JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

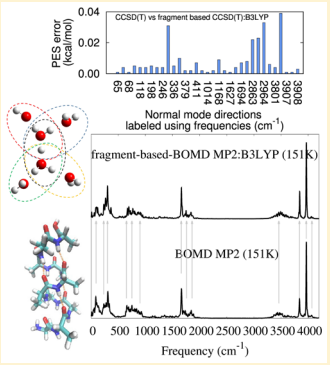
Hybrid Extended Lagrangian, Post-Hartree–Fock Born–Oppenheimer ab Initio Molecular Dynamics Using Fragment-Based Electronic Structure

Junjie Li, Cody Haycraft, and Srinivasan S. Iyengar*

Department of Chemistry and Department of Physics, Indiana University, 800 E. Kirkwood Ave., Bloomington, Indiana 47405, United States

Supporting Information

ABSTRACT: We present a hybrid ab initio molecular dynamics scheme that includes both DFT and Hartree–Fock-based extended Lagrangian and converged post-Hartree–Fock Born–Oppenheimer components, combined within the framework of a molecular fragmentation-based electronic structure. The specific fragmentation algorithm used here is derived from ONIOM but includes multiple, overlapping "model" systems. The interaction between the various overlapping model systems is approximated by invoking the principle of inclusion–exclusion at the chosen higher level of theory and within a "real" calculation performed at the chosen lower level of theory. Furthermore, here, the lower level electronic structure of the full system is propagated through an extended Lagrangian formalism, whereas the fragments, treated using post-Hartree–Fock electronic structure theories, are computed using the normal converged Born–Oppenheimer treatment. This conservative dynamical approach largely reduces the computational cost to approximate on-the-fly dynamics using post-Hartree–Fock electronic structure techniques and can be very beneficial for large systems where SCF convergence may be challenging and time consuming. Benchmarks are provided for medium-sized protonated water clusters, $H_9O_4^+$ and $H_{13}O_6^+$, and polypeptide fragments, including a proline tripeptide fragment, and alanine decamer. Structural features are in excellent agreement between the hybrid approach using an MP2:B3LYP fragment-based electronic structure and BOMD using MP2 for the full system. Vibrational properties derived from dynamical correlation functions do show a small redshift for the extended Lagrangian treatments, especially at higher frequencies. Strategies are discussed to improve this redshift. The computational methodology works in parallel using both MPI and OpenMP and shows good scaling with the processor number. The timing benchmarks are provided for the alanine decamer. A powerful feature of the computational implementation is the fact that it is completely decoupled from the electronic structure package being employed and thus allows for an integrated approach that may include several different packages. These computational aspects will be further probed in future publications.



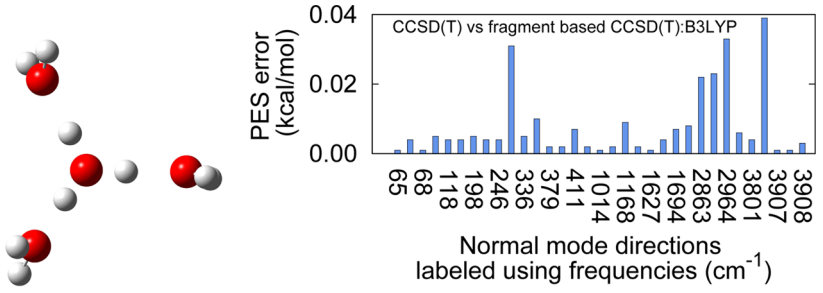
I. INTRODUCTION

Since the seminal work of Karplus[1] and Leforestier[2] and the extended Lagrangian generalization by Car and Parrinello,[3] ab initio molecular dynamics (AIMD) has become a central tool for studying reactive events and vibrational properties beyond the harmonic approximation and for biochemical modeling when combined with hybrid techniques. Here, electronic structure and gradients are simultaneously computed in-step with nuclear motion. Due to this, AIMD is in general much more resource intensive as compared to both classical, force field-based, molecular dynamics methods as well as commonly used electronic structure applications such as geometry optimization and frequency calculations. This limitation has thus far restricted AIMD mostly to DFT-based application for medium-sized systems. Improvements in electronic structure methodology, certainly, have critical impact on AIMD, and one of the ways to reduce the cost of the on-the-fly electronic structure calculations is through the utilization of the recently popularized fragment-based electronic structure schemes,[4−25] where a molecular system is divided into multiple regions and the total energy of the system is assembled from the energy of individual pieces. Detailed discussion of these methods can be found in several recently published review articles.[26−30]

