

# Weighted-Graph-Theoretic Methods for Many-Body Corrections within ONIOM: Smooth AIMD and the Role of High-Order Many-Body Terms

Juncheng Harry Zhang, Timothy C. Ricard, Cody Haycraft, and Srinivasan S. Iyengar\*

Cite This: *J. Chem. Theory Comput.* 2021, 17, 2672–2690

Read Online

ACCESS |

Metrics & More

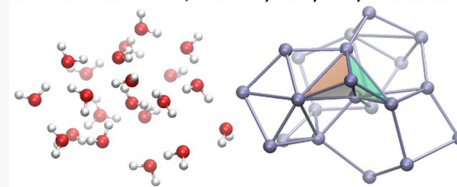
Article Recommendations

**ABSTRACT:** We present a weighted-graph-theoretic approach to adaptively compute contributions from many-body approximations for smooth and accurate post-Hartree–Fock (pHF) *ab initio* molecular dynamics (AIMD) of highly fluxional chemical systems. This approach is ONIOM-like, where the full system is treated at a computationally feasible quality of treatment (density functional theory (DFT) for the size of systems considered in this publication), which is then improved through a perturbative correction that captures local many-body interactions up to a certain order within a higher level of theory (post-Hartree–Fock in this publication) described through graph-theoretic techniques. Due to the fluxional and dynamical nature of the systems studied here, these graphical representations evolve during dynamics. As a result, energetic “hops” appear as the graphical representation deforms with the evolution of the chemical and physical properties of the system. In this paper, we introduce dynamically weighted, linear combinations of graphs, where the transition between graphical representations is smoothly achieved by considering a range of neighboring graphical representations at a given instant during dynamics. We compare these trajectories with those obtained from a set of trajectories where the range of local many-body interactions considered is increased, sometimes to the maximum available limit, which yields conservative trajectories as the order of interactions is increased. The weighted-graph approach presents improved dynamics trajectories while only using lower-order many-body interaction terms. The methods are compared by computing dynamical properties through time-correlation functions and structural distribution functions. In all cases, the weighted-graph approach provides accurate results at a lower cost.

Weighted graph-theoretic methods for smooth fragment-based AIMD. Post-Hartree-Fock accuracy and DFT cost

$$E_{\{G^{\alpha}\}}[\tilde{\mathbf{r}}(t)] = E^{\text{level},0}[\tilde{\mathbf{r}}(t)] + \left( \sum_{r=0}^{\infty} (-1)^r \sum_{\alpha \in \mathcal{V}_{r,\beta}} (E_{\alpha,r}^{\text{level},1}[\tilde{\mathbf{r}}(t)] - E_{\alpha,r}^{\text{level},0}[\tilde{\mathbf{r}}(t)]) \left[ \sum_{m=r}^{\infty} (-1)^m p_{\alpha,\beta}^m \right] \right)$$

Accurate and efficient, on-the-fly many body interactions



## I. INTRODUCTION

*Ab initio* molecular dynamics<sup>1–5</sup> (AIMD) is an important computational tool that is often used to (a) study reactive systems,<sup>6–9</sup> (b) construct ensemble averages that help define expectation values for observables,<sup>10,11</sup> and (c) treat systems beyond the standard harmonic approximation to compute vibrational properties.<sup>12–22</sup> It is generally considered to be a good starting point for the study of anharmonic systems that tend to be fluxional and require adequate sampling of the potential energy surface beyond that described by the harmonic normal vibrational mode coordinates. This dynamic sampling of the local potential energy surface proves to be a critical approach to obtain vibrational density of states (VDOS), through time-correlation functions,<sup>10–12,17–20,23,24</sup> and other experimentally observable properties for fluxional systems.<sup>12,25</sup> In AIMD, the instantaneous electronic energies and gradients are needed to compute the classical trajectories corresponding to the nuclear degrees of freedom. On account of the steep algebraic scaling of post-Hartree–Fock (pHF) methods,<sup>26–28</sup> most practical AIMD studies on moderately sized systems are limited to “on-the-fly” density functional theory (DFT) studies.<sup>29–33</sup> Although DFT

methods have proved to be extremely useful for a wide variety of problems, a number of critical challenges remain.<sup>34–36</sup> To remedy this dilemma, we have developed graph-theoretic methods that, based on the studies in refs 37–45 appear to achieve post-Hartree–Fock accuracy at DFT cost for “on-the-fly” AIMD trajectories and potential surface calculations. The key idea in these studies begins with ONIOM;<sup>46</sup> however, then, the “model” and “real” system energies and gradients are constructed using many-body expansions up to the arbitrary rank. Furthermore, these many-body expansions are obtained in a general fashion using adaptive graph-theoretic techniques, which is time-dependent in AIMD trajectories and degrees of freedom dependent on potential energy surface calculations. The use of accurate analytical gradients associated with the

Received: December 11, 2020

Published: April 23, 2021



resulting energy expression yields both extended Lagrangian<sup>38,39</sup> and Born–Oppenheimer<sup>37–39</sup> (BOMD)-based *ab initio* molecular dynamics simulation protocols that have been shown to provide accuracy comparable to CCSD and MP2 levels of theory, with DFT-computational expense.

The methods in refs 37–45 are strongly influenced by recent, ONIOM-based,<sup>46–52</sup> molecular fragmentation methods<sup>53–65</sup> as well as developments in the many-body theory.<sup>66–77</sup> The key ingredient in the developments discussed in refs 37–45 involves a geometric and graph-theoretic interpretation that efficiently combines molecular fragmentation, ONIOM, and many-body expansions. This results in a powerful algorithm that rapidly converges as a function of many-body rank<sup>43</sup> and allows the simultaneous utilization of multiple electronic structural packages within a single AIMD step and potential energy surface calculation. This graph-theoretic approach partitions the chemical system considered into a set of chemical groups (such as water and hydronium in refs 43 and 42 or a single amino acid group in a polyaniline chain in refs 40 and 41) that are considered as geometric nodes used to provide approximations to local molecular one-body electronic energetic contributions. First-order interactions between these discrete nodes are captured by creating edges, which are the union of two nodes. Once the nodes and edges are defined, the chemical system of interest is now represented as a graph. Higher-order (or *n*-body) terms can then be considered based upon the connectivity of the nodes, and the criteria to determine this connectivity are discussed in detail in refs 40 and 43 and summarized in Appendix A. Since the local interactions and bonding networks may change as the system evolves either in dynamics or while sampling the molecular potential energy surface, the graphical representation should adaptively consider and reweigh local many-body effects. This adaptability requirement leads to an evolving representation of the system as a function of nuclear configuration. These graphical evolutions can occur in two ways: (a) the definition of the nodes may change due to local chemical transformations such as hydrogen/proton transfer or other reactive events or (b) the formation/elimination of edges (which in this work represents hydrogen bonds or higher-order nonbonded interactions) due to the change in local, internal, interactions within the system studied. These two dynamical properties of the graphical representation force this approach to adapt to the changing physics inherent to both dynamics trajectories and potential energy surface calculations. The adaptive nature of this graphical approach complicates its direct application to AIMD trajectories of reactive systems where they will experience “graphical hops” or the associated “hotspots”.<sup>78</sup> Previously, we have developed schemes<sup>42</sup> for using multiple graphical representations simultaneously to construct smooth and differentiable quantum nuclear potential surfaces. Here, we consider all graphical representations arising within a given simulation time window, and the energy surface may be thought of as arising from a linear combination of such “local” graphical representations. To properly evaluate our methods, AIMD trajectories are studied for protonated water clusters due to their highly fluxional characteristics.<sup>12,13,17,79–81</sup> Furthermore, the resultant trajectories are compared with those obtained from progressively growing ranks of many-body theory to analyze the convergence of trajectories with respect to many-body rank contributions.

The paper is organized as follows: in Section II, we provide a brief survey of our graph-theory-based *ab initio* molecular dynamics methodology that has strong connections to many-

body approximations<sup>66–75,77</sup> as well as fragmentation-based techniques<sup>48–52,57,58,60–63,73,75,82–89</sup> and ONIOM.<sup>46,90–92</sup> In Section III, we introduce an approach that provides a weighted combination of multiple graphical representations found during an AIMD trajectory to arrive at a conservative dynamics protocol. In Section IV, we then go on to highlight our results on small- and medium-sized protonated water clusters. One significant highlight of our work is being able to achieve the conservative CCSD level AIMD on the protonated 21-water cluster ( $\text{H}_{43}\text{O}_{21}^+$ ), which has provided a long-standing challenge to theory and experiment.<sup>12,17,79,93,94</sup> The rapid convergence properties of this approach with increasing many-body terms are presented in Section IV.II. Structural properties such as radial distribution functions (RDFs) and dynamical properties such as time-correlation functions are compared for various trajectories in Section IV.III. Conclusions are given in Section V.

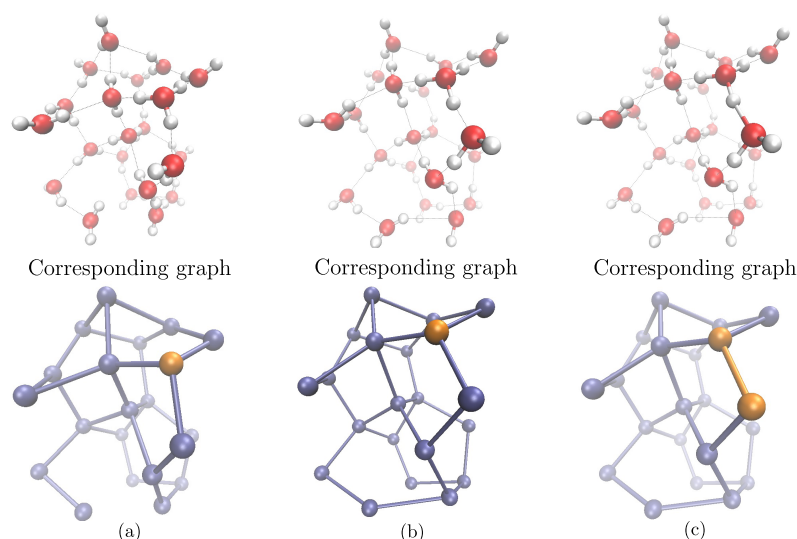
## II. GRAPHICAL REPRESENTATION OF LOCAL MANY-BODY INTERACTIONS FOR *AB INITIO* MOLECULAR DYNAMICS

In a series of publications,<sup>37–44</sup> we have developed graph-theory-based techniques to compute efficient and adaptive many-body expansions that have strong connections to molecular fragmentation and ONIOM. The salient features of this approach are as follows: the molecular assembly is partitioned into a set of nodes or vertices. These nodes are then connected through edges based on a chosen edge length cutoff that captures the extent to which bonded and nonbonded interactions are to be captured. Together, the set of nodes and edges produce a graph,  $\mathcal{G}_{\mathbf{x}} \equiv \{\mathbf{V}_0^{\mathbf{x}}; \mathbf{V}_1^{\mathbf{x}}\}$ , which is obtained from the instantaneous molecular structure,  $\mathbf{x}$ . Here,  $\mathbf{V}_0^{\mathbf{x}}$  is the set of vertices and  $\mathbf{V}_1^{\mathbf{x}}$  is the set of edges for the molecular graph depicting the instantaneous structure,  $\mathbf{x}$ . This graph is now said to represent local interactions, where through the presence of edges, two-body local interactions are captured. However, inherently present in this graph are also higher-order interactions represented by triangles, tetrahedrons, and objects with five or more nodes. The critical aspect here, which makes the connections to many-body theory rigorous, is that these higher-order objects are completely connected. That is, here, all pairs of nodes in the included higher-order objects are connected through edges and such objects are known as affine simplexes.<sup>95–97</sup> Each set of rank-*r* simplexes,  $\{\mathbf{V}_r | r = 0, \dots, \mathcal{R}\}$ , thus arises from a (truncated) power set of the elements within the graph

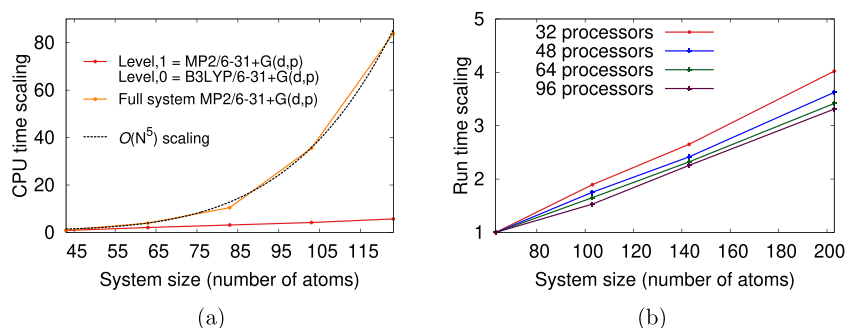
$$\{\mathbf{V}_r | r = 0, \dots, \mathcal{R}\} \equiv \{\mathbf{V}_0, \mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_r, \dots, \mathbf{V}_{\mathcal{R}}\} \quad (1)$$

The elements of such a power set provide a general and robust scheme to construct many-body expansions for arbitrary systems. One part of this publication includes trajectories computed with predefined maximum ranks, ranging from edges ( $\mathcal{R} = 1$ ) to pentahedrons, or “pentatopes”, ( $\mathcal{R} = 4$ ), which determines the maximum many-body contribution considered within the graphical representation.

The above graphical description allows a dynamic and flexible representation of local many-body interactions. The energetic measure we begin within this paper, and considered in refs 37–44, consists of a perturbative, ONIOM-type, correction to a result obtained at a lower level of theory, where the perturbative correction is the difference between two many-body expansions (replacing the standard “model-high” minus “model-low” portion in ONIOM) given by the graphical representation



**Figure 1.** Here, the protonated 21-water cluster is represented as a graph, as required by eq 3, for each structure considered in the AIMD trajectory. As the system evolves, the graphical representation changes to better represent the local interactions and local chemistry of the system. Three structures of the protonated 21-water system, and their respective graphical representations, are considered above. For illustration of this graphical representation change, there is a hydrogen bond formation between (a) and (b), which is captured by the creation of an additional edge in the graphical representation. Here, the orange nodes represent the protonated water; there is a transition from an eigen-like structure (b) with the proton localized to one water to a Zundel-like structure (c) with the proton shared between two waters. These changes in the graphical representation often lead to energetic hops, as shown in Figure 4. Equation 5 is later employed to smoothen the energy and forces of these graphical transitions.



**Figure 2.** Computational scaling and cost studies above are shown for MP2 extrapolations from B3LYP, with a Pople style basis of 6-31+G(d,p) for a series of polyaniline stands with increasing chain lengths. (a) This method demonstrates lower-order scaling than the full-system MP2 calculations and is further expanded using parallel computing resources in (b).

above. Consistent with the notions behind ONIOM,<sup>46</sup> we have introduced<sup>37–44</sup> an energy expression that conveys this general idea

$$E_{\text{MBE,gt}}^{\text{ONIOM}}(\bar{\mathbf{x}}) = E^{\text{level},0}(\bar{\mathbf{x}}) + E_{\text{MBE}}^{\text{level},1}(\bar{\mathbf{x}}) - E_{\text{MBE}}^{\text{level},0}(\bar{\mathbf{x}}) \quad (2)$$

where the left side,  $E_{\text{MBE,gt}}^{\text{ONIOM}}(\bar{\mathbf{x}})$ , denotes the graph-theoretically obtained many-body correction to ONIOM and the term  $E_{\text{MBE}}^{\text{level},1}(\bar{\mathbf{x}})$  on the right side may encompass the full system or a chosen “active site”, and we have considered both options within AIMD<sup>37–41</sup> and quantum nuclear potential surface treatments.<sup>42,45</sup> In addition to the extrapolatory, ONIOM-like form of eq 2, each term in the extrapolation is a many-body expansion, which is now written in a general and computationally robust fashion up to order (or rank)  $\mathcal{R}$  as

$$E_{\text{MBE}}^{\text{level},1} = \sum_{r=0}^{\mathcal{R}} (-1)^r \sum_{\alpha \in \mathbf{V}_r} E_{\alpha,r}^{\text{level},1} \left[ \sum_{m=r}^{\mathcal{R}} (-1)^m p_{\alpha}^{r,m} \right] \quad (3)$$

where  $p_{\alpha}^{r,m}$  is the number of times the  $\alpha$ th  $(r+1)$ -body term (in set  $\mathbf{V}_r$ ) appears in all  $(m+1)$ -body terms (in set  $\mathbf{V}_m$  for  $m \geq r$ )

and consequently  $\left[ \sum_{m=r}^{\mathcal{R}} (-1)^m p_{\alpha}^{r,m} \right]$  is the overcounting correction for the number of times the  $\alpha$ th  $(r+1)$ -body term appears in all objects of rank greater than or equal to  $r$ . It is important to emphasize that eq 3 is essentially identical to standard many-body expressions but presented now using graph theory. An illustration of such a graphical decomposition of a protonated water cluster treated later in this publication is shown in Figure 1; the dynamical nature of such a graphical representation is also shown and addressed in Section III. Thus, the full energy expression, which combines eqs 2 and 3, becomes

$$\begin{aligned} E_{\text{MBE,gt}}^{\text{ONIOM}} &= E^{\text{level},0} + \sum_{r=0}^{\mathcal{R}} (-1)^r \left\{ \sum_{\alpha \in \mathbf{V}_r} (E_{\alpha,r}^{\text{level},1} - E_{\alpha,r}^{\text{level},0}) \right. \\ &\quad \left. \left[ \sum_{m=r}^{\mathcal{R}} (-1)^m p_{\alpha}^{r,m} \right] \right\} \\ &= E^{\text{level},0} + \Delta \text{MBE}_{\mathcal{R}}^{\text{level},1;\text{level},0} \end{aligned} \quad (4)$$



In practice, the individual simplexes in eqs 3 and 4 are computed independently and in parallel (see Figure 2). Furthermore, our implementation also allows the use of separate electronic structural packages for each level of treatment and currently supports the simultaneous use of Gaussian,<sup>98</sup> ORCA,<sup>99</sup> Psi4,<sup>100</sup> Quantum Espresso,<sup>101</sup> and OpenMX.<sup>102</sup> In this paper, electronic structural calculations are solely computed using the Gaussian package. While our previous studies<sup>38–41</sup> achieved parallelism through the use of MPI, here we utilize an unconstrained asynchronous queue system that is implemented in Python.

