Development of reactive force fields using *ab initio* molecular dynamics simulation minimally biased to experimental data

Chen Chen,^{1,2,a)} Christopher Arntsen,^{1,a)} and Gregory A. Voth^{1,b)}

¹*Department of Chemistry, James Franck Institute, and Institute for Biophysical Dynamics, The University of Chicago, Chicago, Illinois 60637, USA*

²*College of Chemistry and Molecular Sciences, Hubei Key Laboratory of Electrochemical Power Sources, Wuhan University, Wuhan 430072, China*

(Received 1 June 2017; accepted 12 July 2017; published online 1 August 2017)

Incorporation of quantum mechanical electronic structure data is necessary to properly capture the physics of many chemical processes. Proton hopping in water, which involves rearrangement of chemical and hydrogen bonds, is one such example of an inherently quantum mechanical process. Standard *ab initio* molecular dynamics (AIMD) methods, however, do not yet accurately predict the structure of water and are therefore less than optimal for developing force fields. We have instead utilized a recently developed method which minimally biases AIMD simulations to match limited experimental data to develop novel multiscale reactive molecular dynamics (MS-RMD) force fields by using relative entropy minimization. In this paper, we present two new MS-RMD models using such a parameterization: one which employs water with harmonic internal vibrations and another which uses anharmonic water. We show that the newly developed MS-RMD models very closely reproduce the solvation structure of the hydrated excess proton in the target AIMD data. We also find that the use of anharmonic water increases proton hopping, thereby increasing the proton diffusion constant. *Published by AIP Publishing.* [<http://dx.doi.org/10.1063/1.4985903>]

I. INTRODUCTION

With the rapid increase in computer power, complemented by more sophisticated and efficient computational algorithms, simulations have become an effective way of understanding and even predicting the behavior of complex chemical systems. Quantum mechanics plays a fundamental role in chemical processes, and for a simulation to be physically meaningful, i.e., for the simulation to accurately describe a given chemical process with the correct underlying physics, it is often necessary to include electronic structure effects. More specifically, the electronic structure of a material can greatly affect its polarization, charge transfer, and chemical reactivity. However, the inclusion of these quantum effects (e.g., chemical reactivity) introduces a significant computational cost which greatly limits the range of applications. It is therefore desirable to use molecular mechanical models, which can increase the breadth and depth of the simulation studies, for applicable systems. Incorporating quantum mechanical effects such as reactivity into such models is a challenging task, especially given the fact that many approximate quantum mechanical methods do not necessarily provide an accurate description of a given system in the first place.

Consider, for example, simulations of liquid water. *Ab initio* molecular dynamics (AIMD) is the most common method for performing dynamics simulations that explicitly treat electronic degrees of freedom. However, when employing

a sufficiently large basis set, AIMD simulations can only handle upwards of several hundred water molecules (or several thousand valence electrons) at the generalized gradient approximation (GGA) level of density functional theory (DFT). Unfortunately, at this particular level of DFT, water is often found to be highly over-structured, with a melting temperature of around 400 K and a self-diffusion constant an order of magnitude smaller than experiment. Indeed, most DFT-based AIMD performs rather poorly in terms of accuracy relative to the experiment for systems of liquid water.

To mitigate these limitations, more sophisticated quantum mechanical techniques have been employed. One recent example is from Willow *et al.*¹ In that work, the authors use an embedded fragment second order Møller-Plesset scheme to perform AIMD. Such simulations promise to yield quite accurate predictions of the water self-diffusion constant, dipole moment, and coordination number relative to the experiment. However, this method is still very expensive and therefore cannot be applied to large systems; simulations were performed on systems of only 32 water molecules.

Another approach utilizes sophisticated quantum mechanical calculations to parameterize classical water models.^{2,3} In terms of force-matching, Wang and co-workers have developed a four-site water model using forces generated using DFT (including Hartree-Fock exchange) and MP2,^{4,5} as well as models fit to post Hartree-Fock methods.^{6,7} Resulting classical models have been shown to reproduce several experimental properties, including density and IR spectra, and improve the water self-diffusion constant relative to AIMD.

As an alternative, recent work in our group^{8–10} has advanced the use of a direct minimal bias of simulation

^{a)}C. Chen and C. Arntsen contributed equally to this work.

^{b)}Author to whom correspondence should be addressed: gavoth@uchicago.edu

models with experimental data, called “experiment directed simulation” (EDS). As one key example, rather than employing a more sophisticated electronic structure method to perform the AIMD simulation, a combined EDS-AIMD approach¹⁰ utilizes limited experimental data to parameterize a bias of the DFT functional in an AIMD simulation. For the example of bulk liquid water, the results of Ref. 10 show that applying a minimal bias to just the interaction between water oxygen atoms (i.e., the O–O radial distribution function or RDF) results in a substantially less structuring of EDS-AIMD water and also an increase in the water self-diffusion constant (a quantity not biased in the simulation).