To explore the application of fragment-based electronic structure in AIMD and potentially ab initio quantum nuclear dynamics,[31] we recently[32] (a) considered a generalization to the ONIOM[5,13,33−41] energy extrapolation scheme using the set-theoretic principle of inclusion and exclusion[42] to treat multiple overlapping model systems[32] (abbreviated in this publication as the PIE-ONIOM energy expression) and (b) applied this method to obtain accurate potential energy surfaces and to compute ab initio dynamical trajectories with correlation functions that agree well with higher levels of theory such as MP2 and coupled cluster. Specifically, mean absolute errors in potential energy surfaces are (a) less than 0.05 kcal/mol when the PIE-ONIOM results that combine B3LYP and MP2 are compared with MP2 calculations on the

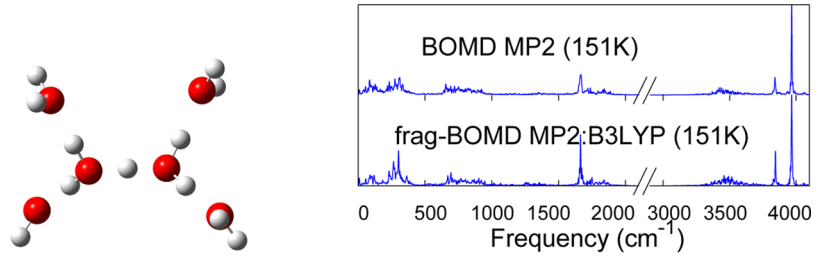
Received: January 1, 2016
Published: May 10, 2016

DOI: 10.1021/acs.jctc.6b00001
J. Chem. Theory Comput. 2016, 12, 2493−2508



(a)$H_9O_4^+$structure (b)PES Error

Figure 1. Figure 1 and Figure 2 briefly summarize the developments from ref 32. Here, the vibrational ground state ($\rho(R)$) weighted mean absolute error in the potential surface is provided for $H_9O_4^+$. The error is computed between the fragment-based calculations and full CCSD(T) calculations as $\int dR\{\rho(R)|E_{CCSD(T)} - E_{CCSD(T):B3LYP}|\}$. The statistics are obtained from 693 different geometries chosen along normal mode coordinates where the horizontal axis in the figure represents mode frequencies. Thus, the figure depicts the error in the critical regions of the surface sampled by the ground state wavepacket and is clearly very small. When the full surface is included (that is, without vibrational ground state weighting), the error rises to the order of 0.1 kcal/mol. All calculations use the 6-31+G(d,p) basis function.



(a)$H_{13}O_6^+$structure (b)VDOS

Figure 2. Solvated Zundel cation, $H_{13}O_6^+$: The vibrational density of states (VDOS) are computed from the fragment-based AIMD trajectories (frag-BOMD MP2:B3LYP), and these agree quite well with full MP2-based AIMD. All calculations use the 6-31+G(d,p) Gaussian basis set.

full system and (b) less than 0.07 kcal/mol when a similar hybrid calculation is compared with CCSD(T) on the full system. This data was obtained in ref 32 by considering 693 different geometries of protonated water clusters sampled along normal mode coordinates, and the associated computational effort is a very small fraction of the computational cost incurred in the full calculations. The AIMD trajectories computed using this approach reproduce vibrational density of states from regular MP2 calculations as well, and these results are summarized in Figures 1 and 2. As seen in Figure 1(b), and noted in the caption, the potential energy surface errors are indeed extremely small and vibrational density of states peaks agree quite well when the hybrid approach is compared to higher level theory as seen from Figure 2(b).

Despite the preliminary success of our approach in ref 32, several challenges remain: (1) During dynamical simulations, the molecular structural variation can lead to changes in the fragmentation topologies that must be handled carefully to maintain conservative dynamics. (2) The computational expense may be dominated by the lower level calculation on the full system since our approach is based on ONIOM. This second hurdle is also seen in other hybrid fragmentation schemes, such as the Generalized Energy Based Fragmentation (GEBF),[11] Molecular Tailoring Approach (MTA),[14] the eXtended ONIOM method,[15] and Molecule-in-Molecule (MIM).[16] (3) Analytical gradient for the embedded charges can be expensive.[43,44] In the current publication, we propose a solution to the second problem listed above by constructing a hybrid method that combines Extended Lagrangian treatment of the full system at the lower level of theory with Born-Oppenheimer dynamics for the fragments computed at high and low levels. Thus, the paper is organized as follows: Section II introduces the new formalism with facilitating technical discussions provided in Appendix A. The conservative properties of the approach are then tested in Section III. The approach is used to compute structural properties of protonated water clusters in Section IV, dynamical trajectories, and correlation functions for protonated water clusters and polypeptide fragments in Section V. In all cases, the trajectories are compared with MP2-based AIMD trajectories, and the results are in very good agreement. We also demonstrate numerical benchmarks in Section VI that show a large reduction in computational effort. Furthermore, the algorithm is written in parallel using MPI (also demonstrated in Section VI) and through a C++ package that invokes external electronic structure software. Concluding remarks are made in Section VII.