### III. WEIGHTED-GRAPHICAL REPRESENTATION FOR ADAPTIVE MANY-BODY CONTRIBUTIONS AND CONSERVATIVE AIMD IN FLUXIONAL SYSTEMS

The manner in which the graphical representation of a given molecular system is constructed is detailed in Appendix A and allows for dynamic definitions of both nodes and edges, thereby influencing the many-body contributions in eq 4. As a system evolves during AIMD (or is sampled in potential energy surface calculations as done in refs 42 and 45), the graphical representation adapts to model the evolving nature of the nuclear framework, as dictated by inter- and intramolecular interactions captured by eq 4. One example of how these graphical connectivities may change during dynamics is shown in Figure 1, where both an edge creation/destruction and node modifications are shown. Hence, when studying reactive and fluxional systems, the application of the graphical approach is complicated by its adaptive nature. This implies that the simplex weights,  $p_{\alpha}^{r,m}$ , change over time (or space in the case of potential energy surface calculations) as local interactions (edges) are included and excluded from the graphical representation and the intrinsic properties of the nodal decomposition also change, as shown in Figure 1. Due to the adaptive definition of the graphical representation, the evolution of the system creates singular, nondifferentiable, hops in the potential energy surface as the graphical representation morphs to capture the changing physics of the system. To properly account for these “graphical hops” or the associated “hotspots”<sup>78</sup> presented in AIMD, in ref 42, we have also constructed a scheme that allows the simultaneous description of multiple graphical representations of molecular structures and assemblies

$$\begin{aligned} E_{\{\mathcal{G}^{\beta}\}}[\bar{\mathbf{x}}(t)] &= \sum_{\beta} E_{\text{MBE,gt}}^{\text{ONIOM}}[\bar{\mathbf{x}}(t); \beta] \\ &= E^{\text{level},0}[\bar{\mathbf{x}}(t)] + \sum_{\beta} v_{\beta}(t) \\ &\quad \times \left( \sum_{r=0}^{\mathcal{R}} (-1)^r \sum_{\alpha \in \mathbf{V}_{r,\beta}} (E_{\alpha,r}^{\text{level},1}[\bar{\mathbf{x}}(t)] \right. \\ &\quad \left. - E_{\alpha,r}^{\text{level},0}[\bar{\mathbf{x}}(t)]) \left[ \sum_{m=r}^{\mathcal{R}} (-1)^m p_{\alpha,\beta}^{r,m} \right] \right) \end{aligned} \quad (5)$$

where we have introduced a linear combination of expressions arising from a family of individual time-dependent, molecular structure adaptive graphs  $\{\mathcal{G}_{\bar{\mathbf{x}}(t)}^{\beta}\}$

$$\mathcal{G}_{\bar{\mathbf{x}}(t)}^{\beta} \equiv \{\mathbf{V}_{0,\beta}^{\bar{\mathbf{x}}}; \mathbf{V}_{1,\beta}^{\bar{\mathbf{x}}}\} \rightarrow E_{\text{MBE,gt}}^{\text{ONIOM}}[\bar{\mathbf{x}}(t); \beta] \quad (6)$$

and each graph,  $\mathcal{G}_{\bar{\mathbf{x}}(t)}^{\beta}$ , has an energy expression as given in eq 4, encoded within the expression  $E_{\text{MBE,gt}}^{\text{ONIOM}}[\bar{\mathbf{x}}(t); \beta]$ . Furthermore, in eq 5, the sum over  $\alpha$  is within the set  $\mathbf{V}_{r,\beta}^{\bar{\mathbf{x}}}$ , which includes all rank- $r$  simplexes in the family of graphs,  $\{\mathcal{G}_{\bar{\mathbf{x}}(t)}^{\beta}\}$ . The sum over  $m$ , as in eqs 3 and 4, is an overcounting correction term, which removes the additional number of appearances of the  $\alpha$ th rank- $r$  fragment inside all higher-rank simplexes (or molecular fragments) within a specific graph  $\mathcal{G}_{\bar{\mathbf{x}}(t)}^{\beta}$ . Note that this double counting factor in eq 5 now includes a  $\beta$  dependence unlike the case in eq 4. Thus, the changes between graphs,  $\mathcal{G}_{\bar{\mathbf{x}}(t)}^{\beta}$ , capture the respective transformation of nodes reflected in  $\mathbf{V}_{0,\beta}^{\bar{\mathbf{x}}}$  and new edge creation and destruction, captured within  $\mathbf{V}_{1,\beta}^{\bar{\mathbf{x}}}$ . The change in edges in turn affects the sets  $\mathbf{V}_{r,\beta}^{\bar{\mathbf{x}}}$  for  $r > 1$  and also affects the multiplicity values in summation of  $m$  in eq 5. The contribution to the energy expressions from these graphs is modulated by a normalized weighting term,  $v_{\beta}(t)$ , acting on the family of graphs that are present inside some AIMD time window,  $[t - \Upsilon, t]$ . Thus

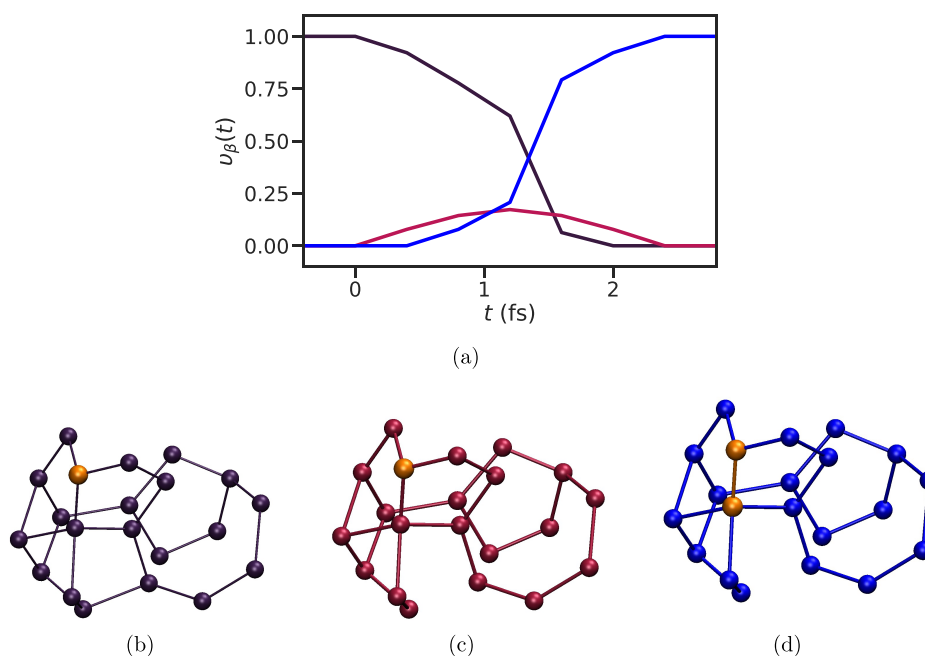
$$\sum_{\beta} v_{\beta}(t) = 1 \quad (7)$$

which was already assumed in moving from the first line in eq 5 to the second line. Gradients for this energy expression in eq 5 are given in Appendix B, and these are responsible for providing conservative dynamics later in the publication.

Equation 5 is similar in spirit to the work in refs 42 and 45 but differs in that, in refs 42 and 45, the weights are obtained from an eigenvalue problem that is constructed using all graphical connectivities appearing inside the potential energy space for a molecule.<sup>42,45</sup> However, AIMD is local since the potential and gradients only depend on the instantaneous nuclear framework. This may be contrasted with quantum nuclear dynamics,<sup>15,42,45</sup> where the potential energy surface may be global. This allows one to only consider the time-dependent, local, neighborhood of graphical representations in the vicinity of the current molecular structure. Thus, for AIMD, instead of considering “all” of the graphical representations available in the system, or those within an energetic window, here we consider the graphical representations accessible to the system inside some time or position window,  $\{\mathcal{G}_{\bar{\mathbf{x}}(t)}^{\beta}\} \in [t - \Upsilon, t]$ , modulated by  $v_{\beta}(t)$ . For example, if there are two graphical connectivities that appear inside the time window,  $[t - \Upsilon, t]$ , then, the function,  $v_{\beta}(t)$ , smoothly interpolates between the potential surfaces represented by the two graphical connectivities. The precise form of  $v_{\beta}(t)$  may be thought of as a smoothened version of the well-known Heaviside function, and for simplicity here, we utilize a cuberoot interpolation scheme. It is, of course, possible that there could be more than two graphical representations inside some specific time window,  $[t - \Upsilon, t]$ . Thus, if there are a set of graphs,  $\{\mathcal{G}_{\bar{\mathbf{x}}(t)}^{\beta}\}$ , encountered inside the time window  $[t - \Upsilon, t]$ , at respective time steps,  $t_{\beta} \in [t - \Upsilon, t]$ , where the index  $\beta$  is an ordered set,  $\beta = 0, \dots, N$ , for  $(N + 1)$  graphical topologies encountered inside the time window,  $[t - \Upsilon, t]$ , then the weights for the individual graphs used in eq 5 are obtained using the expressions

$$v_{\beta}(t) = f\left(\frac{t - t_{\beta} + \Delta t}{\Upsilon}\right); \quad \beta = N \quad (8)$$





**Figure 3.** Illustration of the effect of choice of  $v_\beta(t)$  in eq 10. In (a), we provide a time fragment of the AIMD trajectory for the protonated 21-mer discussed later in the paper. Specifically, the region shown here has multiple graphical representations where the population (eq 10) of each graphical representation is presented using colored traces.  $t = 0$  marks the appearance of the second representation, which is the first change in the graphical representation inside this time series. In (b)–(d), we present the associated graphical representations colored accordingly. As in Figure 1, the protonated nodes are presented as orange.

$$v_\beta(t) = f\left(\frac{t - t_\beta + \Delta t}{\Upsilon} \prod_{\gamma=\beta+1}^N \left[1 - \frac{t - t_\gamma + \Delta t}{\Upsilon}\right]\right); \quad \beta = 1, \dots, N-1 \quad (9)$$

$$v_\beta(t) = 1 - \sum_{\beta'=1}^N v_{\beta'}(t); \quad \beta = 0 \quad (10)$$

Here,  $\Delta t$  is the molecular dynamics time step, and eq 10 explicitly enforces the normalization condition in eq 7 by adapting the weights for the first topology,  $\beta = 0$ , inside the time window based on the subsequent topologies,  $\beta = 1, \dots, N$ , discovered during dynamics. When  $N = 2$ , eq 9 does not apply. In this publication, we have chosen

$$f(x) = 0.5 \times (2x - 1)^{1/3} + 0.5 \quad (11)$$

An illustration of the effect of  $v_\beta(t)$  is provided in Figure 3. Specifically in this illustration, the creation of a new edge leads to the graphical transition black  $\rightarrow$  red as shown in Figure 3a, with graphs provided in Figure 3b,c. Thus, a little after the time of the first hop noted as “0” on the horizontal axis in Figure 1a, the energy is obtained from two graphical representations, black and red, with weights as shown. The edge transition appears to facilitate a proton hop, which leads to the transition (blackred)  $\rightarrow$  blue. The use of eqs 9 and 10 within eq 5 provides for smooth transitions across multiple graphical descriptions, resulting in conservative dynamics, as is probed in Section IV.

This also allows us to reinterpret the instantaneous graphical realization in eq 4 as a special case of eq 5, where the interpolating probabilities,  $v_\beta(t)$ , are chosen to be sharp step functions and only one graphical term dominates the energy expression at each instant in time. The appropriate choice of  $v_\beta(t)$ , within eq 5, allows a natural description through smoothly evolving graphical descriptions of potentials. The choice of

sharp step functions for  $v_\beta(t)$ , in eq 5, yields eq 2 and may be expected to result in “hotspots” (as in ref 78) in AIMD, which are then alleviated through proper choice of  $v_\beta(t)$ , in eq 5.

To study the effectiveness of the algorithm presented here, in Section IV, we first analyze the behavior of a family of trajectories, where each is constructed at a prespecified maximum edge length and global rank. Indeed, we find that at any such specified level of truncation of eq 4, the observed AIMD trajectory hops inside a bundle of trajectories defined at fixed ranks, but the detrimental effect of these hops progressively reduces as the rank increases. Furthermore, as the rank included increases, we also show that the expression in eq 2 converges much more rapidly as compared to the standard many-body expressions written in eq 3 using the graph-theoretic notation. Hence, by then considering a monotonic sequence of AIMD trajectories, each at a progressively higher rank, or each considering higher-order many-body contributions in eq 4, we show that conservative AIMD dynamics is possible for many practical systems. This situation is then compared with the more effective weighted-graph trajectories, as allowed by eq 5, that suitably interpolate between a neighborhood of trajectories and provide good quality energy conservation properties and comparable to those constructed at higher rank but now at a lower cost.

An additional factor that must be noted here is the fact that the time dependence of eq 5, given the weights described by eqs 8–11, automatically makes the AIMD Hamiltonian time-dependent. Since time-correlations are computed here to obtain dynamical properties, and these require a conservative Hamiltonian, the time dependence of eq 5 is potentially problematic from the perspective of dynamics. However, as we will find in Section IV, the numerical effect of this time dependence is essentially negligible. Furthermore, there is also the question of preserving time-reversal symmetry given that the

**Table I. Energy Conservation Properties for Dynamical Simulations (Microcanonical) Using the Adaptive Edge Formation Scheme; These Trajectories Correspond to Figure 4**

system	edge cutoff (Å)	$\mathcal{R}$	$N_f^a$	$\Delta t$ (fs)	simulation time (ps)	$T_{\text{Ave}}$ (K) <sup>b</sup>	$\Delta\mathcal{H}$ (kcal/mol)	$\mathcal{H}_{\text{Drift}}^c$ (kcal/mol)	$\langle F \rangle^d$ ( $10^{-6}$ kcal/Å)
$\text{H}_{13}\text{O}_6^+$	$1.1 \times D_i$	1	11.0	0.2	6.00	$164.16 \pm 24.8$	0.511	1.155	6.7
$\text{H}_{43}\text{O}_{21}^+$	$1.1 \times D_i$	1	44.9	0.4	11.31	$277.17 \pm 21.7$	3.978 <sup>e</sup>	12.917	14.9
$\text{H}_{43}\text{O}_{21}^+$	$1.1 \times D_i$	1	47.2	0.4	11.92	$273.52 \pm 23.3$	2.724 <sup>f</sup>	7.50	14.7

<sup>a</sup>Average number of fragments. <sup>b</sup>By use of the equipartition theorem,  $(3/2)(N - 1)kT$ , we convert the kinetic energy into average and root-mean-square (RMS) temperatures. Here, the temperature is a measure of the available energy to sample the conformational space. The initial kinetic energies were randomly distributed along the nuclear degrees of freedom. <sup>c</sup>The drift in the Hamiltonian (total energy) is computed as the difference between the average for the first 100 fs and last 100 fs. In kcal/mol. <sup>d</sup>Equation 12 over the full trajectory. <sup>e</sup>When “special partners” are not included, this number escalates to 570.202 kcal/mol. <sup>f</sup>Properly accounting for “special partners” as allowed by the availability of a larger number of edges (greater than four) around each water molecule as part of the dynamics calculation.

propagation scheme is classical. We use the velocity Verlet integration scheme,<sup>103</sup> which is a third-order Trotter factorization<sup>104,105</sup> of the classical Liouville propagator<sup>106</sup> and thus formally satisfies time-reversal symmetry under the conditions that (a) the Hamiltonian is not time-dependent and (b) time steps are small enough. However, as noted above, the Hamiltonian in eq 5 is time-dependent. However, we find later, in Section IV, that the total energy is well conserved during dynamics and the deviation in classical forces also well controlled. This numerical result, in conjunction with a Trotter-factorized, time-reversible classical integrator,<sup>103</sup> leads us to expect that any loss in time-reversal symmetry would be insignificant.