While inclusion of electronic degrees of freedom in simulations of water as a baseline allows in principle for a more rigorous description, it is not clearly necessary for accuracy at all: there are several classical empirical water models that can reproduce a number of experimental properties rather well (see, e.g., Refs. 11–15, and in that vein, Ref. 15 represents the current and best state-of-the-art). In the case of hydrated excess proton diffusion, however, a purely classical model is incapable of accurately reproducing the physics of the process. Proton diffusion in water is comprised of both vehicular diffusion and Grotthuss hopping, in which covalent and hydrogen bonds rearrange, significantly increasing the overall proton self-diffusion constant.^{16–21} Because the Grotthuss process involves chemical reactivity, it is inherently quantum mechanical in nature. Traditional force fields, which have fixed bonding topologies, are therefore inadequate to describe this process. Multistate empirical valence bond (MS-EVB) and more recently multiscale reactive molecular dynamics (MS-RMD) methods have been shown to capture the physics of this process and closely reproduce the solvation structure of a hydrated excess proton in water.^{19,22–27} Unlike classical force fields, these reactive MD methods explicitly allow for bond breaking and formation and are thus capable of including the essential physics involved in proton transport in water and aqueous systems.

The most sensible development and parameterization of reactive MD force fields utilize explicit quantum mechanical data, given the quantum mechanical nature of reactivity, but this parameterization is highly nontrivial. MS-EVB models were initially fit to reproduce potential energy surfaces along the proton transfer coordinate in small clusters using MP2.^{19,22–24} Such parameterizations utilized gas-phase data to generate aspects of an MS-EVB model designed for bulk simulations. By contrast, condensed phase force-matching to AIMD data has previously been employed to parameterize MS-RMD models, wherein a variational force residue is minimized.^{25,26} Both these fitting methods were demonstrably successful but also have their challenges and limitations.

In a preliminary communication, we have more recently shown that relative entropy minimization (REM)^{28–31} is an effective and efficient method for generating new MS-RMD force fields from AIMD data.²⁷ In that work, we parameterized our model to bulk phase AIMD data for the excess proton in water. Despite the fact that AIMD water is known to be over-structured due to deficiencies in the underlying level of DFT,^{32,33} the excess proton solvation behavior is actually

similar to the experiment for a number of properties,³⁴ and we were able to generate MS-RMD models with REM which faithfully reproduced both the AIMD RDFs and potential of mean force (PMF) along the proton transfer reaction coordinate. Given the recent development of EDS-AIMD,¹⁰ in which the water over-structuring is significantly reduced and self-diffusion is increased, in this work, we instead develop MS-RMD models fit to EDS-AIMD data. As such, we have parameterized two new MS-RMD models to the EDS-AIMD data using REM. The primary difference between the two models is the underlying water model employed in the MS-RMD: one uses standard harmonic simple point charge/flexible water (SPC/Fw),¹⁴ while the other uses anharmonic aSPC/Fw.³⁵ It is important to note that both the EDS-AIMD and the MS-RMD models in this paper utilize REM in different—and important—ways.

The remainder of this paper is organized as follows: in Sec. II, we describe the MS-RMD force field, REM, and the basic EDS-AIMD scheme. In Sec. III, we discuss the models developed in this paper and give results for the hydrated proton solvation structure, the PMF along the proton transfer coordinate, the excess proton diffusion constant, and the proton hopping behavior. Concluding remarks are provided in Sec. IV.

II. METHODOLOGY

In MS-RMD, the ground state wavefunction is defined in terms of a linear combination of basis states i , each having a unique bonding topology

$$|\Psi\rangle = \sum_{i=1}^N c_i |i\rangle. \quad (1)$$

The coefficients c_i are determined by diagonalizing the RMD Hamiltonian “on the fly” for each configuration of the nuclei $\mathbf{x}$,

$$\mathbf{H}\mathbf{c} = E_0\mathbf{c}, \quad (2)$$

where $E_0(\mathbf{x})$ is understood here to be a diagonal matrix of eigenvalues as a function of all nuclear coordinates. The diagonal matrix elements of the Hamiltonian matrix $\mathbf{H}$ are defined as

$$h_{ii} = V_{\text{H}_3\text{O}^+}^{\text{intra}} + \sum_{k=1}^{N_{\text{H}_2\text{O}}} V_{\text{H}_2\text{O}}^{\text{intra},k} + \sum_{k=1}^{N_{\text{H}_2\text{O}}} V_{\text{H}_3\text{O}^+, \text{H}_2\text{O}}^{\text{inter},k} + \sum_{k < k'}^{N_{\text{H}_2\text{O}}} V_{\text{H}_2\text{O}}^{\text{inter},kk'}. \quad (3)$$