II. HYBRID EXTENDED LAGRANGIAN BORN−OPPENHEIMER DYNAMICS USING FRAGMENT-BASED ELECTRONIC STRUCTURE FROM PIE-ONIOM FOR POST-HARTREE−FOCK TREATMENT

Classical ground state ab initio molecular dynamics primarily has two flavors: one is the traditional Born−Oppenheimer molecular dynamics (BOMD), where the electronic energy is converged, and the other is the extended Lagrangian molecular dynamics (ELMD)[3,45−57] formalism, where the electronic degrees of freedom, either single particle density matrix or molecular orbital coefficients, are assigned fictitious inertia and propagated simultaneously with the classical nuclei through a simple adjustment of time scales.[58] The first ELMD

DOI: 10.1021/acs.jctc.6b00001
J. Chem. Theory Comput. 2016, 12, 2493−2508

method, the Car−Parrinello molecular dynamics (CPMD)[3,45,47] approach was influenced by earlier work from Anderson[59] and Parrinello and Rahman[45,60] and was developed in 1985. In CPMD, the wavefunction is expanded in a plane-wave basis, and pseudopotentials are used to describe the core electrons. In 2001, Atom-centered Density Matrix Propagation (ADMP),[48−54] an ELMD formalism that employs the single particle electronic density matrix written in an all-electron atom-centered Gaussian basis was developed. The ADMP method has been demonstrated through applications to vibrational spectroscopy[61−66] and biological ion-channel studies[54] and the study of atmospheric reactions, where pump−probe techniques have been incorporated into the dynamics.[67−69] Critically, the ADMP formalism has had an important role in the prediction of the amphiphilic nature of the hydrated proton[70,71] that was later confirmed through several experiments.[70−76] The ADMP method has many attractive features. Systems can be simulated by accurately treating all electrons or by using pseudopotentials. Through the use of smaller values for the tensorial fictitious mass, relatively large time steps can be employed, and lighter atoms such as hydrogen nuclei are routinely used. Atom-centered functions can be used with the appropriate physical boundary conditions for molecules, polymers, surfaces, and solids, without the need to treat replicated images to impose 3d periodicity. Consequently, charged systems and QM/MM models of biological systems[54] can be treated as easily as neutral molecules. Deviations from the Born−Oppenheimer surface and the mixing of fictitious and real kinetic energies can also be rigorously controlled on-the-fly in ADMP. ADMP trajectories of the order of picoseconds show stable dynamics, and the adiabaticity can be controlled effectively in these systems without thermostats. The important conceptual and computational differences between ADMP and other Gaussian basis set-based implementations[77−80] of the Car−Parrinello formalism have been discussed in detail in ref 50.

The elimination of SCF in extended Lagrangian MD is a computational advantage, especially for large molecules and systems where SCF convergence may be challenging. However, thus far, a critical restriction for ELMD (Car−Parrinello as well as ADMP) has been that it can only be applied to single particle formalisms such as DFT and Hartree−Fock. Here, we combine ELMD and post-Hartree−Fock-based BOMD through PIE-ONIOM and allow extended Lagrangian (Car−Parrinello and ADMP) techniques to become applicable to higher level electronic structure methods, such as MP2 and potentially CCSD, and also improve the computational efficiency of fragment-based AIMD. However, before we embark into a discussion of this dynamical formalism, we must first summarize the PIE-ONIOM energy extrapolation equations. The overall PIE-ONIOM molecular energy expression[32] is given by

$$E^{\text{PIE-ONIOM}} \equiv E_{\text{low}}(0) + \sum_{i=1}^{n} \mathcal{S}(i) - \sum_{1<=i<j<=n} \mathcal{S}(i \cap j) + \sum_{1<=i<j<k<=n} \mathcal{S}(i \cap j \cap k) - \cdots + (-1)^{n-1} \sum \mathcal{S}(1 \cap \cdots \cap n) \quad (1)$$

where the extrapolation term is defined as

$$\mathcal{S}(i) = E_{\text{high}}(i) - E_{\text{low}}(i) \quad (2)$$

when two levels of theory are used. (This approach is easily generalized to more levels,[32] but this generalization is not considered here.) In eq 1, $\mathcal{S}(i \cap j)$ represents the extrapolation term for the derivative fragment $(i \cap j)$, formed from the intersection of the ith and jth primitive fragments. Every fragment is treated at two levels of theory as written in eq 2, and the entire system (0th fragment) is only considered at the lowest level of theory (with energy $E_{\text{low}}(0)$). This energy extrapolation scheme is pictorially demonstrated in Figure 3. A powerful feature of the algorithm is the fact that