#### IV. RESULTS AND DISCUSSION

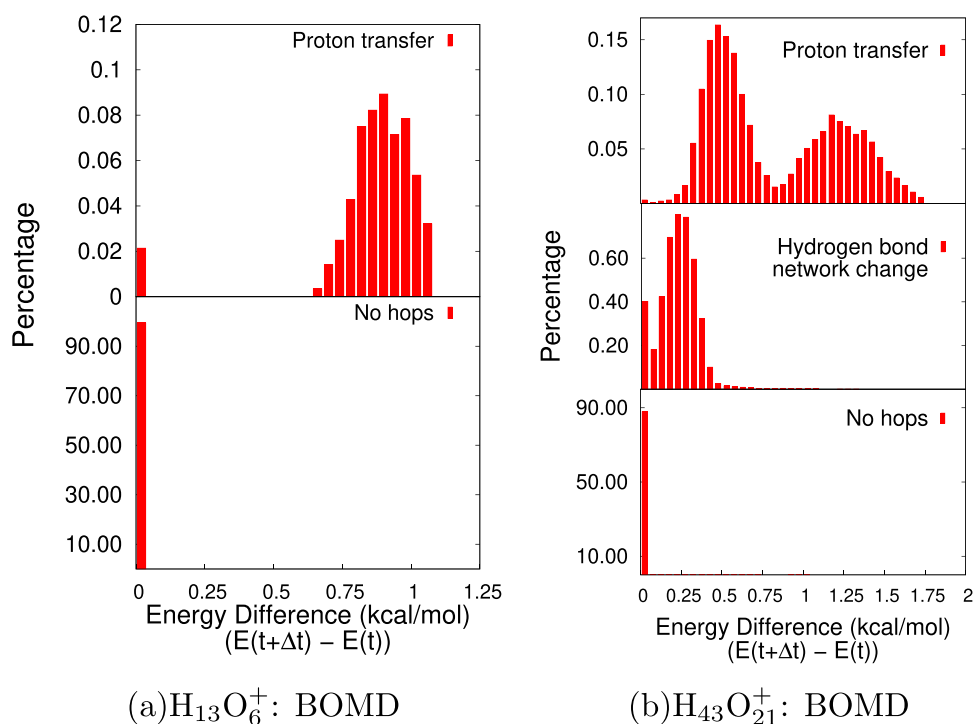
The graph-theoretic methodology, discussed above, is an efficient approach to compute post-Hartree–Fock AIMD trajectories<sup>38,39</sup> and approximations to large basis set AIMD.<sup>41</sup> However, the treatment of highly fluxional chemical structures would inherently lead to energy hops due to “hotspots” created by the evolving nature of the graphical representation in fluxional systems. We examine these “hotspots” and reduce their impact upon AIMD trajectories using the methods presented above. We consider AIMD trajectories for two protonated water clusters: (a) the small-sized solvated Zundel ( $\text{H}_{13}\text{O}_6^+$ ) cation, which is thought to play a significant role in the fundamental study of proton transfer in condensed phase systems and a variety of chemical, biological, and material problems,<sup>14,81,107–116</sup> and (b) the medium-sized protonated 21-water cluster<sup>79,117–125</sup> ( $\text{H}_{43}\text{O}_{21}^+$ ), which has presented a long-standing challenge for theoretical as well as experimental studies.<sup>12,17,117</sup>

The rationale behind the choice of these two sets of systems is clear upon the inspection of eqs 2 and 3. There are two parameters that dictate the accuracy of these approximations. First, a distance cutoff defines the maximum edge length, which connects nodes in the graphical description of the molecular framework. Then, for a chosen local edge length cutoff, one may also consider different maximum ranks of the graph, as indicated by eqs 1 and 3. As the maximum rank increases, it is critical to note that the approximations in eqs 2 and 3 converge to the correct full-system high-level energy. (Please see Section 2.1 and Section 4 in ref 43 for a detailed analysis.) However, as the rank and the edge length increase, the number of fragments increases drastically as we will see and the average fragments get larger. For the case of the solvated Zundel, however, the number and size of fragments still remain manageable. Thus, the choice of the solvated Zundel cation allows us to gauge the convergence of trajectories constructed using eq 2 for a growing family of edge

length cutoff and rank. This also allows us to compare the results from the weighted-graph approximation, eq 5 in the previous section, with the trajectories obtained using higher ranks and edge lengths. Thus, the study on the solvated Zundel allows us to gauge the extent to which the impact of “hotspots” may be alleviated through appropriate modification of the edge length cutoff as well as maximum rank,  $\mathcal{R}$ , along with the use of weighted-graphical representations. While the solvated Zundel cation allows us to probe the convergence of trajectories with increasing rank and edge length cutoff, this kind of analysis is impossible for the protonated 21-water cluster, where the number and complexity of clusters grow tremendously. In both cases, a maximum of pentameric structures ( $\mathcal{R} = 4$ ) are considered and compared with the weighted-graphical approach.

The protonated 21-water cluster is considered to be a “magic” cluster, which displays heightened stability,<sup>79,118,124,125</sup> where argon-tagged and deuterium-tagged vibrational action spectroscopy<sup>79,117,120,126,127</sup> and computational<sup>12,17,79,94,117,122,123,128</sup> approaches have been employed to study its spectroscopic features. Although this cluster has been studied with DFT to capture the essential experimental findings,<sup>12,17,79,128</sup> studying this system at the post-Hartree–Fock level of treatment remains untenable using standard full-system treatments. Thus, in Section IV.III, coupled cluster quality trajectories are considered to further elucidate the nature of the spectral evolution of these OH peaks.

**IV.I. Ascertaining the Impact of “Hotspots” on Reduced Rank,  $\mathcal{R} = 1$ , AIMD Trajectories.** The trajectories for the solvated Zundel clusters are extrapolated from B3LYP/6-31+G(d,p) to CCSD/6-31+G(d,p), and the trajectories for the protonated 21-water clusters are extrapolated from B3LYP/6-31++G(d,p) to CCSD/6-31++G(d,p), using the instantaneous as well as weighted-graph-theoretic methodology in eq 2. Here, the graph structure is dynamically computed at predefined levels of rank,  $\mathcal{R}$ , and maximum edge length. In addition, for the case of water, we find a third critical parameter that is related to the number of edges that are allowed for any given node. It might be natural to only consider a maximum edge capacity cutoff surrounding any given node, based on the expected tetrahedral electron configuration geometry that surrounds a given water molecule. However, this turns out to not be a good approximation because proton hops are thought to be facilitated by “special partners”<sup>129,130</sup> that arise from a selected set of important interactions from the second solvation shell of water molecules surrounding the solvated proton that may instantaneously oversaturate the hydration. Given these prescriptions, we have considered two sets of simulations listed in Table I, where, in one case, the number of local edges are truncated at



**Figure 4.** Hotspots or energetic hops in the two-body ( $\mathcal{R} = 1$  in eq 2) AIMD occur due to a variety of reasons. Here, we present the energetic change between two steps of AIMD. For smaller clusters, a proton transfer is the most likely reason for the lack of energy conservation. However, in the larger clusters, a variety of other events, related to the changing hydrogen bond environment, cause energetic hotspots in two-body ( $\mathcal{R} = 1$  in eq 2) AIMD trajectories. In each case, the vertical axis represents the relative probability of occurrence of each such event.

**Table II.** Energy Conservation Properties for Dynamical Simulations (Microcanonical) Using Expanded Edge Length Cutoffs and Larger-Rank Many-Body Terms

system	cutoff (Å)	$\mathcal{R}$	$N_f^a$	$\Delta t$ (fs)	simulation time (ps)	$T_{\text{Ave}}$ (K) <sup>b</sup>	$\Delta\mathcal{H}$ (kcal/mol)	$\mathcal{H}_{\text{Drift}}^c$ (kcal/mol)	$\langle F \rangle^d$ ( $10^{-6}$ kcal/Å)
$\text{H}_{13}\text{O}_6^+$	$1.1 \times D_i$	1	11.0	0.2	6.00	$164.16 \pm 24.8$	0.511	1.155	6.7
	6.00	1	18.8	0.2	6.00	$181.68 \pm 33.3$	1.921	4.989	13.8
	6.00	2	29.9	0.2	6.00	$160.22 \pm 23.0$	0.436	0.966	18.0
	6.00	3	35.2	0.2	6.00	$156.10 \pm 23.0$	0.063	0.078	9.4
	6.00	4	34.6	0.2	6.00	$154.73 \pm 22.4$	0.049	−0.044	7.1
$\text{H}_{43}\text{O}_{21}^+$	$1.1 \times D_i$	1	47.0	0.4	7.00	$277.23 \pm 22.5$	1.907	6.283	14.8
	4.00	3	74.0	0.4	7.00	$271.24 \pm 21.1$	1.092	0.109	14.5
	4.50	4	173.0	0.4	6.49	$273.95 \pm 22.5$	0.794	2.092	18.4

<sup>a</sup>Average number of fragments. <sup>b</sup>By use of the equipartition theorem,  $(3/2)(N - 1)kT$ , we convert the kinetic energy into average and RMS temperatures. Here, the temperature is a measure of the available energy to sample the conformational space. The initial kinetic energies were randomly distributed along the nuclear degrees of freedom. <sup>c</sup>The drift in the Hamiltonian (total energy) is computed as the difference between the average for the first 100 fs and last 100 fs. In kcal/mol. <sup>d</sup>Equation 12 over the full trajectory.

four and a second set of simulations where these are dynamically determined to allow additional second solvation shell partners that may appear inside the edge length cutoff, perhaps to facilitate proton transfer.

The evolution of the total energy for these trajectories shows hotspots that correspond to instantaneous changes in the overall graphical representation of the clusters, and the cumulative conservative properties for these trajectories are shown in Table I. To ascertain the quality of dynamics, we present three different parameters in Table I and in all further studies in this paper. The quantities  $\Delta\mathcal{H}$  and  $\mathcal{H}_{\text{Drift}}$  represent the standard deviation of total energy and the drift in total energy during the AIMD trajectories. It is generally desirable to have these quantities remain in the sub-kcal/mol range, and clearly, the dynamics of the larger system violates these criteria in Table I. In addition to

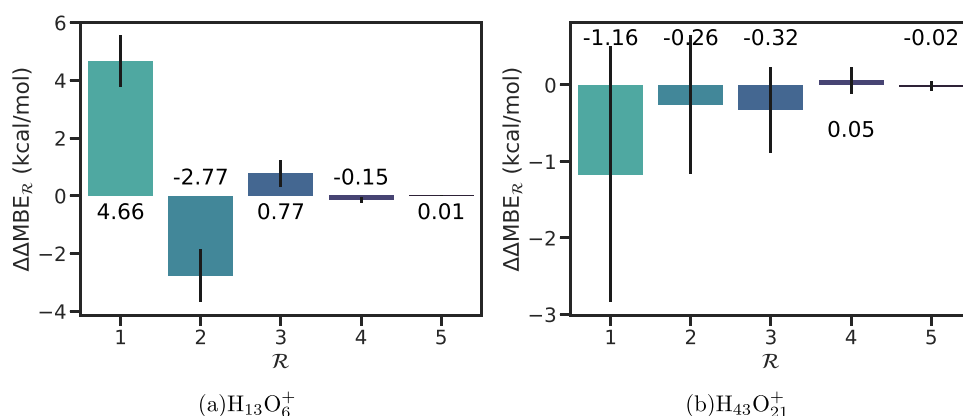
the energy criteria, we also inspect the forces computed during AIMD trajectories. Thus, we compute

$$\langle F \rangle = \frac{1}{\mathcal{T}} \int_0^{\mathcal{T}} dt \left\| \sum_{i=1}^{N_{\text{Atoms}}} \vec{F}_i(t) \right\|_2 \quad (12)$$

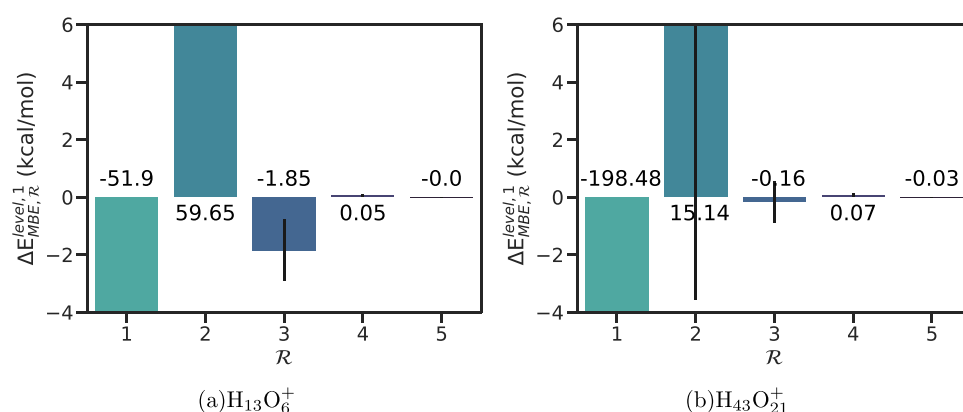
where  $\vec{F}_i(t)$  is the atomic force on atom  $i$  at time  $t$ . For a strictly conservative system, there are no external forces and hence the time evolution of the total force must be negligible. Thus, eq 12 provides a measure of the accumulation of residual forces over the entire trajectory.

Clearly, the energy parameters in Table I show insufficient conservation due to the change in graphical connectivity during the dynamics. Resulting from these hops, in Figure 4, we present a distribution of the interstep energetic change, given by  $E(t + \Delta t) - E(t)$ . This quantity, when large, defines an energetic hop,





**Figure 5.** Term-by-term convergence of eq 2 for all rank- $r$  objects, as gauged by eq 13. Specifically, the average value of  $\Delta\Delta\text{MBE}_R$ , over all structures found in the AIMD trajectory in Table I, is shown here along with standard deviations shown using vertical lines. Also see Figure 6.



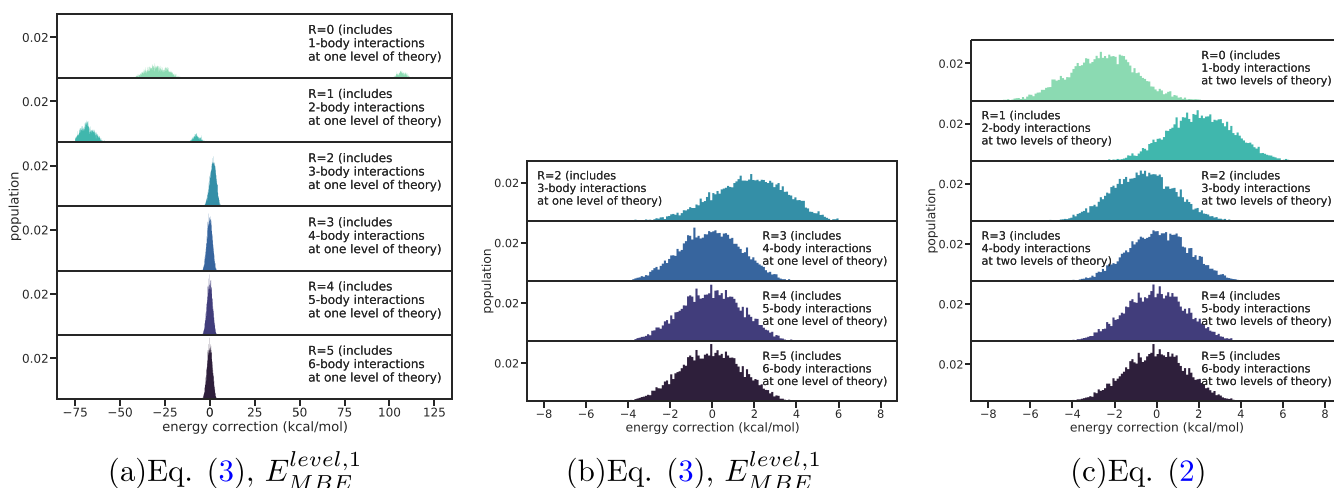
**Figure 6.** Term-by-term convergence of eq 3 for all rank- $r$  objects, as gauged by eq 14. Specifically, the average value of  $\Delta E_{\text{MBE},R}^{\text{level},1}$ , over all structures found in the AIMD trajectory in Table I, is shown here along with standard deviations shown using vertical lines. Clearly, not only is the convergence better with eq 2 (shown in Figure 5), critical many-body contributions are also captured efficiently using eq 2.

presumably between multiple graphical representations. It is clear from the scale of the vertical axes in Figure 4 that only a very small percentage of the trajectory steps show symptoms of changes in graphical representation, and these capture two different kinds of physical situations: (i) changes in the hydrogen bond network (for example, the change between Figure 1a,b) and (ii) changes in proton charge center (the change between Figure 1b,c). Additionally, the formation of “special partner” edges may statistically influence a proton hop. Failure to properly address these events within an adaptive graphical framework causes nonconservative dynamics. Furthermore, the effect on forces due to these hops, as captured by eq 12 in Table I, appears to be acceptable because, as seen from Figure 4, the percentage of topology change is low during the trajectories. Thus, the energy conservation criterion in Table I provides a more rigorous estimate of the perturbation to the dynamics due to graphical hops.