The interatomic interactions between the hydronium ion and surrounding waters are defined as

$$\begin{aligned} V_{\text{H}_3\text{O}^+, \text{H}_2\text{O}}^{\text{inter}} = & 4\epsilon_{\text{O}^*\text{O}_w} \left[\left(\frac{\sigma_{\text{O}^*\text{O}_w}}{R_{\text{O}^*\text{O}_w}} \right)^{12} - \left(\frac{\sigma_{\text{O}^*\text{O}_w}}{R_{\text{O}^*\text{O}_w}} \right)^6 \right] \\ & + 4\epsilon_{\text{HO}_w} \left[\left(\frac{\sigma_{\text{H}^*\text{O}_w}}{R_{\text{H}^*\text{O}_w}} \right)^{12} - \left(\frac{\sigma_{\text{H}^*\text{O}_w}}{R_{\text{H}^*\text{O}_w}} \right)^6 \right] \\ & + \sum_m^4 \sum_{n_k}^3 \frac{q_m^{\text{H}_3\text{O}^+} q_{n_k}^{\text{H}_2\text{O}}}{R_{mn_k}} + V_{\text{OO}_k}^{\text{rep}} + V_{\text{HO}_k}^{\text{rep}}. \end{aligned} \quad (4)$$

The oxygen and hydrogen atoms denoted with an asterisk are for hydronium oxygen and hydrogen nuclei. The subscript “w” denotes a water molecule. The last two terms in the above equation, $V_{\text{OO}_k}^{\text{rep}}$ and $V_{\text{HO}_k}^{\text{rep}}$, are added corrections to the diabatic energies of the underlying classical force field and have the effect of correcting for an over-attraction between classical hydronium and water. They are defined as

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

and

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

The off-diagonal matrix elements h_{ij} , which describe mixing between states $|i\rangle$ and $|j\rangle$, are defined by

$$h_{ij} = \left(V_{\text{const}}^{ij} + V_{\text{ex}}^{ij} \right) \cdot A \left(R_{\text{OO}}, \mathbf{q} \right). \quad (7)$$

The lattermost term on the right hand side of the equation, $A(R_{\text{OO}}, \mathbf{q})$, is a geometry dependent scaling factor

$$A \left(R_{\text{OO}}, \mathbf{q} \right) = \exp \left(-\gamma q^2 \right) \cdot \left\{ 1 + P \exp \left[-k \left(R_{\text{OO}} - D_{\text{OO}} \right)^2 \right] \right\} \cdot \left\{ \frac{1}{2} \left\{ 1 - \tanh \left[\beta \left(R_{\text{OO}} - R_{\text{OO}}^0 \right) \right] \right\} + P' \exp \left[-\alpha \left(R_{\text{OO}} - r_{\text{OO}}^0 \right) \right] \right\} \quad (8)$$

and is a function of the interatomic distance between the hydronium and water oxygen and $\mathbf{q}$, which is defined by

$$\mathbf{q} = \frac{\mathbf{r}_{\text{O}} + \mathbf{r}_{\text{O}'}}{2} - \mathbf{r}_{\text{H}^*}. \quad (9)$$

In Eq. (7), V_{const}^{ij} is a constant, and V_{ex}^{ij} describes the electrostatic interaction between the Zundel complex, H_5O_2^+ , and surrounding waters in the complex

$$V_{\text{ex}}^{ij} = \sum_m^7 \sum_k^{N_{\text{H}_2\text{O}}-1} \sum_{n_k}^3 \frac{q_{n_k}^{\text{H}_2\text{O}} q_m^{\text{ex}}}{R_{mn_k}}. \quad (10)$$

For the two MS-RMD models fit in this work, the parameters are given in the [supplementary material](#). We note that the underlying classical force field parameters were not fit and were taken from MS-EVB 3.2.²⁴

In order to parameterize the models presented in this paper, we employed the REM approach. The relative entropy is defined by

$$S_{\text{rel}} = \sum_i p_{\text{T}}(i) \ln \frac{p_{\text{T}}(i)}{p_{\text{M}}(i)}, \quad (11)$$

where $p_{\text{T}}(i)$ and $p_{\text{M}}(i)$ are the probabilities of the target and model ensembles being in state i , respectively. In the constant NVT ensemble, this term can be rearranged to give

$$S_{\text{rel}} = \beta \langle U_{\text{M}} - U_{\text{T}} \rangle_{\text{T}} - \beta (A_{\text{M}} - A_{\text{T}}) + \langle S_{\text{map}} \rangle_{\text{T}}, \quad (12)$$

where U is the potential energy, A is the configurational part of the Helmholtz free energy, and S_{map} is the mapping entropy which represents entropy arising from degeneracies in the model. The notation $\langle \dots \rangle$ indicates averages over the ensembles, labeled by the subscript. The relative entropy is minimized upon the condition that the average derivative of the model potential energy with respect to model parameters λ is equal over the model and target ensembles,

$$\left\langle \frac{\partial U_{\text{M}}}{\partial \lambda_i} \right\rangle_{\text{M}} = \left\langle \frac{\partial U_{\text{M}}}{\partial \lambda} \right\rangle_{\text{T}}. \quad (13)$$

