accurate density functional theory based estimation of pK(a) values of polar residues combined with experimental data: from amino acids to minimal proteins. *Phys. Chem. Chem. Phys.* **2012**, *14* (12), 4181–4187.

(6) Mangold, M.; Rolland, L.; Costanzo, F.; Sprik, M.; Sulpizi, M.; Blumberger, J. Absolute pKaValues and Solvation Structure of Amino Acids from Density Functional Based Molecular Dynamics Simulation. *J. Chem. Theory Comput.* **2011**, *7* (6), 1951–1961.

(7) Kelley, C. P.; Cramer, C. J.; Truhlar, D. G. Adding Explicit Solvent Molecules to Continuum Solvent Calculations for the Calculation of Aqueous Acid Dissociation Constants. *J. Phys. Chem. A* **2006**, *110*, 2493–2499.

(8) Zhang, S. A reliable and efficient first principles-based method for predicting pK(a) values. III. Adding explicit water molecules: can the theoretical slope be reproduced and pK(a) values predicted more accurately? *J. Comput. Chem.* **2012**, *33* (5), 517–526.

(9) Ivanov, I.; Chen, B.; Raugei, S.; Klein, M. L. Relative pKa Values from First-Principles Molecular Dynamics: The Case of Histidine Deprotonation. *J. Phys. Chem. B* **2006**, *110*, 6365–6371.

(10) Ivanov, I.; Klein, M. L. Deprotonation of a Histidine Residue in Aqueous Solution Using Constrained Ab Initio Molecular Dynamics. *J. Am. Chem. Soc.* **2002**, *124*, 13380–13381.

(11) Swanson, J. M.; Maupin, C. M.; Chen, H.; Petersen, M. K.; Xu, J.; Wu, Y.; Voth, G. A. Proton Solvation and Transport in Aqueous and Biomolecular Systems: Insights from Computer Simulations. *J. Phys. Chem. B* **2007**, *111*, 4300–4314.

(12) Park, J. M.; Laio, A.; Iannuzzi, M.; Parrinello, M. Dissociation Mechanism of Acetic Acid in Water. *J. Am. Chem. Soc.* **2006**, *128*, 11318–11319.

(13) Uddin, N.; Choi, T. H.; Choi, C. H. Direct absolute pKa predictions and proton transfer mechanisms of small molecules in aqueous solution by QM/MM-MD. *J. Phys. Chem. B* **2013**, *117* (20), 6269–6275.

(14) Warshel, A.; Weiss, R. M. An Empirical Valence Bond Approach for Comparing Reactions in Solution and in Enzymes. *J. Am. Chem. Soc.* **1980**, *102*, 6218–6226.

(15) Schmitt, U. W.; Voth, G. A. Multistate Empirical Valence Bond Model for Proton Transport in Water. *J. Phys. Chem. B* **1998**, *102*, 5547–5551.

(16) Bellucci, M. A.; Coker, D. F. Empirical valence bond models for reactive potential energy surfaces: a parallel multilevel genetic program approach. *J. Chem. Phys.* **2011**, *135* (4), 044115.

(17) Park, K.; Lin, W.; Paesani, F. A refined MS-EVB model for proton transport in aqueous environments. *J. Phys. Chem. B* **2012**, *116* (1), 343–352.

(18) Lee, H. J.; Svahn, E.; Swanson, J. M.; Lepp, H.; Voth, G. A.; Brzezinski, P.; Gennis, R. B. Intricate role of water in proton transport through cytochrome c oxidase. *J. Am. Chem. Soc.* **2010**, *132* (45), 16225–16239.

(19) Maupin, C. M.; McKenna, R.; Silverman, D. N.; Voth, G. A. Elucidation of the Proton Transport Mechanism in Human Carbonic Anhydrase II. *J. Am. Chem. Soc.* **2009**, *131*, 7598–7608.

(20) Yamashita, T.; Voth, G. A. Insights into the mechanism of proton transport in cytochrome c oxidase. *J. Am. Chem. Soc.* **2012**, *134* (2), 1147–1152.

(21) Li, H.; Chen, H.; Steinbronn, C.; Wu, B.; Beitz, E.; Zeuthen, T.; Voth, G. A. Enhancement of proton conductance by mutations of the selectivity filter of aquaporin-1. *J. Mol. Biol.* **2011**, *407* (4), 607–620.

(22) Feng, S.; Savage, J.; Voth, G. A. Effects of Polymer Morphology on Proton Solvation and Transport in Proton-Exchange Membranes. *J. Phys. Chem. C* **2012**, *116* (36), 19104–19116.

(23) Tse, Y.-L. S.; Herring, A. M.; Kim, K.; Voth, G. A. Molecular Dynamics Simulations of Proton Transport in 3M and Nafion Perfluorosulfonic Acid Membranes. *J. Phys. Chem. C* **2013**, *117* (16), 8079–8091.

(24) Liang, R.; Swanson, J. M. J.; Voth, G. A. Benchmark Study of the SCC-DFTB Approach for a Biomolecular Proton Channel. *J. Chem. Theory Comput.* **2014**, *10* (1), 451–462.

(25) Xu, J.; Yamashita, T.; Agmon, N.; Voth, G. A. On the origin of proton mobility suppression in aqueous solutions of amphiphiles. *J. Phys. Chem. B* **2013**, *117* (49), 15426–15435.

(26) Knight, C.; Lindberg, G. E.; Voth, G. A. Multiscale reactive molecular dynamics. *J. Chem. Phys.* **2012**, *137* (22), 22A525.

(27) Knight, C.; Voth, G. A. Coarse-graining away electronic structure: a rigorous route to accurate condensed phase interaction potentials. *Mol. Phys.* **2012**, *110* (9–10), 935–944.