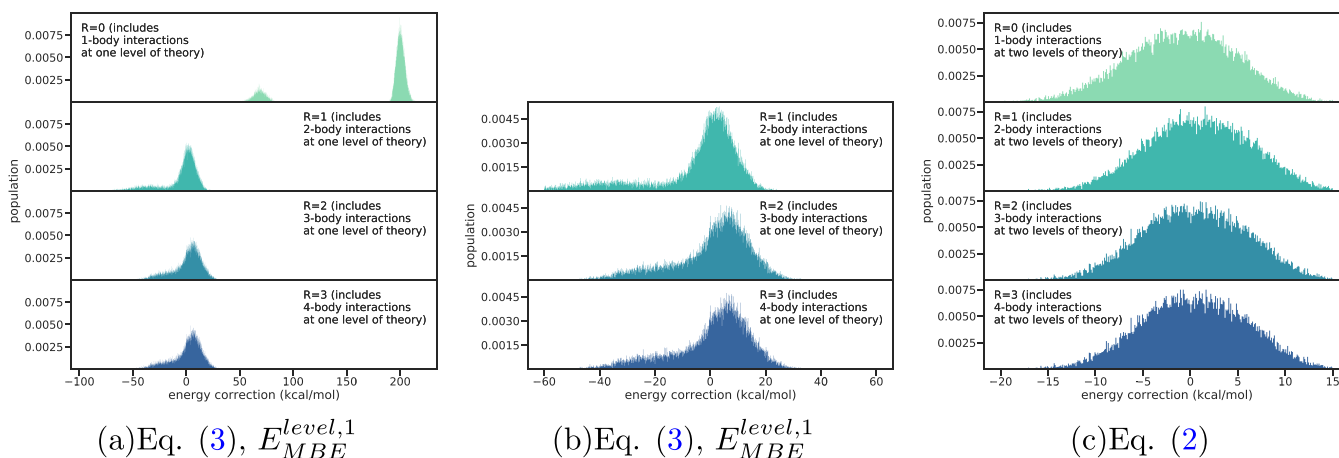
Given these energetic hops encountered using these adaptive cutoffs, as seen in Figure 4, one may be inclined to utilize a complete graph to describe the system. A complete graph is one where given a set of nodes, all pairs of nodes are connected through edges. As one can easily see, for a complete graph, the number of simplexes grows exponentially with system size. (See, for example, Table II in the next section, where we present the number of molecular fragments created from simplexes for both the solvated Zundel and the protonated 21-water cluster.) As a result, in Appendix A, we discuss our approach to obtain a family

of graphs that capture the relevant molecular interactions. However, as is also clear from Table II in the paper, even in these cases, the number of fragments grow dramatically for the protonated 21-water cluster. Second, during dynamics, the description of nodes in a graph may also change when protons migrate (as illustrated in Figure 1b,c), which, as shown in Figure 4, produces larger energetic hops than the edge deformations. In these cases, a complete graph is considered, which will still require multiple graphs with different nodal definitions to correctly represent the dynamics. The goal of the current paper is to ask how to construct a numerically conservative scheme, which both accounts for proton transfer events and avoids the admittedly intractable, exponential scaling number of edges and higher-order simplexes available in a complete graph. In the next section, energy convergence with edge cutoff and maximum rank, as well as the use of eq 5 for weighted-graph-based dynamics, are considered to address these goals.

**IV.II. Rapid Convergence of the Expression in Equation 4 as a Function of  $R$ , Relative to Many-Body Expansions, and the Associated Improvement in Conservative Properties of AIMD Trajectories.** We next expand the level of graphical approximation by increasing  $R$  and edge length for every geometry found for the trajectories in Table I. This allows us to gauge the convergence properties of eq 4, as a function of  $R$ , following which we construct trajectories with progressively increasing values for  $R$  and edge length cutoff parameters to study the effect on energy conservation. The



**Figure 7.** Term-by-term convergence of eq 2 for  $H_{13}O_6^+$  (c) is compared with that from standard many-body expansion (a, b). In (b), the  $R=0$  and 1 terms are eliminated for clarity. The graphical representation was constructed using a maximum edge length of 7 Å.



**Figure 8.** Term-by-term convergence of eq 2 for  $H_{43}O_{21}^+$  (c) is compared with that from standard many-body expansion (a, b). In (b), the  $R=0$  term is eliminated for clarity. The graphical representation was constructed using a maximum edge length of 4 Å.

convergence properties of eq 4 are presented in Figures 5–8. Specifically, in Figure 5, we consider the cumulative contribution from all rank- $R$  objects toward the expression in eq 4, given by the quantity

$$\Delta\Delta MBE_R \equiv \Delta MBE_R^{level,1;level,0} - \Delta MBE_{R-1}^{level,1;level,0} \quad (13)$$

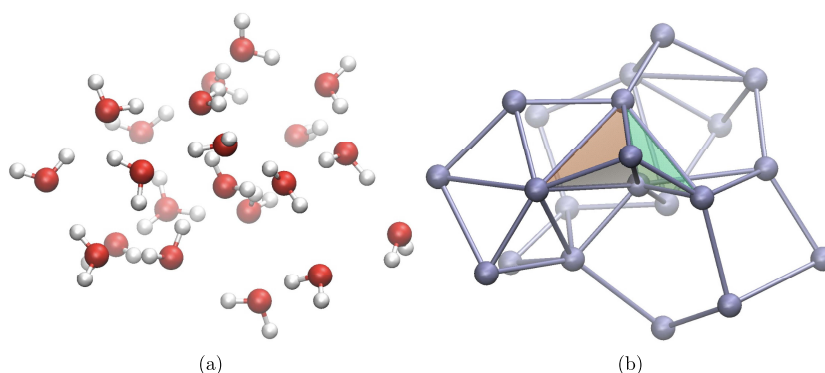
This measure informs us of the net contribution of a specific rank many-body correction to the full-system lower level of theory energy, as in eq 2, and allows careful analysis of the convergence properties of eq 4. As seen in Figure 5, the contributions from each rank progressively reduce in absolute contribution, which converges to less than 1 kcal/mol at  $R = 3$ . In fact, even with only two-body interactions, the average error is acceptable for the larger protonated 21-water cluster and within 5 kcal/mol for the solvated Zundel. This aspect is in sharp contrast to the situation in Figure 6, where a similar quantity to that in eq 13 is computed, but for eq 3, the regular many-body result for a given level (level,1, here) of theory. That is

$$\Delta E_{MBE,R}^{level,1} \equiv E_{MBE,R}^{level,1} - E_{MBE,R-1}^{level,1} \quad (14)$$

where  $E_{MBE,R}^{level,1}$  on the right side is the expression in eq 3 at level  $R$ . In contrast to eq 13, this measure provides the additional many-body contributions that arises from increasing-order many-body

terms. Very clearly, the many-body contributions from level,1 are much larger in magnitude, as seen by a comparison of Figures 5 and 6, resulting in slower convergence. Figure 6 is also consistent with the studies in ref 65, where it is noted that three-body contributions are essential in the treatment of water. While this is true for the many-body contributions in eq 3, as evaluated by eq 14, the ONIOM-based idea that is graph-theoretically implemented here converges much faster since (a) many-body contributions are already included through a base, level,0 calculation and (b) additional many-body contributions are only needed through the difference in eq 4. This aspect is consistent with the numerical demonstrations on a set of equilibrium structures in ref 43 and is also consistent with discussions in refs 58 and 63. However, here, we considered a large library of structures that are obtained from AIMD trajectories, as noted in Table I. For example, Figures 5 and 6 include the behavior from a total of 9000 structures for  $(H_2O)_6H^+$  and 28 000 structures for  $(H_2O)_{21}H^+$ . Thus, while Figure 5 indicates that four-body post-Hartree–Fock contributions may be necessary to recover accurate CCSD energies to sub-kcal/mol accuracy, these calculations are already acceptable at  $R=1$  and 2.

The effect of larger-rank objects is also seen from Figures 7c and 8c, where the entire distribution of energies corresponding to eq 4 is provided for increasing levels of  $R$ . The horizontal axis



**Figure 9.** Higher-order interactions captured as part of the  $\mathcal{R} = 3$  calculations in Table II: atomistic representation (a) and graphical representation (b). The tetrahedral simplexes (orange and green) capture four-body interactions, whereas the colored faces capture three-body interactions. These simplexes are determined on-the-fly during the dynamics trajectories discussed in Table II.

**Table III.** Energy Conservation Properties for Dynamical Simulations (Microcanonical) Performed Using Equation 5

system	cutoff (Å)	$\mathcal{R}$	$N_f^a$	$\Delta t$ (fs)	simulation time (ps)	$T_{\text{Ave}}$ (K) <sup>b</sup>	$\Delta\mathcal{H}$ (kcal/mol)	$\mathcal{H}_{\text{Drift}}^c$ (kcal/mol)	$\langle F \rangle^d$ ( $10^{-6}$ kcal/Å)
$\text{H}_{13}\text{O}_6^+$	$1.1 \times D_i$	1	11.0	0.2	6.00	$168.12 \pm 25.3$	0.036	0.041	6.7
$\text{H}_{43}\text{O}_{21}^+$	$1.1 \times D_i$	1	47.1	0.4	7.00	$272.38 \pm 21.4$	0.695	-2.117	14.9

<sup>a</sup>Average number of fragments. <sup>b</sup>By use of the equipartition theorem,  $(3/2)(N - 1)kT$ , we convert the kinetic energy into average and RMS temperatures. Here, the temperature is a measure of the available energy to sample the conformational space. The initial kinetic energies were randomly distributed along the nuclear degrees of freedom. <sup>c</sup>The drift in the Hamiltonian (total energy) is computed as the difference between the average for the first 100 fs and last 100 fs. In kcal/mol. <sup>d</sup>Equation 12 over the full trajectory.

**Table IV.** Deviation of the Total Force for the Dynamical Simulations

system	cutoff	$\mathcal{R}$	trajectory Source <sup>a</sup>	$\langle F \rangle_{\text{hops}}$ ( $10^{-6}$ kcal/Å)	$\langle F \rangle_{\text{non-hops}}$ ( $10^{-6}$ kcal/Å)	$\langle F \rangle$ ( $10^{-6}$ kcal/Å)
$\text{H}_{13}\text{O}_6^+$	$1.1 \times D_i$	1	Table III	$6.6 \pm 2.8$	$6.7 \pm 2.8$	$6.7 \pm 2.8$
$\text{H}_{43}\text{O}_{21}^+$	$1.1 \times D_i$	1	Table III	$14.8 \pm 6.4$	$15.0 \pm 6.3$	$14.9 \pm 6.4$
$\text{H}_{13}\text{O}_6^+$	$1.1 \times D_i$	1	Table II	$6.6 \pm 2.8$	$6.7 \pm 2.9$	$6.7 \pm 2.8$
$\text{H}_{13}\text{O}_6^+$	6.0 Å	1	Table II	$13.4 \pm 5.8$	$13.8 \pm 5.9$	$13.8 \pm 5.9$
$\text{H}_{13}\text{O}_6^+$	6.0 Å	2	Table II	$18.1 \pm 7.9$	$18.0 \pm 7.8$	$18.0 \pm 7.8$
$\text{H}_{13}\text{O}_6^+$	6.0 Å	3	Table II	$9.4 \pm 5.8$	$9.4 \pm 5.6$	$9.4 \pm 5.6$
$\text{H}_{13}\text{O}_6^+$	6.0 Å	4	Table II	$6.9 \pm 3.1$	$7.1 \pm 3.2$	$7.1 \pm 3.2$
$\text{H}_{43}\text{O}_{21}^+$	$1.1 \times D_i$	1	Table II	$14.7 \pm 6.3$	$14.9 \pm 6.3$	$14.8 \pm 6.3$
$\text{H}_{43}\text{O}_{21}^+$	4.0 Å	3	Table II	$14.6 \pm 6.2$	$14.5 \pm 6.1$	$14.5 \pm 6.1$
$\text{H}_{43}\text{O}_{21}^+$	4.5 Å	4	Table II	$18.5 \pm 8.1$	$18.3 \pm 8.0$	$18.4 \pm 8.0$

<sup>a</sup>Corresponding table number where the trajectory is presented.

for each case is shifted such that the average value obtained from the highest rank for that system, that is, the ensemble average of the entire distribution, is set to zero. In both Figures 7c and 8c, the distribution converges at  $\mathcal{R}=3$ , as is consistent with Figure 5. In Figures 7a,b and 8a,b, we present a similar analysis of the convergence properties when the standard many-body expansion is used, as depicted in eq 3. As shown in Figures 7a,b and 8a,b, the convergence properties from eq 3 are slow as compared to that arising from eq 4 and this is consistent with the results in ref 43 and also in Figures 5 and 6. This observation is also consistent with previous discussions in refs 58 and 63 of a quicker and more reliable convergence for “composite methods” as compared to a standard many-body expansion.<sup>63,65,131,132</sup> Equation 2 appears to provide a convergent trend around  $\mathcal{R} = 3$ .

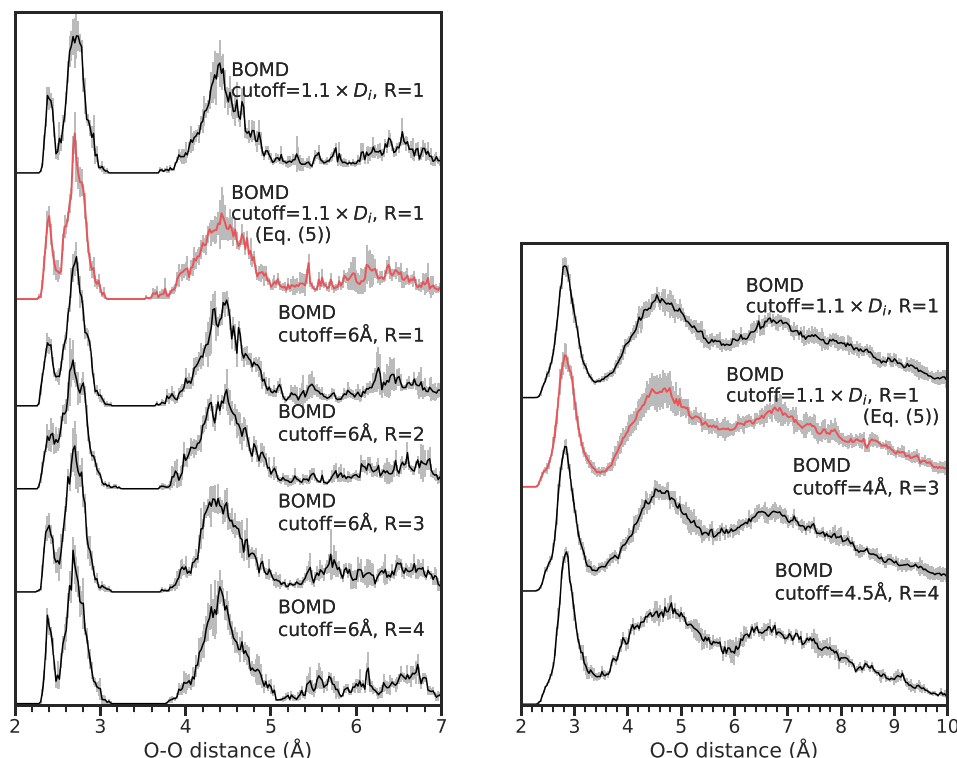
In Table II, the dynamics calculations for larger ranks include the appropriate gradients appearing from higher-rank calculations, and it is clear that including the gradients from higher ranks has a critical impact on the conservative properties of the trajectories. An example of the kind of interactions captured in these simulations is shown in Figure 9. The energy conservation

obtained from these higher-rank studies is much improved as seen by comparison of Tables I and II. However, using eq 5, as we see in Table III, leads to far better energy conservation as compared to Tables I and II. Thus, the weighted-graph expression from eq 5 provides good energy conservation with a negligible increase in the number of fragments, as can be seen from Tables I–III.