The simplest implementation of such a minimization is the Newton-Raphson iterative scheme, where λ_i is updated according to

$$\begin{aligned} \lambda_{i+1} &= \lambda_i - \frac{(\partial S_{\text{rel}} / \partial \lambda)}{(\partial^2 S_{\text{rel}} / \partial \lambda^2)} \\ &= \lambda_i - \chi \left[\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_{\text{T}} - \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_{\text{M}} \right] \\ &\quad \times \left[\left\langle \frac{\partial^2 U}{\partial \lambda^2} \right\rangle_{\text{T}} - \left\langle \frac{\partial^2 U}{\partial \lambda^2} \right\rangle_{\text{M}} + \beta \left\langle \left(\frac{\partial U}{\partial \lambda} \right)^2 \right\rangle_{\text{M}} - \beta \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_{\text{M}}^2 \right]^{-1}. \end{aligned} \quad (14)$$

In this equation, χ is just a mixing term.

As we noted in Ref. 27, the Newton-Raphson scheme is a local minimization and can therefore result in degenerate solutions, with no clear guarantee that one reaches the global minimum of S_{rel} . In fact, because of this shortcoming, the converged set of parameters may be dependent on the initial parameter set. The primary reason in the present case for there being so many solutions is the large number of parameters, the majority of which are nonlinear, that are being simultaneously fit. Interestingly, the extraordinary flexibility of such a fitting scheme actually appears to be advantageous in that some parameters can effectively compensate for others. On the other hand, a large number of flexible parameters can potentially lead to overfitting. We have found that all models fit using REM have successfully reproduced the proton solvation structure. The location of the proton transfer PMF minimum is also always reproduced, though the barrier height can differ slightly. We also find some variation in the resulting proton self-diffusion constant, though the values are usually within 30% of those reported here. One can certainly imagine a more sophisticated implementation of REM which does not use the Newton-Raphson scheme and is therefore not subject to getting trapped in local minima, but within the context of the work presented here, we find such an implementation to be suitable. In addition, the simplicity of a Newton-Raphson implementation certainly adds to the appeal of REM.

REM is indifferent to target ensemble, so as higher level quantum chemistry methods become computationally feasible for dynamics simulations, REM can be used to further refine models. As an alternative to a higher level (but much more costly) of quantum mechanics in the AIMD, which in theory would more accurately describe the system relative to experiment, we have chosen here to parameterize our model using quantum data generated using the recently developed EDS-AIMD method.¹⁰ In EDS-AIMD, the interactions between water oxygen atoms are biased to reproduce the experimental $\text{O}_{\text{w}}-\text{O}_{\text{w}}$ RDF in AIMD simulations of water. In order to do so, the bias specifically targets the oxygen coordination number and its moments. The bias $V(r_i)$ is defined as

$$V(r_i) = \sum_{k=0}^3 \frac{\alpha_k}{\hat{f}_k} \sum_{j \neq i}^N r_{ij}^k \left[1 - u(r_0 - r_{ij}) \right], \quad (15)$$

where N is the number of oxygen neighbors of given water oxygen, r^k is the k th power of distance to said coordinating atoms, r_{ij} is the pairwise distance of oxygen neighbors, u is a step function, and r_0 is the *experimental* distance of the first solvation peak of the O–O radial distribution function. The linear bias α_k is determined using the EDS algorithm to match experimental data, and $\hat{f}_k$ is the desired ensemble average of the k th scalar calculated from the experimental RDF. We direct the reader to Ref. 10 for more details on the EDS method used to generate the target ensemble used here.

The target EDS-AIMD configurations in this model utilize the BLYP functional,^{36,37} with the D3 Grimme dispersion correction^{38,39} and TZVP basis set. The systems used to parameterize the models consist of 128 water molecules plus an excess proton. Production runs of the MS-RMD models use 256 water molecules plus an excess proton.

III. RESULTS AND DISCUSSION

We begin with a brief discussion of the structural properties of an excess proton in water simulated using EDS-AIMD, particularly in comparison to unbiased AIMD (BLYP+D3 level). The relevant RDFs for EDS-AIMD and AIMD are shown in Fig. 1. Interestingly, despite the significant change in the water O_w–O_w RDFs between EDS-AIMD and AIMD (see Ref. 10), the hydrated proton solvation structure is only marginally altered. First, consider the O*–O_w RDF, which perhaps is the most telling in terms of the proton solvation structure (where the asterisk denotes a hydronium nucleus). The first solvation shell peak for both EDS-AIMD and AIMD is located at 2.53 Å; the first solvation shell peak heights are 4.22 for AIMD and EDS-AIMD. This is consistent with previous results for the hydrated excess proton comparing EDS-AIMD and AIMD,¹⁰ which did not include the D3 dispersion correction. Interestingly, the height of the second solvation shell