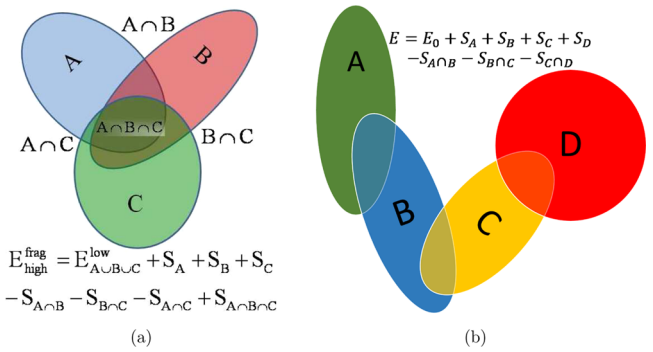


Figure 3. Illustration for the energy expression in eq 1. The intersections between primitive fragments are subtracted out to remove overcounting according to the principle of inclusion and exclusion.

the identity of the overlapping model systems or fragments are efficiently evaluated either (i) through an automated bit-manipulation algorithm presented in ref 32 that recomputes fragments at each instant in time during dynamics or (ii) through user specification. In the bit-manipulation algorithm, fragments are represented as bits in an integer, and overlapping regions are computed using binary AND operations. The use of bitwise arithmetic improves efficiency, but a large integer library may be required, depending on the number of fragments involved and the platform being used for the calculation. One such library is available at http://www.boost.org/.

Clearly, for a large enough system the computational expenses may be dominated by the lower level calculation, $E_{\text{low}}(0)$. Thus, in the scheme introduced here, it is electronic degrees of freedom of the entire system, $E_{\text{low}}(0)$ that we will propagate using an extended Lagrangian technique. The rest of the fragments, where more accurate calculations maybe needed, will be treated using converged Born−Oppenheimer treatments with post-HF electronic structure. While we choose to demonstrate this approach using the ADMP Lagrangian[48,49] as the choice for ELMD, extensions to employ CPMD (that is, using molecular orbital coefficients instead of the single particle density matrices with plane wave basis sets instead of the atom-centered Gaussian basis functions employed here) can be done in a similar fashion. The Lagrangian of the entire system is

$$\mathcal{L} = \frac{1}{2}\text{Tr}[\mathbf{V}^{\mathbf{T}}\mathbf{M}\mathbf{V}] + \frac{1}{2}\text{Tr}[(\boldsymbol{\mu}^{1/4}_{\text{low},\mathbf{0}}\mathbf{W}_{\text{low},\mathbf{0}}\boldsymbol{\mu}^{1/4}_{\text{low},\mathbf{0}})^2] - E^{\text{PIE-ONIOM}}(\mathbf{R}, \mathbf{P}_{\text{low},\mathbf{0}}) - \text{Tr}[\boldsymbol{\Lambda}_{\text{low},\mathbf{0}}(\mathbf{P}^{\mathbf{2}}_{\text{low},\mathbf{0}} - \mathbf{P}_{\text{low},\mathbf{0}})] \quad (3)$$

In this equation, **R**, **V**, and **M** are the nuclear geometry, velocity, and mass, respectively, and the single particle density

DOI: 10.1021/acs.jctc.6b00001
J. Chem. Theory Comput. 2016, 12, 2493−2508

matrix $\mathbf{P}_{\mathbf{low,0}}$ that determines $E_{\mathrm{low}}(0)$ has fictitious inertia tensor $\boldsymbol{\mu}_{\mathrm{low,0}}$, and velocity $\mathbf{W}_{\mathbf{low,0}}$. The quantity $\boldsymbol{\Lambda}_{\mathrm{low,0}}$ is a Lagrangian multiplier that maintains the N-representability of $\mathbf{P}_{\mathbf{low,0}}$. The energy functional $E^{\mathrm{PIE\text{-}ONIOM}}(\mathbf{R},\mathbf{P}_{\mathbf{low,0}})$ is identical to that in eq 1 but differs with the additional dependence of $E_{\mathrm{low}}(0) \rightarrow E_{\mathrm{low}}(0)(\mathbf{R},\mathbf{P}_{\mathbf{low,0}})$ since the lower level full system electronic structure is now to be propagated rather than converged. This leads to the equations of motion for nuclei and density matrix

$$\mathbf{M}\frac{\partial^2\mathbf{R}}{\partial t^2} = -\left.\frac{\partial E^{\mathrm{PIE\text{-}ONIOM}}(\mathbf{R},\,\mathbf{P}_{\mathrm{low,0}})}{\partial \mathbf{R}}\right|_{\mathbf{P}_{\mathrm{low,0}}} \tag{4}$$

and

$$\begin{aligned}\mu_{\mathrm{low,0}}^{1/2}\frac{\partial^2\mathbf{P}_{\mathrm{low,0}}}{\partial t^2}\mu_{\mathrm{low,0}}^{1/2} &= -\left.\frac{\partial E^{\mathrm{PIE\text{-}ONIOM}}(\mathbf{R},\,\mathbf{P}_{\mathrm{low,0}})}{\partial \mathbf{P}_{\mathrm{low,0}}}\right|_{\mathbf{R}} - \boldsymbol{\Lambda}_{\mathrm{low,0}}\mathbf{P}_{\mathrm{low,0}} \\ &\quad - \mathbf{P}_{\mathrm{low,0}}\boldsymbol{\Lambda}_{\mathrm{low,0}} + \boldsymbol{\Lambda}_{\mathrm{low,0}}\end{aligned} \tag{5}$$