To further probe the effect of the time dependence of the potential introduced in eqs 8–11, we present in Table IV the average and standard deviation of  $\langle F \rangle$  for the section of the trajectory where the dynamics is determined by a single graph ( $\langle F \rangle_{\text{non-hops}}$  in eq 16) and the section of the trajectory that involves a transition among multiple graphs, with weights given in eqs 8–11 ( $\langle F \rangle_{\text{hops}}$  in eq 15). The latter is obtained using a 10-step window surrounding each step that shows a change in graphical representation (or a “hotspot” step),  $t_w$  according to

$$\langle F \rangle_{\text{hops}} = \sum_{\alpha} \frac{1}{10 \times \Delta t} \int_{t_{\alpha}-5 \times \Delta t}^{t_{\alpha}+5 \times \Delta t} dt \left\| \sum_{i=1}^{N_{\text{Atoms}}} F_i(t) \right\|_2 \quad (15)$$





**Figure 10.** Radial distribution functions for oxygen–oxygen distances computed from the  $(\text{H}_2\text{O})_6\text{H}^+$  (a) and  $(\text{H}_2\text{O})_{21}\text{H}^+$  (b) trajectories. See the text for details.

where  $\{t_\alpha\}$  are assumed to be the points where a proton transfer or hydrogen bond network change occurs. The time-dependent weight, described in eqs 8–11, smoothens the energy and forces from the step of topology change to five steps after. Similarly, the force conservation is evaluated for the duration of trajectory devoid of hops using

$$\begin{aligned} \langle F \rangle_{\text{non-hops}} &= \frac{1}{\mathcal{T} - N_{\text{H}} \times 10 \times \Delta t} \left[ \int_0^{\mathcal{T}} dt \left\| \sum_{i=1}^{N_{\text{Atoms}}} \vec{F}_i(t) \right\|_2 \right. \\ &\quad \left. - \sum_{\alpha} \int_{t_{\alpha}-5 \times \Delta t}^{t_{\alpha}+5 \times \Delta t} dt \left\| \sum_{i=1}^{N_{\text{Atoms}}} \vec{F}_i(t) \right\|_2 \right] \\ &= \frac{1}{\mathcal{T} - N_{\text{H}} \times 10 \times \Delta t} [\mathcal{T} \times \langle F \rangle - 10 \times \Delta t \times \langle F \rangle_{\text{hops}}] \end{aligned} \quad (16)$$

or

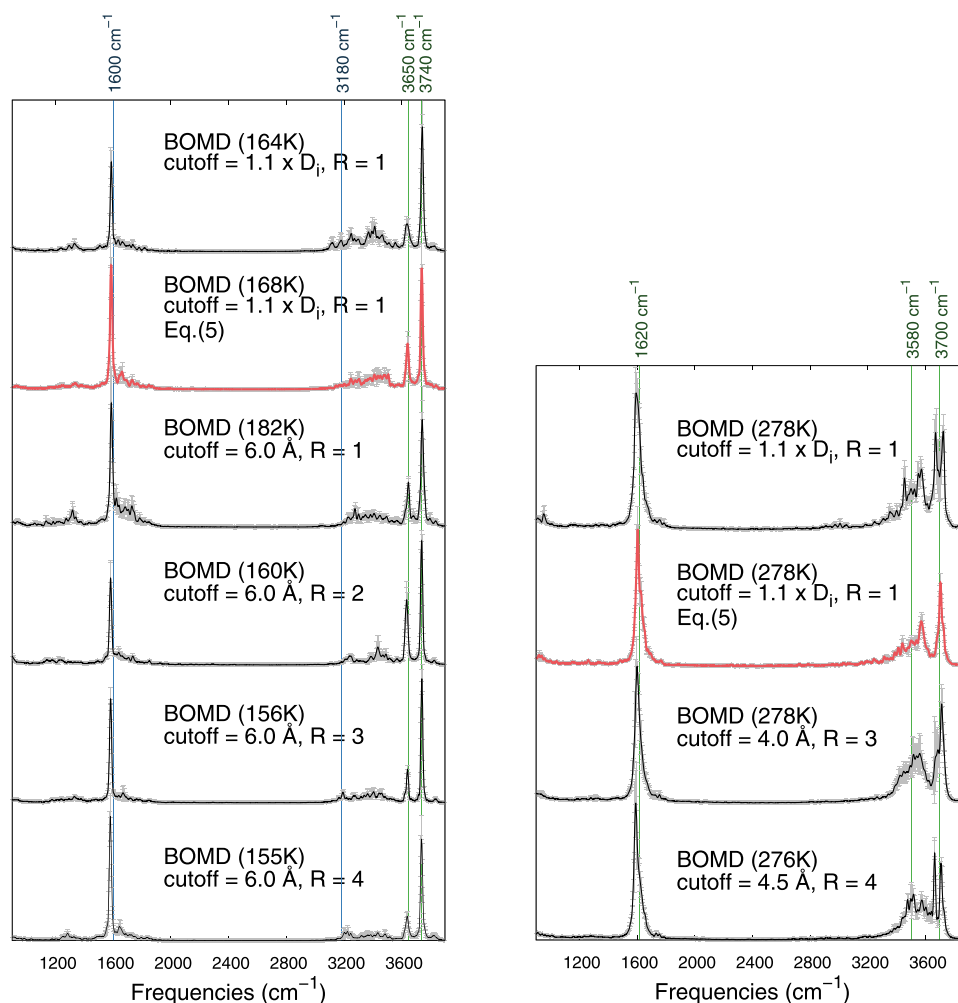
$$\begin{aligned} \langle F \rangle &= \frac{1}{\mathcal{T}} [(\mathcal{T} - N_{\text{H}} \times 10 \times \Delta t) \times \langle F \rangle_{\text{non-hops}} \\ &\quad + 10 \times \Delta t \times \langle F \rangle_{\text{hops}}] \end{aligned} \quad (17)$$

where  $N_{\text{H}}$  is the number of hops. As is clear from Table IV, both  $\langle F \rangle_{\text{non-hops}}$  and  $\langle F \rangle_{\text{hops}}$  are small in magnitude. However, it is also clear from the equations above that these are both positive definite and  $\langle F \rangle$  is a weighted average of  $\langle F \rangle_{\text{non-hops}}$  and  $\langle F \rangle_{\text{hops}}$  and hence bounded by these quantities. Furthermore, for all cases, weighted and nonweighted trajectories, where the former contains time-dependent potentials and the latter does not,  $\langle F \rangle_{\text{non-hops}}$  and  $\langle F \rangle_{\text{hops}}$  are close in value and well within the acceptable range. We have also analyzed the time evolution of  $\langle F \rangle$ , as well as its distribution, and found that there is no adverse numerical effect associated with this time dependence.

#### IV.III. Comparison of Dynamical Properties from all AIMD Simulations.

In this section, we inspect dynamical properties obtained from the simulations summarized in Tables I, II, and III to ascertain the accuracy of results obtained from these simulations. These AIMD trajectories were calculated using the Born–Oppenheimer molecular dynamics<sup>1</sup> (BOMD) with gradients in Appendix B. Our baseline for comparison will be the larger-rank AIMD simulations provided in Table II. Results from all other simulations will be compared with these to check accuracy. Two principal dynamical properties are examined: (a) structural features arising from dynamics, such as the oxygen–oxygen radial distribution functions (RDFs), which convey the overall solvation structure sampled during dynamics; and (b) the vibrational density of states obtained from the Fourier transform of the velocity autocorrelation (FT-VAC) function, which is the Fourier space analogue of a trajectory and captures the extent to which anharmonic effects have a role on the vibrational properties of the system.<sup>12,18–22</sup>

Considering the significance of these systems, as discussed on top of Section IV, we also undertake a detailed analysis of the associated vibrational properties. As noted in the previous section, and in Tables I and II, the lower-rank trajectories offered less conservative dynamics due to energetic hops, which, although rare, appear to provide significant perturbations to the overall trajectories, as seen in Figure 4. However, as we will find in the discussions below, these lower-rank trajectories still result in dynamical properties that are in sufficient agreement with the higher-rank trajectories. Additionally, based on Table III, and the discussion in the previous section, it is clear that the weighted-graph scheme from eq 5 provides conservative dynamics, and in the following discussion, it is found that this scheme produces dynamical properties consistent with the higher-rank AIMD simulations. Thus, in summary, the



**Figure 11.** Vibrational density of states calculated from dynamical trajectories using eq 20. The results from the dynamics trajectory computed using eq 5 are shown in red. Critical vibrational frequencies corresponding to experimental features<sup>79,93,94,117,127,134,135</sup> are indicated with vertical lines and the frequencies listed on top of the figure. In (a), blue features are from ref 134 and green features are from ref 135 in (b), the green features are from refs 79 and 127.

weighted-graph approach may be considered as the method of choice for future AIMD trajectories.

**IV.III.I. Structural Distribution.** The oxygen–oxygen radial distribution functions (RDFs) for the trajectories in Tables I, II, and III are presented in Figure 10. The distribution functions and associated standard deviations were computed as follows. First, a slice of length 2 ps (0–2 ps) from each trajectory was used to obtain an O–O RDF. Then, the time slice was incremented by 25 fs to create an O–O RDF for the trajectory slice 0.025–2.025 ps. Following this, all such moving time-windowed RDFs were averaged to compute the overall RDF; standard deviations from such time-windowed RDFs were also computed, and the results are shown in Figure 10. Such a statistical treatment allows for reduction in variance in the O–O RDF. This procedure has even greater utility in the next subsection, where we compute the Fourier transform of the velocity autocorrelation function where such a time-windowing procedure is known<sup>23</sup> to reduce the variance and reduce “leakage” in frequency space, as will be discussed later.

In Figure 10, the peaks at approximately 2.5 Å are indicative of strongly hydrogen-bonded Zundel-like substructures, while the peak at around 2.7 Å indicates the classical Eigen-like cationic formation in trajectories. While the solvated Zundel shows the propensity to display both Zundel and Eigen substructure

characteristics, for all trajectories, the protonated 21-water cluster shows a broad distribution in the range of 2.4–3.2 Å, which is indicative of significant contributions from the Zundel-like and Eigen-like counterparts. The second solvation shell surrounding the oxygenic center of excess positive charge can be seen at roughly 4.5 Å for both clusters. The spread of this second solvation shell in both cases is roughly 1 Å that encodes within it the process of Grotthuss proton shuttle<sup>133</sup> in these clusters. This second solvation shell captures one to three interactions between the most-protonated oxygen center, its nearest “Zundel-like-partner” involved in the Grotthuss shuttle, and a third water molecule that is hydrogen-bonded to the “Zundel-like-partner”. The broad distribution (with width 1 Å) is indicative of the one to three interactions as these adapt to the proton delocalization between the most-protonated oxygen center and the nearest “Zundel-like-partner”. It is also clear that while such secondary participation in proton transfer must necessarily be truncated at the second solvation shell for the solvated Zundel, as can also be seen from the drop in RDF at 5 Å, this is not the case for the larger cluster. For the protonated 21-water cluster, we see what essentially looks like a solvation continuum following the 5.5 Å shoulder, or dip, that appears to separate the second solvation shell from the remaining solvation data, but this separation is certainly not as abrupt as that for the

solvated Zundel case. Hence, additional solvation shell participation in the Grotthuss shuttle process is certainly captured in all calculations in Figure 10b. However, at this stage, the finite size of the cluster limits the analysis in terms of probing the broader participation of water molecules in the proton shuttle process beyond the second solvation shell.

The overall structural characteristics demonstrated by these RDFs remain consistent, independent of the choice of graphical parameters within the standard deviations found for the O–O distances. This is despite the fact that the energy conservation properties are significantly dependent upon the rank and edge length cutoff and are perhaps related to the fact that the population of hops is extremely small, as already noted in Figure 4. Meanwhile, the use of eq 5 to compute smoothened, weighted-graph-theoretically obtained dynamics trajectories provided RDF populations similar to those found from the trajectories, which use a higher-rank, single adaptive graphical representation. The use of weighted graphs affords costs equivalent to the lowest-rank trajectories, but with the energy conservation and structural characteristics equivalent to the largest rank treatments. In other applications, it may, of course, be necessary for the weighted-graph representation itself to include some higher-rank terms, and the theory presented here certainly allows for such computations to be adaptively performed.

**IV.III.II. Time-Correlation Functions.** Dynamics trajectories are compared using the vibrational density of states (VDOS) to determine the influence of the choice of graphical treatments upon the quality of the sampled potential energy surface. The dynamically averaged vibrational density of states were computed using the Fourier transform of the nuclear velocity autocorrelation function (FT-VAC). The details on how these computations are performed are given in Appendix C. The resultant spectra provide a broad metric for the effectiveness of the adaptive graphical representation of the potential energy surface sampled to construct the AIMD trajectories. The VDOS spectra for AIMD with increasing maximum rank,  $\mathcal{R}$ , are presented for the solvated Zundel and the protonated 21-water cluster in Figure 11. A few critical experimental features are indicated by colored lines in Figure 11, with the frequencies labeled above the figure. Before we undertake a detailed comparison with the experimental results, it is critical to note that the data in Figure 11 represents the velocity distribution in frequency space. Hence, it does not capture the dipole intensities that are needed to carry out a comparison with the experimental vibrational spectral results. However, Figure 11 does allow us to probe the agreement across all trajectories computed here and also evaluate the Fourier space activity, as detected by nuclear motion, in a certain spectral range.

The peaks marked on top of Figure 11 were selected from a set of gas-tagged spectroscopy experimental results.<sup>79,94,127,134</sup> The solvated Zundel peaks were selected from ref 134, which used a laser photodissociation in a time-of-flight mass photofragmentation spectrometer, and ref 135 offered more refined spectral peaks for the higher-frequency peaks using photodissociation spectroscopy of clusters produced by pulse-discharge supersonic expansion with mass selection. The protonated 21-water cluster features chosen from ref 127, which offers an expanded IR window as compared to previous studies,<sup>79,117</sup> and assignment were further clarified by ref 94. For the solvated Zundel, four critical peaks are highlighted: the peak at 1600  $\text{cm}^{-1}$  is the intramolecular HOH bend modes, the peak at 3180  $\text{cm}^{-1}$  captures the OH stretch of the Zundel waters with the first

solvation shell, and the symmetric bending and asymmetric stretching modes are measured to be at 3650 and 3740  $\text{cm}^{-1}$ , respectively.<sup>134–136</sup> Likewise, three features are labeled in Figure 11 for the protonated 21-water cluster: 1620  $\text{cm}^{-1}$ , which is the HOH intramolecular bends of dangling water molecules, 3700  $\text{cm}^{-1}$  peak capturing the free OH stretch, and a peak at 3580  $\text{cm}^{-1}$ , which is the maximum of the standard acceptor–acceptor–donor and donor–donor–acceptor stretch region for the internal neutral water species.<sup>79,117,127</sup>

As observed for the RDFs in the previous section, Figure 11a demonstrates that the lower-rank trajectories produced vibrational density of states consistent with the large-rank trajectories. As noted by the structural properties, the smoothened dynamics afforded by the weighted-graph trajectory reproduced the vibrational features in good agreement with the higher-rank trajectories. In Figure 11b, we observe trajectories of three different graphical representations of protonated 21-water cluster, as shown in Table II, along with the weighted-graph trajectories using the adaptive cutoff from Tables I and III, respectively. These trajectories produced spectra with similar features, although the intensities vary due to a difference in simulation time; this variance in intensities is especially true for the OH stretches in the 3500–3900  $\text{cm}^{-1}$  region, which are significantly impacted by the conformational sampling by the given trajectories.<sup>12,17,93,137</sup> Consistent with the previous findings in Figure 10b, although the energy conservation significantly improves with increasing rank terms, the spectral results for these protonated 21-water clusters exhibit the same features independent of graphical representation due to the low probability of such hops, as seen in Figure 4. The weighted-graph trajectory, as before, offered both lower cost and energy-conserving approach while matching the vibrational density of states of the higher-rank trajectory.

## V. CONCLUSIONS

We introduce a weighted-graph-theoretic formalism to obtain smooth, post-Hartree–Fock AIMD trajectories for fluxional systems at DFT cost. The approach is based on ONIOM, and here, a full system is divided into individual units that are treated as nodes in a graph. These nodes are then connected to form edges, faces, and higher-order simplexes as a result of the graph-theoretic decomposition of the molecular system. The energy and gradients obtained from the molecular fragments associated with these simplexes are then computed at two levels of theory, as in ONIOM, and combined with the full-system energy (and gradients) obtained at a lower level of theory (DFT here) to obtain a graph-theoretic expression that has close connections to several molecular fragmentations as well as many-body approximations.

However, for fluxional systems, such as protonated water clusters treated here, the graph-theoretic decomposition changes during dynamics due to proton hops, formation, and cleavage of hydrogen bonds and due to the change in hydrogen-bonding environments that accompany such a dynamical evolution. To address this, we have introduced here a weighted-graph-theoretic method where the graph-theoretic depiction of molecular structure smoothly interpolates inside a subspace of graph representations that are each encountered inside a time window during AIMD. The resultant weighted graphs smoothly interpolate and produce a conservative dynamics method that is used to compute AIMD trajectories for protonated water clusters. The approach is tested by comparing with trajectories that are computed by using high-



rank simplexes that come close to including all possible many-body interactions in the smaller water cluster in a full-rank AIMD simulation. (This is impossible to do for the larger water cluster treated here.) The resultant structural and time-correlation functions are in extremely good agreement between the weighted-graph approach and the full-rank AIMD trajectories. We further demonstrate the rapid convergence of our graph-theoretic formalism with increasing many-body rank. We numerically demonstrate that, due to many-body approximations already being present in the full-system calculation, and additional contributions being included based on the difference in energy for any given rank, as allowed by the ONIOM description, the energy converges rapidly for a wide range of thousands of stable and unstable nuclear configurations found as part of AIMD trajectories.