peak in EDS-AIMD is slightly lower in AIMD. This solvation peak is largely dominated by water–water interactions, which means the EDS bias is more able to reduce the over-structuring representative of AIMD water. The H*–O_w RDFs for both systems have very similar peak heights and locations for both the first and second solvation shells. The O*–H_w RDFs showed slightly more variability. While the first solvation shell peak location and height are nearly identical in both systems, the second solvation shell in EDS-AIMD is less pronounced and at a slightly longer distance (5.10 Å compared to 4.94 Å), again a result of reduced water–water over-structuring. Importantly, the so-called pre-solvation shell,²⁴ located at around 2 Å, is less prominent in EDS-AIMD than in AIMD with just BLYP+D3 DFT.

We also compared the PMF along the proton transfer coordinate between the AIMD, EDS-AIMD, and the two RMD models developed in this paper. We define the proton transfer coordinate δ ,

$$\delta = |r_{\text{O}^*\text{H}^*} - r_{\text{O}_{1\text{x}}\text{H}^*}|, \quad (16)$$

where $r_{\text{O}^*\text{H}^*}$ is the distance between hydronium oxygen and hydronium hydrogen, and $r_{\text{O}_{1\text{x}}\text{H}^*}$ is the distance between hydronium hydrogen and nearest water oxygen. As was seen in previous work which did not include the D3 dispersion correction to BLYP,¹⁰ the AIMD and EDS-AIMD PMFs are virtually identical, both having a minimum at 0.33 Å and a barrier of 0.60 kcal/mol.

We next turn to the hydrated proton solvation structure, comparing the MS-RMD fit models with the target EDS-AIMD results as shown in Fig. 1. First, we compare the O*–O_w RDF, as it is most illustrative of solvation structure around hydrated protons (“hydronium” cation). The MS-RMD models with water solvent having harmonic and anharmonic internal vibrations both reproduce the location of the first solvation peak, at 2.53 Å, with peak heights of 4.23 and 4.40 for MS-RMD models compared to 4.22 for EDS-AIMD. The

FIG. 1. Radial distribution functions for AIMD, EDS-AIMD, MS-RMD 4, and MS-RMD 5. The oxygen and hydrogen atoms denoted with an asterisk are for hydronium oxygen and hydrogen nuclei. The subscript “w” denotes a water molecule.

MS-RMD models do not entirely reproduce the depth of the first valley, though the density in the regions between the first and second solvation shells is similar for EDS-AIMD and RMD. Importantly, there is greater overlap in the second solvation between EDS-AIMD and MS-RMD models. This is actually less a function of a well-fit model than it is an improved AIMD target via EDS-AIMD. As was mentioned earlier, typical GGA DFT-based AIMD water is over-structured, and this second solvation peak is mostly governed by water-water interactions. Because of this, our preliminary efforts²⁷ to parameterize new MS-RMD models to AIMD were unable to completely reproduce the O^*-O_w second solvation shell peak. Still, as seen in Fig. 1, there is not perfect overlap of the second solvation shell peaks between the EDS-AIMD and the MS-RMD models.

Comparing the O^*-H_w RDFs, the MS-RMD models cannot exactly reproduce the first solvation shell peak of the EDS-AIMD data, but the agreement is still very good. In EDS-AIMD, the first intermolecular solvation peak is centered at 3.05 Å, with a height of 2.13; in MS-RMD, the peak is centered at 3.15 Å with a height of 2.05. When fitting to EDS-AIMD, the second solvation peak of the new MS-RMD models has a greater overlap with the target: EDS-AIMD has a peak centered at 5.00 Å with a height of 1.09, while MS-RMD has a peak centered at 5.12 Å with a height of 1.11. Again, this is due to the improved water-water interactions of the target.

The H^*-O_w RDF in Fig. 1 also shows a good overlap between the EDS-AIMD and MS-RMD results. The first solvation shell for EDS-AIMD is centered at 1.50 Å with a height of 2.74; the RMD models are centered at 1.47 Å with a height of 2.88 and 3.30 and for MS-RMD 4 and MS-RMD 5 models, respectively. The second solvation shell for EDS-AIMD is centered at 3.08 Å with a height of 1.75; the MS-RMD models are centered at 3.05 Å with a height of 1.79.

There is also a good overlap of the H^*-H_w EDS-AIMD and MS-RMD RDFs; however, there is no real improvement in the overlap between new MS-RMD models with EDS-AIMD compared with previous models.²⁷ This is a result of the weak contact between hydronium hydrogens and water hydrogens, which means such interactions contribute less to the total potential energy and therefore contribute less to the relative entropy.