where

$$\begin{aligned}\left.\frac{\partial E^{\mathrm{PIE\text{-}ONIOM}}(\mathbf{R},\,\mathbf{P}_{\mathrm{low,0}})}{\partial \mathbf{R}}\right|_{\mathbf{P}_{\mathrm{low,0}}} &\equiv \left.\frac{\partial E_{\mathrm{low}}(0)}{\partial \mathbf{R}}\right|_{\mathbf{P}_{\mathrm{low,0}}} \\ &\quad + \sum_{i=1}^{n}\frac{\partial \mathcal{S}(i)}{\partial \mathbf{R}} - \sum_{1<=i<j<=n}\frac{\partial \mathcal{S}(i\cap j)}{\partial \mathbf{R}} + \cdots\end{aligned} \tag{6}$$

and

$$\left.\frac{\partial E^{\mathrm{PIE\text{-}ONIOM}}(\mathbf{R},\,\mathbf{P}_{\mathrm{low,0}})}{\partial \mathbf{P}_{\mathrm{low,0}}}\right|_{\mathbf{R}} \equiv \left.\frac{\partial E_{\mathrm{low}}(0)}{\partial \mathbf{P}_{\mathrm{low,0}}}\right|_{\mathbf{R}} \tag{7}$$

In eq 6, the constraint, "$|\mathbf{P}_{\mathbf{low,0}}$", is only included on the first term on the right side since that is the only term that is a function of $\mathbf{P}_{\mathbf{low,0}}$. The terms $\frac{\partial \mathcal{S}(i)}{\partial \mathbf{R}}$, $\frac{\partial \mathcal{S}(i\cap j)}{\partial \mathbf{R}}$, etc., involve standard electronic structure gradients that are already present in most electronic structure packages. This is because, as noted in eq 2, these terms involve gradients from higher level electronic structure methods (MP2 for the calculations presented in here), and when DFT is used for the lower level in eq 2, the SCF is converged to a high degree of accuracy. In this publication, all SCF calculations involved in $\frac{\partial \mathcal{S}(i)}{\partial \mathbf{R}}$, $\frac{\partial \mathcal{S}(i\cap j)}{\partial \mathbf{R}}$, are well converged requiring the RMS change in density matrix to be less than 10^{-8} *and* the maximum change in the density matrix to be less that 10^{-6}. These are the default threshold conditions used in the version of the Gaussian quantum chemistry package[81] used here to compute the electronic energy and nuclear gradients for the fragments. The energy change during SCF is not used to test convergence, but a 10^{-N} RMS change in density matrix typically corresponds to a 10^{-2N} change in energy in atomic units. While in principle this translates to an energy convergence threshold of 10^{-16} Hartree, in practice we find that SCF convergence is always better than 10^{-12} Hartree in most calculations presented here, with very few at 10^{-10}. (A more detailed analysis on this aspect is presented in Section III.) As a result of this, there is practically no drift in these calculations, which appears to be a problem for Born−Oppenheimer-type dynamics calculations in ref 82.

Proceeding further with the discussion of the equations of motion, since $\mathbf{P}_{\mathbf{low,0}}$ is propagated according to eq 5, the gradient term $\left.\frac{\partial E_{\mathrm{low}}(0)}{\partial \mathbf{R}}\right|_{\mathbf{P}_{\mathrm{low,0}}}$ in eq 6 contains forces from the unconverged density matrix, $\mathbf{P}_{\mathbf{low,0}}$, in addition to the standard Born−Oppenheimer gradients, and these aspects have been well described in refs 48, 49, and 51. We summarize these gradients in Appendix A since the major portion of these forces have already been discussed in other publications noted above. It is only the interpretation of these gradients with respect to the fragment-based technology that is new here. The equations of motion are integrated using the velocity Verlet algorithm,[83] while the Lagrangian multiplier matrices are determined through an iterative procedure.[48,49] In the sections below, we probe the accuracy and efficiency of this hybrid scheme by studying protonated water clusters and polypeptides fragments. In all cases, consistency of the dynamics is checked by computing the conserved quantities, and these conserved quantities (the conjugate Hamiltonian for the Lagrangian in eq 3) are outlined in Appendix B.