## ■ APPENDIX A: GRAPHICAL REPRESENTATION OF MOLECULAR SYSTEMS YIELDS MANY-BODY CONTRIBUTION UP TO ARBITRARY RANKS IN EQUATION 2

To apply eq 2 to the protonated water clusters, these water clusters are partitioned into a field of vertices or nodes. Once these nodal partitions are defined, then as mentioned in the paper, there are two specific control parameters in eq 3 that define the construction and evaluation of the graph, which represents the many-body interactions inherent to the system. These graphical elements, or simplexes, yield the necessary set of electronic structural calculations used to evaluate the extrapolated energy. First, the local family of edges, which represents two-body interactions, are assembled based upon an edge length cutoff. Next, the maximum order of many-body interactions considered is truncated by the selection of “ $\mathcal{R}$ ” in eq 3.

The protonated water clusters are partitioned by designating each water molecule as a node within the graphical representation. As our method is applied to dynamic, protonated systems, each “water” node is defined as an oxygen with all hydrogens within 1.4 Å included within the node. This definition allows for the graphical representation to capture and adapt to proton migration events, and in effect, some of these “water” nodes are hydronium species, where the center for excess positive charge changes throughout the trajectory (as may be clear from Figure 1b,c). For each geometry, the atomic coordinate of the oxygen defines the coordinates of the node, which represents the water/hydronium. Two different flavors of edge cutoffs are considered for the generation of edges for the graphical representation. Principally, we consider an adaptive envelope scheme,<sup>37,43</sup> where a cutoff based on the minimum node–node distance ( $D_i$ ) for a chosen node (index  $i$ ) is used; in this scheme, edges are generated within a cutoff of  $1.1 \times D_i$  of node- $i$ . This choice of an adaptive envelope was previously performed well for water and protonated water systems with treatment involving up to two-body interactions.<sup>37,43</sup> In addition, we also consider a fixed distance cutoff, which connects nodes by edges when the mutual distance is within the designated Cartesian distance. Similar distance-based approaches can be found in a variety of fragmentation-based methods.<sup>57,59,65,87,138–143</sup>

The networks generated by either edge cutoff scheme allow the selection of local many-body terms: one-body ( $r = 0$ , nodal contributions), two-body ( $r = 1$ , edge contributions), three-body ( $r = 2$ , triangular contributions), and higher-order contributions toward eq 3. This methodology is general enough

that it can be applied to both bonded and nonbonded interactions within target molecular systems and has been shown for AIMD simulations in refs 37–41. Due to the requirement to only consider objects defined based upon the closed convex-hull simplex condition (eq A1), the adaptive envelop scheme generally allows only the inclusion of one-body and two-body terms, while a high enough fixed distance cutoff may include three-body or higher-order contributions.

We impose the requirement that all higher-order interactions treated by our method must arise from a simplex found within the graphical representation. Since simplexes are affine convex hulls,<sup>95–97</sup> this leads to the treatment of only “closed” objects in the power set defined in eq 1. This condition requires the restriction that for an  $r$ -rank simplex to be included in the set  $V_r$ , all of the constituent  $(r - 1)$ -rank simplexes would already be included in  $V_{r-1}$  of the power set shown in eq 1. Thus, the following condition must be satisfied for an  $r$ -rank object ( $\nu_r$ ) to be included in the above power set

$$\begin{aligned} \forall \nu_r \in V_r \\ \alpha_{r-1} \in \nu_r \Rightarrow \{\alpha_{r-1}\} \subset V_{r-1} \end{aligned} \quad (\text{A1})$$

What this implies is that, for all  $\nu_r$  in  $V_r$ , all of the constituent lower-rank simplexes ( $\{V_{r-1}, \dots, V_0\} \equiv \{\{\alpha_{r-1}\}, \dots, \{\alpha_0\}\}$ ) are also included in the power set shown in eq 1. For example, each face (triangle, two-simplex) is only included in  $V_2$  when its three nodes are all mutually connected by edges. This condition enforces the requirement that the higher-order terms are simplexes and gives a systematic approach for describing local  $n$ -body terms.<sup>43,44</sup> Note because of the condition imposed by eq A1, the effective local rank is determined by connectivity provided by the edge length cutoff; the parameter  $\mathcal{R}$ , in eq 4, only functions as a further cap to the extent of the many-body interactions considered.

Due to both the condition imposed by eq A1 and the automatic creation/destruction of edges dictated by distance cutoffs, the effective local ranks of the many-body interaction evolve with the system configuration. In our case, arising from the use of graphs, such optimal local ranks capture local many-body interactions and, within our scheme, may be adaptively determined through the local simplexes<sup>97</sup> surrounding any given node. In regions with significant nodal interactions the graphical representation includes higher-rank simplexes and hence higher-rank many-body interactions, whereas sparser, less interacting regions would naturally be well served by lower-rank simplexes. Of course, we could artificially limit the ranks surrounding certain nodes.

## ■ APPENDIX B: GRADIENTS FROM EQ 5 FOR AB INITIO MOLECULAR DYNAMICS

The expression for nuclear gradients of the energy from eq 4 can be found in ref 44. These gradients are modified in light of eq 5 and may be written as

$$\begin{aligned} \frac{\partial E_{\text{MBE,gt}}^{\text{ONIOM}}[\bar{\mathbf{x}}(t)]}{\partial \bar{\mathbf{x}}} &= \frac{\partial E^{\text{level},0}[\bar{\mathbf{x}}(t)]}{\partial \bar{\mathbf{x}}} + \sum_{\beta} v_{\beta}(t) \times \\ &\quad \sum_{r=0}^R (-1)^r \left\{ \sum_{\alpha \in \mathbf{V}_{r,\beta}} \left( \frac{\partial E^{\text{level},1}_{\alpha,r}[\bar{\mathbf{x}}(t)]}{\partial \bar{\mathbf{x}}_{\alpha,r}} \right) \right. \\ &\quad \left. - \frac{\partial E^{\text{level},0}_{\alpha,r}[\bar{\mathbf{x}}(t)]}{\partial \bar{\mathbf{x}}_{\alpha,r}} \right\} \left( \frac{\partial \bar{\mathbf{x}}_{\alpha,r}}{\partial \bar{\mathbf{x}}} \right) \\ &\quad \left[ \sum_{m=r}^R (-1)^m p_{\alpha,\beta}^{r,m} \right] \end{aligned} \quad (\text{B1})$$

This expression allows for a gradual transition between gradients of the time-dependent graphical representations,  $\{\mathcal{G}_{\bar{\mathbf{x}}(t)}^{\beta}\}$ . If bonds were broken by the partitioning of the system into nodes, link atoms<sup>144,145</sup> may have been included within the nodal definitions but would not be part of the system coordinate variable, as treated by ONIOM.<sup>46</sup> This would imply that the nuclear coordinates for a subsystem  $\alpha$ ,  $\bar{\mathbf{x}}_{\alpha,r}$  may not be entirely a subset of the system coordinate variable,  $\bar{\mathbf{x}}(t)$ . A Jacobian,  $\left[\frac{\partial \bar{\mathbf{x}}_{\alpha,r}}{\partial \bar{\mathbf{x}}}\right]$ , can be employed to transform the  $r$ -rank many-body gradients back to the full-system gradients.<sup>37,44,146</sup> These gradients allow the direct use of the Born–Oppenheimer molecular dynamics<sup>1</sup> with smooth gradient transitions between the graphical representation when using eq 5. Such molecular dynamics trajectories, using fixed graphical representations using eq 4, have been previously demonstrated in refs 37–41.

### ■ APPENDIX C: VIBRATIONAL DENSITY OF STATES IN FIGURE 11: WINDOWED FOURIER TRANSFORMS AND FREQUENCY SCALING TO OVERCOME THE NEGLECT OF CRITICAL QUANTUM NUCLEAR EFFECTS

The dynamically averaged vibrational density of states were computed by using the Fourier transform of the nuclear velocity autocorrelation function (FT-VAC), defined as

$$\begin{aligned} I_V(\omega) &= \lim_{T \rightarrow \infty} \int_{t=0}^{t=T} dt \exp(-i\omega t) \langle \mathbf{V}(0) \cdot \mathbf{V}(t) \rangle \\ &= \lim_{T \rightarrow \infty} \sum_{i=1}^{N_{\text{Atoms}}} \sum_{j=1}^3 \int_{t=0}^{t=T} dt \exp(-i\omega t) \\ &\quad \int_{t'=0}^{t'=T} dt' V_{i,j}(t') \times V_{i,j}(t' + t) \\ &= \lim_{T \rightarrow \infty} \sum_{i=1}^{N_{\text{Atoms}}} \sum_{j=1}^3 \left| \int_{t=0}^{t=T} dt \exp(-i\omega t) V_{i,j}(t) \right|^2 \end{aligned} \quad (\text{C1})$$

where the term,  $\langle \mathbf{V}(0) \cdot \mathbf{V}(t) \rangle$ , in the first equation, indicates the ensemble average and is equal to the  $t'$ -integral in the second equation under the ergodicity condition. To reduce the Fourier transform of the velocity autocorrelation function to the summation of the power spectral densities of the individual nuclear velocities in the third equation, we have used the Wiener–Khinchin theorem.<sup>23</sup> The Wiener–Khinchin theorem<sup>23</sup> is a special case of the correlation theorem applied to autocorrelation functions. Equation C1 has been used on each

trajectory, and we have presented the respective  $I_V(\omega)$  for each trajectory in Figure 11.

In computing the spectra in Figure 11, we have used the following steps. As in the case of the RDF plots in the previous section, vibrational density of states in Figure 11 (eq C1) are computed for 2 ps moving time window slices, with 25 fs increments. The vibrational density of states for these windows are averaged and shown in bold in Figure 11. The standard deviations are computed from these moving time windows and are presented as gray traces. The idea of introducing a 2 ps moving window is related to two aspects in the theory of computing accurate Fourier transforms that (i) reduce “leakage”<sup>23</sup> and (ii) reduce variance in the Fourier transform. Specifically, whenever a finite sample of frequencies is constructed, there is always a leakage (or smudging phenomenon) where information from one frequency leaks to the neighboring frequency regions of the spectrum. Such leakage between frequencies is due to the projection of the velocity autocorrelation function onto an incomplete trigonometric basis spanning the finite length of the trajectory.<sup>23,147</sup> The extent of the leakage domain is generally quite substantial and falls off as  $1/\omega^2$ . Data windowing reduces the leakage of information in the frequency domain, and the recommended form of data windowing is one that includes overlapping segments. The second reason for computing the Fourier transforms of these segments and subsequent averaging of these is to reduce variance in the Fourier transform, as explained in ref 23. Finally, the resulting frequencies are scaled to match the highest frequency peak to the respective experimental peak within the considered range.

There is a significant blue shift in the frequencies of the vibrational peaks of both the solvated Zundel and the protonated 21-water cluster, calculated by eq C1, as compared to the annotated experimental peaks.<sup>79,94,135,136</sup> Thus, to properly gauge the nature of the vibrational density of states, the frequencies are scaled for the solvated Zundel to match the 3740  $\text{cm}^{-1}$  peak, and for the protonated 21-water cluster, it is scaled to match the free OH stretch (3700  $\text{cm}^{-1}$ ), which resulted in a spectral fingerprint closer to the experimental peaks. This scaling is due to the underlying quantum nuclear effects, which are especially significant in hydrogen-bonded systems such as water.<sup>148–154</sup> While classical AIMD allows the exploration of the fully anharmonic potential energy surface available to all of the nuclear degrees of freedom, it remains restricted to the classical nuclear description and does not implicitly include zero point effects and the discrete nature of nuclear eigenstates. Hence, the frequencies obtained from classical AIMD still neglect critical quantum nuclear effects, which can lead, in many cases, to significant shifts in the frequencies obtained from the trajectory velocities. Thus, to match the experimental spectral features, a frequency scaling factor of 0.941 was applied to solvated Zundel spectra and a scaling factor of 0.939 was applied to the protonated 21-water cluster. The peaks for the OH stretch of the Zundel waters (3180  $\text{cm}^{-1}$ ) are significantly blue-shifted, although this shift is likely due to the neglect of the non-negligible quantum nuclear effect for the mode.<sup>148,149,151,153</sup>

### ■ AUTHOR INFORMATION

#### Corresponding Author

Srinivasan S. Iyengar – Department of Chemistry and Department of Physics, Indiana University, Bloomington, Indiana 47405, United States; [orcid.org/0000-0001-6526-2907](https://orcid.org/0000-0001-6526-2907); Email: [iyengar@indiana.edu](mailto:iyengar@indiana.edu)

## Authors

Juncheng Harry Zhang – Department of Chemistry and Department of Physics, Indiana University, Bloomington, Indiana 47405, United States

Timothy C. Ricard – Department of Chemistry and Department of Physics, Indiana University, Bloomington, Indiana 47405, United States

Cody Haycraft – Department of Chemistry and Department of Physics, Indiana University, Bloomington, Indiana 47405, United States

Complete contact information is available at:  
<https://pubs.acs.org/10.1021/acs.jctc.0c01287>

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This research was supported by the National Science Foundation grants CHE-1665336 and OMA-1936353 to S.S.I.

## REFERENCES

- (1) Wang, I. S. Y.; Karplus, M. Dynamics of Organic Reactions. *J. Am. Chem. Soc.* **1973**, *95*, 8160.
- (2) Leforestier, C. Classical Trajectories Using the Full *Ab Initio* Potential Energy Surface  $H^+ + CH_4 \rightarrow CH_3 + H^+$ . *J. Chem. Phys.* **1978**, *68*, 4406.
- (3) Car, R.; Parrinello, M. Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.* **1985**, *55*, 2471.
- (4) Schlegel, H. B.; Millam, J. M.; Iyengar, S. S.; Voth, G. A.; Daniels, A. D.; Scuseria, G. E.; Frisch, M. J. *Ab initio* molecular dynamics: Propagating the density matrix with Gaussian orbitals. *J. Chem. Phys.* **2001**, *114*, 9758.
- (5) Crespo-Otero, R.; Barbatti, M. Recent Advances and Perspectives on Nonadiabatic Mixed Quantum-Classical Dynamics. *Chem. Rev.* **2018**, *118*, 7026.
- (6) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562.
- (7) Iannuzzi, M.; Laio, A.; Parrinello, M. Efficient Exploration of Reactive Potential Energy Surfaces Using Car-Parrinello Molecular Dynamics. *Phys. Rev. Lett.* **2003**, *90*, No. 238302.
- (8) Phatak, P.; Sumner, I.; Iyengar, S. S. Gauging the Flexibility of the Active Site in Soybean Lipoxygenase-1 (SLO-1) Through an Atom-Centered Density Matrix Propagation (ADMP) Treatment That Facilitates the Sampling of Rare Events. *J. Phys. Chem. B* **2012**, *116*, 10145.
- (9) Phatak, P.; Venderley, J.; Debrot, J.; Li, J.; Iyengar, S. S. Active Site Dynamical Effects That Facilitate the Hydrogen Transfer Process in Soybean Lipoxygenase-1 (SLO-1): Isotope Effects. *J. Phys. Chem. B* **2015**, *119*, 9532.
- (10) Gordon, R. G. Correlation Functions for Molecular Motion. *Adv. Magn. Opt. Reson.* **1968**, *3*, 1.
- (11) McQuarrie, D. A. *Statistical Mechanics*; University Science Books: Sausalito, CA, 2000.
- (12) Iyengar, S. S.; Petersen, M. K.; Day, T. J. F.; Burnham, C. J.; Teige, V. E.; Voth, G. A. The Properties of Ion-Water Clusters. I. The Protonated 21-Water Cluster. *J. Chem. Phys.* **2005**, *123*, No. 084309.
- (13) Iyengar, S. S.; Day, T. J. F.; Voth, G. A. On the Amphiphilic Behavior of the Hydrated Proton: An *Ab Initio* Molecular Dynamics Study. *Int. J. Mass Spectrom.* **2005**, *241*, 197.
- (14) Iyengar, S. S. Dynamical Effects on Vibrational and Electronic Spectra of Hydroperoxyl Radical Water Clusters. *J. Chem. Phys.* **2005**, *123*, No. 084310.
- (15) Sumner, I.; Iyengar, S. S. Quantum Wavepacket *Ab Initio* Molecular Dynamics: An Approach for Computing Dynamically Averaged Vibrational Spectra Including Critical Nuclear Quantum Effects. *J. Phys. Chem. A* **2007**, *111*, 10313.
- (16) Li, X.; Teige, V. E.; Iyengar, S. S. Can the Four-Coordinated, Penta-Valent Oxygen in Hydroxide Water Clusters Be Detected Through Experimental Vibrational Spectroscopy. *J. Phys. Chem. A* **2007**, *111*, 4815.
- (17) Iyengar, S. S. Further Analysis of the Dynamically Averaged Vibrational Spectrum for the “magic” Protonated 21-Water Cluster. *J. Chem. Phys.* **2007**, *126*, No. 216101.
- (18) Li, X.; Moore, D. T.; Iyengar, S. S. Insights from First Principles Molecular Dynamics Studies Towards Infra-Red Multiple-Photon and Single-Photon Action Spectroscopy: Case Study of the Proton-Bound Di-Methyl Ether Dimer. *J. Chem. Phys.* **2008**, *128*, No. 184308.
- (19) Li, X.; Oomens, J.; Eyler, J. R.; Moore, D. T.; Iyengar, S. S. Isotope Dependent, Temperature Regulated, Energy Repartitioning in a Low-Barrier, Short-Strong Hydrogen Bonded Cluster. *J. Chem. Phys.* **2010**, *132*, No. 244301.
- (20) Dietrick, S. M.; Iyengar, S. S. Constructing Periodic Phase Space Orbits from *Ab Initio* Molecular Dynamics Trajectories to Analyze Vibrational Spectra: Case Study of the Zundel ( $H_5O_2^+$ ) Cation. *J. Chem. Theory Comput.* **2012**, *8*, 4876.
- (21) Li, J.; Pacheco, A. B.; Raghavachari, K.; Iyengar, S. S. A Grotthuss-like proton shuttle in the anomalous  $C_2H_3^+$  carbocation: energetic and vibrational properties for isotopologues. *Phys. Chem. Chem. Phys.* **2016**, *18*, 29395.
- (22) Sager, L. M.; Iyengar, S. S. Proton relays in anomalous carbocations dictate spectroscopy, stability, and mechanisms: case studies on  $C_2H_5^+$  and  $C_3H_3^+$ . *Phys. Chem. Chem. Phys.* **2017**, *19*, 27801.
- (23) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in C*; Cambridge University Press: New York, 1992.
- (24) Li, X.; Teige, V. E.; Iyengar, S. S. Can the four-coordinated, penta-valent oxygen in hydroxide water clusters be detected through experimental vibrational spectroscopy. *J. Phys. Chem. A* **2007**, *111*, 4815.
- (25) Phatak, P.; Venderley, J.; Debrot, J.; Li, J.; Iyengar, S. S. Active Site Dynamical Effects in the Hydrogen Transfer Rate-limiting Step in the Catalysis of Linoleic Acid by Soybean Lipoxygenase-1 (SLO-1): Primary and Secondary Isotope Contributions. *J. Phys. Chem. B* **2015**, *119*, 9532.
- (26) Head-Gordon, M.; Pople, J. A.; Frisch, M. J. MP2 energy evaluation by direct methods. *Chem. Phys. Lett.* **1988**, *153*, 503.
- (27) Schlegel, H. B.; Frisch, M. J. Computational Bottlenecks in Molecular Orbital Calculations. *Theoretical and Computational Models for Organic Chemistry*; Springer, 1991; pp 5–33.
- (28) Christiansen, O.; Koch, H.; Jorgensen, P. The 2nd-Order Approximate Coupled-Cluster Singles and Doubles Model CC2. *Chem. Phys. Lett.* **1995**, *243*, 409.
- (29) Gerber, R. B.; Varner, M. E.; Hammerich, A. D.; Riikonen, S.; Murdachaew, G.; Shemesh, D.; Finlayson-Pitts, B. J. Computational Studies Of Atmospherically-Relevant Chemical Reactions In Water Clusters And On Liquid Water And Ice Surfaces. *Acc. Chem. Res.* **2015**, *48*, 399.
- (30) Gerber, R. B.; Sebek, J. Dynamics Simulations of Atmospherically Relevant Molecular Reactions. *Int. Rev. Phys. Chem.* **2009**, *28*, 207.
- (31) Koitz, R.; Hutter, J.; Iannuzzi, M. Formation and properties of a terpyridine-based 2D MOF on the surface of water. *2D Mater.* **2016**, *3*, No. 025026.
- (32) Hutter, J. Car-Parrinello Molecular Dynamics. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 604–612.
- (33) Kirchner, B.; Di Dio, P. J.; Hutter, J. In Real-World Predictions From *Ab Initio* Molecular Dynamics Simulations. In *Multiscale Molecular Methods in Applied Chemistry*; Kirchner, B.; Vrabec, J., Eds.; Topics in Current Chemistry; Springer, 2012; Vol. 307.
- (34) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Challenges for Density Functional Theory. *Chem. Rev.* **2012**, *112*, 289.
- (35) Klimeš, J.; Michaelides, A. Perspective: Advances and challenges in treating van der Waals dispersion forces in density functional theory. *J. Chem. Phys.* **2012**, *137*, No. 120901.
- (36) Peverati, R.; Truhlar, D. The Quest for a Universal Density Functional: The Accuracy of Density Functionals Across a Broad