Next, we compare the PMF along the proton transfer coordinate between EDS-AIMD and MS-RMD 4 and 5 as seen in Fig. 2. The REM generates MS-RMD models that closely reproduce the PMF along the proton transfer coordinate of the target, though the MS-RMD models diverge most significantly in regions where the free energy is higher. In the present work, both MS-RMD 4 and MS-RMD 5 actually quite closely reproduce the high free energy regions where δ is large but do not perfectly reproduce the proton transfer barrier. MS-RMD 4 has a barrier of 0.78 kcal/mol, which is larger but still close to that of EDS-AIMD, 0.60 kcal/mol; the free energy minimum for MS-RMD 4 is located at 0.32 Å, whereas that of EDS-AIMD is 0.34 Å. The free energy barrier of MS-RMD 5, 0.67 kcal/mol, agrees better with that of EDS-AIMD, and the free energy minimum 0.32 Å is the same as that of MS-RMD 4. We have shown in previous work²⁷ that REM

FIG. 2. Potential of mean force along the proton transfer coordinate defined in Eq. (16) for the MS-RMD models discussed in this paper, as well as for AIMD and EDS-AIMD.

is far better at reproducing the region of PMF with lower free energy. This is due to the fact that for a given ensemble, more configurations are from lower free energies, which means that lower free energy regions have a greater contribution to the average potential energy derivatives in Eqs. (12)–(14).

One of the objectives of the present work is to also improve the proton self-diffusion constants of MS-RMD. Previous MS-EVB and MS-RMD models have underestimated the experimental excess proton self-diffusion constant. For example, the proton self-diffusion constant for MS-EVB 3.2 is $0.37 \pm 0.01 \text{ Å}^2/\text{ps}$ and for MS-RMD 3 is $0.41 \pm 0.05 \text{ Å}^2/\text{ps}$,²⁷ compared to the experimental value of $0.94 \pm 0.01 \text{ Å}^2/\text{ps}$. This disagreement is likely due to a number of factors: primarily that MS-EVB and MS-RMD do not include nuclear quantum effects; finite size effects can also play a role. EDS-AIMD, on the other hand, has a higher diffusion constant of $0.58 \pm 0.02 \text{ Å}^2/\text{ps}$ compared to experiment. The MS-RMD models presented herein differ in their improvement of the calculated self-diffusion constants. For MS-RMD 4, the calculated self-diffusion constant is $0.41 \pm 0.07 \text{ Å}^2/\text{ps}$, which is the same as previously reported models. For MS-RMD 5, however, it is $0.47 \pm 0.05 \text{ Å}^2/\text{ps}$, which makes this model the most accurate MS-RMD hydrated excess proton model to date in terms of the self-diffusion constant. As such, we find that inclusion of anharmonic water, aSPC/Fw, is important for increasing the proton self-diffusion constant. This is also consistent with the work of Paesani and co-workers.³⁵

We also note that accounting for finite size effects and nuclear quantum effects will increase the self-diffusion constant value by as much as a factor of 1.56.²⁴ If we multiply the MS-RMD 5 result by this correction factor, we arrive at an excess proton self-diffusion constant estimate of $0.7 \text{ Å}^2/\text{ps}$, which is quite close to the experimental value. We note that the EDS-AIMD excess proton diffusion result with this same correction becomes $0.9 \text{ Å}^2/\text{ps}$, which is in essentially perfect agreement with experiment.

One of the most important properties of the hydrated excess proton in water is that it undergoes Grotthuss hopping; quantifying such hops is therefore an important way of analyzing any excess proton model. We therefore performed an analysis of the concerted hopping behavior.³⁴ The most natural way of quantifying hopping events is to associate them with a time scale τ . We then define a hopping function $h_n(\tau)$, which

is a function of the number of water molecules to which the excess proton is associated; the value n is the number of distinct water molecules that the excess proton is associated with during n intervals of length τ . For example, if an excess proton were initially associated with water A, and after a time τ is still associated with A, then $h_1(\tau)$ is incremented. If after a second interval τ the proton had hopped to another water B, then $h_2(\tau)$ would be incremented. In Fig. 3, we show the concerted hopping for a number of models: AIMD, EDS-AIMD, MS-EVB 3.2, MS-RMD 3, MS-RMD 4, and MS-RMD 5. While we do not discuss all of these models in depth in this paper, we show them here in order to emphasize the improvement in hopping in the anharmonic water (MS-RMD 5) model presented herein. First, for $n=1$ and $n=2$, the concerted hopping across all models is similar. That is, for all models, the likelihood that an excess proton is associated with 1 or with 2 water molecules is nearly identical. The primary differences emerge when one looks at $h_3(\tau)$. First, for both AIMD and EDS-AIMD, there is significantly more hopping at short times than in any of the EVB/RMD models. This means that concerted and/or rapid sequential hopping events are much more likely to occur in AIMD and EDS-AIMD, which partially explains the larger self-diffusion constants. The concerted hopping for our harmonic water model, MS-RMD 4, is similar to that of previous MS-RMD and MS-EVB models. The anharmonic MS-RMD 5, on the other hand, exceeds those of previous models, which contributes to the enhancement in the self-diffusion constant. We also see that the differences in concerted hopping are most pronounced at short τ values and that as τ gets large enough, all the models are similar. Effectively, the concerted hopping in short time intervals provides an enhancement to the self-diffusion constant, which none of the MS-RMD models

FIG. 3. Schematic of concerted proton hopping (above) and $h_n(\tau)$ for $n = 1, 2, 3$.