However, before we proceed further, we must note and distinguish our approach from two other recent developments in Born−Oppenheimer and extended Lagrangian dynamics. In ref 84, the author has carefully analyzed the time scale of oscillations of the different constituents of energy in a post-Hartree−Fock (MP2) gradient expression. The author infers through many computational studies that the post-Hartree−Fock portion of the gradient is in fact a slowly varying quantity as a function of nuclear motion, which allows the authors to construct a multiple time step approach that appears to be quite powerful. This extension is certainly an option available to us here and will be considered in future publications. In ref 82, the authors address the critical drift problem in standard Born−Oppenheimer dynamics by propagating a surrogate electronic density through an extended Lagrangian formalism, where such a density is harmonically bound to a converged density. This latter approach has extended Lagrangian and Born−Oppenheimer components, much like the current formalism, but differs in that the ideas in ref 82 are thus far implemented for DFT and DFTB, whereas the approach here is specifically geared toward computing post-Hartree−Fock dynamics. Thus, to differentiate between the two approaches, from hereon, the method developed in this paper has both extended Lagrangian and post-Hartree−Fock Born−Oppenheimer dynamics components funneling from the ADMP Lagrangian in Eq. (3), and is hence referred to as ADMP-pHF, and the PIE-ONIOM energy functional, eq 1, is implied. When the extended Lagrangian treatment is not invoked and Born−Oppenheimer treatment is used for all components of the energy expression in eq 1, the associated method is referred to here as frag-BOMD.

III. ACCURACY AND EFFICIENCY OF THE HYBRID EXTENDED LAGRANGIAN, POST-HARTREE−FOCK BORN−OPPENHEIMER DYNAMICS (ADMP-PHF) TREATMENT

We benchmark the dynamics formalism by studying (small- and medium-sized) protonated water clusters and two different polypeptide fragments. We choose two challenging

DOI: 10.1021/acs.jctc.6b00001
J. Chem. Theory Comput. 2016, 12, 2493−2508

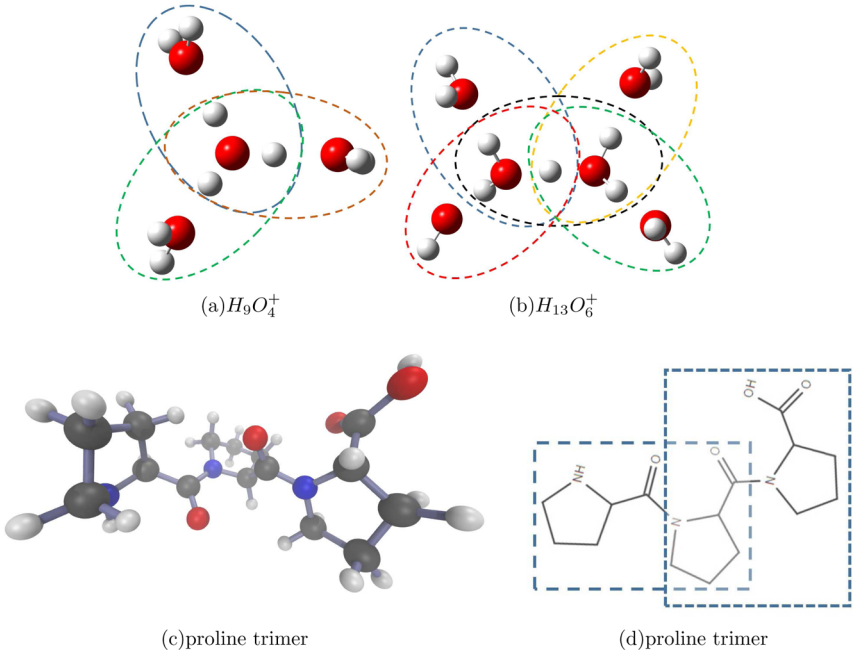


Figure 4. Protonated water clusters shown in Figures 1(a) and 2(b) are fragmented as depicted here, and the electronic energy is assembled using eq 1. The proline tripeptide (c) is similarly fragmented using overlapping dipeptides (d). Dangling bonds are saturated with link atoms (hydrogen) as in standard ONIOM.

protonated water clusters, the Eigen cation ($H_9O_4^+$) and the solvated Zundel cation ($H_{13}O_6^+$), as our benchmark systems with their structures shown in Figure 4(a) and (b). These type of systems have significance in biological,[85−92] atmospheric,[69,93−95] and condensed phase chemistry and have been studied extensively by both experiments[96−104] and by theories.[61,62,64,70,71,97,105−120] Here, accurate treatment of hydrogen bonding is crucial to model structural and dynamical behavior. In addition, polypeptide fragments have recently been studied in some detail using mass spectrometry with a goal to elucidate conformational transformations and kinetics.[121,122] For example, in refs 121 and 123, the authors observe a step-by-step conformational transition from an all-cis polyproline helix chain to an all-trans polyproline helix chain. While our primary benchmarks here that test the accuracy of ADMP-pHF and frag-BOMD through comparison with full system MP2 calculations are performed for the protonated water clusters stated above, we also present preliminary benchmarks on an all-trans proline tripeptide fragment (Figure 4(c)) to show the consistency of our approach and its potential for future applications to the study of biomolecules. In addition, we also present an analysis of the computational gain from the fragment-based ADMP-pHF and frag-BOMD formalisms by computing trajectories for an all-trans polyalanine fragment.