(28) Wu, Y.; Chen, H.; Wang, F.; Paesani, F.; Voth, G. A. An Improved Multistate Empirical Valence Bond Model for Aqueous Proton Solvation and Transport. *J. Phys. Chem. B* **2008**, *112*, 467–482.

(29) Cuma, M.; Schmitt, U. W.; Voth, G. A. A multi-state empirical valence bond model for acid-base chemistry in aqueous solution. *Chem. Phys.* **2000**, *258*, 187–199.

(30) 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.

(31) 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.

(32) Wang, F.; Akin-Ojo, O.; Pinnick, E.; Song, Y. Approaching post-Hartree−Fock quality potential energy surfaces with simple pair-wise expressions: parameterising point-charge-based force fields for liquid water using the adaptive force matching method. *Mol. Simul.* **2011**, *37* (7), 591−605.

(33) Wang, L.-P.; Chen, J.; Van Voorhis, T. Systematic Parametrization of Polarizable Force Fields from Quantum Chemistry Data. *J. Chem. Theory Comput.* **2013**, *9* (1), 452−460.

(34) Schmitt, U. W.; Voth, G. A. The computer simulation of proton transport in water. *J. Chem. Phys.* **1999**, *111* (20), 9361.

(35) Voth, G. A. Computer Simulation of Proton Solvation and Transport in Aqueous and Biomolecular Systems. *Acc. Chem. Res.* **2006**, *39*, 143−150.

(36) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.

(37) Akin-Ojo, O.; Wang, F. Improving the Point-Charge Description of Hydrogen Bonds by Adaptive Force Matching. *J. Phys. Chem. B* **2009**, *113* (5), 1237−1240.

(38) Knight, C.; Maupin, C. M.; Izvekov, S.; Voth, G. A. Defining Condensed Phase Reactive Force Fields from ab Initio Molecular Dynamics Simulations: The Case of the Hydrated Excess Proton. *J. Chem. Theory Comput.* **2010**, *6*, 3223−3232.

(39) Akin-Ojo, O.; Song, Y.; Wang, F. Developing ab initio quality force fields from condensed phase quantum-mechanics/molecular-mechanics calculations through the adaptive force matching method. *J. Chem. Phys.* **2008**, *129* (6), 064108.

(40) Muhlenbein, H.; Schlierkamp-Voosen, D. Predictive Models for the Breeder Genetic Algorithm I. Continuous Parameter Optimization. *Evol. Comput.* **1993**, *1* (1), 25−49.

(41) Plimpton, S. Fast Parallel Algorithms for Short−Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117*, 1−19.

(42) Yamashita, T.; Peng, Y.; Knight, C.; Voth, G. A. Computationally Efficient Multiconfigurational Reactive Molecular Dynamics. *J. Chem. Theory Comput.* **2012**, *8* (12), 4863−4875.

(43) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. The Weighted Histogram Analysis Method for Free Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13* (8), 1011−1021.

(44) Grossfield, A. *WHAM: an implementation of the weighted histogram analysis method*, version 2.0.7. http://membrane.urmc.rochester.edu/content/wham/ (accessed 8/28/2013).

(45) Chen, J.; Brooks, C. L.; Scheraga, H. A. Revisiting the Carboxylic Acid Dimers in Aqueous Solution: Interplay of Hydrogen Bonding, Hydrophobic Interactions, and Entropy. *J. Phys. Chem. B* **2008**, *112*, 242−249.

(46) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The Statistical-Thermodynamic Basis for Computation of Binding Affinities: A Critical Review. *Biophys. J.* **1997**, *72*, 1047−1069.

(47) 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.

(48) Goedecker, S.; Teter, M.; Hutter, J. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B* **1996**, *54* (3), 1703−1710.

(49) Krack, M. Pseudopotentials for H to Kr optimized for gradient-corrected exchange-correlation functionals. *Theor. Chem. Acc.* **2005**, *114* (1−3), 145−152.

(50) Sobolewski, A. L.; Domcke, W. Ab Initio Investigation of the Structure and Spectroscopy of Hydronium-Water Clusters. *J. Phys. Chem. A* **2002**, *106*, 4158−4167.

(51) Park, M.; Shin, I.; Singh, N. J.; Kim, K. S. Eigen and Zundel Forms of Small Protonated Water Clusters: Structures and Infrared Spectra. *J. Phys. Chem. A* **2007**, *111*, 10692−10702.

(52) Lide, D. R. Properties of Amino Acids. In *CRC Handbook of Chemistry and Physics, Internet Version*; Lide, D. R., Ed.; CRC Press: Boca Raton, FL, 2005; pp 7−1.

(53) Lide, D. R. Dissociation Constants of Organic Acids and Bases. In *CRC Handbook of Chemistry and Physics, Internet Version*; Lide, D. R., Ed.; CRC Press: Boca Raton, FL, 2005; pp 8−46.

(54) Thomas, V.; Maurer, P.; Iftimie, R. On the Formation of Proton-Shared and Contact Ion Pair Forms during the Dissociation of Moderately Strong Acids: An Ab Initio Molecular Dynamics Investigation. *J. Phys. Chem. B* **2010**, *114*, 8147−8155.

(55) Thomas, V.; Rivard, U.; Maurer, P.; Bruhács, A.; Siwick, B. J.; Iftimie, R. Concerted and Sequential Proton Transfer Mechanisms in Water-Separated Acid−Base Encounter Pairs. *J. Phys. Chem. Lett.* **2012**, *3* (18), 2633−2637.

(56) Swanson, J. M.; Simons, J. Role of charge transfer in the structure and dynamics of the hydrated proton. *J. Phys. Chem. B* **2009**, *113* (15), 5149−5161.