FIG. 4. Forward hopping for EDS-AIMD, MS-RMD 4, and MS-RMD 5.

completely capture. Future developments in MS-RMD modeling for this problem will likely require a revised functional form over Eqs. (2)–(10) (the MS-RMD “basis set”) in order to better capture this concerted proton hopping behavior. We emphasize, however, that concerted proton hopping is not the dominant effect defining the solvation structure nor the transport of the hydrated excess proton in water.

Finally, we consider the forward hopping.^{23,24} We define forward hopping as hopping that results in the proton being associated with a new water molecule. Therefore, if a proton is initially associated with water A, and then hops to water B, the forward hopping increases by 1. If the proton subsequently returns to water A, the forward hopping decreases by 1. The forward hopping results for EDS, MS-RMD 4, and MS-RMD 5 are shown in Fig. 4. For the sake of clarity and due to the limited AIMD data, we show only a 20 ps window. First, notice that for EDS-AIMD, the forward hopping far surpasses that of MS-RMD 4 and MS-RMD 5. This explains some of the increase in self-diffusion in EDS-AIMD. Also note that in the region around $t = 18$ ps, forward hopping decreases quickly, followed by a rapid increase. The decrease in forward hopping occurs when the proton hops along a chain of waters across which it had already hopped. Another important property illustrated in this plot is that while forward hopping events occur fairly frequently in the MS-RMD models, they suffer from the phenomenon of “rattling,” which occurs when a proton rapidly hops between the same two water molecules. The rapid rattling processes indeed result in a decrease in the overall diffusion constant, as while proton transfer is occurring, the overall excess proton displacement is not substantial.

IV. CONCLUDING REMARKS

We have developed two new MS-RMD models—models that are fit for the first time to EDS-AIMD data. While the excess proton solvation structure with EDS-AIMD is not substantially altered from more standard AIMD utilizing the BLYP+D3 level of DFT, it does have the advantage of having removed the effects of over-structured water solvent present in virtually all AIMD simulations of water to date. Of the two MS-RMD models we developed in this paper, one uses an underlying harmonic water model (MS-RMD 4), whereas one uses an anharmonic water model (MS-RMD 5). While

both models have similar proton solvation structures, the anharmonic model has a slightly lower internal proton transfer barrier and about a 15% higher proton self-diffusion constant. We ascribe the increase in proton self-diffusion to the increase in concerted and forward hopping. This is a reasonably important property of proton transport in water so is therefore valuable to capture it in an MS-RMD model.

As we have previously shown in the preliminary work,²⁷ REM is a very effective fitting scheme in the development of MS-RMD models, and it is capable of capturing some of the chemically reactive nature of proton transfer into a reactive force field (in this case, MS-RMD). The development of such reactive MD force fields then allows one to study proton solvation and transport in systems on length and time scales that are much too large for AIMD methods that explicitly treat electronic degrees of freedom. As such, the work presented in this paper, along with the EDS-AIMD approach, provides a promising route for accurately parameterizing novel reactive force fields from quantum mechanical simulation data.

SUPPLEMENTARY MATERIAL

See [supplementary material](#) for parameters of the MS-RMD 4 and MS-RMD 5 models.

ACKNOWLEDGMENTS

This research was supported by the Department of Energy (DOE), Office of Basic Energy Sciences (BES), Division of Chemical Sciences, Geosciences, and Biosciences (Award No. DE-SC0005418). C.C. is grateful for the financial support from National Natural Science Foundation of China (Grant Nos. 21303123 and 21303124) and the China Scholarship Council (No. 201306275019). The computational resources in this work were provided by the University of Chicago Research Computing Center (RCC).

¹S. Y. Willow, M. A. Salim, K. S. Kim, and S. Hirata, *Sci. Rep.* **5**, 14358 (2015).

²H. Hu, Z. Ma, and F. Wang, in *Annual Reports in Computational Chemistry*, edited by R. Wheeler (Elsevier, 2014), Vol. 10, pp. 25–43.

³V. Babin, C. Leforestier, and F. Paesani, *J. Chem. Theory Comput.* **9**(12), 5395–5403 (2013).

⁴O. Akin-Ojo and F. Wang, *J. Comput. Chem.* **32**(3), 453–462 (2011).