At this point, it is critical to note that for the peptide systems, chemical bonds intersect the boundary between two fragments, and hence, link atoms are added to saturate the dangling valencies as in standard ONIOM.[31,54,124] The positions of link atoms are uniquely determined based on the connectivity of the system. Specifically, the placement of link atoms is constrained by the positions of the two atoms on either ends of the bond intersecting the fragment boundary according to $\mathbf{r}_{link} = \mathbf{r}_{bond} + g(\mathbf{r}_{sub}-\mathbf{r}_{bond})$, where $\mathbf{r}_{bond}$ is the position of the bonded atom (B) inside the fragment in question to which the link is bound to substitute the atom (S) belonging to the surrounding region and located at $\mathbf{r}_{sub}$. The quantity g is a scale factor normally set to $R(B-L)/R(B-S)$, where $R(B-L)$ is a typical bond length between the bonded atom within the fragment and the link atom (hydrogen in all cases in this work), and $R(B-S)$ is a typical bond length for the pair of atoms describing the bond being dissected to create the fragment. Careful tests have shown that the ONIOM results are quite insensitive to the value of g.[37] It must be stressed that since the link atom positions are fully defined by the coordinates of the full system, as in standard ONIOM, the gradients with respect to link atom coordinates can be transferred to coordinates of the full system for this fragment-based methodology.[54]

All calculations are performed using a C++ package, that is, MPI parallelized across nodes and OpenMP parallelized within each node. For the current publication, the package invokes a development version of the Gaussian quantum chemistry program[81] to compute the electronic energy and nuclear gradients for the fragments and the full system. However, our software package can be easily generalized to include other electronic structure software or combinations of these. These computational aspects and associated generalizations will be discussed in future publications.

The dynamical simulations are performed five different ways for each system with all calculations using the 6-31+g(d,p) atom-centered Gaussian basis set. The five different sets of trajectories are (1) BOMD using pure B3LYP (abbreviated as BOMD B3LYP), (2) ADMP using pure B3LYP (abbreviated as ADMP B3LYP), (3) BOMD using pure MP2 (abbreviated as BOMD MP2), and (4) BOMD using two-layer PIE-ONIOM fragment-based electronic structure. MP2 is used for the high level electronic structure, while the low level is calculated at B3LYP. This calculation is abbreviated as frag-BOMD MP2:B3LYP and was fully benchmarked in ref 32. Also see Figures 1 and 2 where we present the accuracy of this calculation. Finally, (5) the two-

DOI: 10.1021/acs.jctc.6b00001
J. Chem. Theory Comput. 2016, 12, 2493−2508

Table 1. Energy Conservation Properties for Dynamical Simulations (NVE Simulations)

system	level of theory[a]	time[b]	ave temp (K)	$\Delta\mathcal{H}$[c] (in kcal/mol)	drift[d] (in kcal/mol)	ΔE_{SCF}[e] (in Hartree)
$H_9O_4^+$	BOMD B3LYP	20.0 ps	150.90 ± 27.0	0.022	0.062	1.74×10^{-11}
	ADMP B3LYP	20.0 ps	140.34 ± 23.7	0.021	0.018	
	BOMD MP2	20.0 ps	152.10 ± 26.5	0.002	0.004	6.38×10^{-12}
	frag-BOMD MP2:B3LYP	5.2 ps	149.64 ± 32.3	0.014	0.016	4.08×10^{-12}
	ADMP-pHF MP2:B3LYP	8.6 ps	142.80 ± 24.7	0.027	0.055	3.34×10^{-12}
$H_{13}O_6^+$	BOMD B3LYP	20.0 ps	151.92 ± 20.8	0.010	0.011	2.75×10^{-11}
	ADMP B3LYP	13.4 ps	154.00 ± 21.0	0.040	0.021	
	BOMD MP2	5.2 ps	150.52 ± 22.6	0.004	0.001	1.32×10^{-11}
	frag-BOMD MP2:B3LYP	6.0 ps	151.22 ± 23.4	0.042	0.122	6.05×10^{-12}
	ADMP-pHF MP2:B3LYP	5.0 ps	156.44 ± 20.0	0.026	0.033	2.93×10^{-12}
proline trimer[f]	ADMP B3LYP	4.33 ps	237.96 ± 24.8	0.069	0.148	
	BOMD B3LYP	2.18 ps	238.09 ± 23.7	0.016	0.011	8.41×10^{-12}
	frag-BOMD MP2:B3LYP	2.97 ps	246.30 ± 23.5	0.012	0.002	2.96×10^{-12}
	ADMP-pHF MP2:B3LYP	2.90 ps	246.07 ± 22.5	0.047	0.088	1.64×10^{-11}