Spectrum of Databases in Chemistry and Physics. *Philos. Trans. R. Soc., A* **2014**, 372, No. 10120476.

(37) Li, J.; Iyengar, S. S. Ab initio Molecular Dynamics using Recursive, Spatially Separated, Overlapping Model Subsystems Mixed Within an ONIOM Based Fragmentation Energy Extrapolation Technique. *J. Chem. Theory Comput.* **2015**, 11, 3978.

(38) Li, J.; Haycraft, C.; Iyengar, S. S. Hybrid extended Lagrangian, post-Hartree-Fock Born-Oppenheimer ab initio molecular dynamics using fragment-based electronic structure. *J. Chem. Theory Comput.* **2016**, 12, 2493.

(39) Haycraft, C.; Li, J.; Iyengar, S. S. Efficient, "On-the-Fly", Born-Oppenheimer and Car-Parrinello-type Dynamics with Coupled Cluster Accuracy through Fragment Based Electronic Structure. *J. Chem. Theory Comput.* **2017**, 13, 1887.

(40) Ricard, T. C.; Haycraft, C.; Iyengar, S. S. Adaptive, geometric networks for efficient coarse-grained ab initio molecular dynamics with post-Hartree-Fock accuracy. *J. Chem. Theory Comput.* **2018**, 14, 2852.

(41) Ricard, T. C.; Iyengar, S. S. Efficiently capturing weak interactions in ab initio molecular dynamics through "on-the-fly" basis set extrapolation. *J. Chem. Theory Comput.* **2018**, 14, 5535.

(42) Kumar, A.; Iyengar, S. S. Fragment-based electronic structure for potential energy surfaces using a superposition of fragmentation topologies. *J. Chem. Theory Comput.* **2019**, 15, 5769.

(43) Ricard, T. C.; Iyengar, S. S. An efficient and accurate approach to estimate hybrid functional and large basis set contributions to condensed phase systems and molecule-surface interactions. *J. Chem. Theory Comput.* **2020**, 16, 4790.

(44) Ricard, T. C.; Kumar, A.; Iyengar, S. S. Embedded, graph-theoretically defined many-body approximations for wavefunction-in-DFT and DFT-in-DFT: applications to gas- and condensed-phase AIMD, and potential surfaces for quantum nuclear effects. *Int. J. Quantum Chem.* **2020**, 120, No. e26244.

(45) Kumar, A.; DeGregorio, N.; Zhang, J.; Iyengar, S. S. A Multi-topology Molecular Fragmentation Procedure for Efficient and Accurate Potential Surface Calculations. *J. Chem. Theory Comput.*, submitted for publication, 2021.

(46) Maseras, F.; Morokuma, K. IMOMM: A new integrated ab initio + molecular mechanics geometry optimization scheme of equilibrium structures and transition states. *J. Comput. Chem.* **1995**, 16, 1170.

(47) Kerdcharoen, T.; Morokuma, K. ONIOM-XS: An Extension of the ONIOM Method for Molecular Simulation in Condensed Phase. *Chem. Phys. Lett.* **2002**, 355, 257.

(48) Hopkins, B. W.; Tschumper, G. S. A multicentered approach to integrated QM/QM calculations. Applications to multiply hydrogen bonded systems. *J. Comput. Chem.* **2003**, 24, 1563.

(49) Guo, W.; Wu, A.; Xu, X. XO: An Extended ONIOM Method for Accurate and Efficient Geometry Optimization of Large Molecules. *Chem. Phys. Lett.* **2010**, 498, 203.

(50) Mayhall, N. J.; Raghavachari, K. Molecules-In-Molecules: An Extrapolated Fragment-Based Approach for Accurate Calculations on Large Molecules and Materials. *J. Chem. Theory Comput.* **2011**, 7, 1336.

(51) Mayhall, N. J.; Raghavachari, K. Many-Overlapping-Body (MOB) Expansion: A Generalized Many Body Expansion for Nondisjoint Monomers in Molecular Fragmentation Calculations of Covalent Molecules. *J. Chem. Theory Comput.* **2012**, 8, 2669.

(52) Červinka, C.; Beran, G. J. O. Ab initio prediction of the polymorph phase diagram for crystalline methanol. *Chem. Sci.* **2018**, 9, 4622.

(53) Hopkins, B. W.; Tschumper, G. S. Integrated quantum mechanical approaches for extended  $\pi$  systems: Multicentered QM/QM studies of the cyanogen and diacetylene trimers. *Chem. Phys. Lett.* **2005**, 407, 362.

(54) Fedorov, D. G.; Ishida, T.; Kitaura, K. Multilayer Formulation of the Fragment Molecular Orbital Method (FMO). *J. Phys. Chem. A* **2005**, 109, 2638.

(55) Tsuzuki, S.; Orita, H.; Honda, K.; Mikami, M. First-principles lattice energy calculation of urea and hexamine crystals by a combination of periodic DFT and MP2 two-body interaction energy calculations. *J. Phys. Chem. B* **2010**, 114, 6799.

(56) Sahu, N.; Yeole, S. D.; Gadre, S. R. Appraisal of molecular tailoring approach for large clusters. *J. Chem. Phys.* **2013**, 138, No. 104101.

(57) Li, S.; Li, W.; Ma, J. Generalized Energy-Based Fragmentation Approach and Its Applications to Macromolecules and Molecular Aggregates. *Acc. Chem. Res.* **2014**, 47, 2712.

(58) Raghavachari, K.; Saha, A. Accurate Composite and Fragment-Based Quantum Chemical Models for Large Molecules. *Chem. Rev.* **2015**, 115, 5643.

(59) Saha, A.; Raghavachari, K. Analysis of different fragmentation strategies on a variety of large peptides: Implementation of a low level of theory in fragment-based methods can be a crucial factor. *J. Chem. Theory Comput.* **2015**, 11, 2012.

(60) Collins, M. A.; Bettens, R. P. A. Energy-Based Molecular Fragmentation Methods. *Chem. Rev.* **2015**, 115, 5607.

(61) Willow, S. Y.; Salim, M. A.; Kim, K. S.; Hirata, S. Ab initio molecular dynamics of liquid water using embedded-fragment second-order many-body perturbation theory towards its accurate property prediction. *Sci. Rep.* **2015**, 5, No. 14358.

(62) Liu, J.; Qi, L.-W.; Zhang, J. Z. H.; He, X. Fragment Quantum Mechanical Method for Large-Sized Ion -Water Clusters. *J. Chem. Theory Comput.* **2017**, 13, 2021.

(63) Herbert, J. M. Fantasy versus reality in fragment-based quantum chemistry. *J. Chem. Phys.* **2019**, 151, No. 170901.

(64) Sontising, W.; Beran, G. J. O. Theoretical assessment of the structure and stability of the  $\lambda$  phase of nitrogen. *Phys. Rev. Mater.* **2019**, 3, No. 095002.

(65) Liu, K.-Y.; Herbert, J. M. Energy-Screened Many-Body Expansion: A Practical Yet Accurate Fragmentation Method for Quantum Chemistry. *J. Chem. Theory Comput.* **2020**, 16, 475.

(66) Varandas, A. J.; Murrell, J. N. A many-body expansion of polyatomic potential energy surfaces: application to  $H_n$  systems. *Faraday Discuss. Chem. Soc.* **1977**, 62, 92.

(67) Murrell, J.; Carter, S.; Farantos, S.; Huxley, P.; Varandas, A. *Molecular Potential Energy Functions*; Wiley: New York, 1984.

(68) Varandas, A.; Pais, A. A realistic double many-body expansion (DMBE) potential energy surface for ground-state  $O_3$  from a multiproperty fit to ab initio calculations, and to experimental spectroscopic, inelastic scattering, and kinetic isotope thermal rate data. *Mol. Phys.* **1988**, 65, 843.

(69) Xantheas, S. S. Ab Initio Studies of Cyclic Water Clusters  $(H_2O)_n$ ,  $N=1-6$ . II. Analysis of Many-body Interactions. *J. Chem. Phys.* **1994**, 100, 7523.

(70) Xantheas, S. S. Ab Initio Studies of Cyclic Water Clusters  $(H_2O)_n$ ,  $N=1-6$ . III. Comparison of Density Functional with MP2 Results. *J. Chem. Phys.* **1995**, 102, 4505.

(71) Dahlke, E. E.; Truhlar, D. G. Electrostatically Embedded Many Body Expansion for Large Systems, with Applications to Water Clusters. *J. Chem. Theory Comput.* **2007**, 3, 46.

(72) Dahlke, E. E.; Truhlar, D. G. Electrostatically Embedded Many Body Expansion for Simulations. *J. Chem. Theory Comput.* **2008**, 4, 1.

(73) Hirata, S. Fast Electron-Correlation Methods for Molecular Crystals: an Application to the  $\alpha$ ,  $\beta(1)$ , and  $\beta(2)$  Modifications of Solid Formic Acid. *J. Chem. Phys.* **2008**, 129, No. 204104.

(74) Jacobson, L. D.; Herbert, J. M. An Efficient, Fragment-Based Electronic Structure Method for Molecular Systems: Self-Consistent Polarization with Perturbative Two-Body Exchange and Dispersion. *J. Chem. Phys.* **2011**, 134, No. 094118.

(75) Richard, R. M.; Herbert, J. M. A Generalized Many-Body Expansion and a Unified View of Fragment-Based Methods in Electronic Structure Theory. *J. Chem. Phys.* **2012**, 137, No. 064113.

(76) Yu, Q.; Bowman, J. M. Communication: VSCF/VCI vibrational spectroscopy of  $H_7O_3^+$  and  $H_9O_4^+$  using high-level, many-body potential energy surface and dipole moment surfaces. *J. Chem. Phys.* **2017**, 146, No. 121102.

(77) Yang, J.; Hu, W.; Usvyat, D.; Matthews, D.; Schuetz, M.; Chan, G. K.-L. Ab initio determination of the crystalline benzene lattice energy to sub-kilojoule/mole accuracy. *Science* **2014**, 345, 640.