⁵O. Akin-Ojo and F. Wang, *Chem. Phys. Lett.* **513**(1–3), 59–62 (2011).

⁶F. Wang, O. Akin-Ojo, E. Pinnick, and Y. Song, *Mol. Simul.* **37**(7), 591–605 (2011).

⁷E. R. Pinnick, S. Erramilli, and F. Wang, *J. Chem. Phys.* **137**(1), 014510 (2012).

⁸A. D. White and G. A. Voth, *J. Chem. Theory Comput.* **10**(8), 3023–3030 (2014).

⁹T. Dannenhoffer-Lafage, A. D. White, and G. A. Voth, *J. Chem. Theory Comput.* **12**(5), 2144–2153 (2016).

¹⁰A. D. White, C. Knight, G. M. Hocky, and G. A. Voth, *J. Chem. Phys.* **146**(4), 041102 (2017).

¹¹W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *J. Chem. Phys.* **79**(2), 926–935 (1983).

¹²H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, *J. Phys. Chem.* **91**(24), 6269–6271 (1987).

¹³J. L. Abascal and C. Vega, *J. Chem. Phys.* **123**(23), 234505 (2005).

¹⁴Y. Wu, H. L. Tepper, and G. A. Voth, *J. Chem. Phys.* **124**(2), 024503 (2006).

¹⁵G. R. Medders, V. Babin, and F. Paesani, *J. Chem. Theory Comput.* **10**, 2906–2910 (2014).

¹⁶N. Agmon, *Chem. Phys. Lett.* **244**(5–6), 456–462 (1995).

¹⁷M. Tuckerman, K. Laasonen, M. Sprik, and M. Parrinello, *J. Chem. Phys.* **103**(1), 150–161 (1995).

¹⁸J. Lobaugh and G. A. Voth, *J. Chem. Phys.* **104**(5), 2056–2069 (1996).

¹⁹U. W. Schmitt and G. A. Voth, *J. Chem. Phys.* **111**(20), 9361–9381 (1999).

²⁰D. Marx, M. E. Tuckerman, J. Hutter, and M. Parrinello, *Nature* **397**(6720), 601–604 (1999).

²¹C. Knight and G. A. Voth, *Acc. Chem. Res.* **45**(1), 101–109 (2012).

²²T. J. F. Day, A. V. Soudackov, M. Cuma, U. W. Schmitt, and G. A. Voth, *J. Chem. Phys.* **117**(12), 5839–5849 (2002).

²³Y. Wu, H. Chen, F. Wang, F. Paesani, and G. A. Voth, *J. Phys. Chem. B* **112**(2), 467–482 (2008).

²⁴R. Biswas, Y. L. Tse, A. Tokmakoff, and G. A. Voth, *J. Phys. Chem. B* **120**(8), 1793–1804 (2016).

²⁵C. Knight, C. M. Maupin, S. Izvekov, and G. A. Voth, *J. Chem. Theory Comput.* **6**(10), 3223–3232 (2010).

²⁶C. Knight, G. E. Lindberg, and G. A. Voth, *J. Chem. Phys.* **137**(22), 22A525 (2012).

²⁷C. Arntsen, C. Chen, and G. A. Voth, *Chem. Phys. Lett.* **683**, 573–578 (2017).

²⁸M. S. Shell, *J. Chem. Phys.* **129**(14), 144108 (2008).

²⁹A. Chaimovich and M. S. Shell, *Phys. Chem. Chem. Phys.* **11**(12), 1901–1915 (2009).

³⁰A. Chaimovich and M. S. Shell, *Phys. Rev. E* **81**(6), 060104 (2010).

³¹A. Chaimovich and M. S. Shell, *J. Chem. Phys.* **134**(9), 094112 (2011).

³²G. A. Cisneros, K. T. Wikfeldt, L. Ojamae, J. Lu, Y. Xu, H. Torabifard, A. P. Bartok, G. Csanyi, V. Molinero, and F. Paesani, *Chem. Rev.* **116**(13), 7501–7528 (2016).

³³M. J. Gillan, D. Alfe, and A. Michaelides, *J. Chem. Phys.* **144**(13), 130901 (2016).

³⁴Y. L. Tse, C. Knight, and G. A. Voth, *J. Chem. Phys.* **142**(1), 014104 (2015).

³⁵K. Park, W. Lin, and F. Paesani, *J. Phys. Chem. B* **116**(1), 343–352 (2012).

³⁶A. D. Becke, *Phys. Rev. A* **38**(6), 3098–3100 (1988).

³⁷C. Lee, W. Yang, and R. G. Parr, *Phys. Rev. B* **37**(2), 785–789 (1988).

³⁸S. Grimme, *J. Comput. Chem.* **25**(12), 1463–1473 (2004).

³⁹S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, *J. Chem. Phys.* **132**(15), 154104 (2010).