- (78) Kerdcharoen, T.; Liedl, K. R.; Rode, B. M. A QM/MM simulation method applied to the solution of  $\text{Li}^+$  in liquid ammonia. *Chem. Phys.* **1996**, *211*, 313.
- (79) Shin, J.-W.; Hammer, N. I.; Diken, E. G.; Johnson, M. A.; Walters, R. S.; Jaeger, T. D.; Duncan, M. A.; Christie, R. A.; Jordan, K. D. Infrared Signature of Structures Associated with the  $\text{H}(\text{H}_2\text{O})_n$  ( $N = 6$  to 27) Clusters. *Science* **2004**, *304*, 1137.
- (80) Marx, D.; Tuckerman, M. E.; Hutter, J.; Parrinello, M. The Nature of the Hydrated Excess Proton in Water. *Nature* **1999**, *397*, 601.
- (81) Pomès, R.; Roux, B. Structure and Dynamics of a Proton Wire: A Theoretical Study of  $\text{H}^+$  Translocation Along the Single-File Water Chain in the Gramicidin a Channel. *Biophys. J.* **1996**, *71*, 19.
- (82) Huang, L.; Massa, L.; Karle, J. Kernel energy method: Application to DNA. *Biochemistry* **2005**, *44*, 16747.
- (83) Ganesh, V.; Dongare, R. K.; Balanarayan, P.; Gadre, S. R. Molecular Tailoring Approach for Geometry Optimization of Large Molecules: Energy Evaluation and Parallelization Strategies. *J. Chem. Phys.* **2006**, *125*, No. 104109.
- (84) Dolgonos, G. A.; Loboda, O. A.; Boese, A. D. Development of Embedded and Performance of Density Functional Methods for Molecular Crystals. *J. Phys. Chem. A* **2018**, *122*, 708.
- (85) Zhang, D. W.; Zhang, J. Z. H. Molecular Fractionation with Conjugate Caps for Full Quantum Mechanical Calculation of Protein–molecule Interaction Energy. *J. Chem. Phys.* **2003**, *119*, 3599.
- (86) Le, H.-A.; Tan, H.-J.; Ouyang, J. F.; Bettens, R. P. A. Combined Fragmentation Method: A Simple Method for Fragmentation of Large Molecules. *J. Chem. Theory Comput.* **2012**, *8*, 469.
- (87) Gordon, M.; Mullin, J.; Pruitt, S.; Roskop, L.; Slipchenko, L.; Boatz, J. Accurate Methods for Large Molecular Systems. *J. Phys. Chem. B* **2009**, *113*, 9646.
- (88) Collins, M. A. Systematic Fragmentation of Large Molecules by Annihilation. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7744.
- (89) Han, J.; Mazack, M. J. M.; Zhang, P.; Truhlar, D. G.; Gao, J. Quantum Mechanical Force Field for Water with Explicit Electronic Polarization. *J. Chem. Phys.* **2013**, *139*, No. 054503.
- (90) Svensson, M.; Humbel, S.; Froese, R. D.; Matsubara, T.; Sieber, S.; Morokuma, K. ONIOM: a multilayered integrated MO+ MM method for geometry optimizations and single point energy predictions. A test for Diels–Alder reactions and  $\text{Pt}(\text{t-Bu})_3)_2 + \text{H}_2$  oxidative addition. *J. Phys. Chem. A* **1996**, *100*, 19357.
- (91) Chung, L. W.; Hirao, H.; Li, X.; Morokuma, K. The ONIOM Method: Its Foundation and Applications to Metalloenzymes and Photobiology. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 327.
- (92) Chung, L. W.; Sameera, W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F.; Li, H.-B.; Ding, L.; Morokuma, K. The ONIOM Method and Its Applications. *Chem. Rev.* **2015**, *115*, 5678.
- (93) Wolke, C. T.; Fournier, J. A.; Dzugas, L. C.; Fagiani, M. R.; Odbadrakh, T. T.; Knorke, H.; Jordan, K. D.; McCoy, A. B.; Asmis, K. R.; Johnson, M. A. Spectroscopic snapshots of the proton-transfer mechanism in water. *Science* **2016**, *354*, 1131.
- (94) Fournier, J. A.; Wolke, C. T.; Johnson, M. A.; Odbadrakh, T. T.; Jordan, K. D.; Kathmann, S. M.; Xantheas, S. S. Snapshots of Proton Accommodation at a Microscopic Water Surface: Understanding the Vibrational Spectral Signatures of the Charge Defect in Cryogenically Cooled  $\text{H}^+(\text{H}_2\text{O})_n$  ( $n=2-28$ ) Clusters. *J. Phys. Chem. A* **2015**, *119*, 9425.
- (95) Dey, T. K.; Shah, N. R. On the number of simplicial complexes in  $\mathbb{R}^d$ . *Comput. Geom.* **1997**, *8*, 267.
- (96) Adams, C. C.; Franzosa, R. D. *Introduction to Topology: Pure and Applied*; Pearson, 2008.
- (97) Berger, M.; Pansu, P.; Berry, J.-P.; Saint-Raymond, X. Affine Spaces. *Problems in Geometry*; Springer, 1984; p 11.
- (98) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16*, revision B.01; Gaussian Inc.: Wallingford, CT, 2016.
- (99) Neese, F. The ORCA program system. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 73.
- (100) Parrish, R. M.; Burns, L. A.; Smith, D. G. A.; Simmonett, A. C.; DePrince, A. E., III; Hohenstein, E. G.; Bozkaya, U.; Sokolov, A. Y.; Di Remigio, R.; Richard, R. M.; Gonthier, J. F.; James, A. M.; McAlexander, H. R.; Kumar, A.; Saitow, M.; Wang, X.; Pritchard, B. P.; Prakash, V.; Schaefer, H. F., III; Patkowski, K.; King, R. A.; Valeev, E. F.; Evangelista, F. A.; Turney, J. M.; Crawford, T. D.; Sherrill, C. D. PSI4 1.1: An Open-Source Electronic Structure Program Emphasizing Automation, Advanced Libraries, and Interoperability. *J. Chem. Theory Comput.* **2017**, *13*, 3185.
- (101) Giannozzi, P.; Baroni, S.; Bonini, N.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Chiarotti, G. L.; Cococcioni, M.; Dabo, I.; et al. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys.: Condens. Matter.* **2009**, *21*, No. 395502.
- (102) Ozaki, T.; Kino, H.; Yu, J.; Han, M.; Ohfuchi, M.; Ishii, F.; Sawada, K.; Ohwaki, T.; Weng, H.; Toyoda, M.; Okuno, Y.; Perez, R.; Bell, P.; D, T.; Xiao, Y.; Ito, A.; Terakura, K. *User's manual of OpenMX*, version 3.8; OpenMX, 2016.
- (103) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. A Computer-Simulation Method for the Calculation of Equilibrium-Constants for the Formation of Physical Clusters of Molecules - Application to Small Water Clusters. *J. Chem. Phys.* **1982**, *76*, 637.
- (104) Trotter, H. F. On the Product of Semi-Groups of Operators. *Proc. Am. Math. Soc.* **1959**, *10*, 545.
- (105) Nelson, E. Feynman Integrals and the Schrödinger Equation. *J. Math. Phys.* **1964**, *5*, 332.
- (106) Tuckerman, M.; Berne, B. J.; Martyna, G. J. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.* **1992**, *97*, 1990.
- (107) Pomès, R.; Roux, B. Theoretical Study of  $\text{H}^+$  Translocation Along a Model Proton Wire. *J. Phys. Chem. B* **1996**, *100*, 2519.
- (108) Decornez, H.; Drukker, K.; Hammes-Schiffer, S. Solvation and Hydrogen-Bonding Effects on Proton Wires. *J. Phys. Chem. A* **1999**, *103*, 2891.
- (109) Brewer, M. L.; Schmitt, U. W.; Voth, G. A. The Formation and Dynamics of Proton Wires in Channel Environments. *Biophys. J.* **2001**, *80*, 1691.
- (110) Teeter, M. M. Water Structure of a Hydrophobic Protein at Atomic Resolution: Pentagon Rings of Water Molecules in Crystals of Crambin. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 6014.
- (111) Neidle, S.; Berman, H. M.; Shieh, H. S. Highly Structured Water Network in Crystals of a Deoxydinucleoside-Drug Complex. *Nature* **1980**, *288*, 129.
- (112) Lipscomb, L. A.; Peek, M. E.; Zhou, F. X.; Bertrand, J. A.; VanDerveer, D.; Williams, L. D. Water Ring Structure at Dna Interfaces - Hydration and Dynamics of Dna Anthracycline Complexes. *Biochemistry* **1994**, *33*, 3649.
- (113) Tu, C.; Rowlett, R. S.; Tripp, B. C.; Ferry, J. G.; Silverman, D. N. Chemical Rescue of Proton Transfer in Catalysis by Carbonic Anhydrases in the Beta- and Gamma-Class. *Biochemistry* **2002**, *41*, 15429.
- (114) McEwan, M. J.; Phillips, L. F. *Chemistry of the Atmosphere*; Edward Arnold: London, 1975.
- (115) Wayne, R. P. *Chemistry of the Atmosphere*; Clarendon Press: Oxford, 1994.
- (116) Dietrick, S. M.; Pacheco, A. B.; Phatak, P.; Stevens, P. S.; Iyengar, S. S. The Influence of Water on Anharmonicity, Stability and Vibrational Energy Distribution of Hydrogen-Bonded Adducts in Atmospheric Reactions: Case Study of the  $\text{OH} + \text{Isoprene}$  Reaction



Intermediate Using *Ab-Initio* Molecular Dynamics. *J. Phys. Chem. A* **2012**, *116*, 399.

(117) Fournier, J. A.; et al. Snapshots of Proton Accommodation at a Microscopic Water Surface: Understanding the Vibrational Spectral Signatures of the Charge Defect in Cryogenically Cooled  $\text{H}^+(\text{H}_2\text{O})_n^+$ ,  $n=2-28$  Clusters. *J. Phys. Chem. A* **2015**, *119*, 9425.

(118) Wei, S.; Shi, Z.; Castleman, A. W., Jr. Mixed Cluster Ions As a Structure Probe: Experimental Evidence for Clathrate Structure of  $(\text{H}_2\text{O})_{20}\text{H}^+$  and  $(\text{H}_2\text{O})_{21}\text{H}^+$ . *J. Chem. Phys.* **1991**, *94*, 3268.

(119) Yang, X.; Zhang, X.; Castleman, A. W., Jr. Kinetics and Mechanism Studies of Large Protonated Water Clusters,  $\text{H}^+(\text{H}_2\text{O})_N$ ,  $N=1-60$ , at Thermal Energy. *Int. J. Mass Spectrom. Ion Processes* **1991**, *109*, 339.

(120) Douberly, G. E.; Ricks, A. M.; Duncan, M. A. Infrared Spectroscopy of Perdeuterated Protonated Water Clusters in the Vicinity of the Clathrate Cage. *J. Phys. Chem. A* **2009**, *113*, 8449.

(121) Fournier, J. A.; Wolke, C. T.; Johnson, C. J.; Johnson, M. A.; Heine, N.; Gewinner, S.; Schoellkopf, W.; Esser, T. K.; Fagiani, M. R.; Knorke, H.; Asmis, K. R. Site-specific vibrational spectral signatures of water molecules in the magic  $\text{H}_3\text{O}^+(\text{H}_2\text{O})_{20}$  and  $\text{Cs}^+(\text{H}_2\text{O})_{20}$  clusters. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 18132.

(122) Wu, C.-C.; Lin, C.-K.; Chang, H.-C.; Jiang, J.-C.; Kuo, J.-L.; Klein, M. L. Protonated Clathrate Cages Enclosing Neutral Water Molecules:  $\text{H}^+(\text{H}_2\text{O})_{21}$  and  $\text{H}^+(\text{H}_2\text{O})_{28}$ . *J. Chem. Phys.* **2005**, *122*, No. 074315.

(123) Korchagina, K.; Simon, A.; Rapacioli, M.; Spiegelman, F.; L'Hermite, J.-M.; Braud, I.; Zamith, S.; Cuny, J. Theoretical investigation of the solid-liquid phase transition in protonated water clusters. *Phys. Chem. Chem. Phys.* **2017**, *19*, 27288.

(124) Hansen, K.; Ryding, M. J.; Uggerud, E. Magic numbers and stabilities of heavy water clusters,  $(\text{D}_2\text{O})_N(\text{D})^+$ ,  $N=3-48$ . *Int. J. Mass Spectrom.* **2019**, *440*, 14.

(125) Yu, Q.; Bowman, J. M. Tracking Hydronium/Water Stretches in Magic  $\text{H}_3\text{O}^+(\text{H}_2\text{O})_{20}$  Clusters through High-level Quantum VSCF/VCI Calculations. *J. Phys. Chem. A* **2020**, *124*, 1167.

(126) Miyazaki, M.; Fujii, A.; Ebata, T.; Mikami, N. Infrared Spectroscopic Evidence for Protonated Water Clusters Forming Nanoscale Cages. *Science* **2004**, *304*, 1134.

(127) Fournier, J. A.; Johnson, C. J.; Wolke, C. T.; Weddle, G. H.; Wolk, A. B.; Johnson, M. A. Vibrational spectral signature of the proton defect in the three-dimensional  $\text{H}^+(\text{H}_2\text{O})_{21}$  cluster. *Science* **2014**, *344*, 1009.

(128) Yu, H. B.; Cui, Q. The Vibrational Spectra of Protonated Water Clusters: A Benchmark for Self-Consistent-Charge Density-Functional Tight Binding. *J. Chem. Phys.* **2007**, *127*, No. 234504.

(129) Markovitch, O.; Chen, H.; Izvekov, S.; Paesani, F.; Voth, G. A.; Agmon, N. Special pair dance and partner selection: Elementary steps in proton transport in liquid water. *J. Phys. Chem. B* **2008**, *112*, 9456.

(130) Knight, C.; Voth, G. A. The Curious Case of the Hydrated Proton. *Acc. Chem. Res.* **2012**, *45*, 101.

(131) Ouyang, J. F.; Bettens, R. P. A. When are Many-Body Effects Significant. *J. Chem. Theory Comput.* **2016**, *12*, 5860.

(132) Liu, J.; Rana, B.; Liu, K.-Y.; Herbert, J. M. Variational Formulation of the Generalized Many-Body Expansion with Self-Consistent Charge Embedding: Simple and Correct Analytic Energy Gradient for Fragment-Based *ab Initio* Molecular Dynamics. *J. Phys. Chem. Lett.* **2019**, *10*, 3877.

(133) Agmon, N. The Grotthuss Mechanism. *Chem. Phys. Lett.* **1995**, *244*, 456.

(134) Headrick, J. M.; Diken, E. G.; Walters, R. S.; Hammer, N. I.; Christie, R. A.; Cui, J.; Myshakin, E. M.; Duncan, M. A.; Johnson, M. A.; Jordan, K. Spectral Signatures of Hydrated Proton Vibrations in Water Clusters. *Science* **2005**, *308*, 1765.

(135) McDonald, D. C., II; Wagner, J. P.; McCoy, A. B.; Duncan, M. A. Near-Infrared Spectroscopy and Anharmonic Theory of Protonated Water Clusters: Higher Elevations in the Hydrogen Bonding Landscape. *J. Phys. Chem. Lett.* **2018**, *9*, 5664.

(136) Headrick, J. M.; Diken, E. G.; Walters, R. S.; Hammer, N. I.; Christie, R. A.; Cui, J.; Myshakin, E. M.; Duncan, M. A.; Johnson, M. A.

Jordan, K. D. Spectral signatures of hydrated proton vibrations in water clusters. *Science* **2005**, *308*, 1765.

(137) Shin, J.-W.; Hammer, N. I.; Diken, E. G.; Johnson, M. A.; Walters, R. S.; Jaeger, T. D.; Duncan, M. A.; Christie, R. A.; Jordan, K. D. Infrared Signature of Structures Associated with the  $\text{H}^+(\text{H}_2\text{O})_n$  ( $n = 6$  to  $27$ ) Clusters. *Science* **2004**, *304*, 1137.

(138) Pruitt, S. R.; Steinmann, C.; Jensen, J. H.; Gordon, M. S. Fully Integrated Effective Fragment Molecular Orbital Method. *J. Chem. Theory Comput.* **2013**, *9*, 2235.

(139) Steinmann, C.; Fedorov, D. G.; Jensen, J. H. Effective Fragment Molecular Orbital Method: A Merger of the Effective Fragment Potential and Fragment Molecular Orbital Methods. *J. Phys. Chem. A* **2010**, *114*, 8705.

(140) Komeiji, Y.; Mochizuki, Y.; Nakano, T. Three-Body Expansion and Generalized Dynamic Fragmentation Improve the Fragment Molecular Orbital-Based Molecular Dynamics (FMO-MD). *Chem. Phys. Lett.* **2010**, *484*, 380.

(141) Richard, R. M.; Herbert, J. M. Many-Body Expansion with Overlapping Fragments: Analysis of Two Approaches. *J. Chem. Theory Comput.* **2013**, *9*, 1408.

(142) Beran, G. J. O.; Wen, S.; Nanda, K.; Huang, Y.; Heit, Y. Accurate and Robust Molecular Crystal Modeling Using Fragment-Based Electronic Structure Methods. In *Prediction and Calculation of Crystal Structures: Methods and Applications*; Atahan-Evrenk, S.; Aspuru-Guzik, A., Eds.; Topics in Current Chemistry-Series; Springer, 2014; Vol. 345, p 59.

(143) Hua, S.; Hua, W.; Li, S. An Efficient Implementation of the Generalized Energy-Based Fragmentation Approach for General Large Molecules. *J. Phys. Chem. A* **2010**, *114*, 8126.

(144) Bakowies, D.; Thiel, W. Hybrid Models for Combined Quantum Mechanical and Molecular Mechanical Approaches. *J. Phys. Chem. C* **1996**, *100*, 10580.

(145) Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. *J. Mol. Struct.: THEOCHEM* **1999**, *461*, 1–21.

(146) Rega, N.; Iyengar, S. S.; Voth, G. A.; Schlegel, H. B.; Vreven, T.; Frisch, M. J. Hybrid *Ab-Initio*/Empirical Molecular Dynamics: Combining the ONIOM Scheme with the Atom-Centered Density Matrix Propagation (ADMP) Approach. *J. Phys. Chem. B* **2004**, *108*, 4210.

(147) Harris, F. J. On the use of windows for harmonic analysis with the discrete Fourier transform. *Proc. IEEE* **1978**, *66*, 51.

(148) Vendrell, O.; Gatti, F.; Meyer, H.-D. Dynamics and Infrared Spectroscopy of the Protonated Water Dimer. *Angew. Chem., Int. Ed.* **2007**, *46*, 6918.

(149) Vendrell, O.; Gatti, F.; Meyer, H.-D. Full Dimensional (15-Dimensional) Quantum-Dynamical Simulation of the Protonated Water-Dimer II: Infrared Spectrum and Vibrational Dynamics. *J. Chem. Phys.* **2007**, *127*, No. 184303.

(150) Morrone, J. A.; Car, R. Nuclear quantum effects in water. *Phys. Rev. Lett.* **2008**, *101*, No. 017801.

(151) Vendrell, O.; Gatti, F.; Meyer, H.-D. Strong Isotope Effects in the Infrared Spectrum of the Zundel Cation. *Angew. Chem., Int. Ed.* **2009**, *48*, 352.

(152) Li, X.-Z.; Walker, B.; Michaelides, A. Quantum nature of the hydrogen bond. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 6369.

(153) Ceriotti, M.; Fang, W.; Kusalik, P. G.; McKenzie, R. H.; Michaelides, A.; Morales, M. A.; Markland, T. E. Nuclear quantum effects in water and aqueous systems: Experiment, theory, and current challenges. *Chem. Rev.* **2016**, *116*, 7529.

(154) Schran, C.; Marx, D. Quantum nature of the hydrogen bond from ambient conditions down to ultra-low temperatures. *Phys. Chem. Chem. Phys.* **2019**, *21*, 24967.