[a]Water clusters: 6-31+g(d,p) basis is used for B3LYP and MP2. Proline trimer: 6-31+g(d) basis is used for B3LYP and MP2. [b]Time step 0.2 fs for all simulations. [c]Root mean square deviation of the total energy in kcal/mol. The total energy corresponds to the Hamiltonian in eq B4. [d]Drift in total energy is computed as a difference between the average total energies for the first and last 100 fs of dynamics (in kcal/mol). In addition, in Figure SI-1 of the Supporting Information, we provide the evolution of total energy. This is provided in eV to facilitate comparison with results from other work such as ref 82. [e]While SCF convergence thresholds are placed on the density matrix (see text for details) rather than energy, in this column we note the average absolute change in energy (in Hartrees) during the last interation before convergence is achieved for all steps in the trajectory. [f]C-terminus is capped with a carboxylic acid group (see Figure 4(c)).

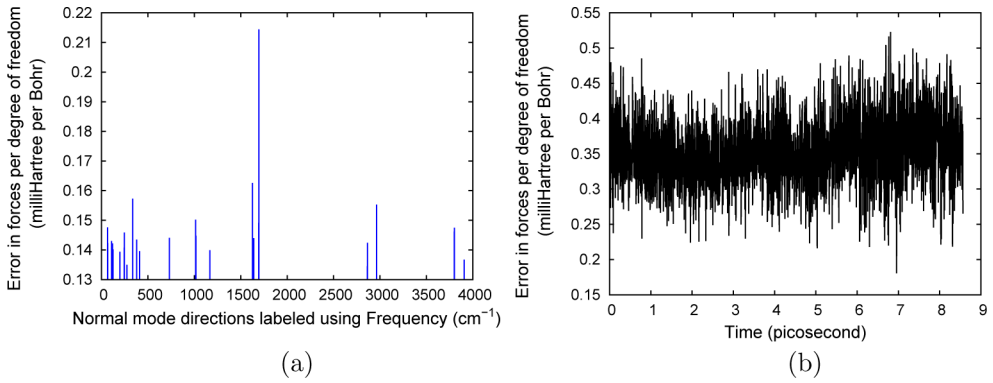


Figure 5. Comparison of forces around an optimized geometry (a) and during dynamics (b) for $H_9O_4^+$. See text for details.

layer PIE-ONIOM fragment-based ADMP-pHF scheme proposed in this publication. The methods MP2 and B3LYP are used for high and low level of theory, respectively, and the calculation is abbreviated as ADMP-pHF MP2:B3LYP.

The stability and consistency of the dynamics trajectories can be gauged from the total energy conservation data for these classical trajectories. The total energy corresponds to the conjugate Hamiltonian in eq B4 and contains the nuclear kinetic energy, the fictitious density matrix kinetic energy, $E^{\text{PIE-ONIOM}}$, and any energy penalty that is added as a result of residual loss of N-representability. As shown in Table 1, all trajectories have the total energy well conserved to within 2−3 orders of magnitude less than 1 kcal/mol. The ADMP-pHF scheme produces energy conservation comparable to the BOMD simulations. We also inspect the drift in the conjugate Hamiltonian in eq B4, and as noted in Table 1, this quantity is also extremely small and well within the acceptable range.[82] In addition, in Table 1, we also present the change in SCF energy in the last iteration before convergence is achieved. This quantity in some sense represents a convergence threshold of the electronic energy. As noted previously, convergence is enforced on the density matrix and not on the SCF energy. In this case, it is required that the RMS change in density matrix be less than 10^{-8} *and* the maximum change in the density matrix be less that 10^{-6}. The corresponding SCF energy convergence is very tight and is noted in Table 1.

In addition to gauging conserved quantities, we also compare forces in Figure 5. Specifically, in Figure 5(a) we compute the forces for 693 different geometries chosen along normal mode coordinates where the horizontal axis in the figure represents mode frequencies. The forces are computed at the level of MP2 and also using the MP2:B3LYP fragment method. The difference $\frac{\| F_{\text{MP2:B3LYP}} - F_{\text{MP2}} \|_2}{3N_{\text{Atoms}}}$ is presented as a function of harmonic frequency. In addition, for the analysis in Figure 5(b), we have retrieved geometries at fixed intervals from the extended Lagrangian Born−Oppenheimer trajectory for $H_9O_4^+$ and computed forces using MP2. The vertical axis in Figure 5(b) is the mean absolute deviation between these forces. As shown, the forces in both figures are well within that expected for chemical accuracy.

IV. STRUCTURAL FEATURES FROM DYNAMICS SIMULATIONS

Next, we examine the structural properties of these trajectories by calculating radial distribution functions for

DOI: 10.1021/acs.jctc.6b00001
J. Chem. Theory Comput. 2016, 12, 2493−2508

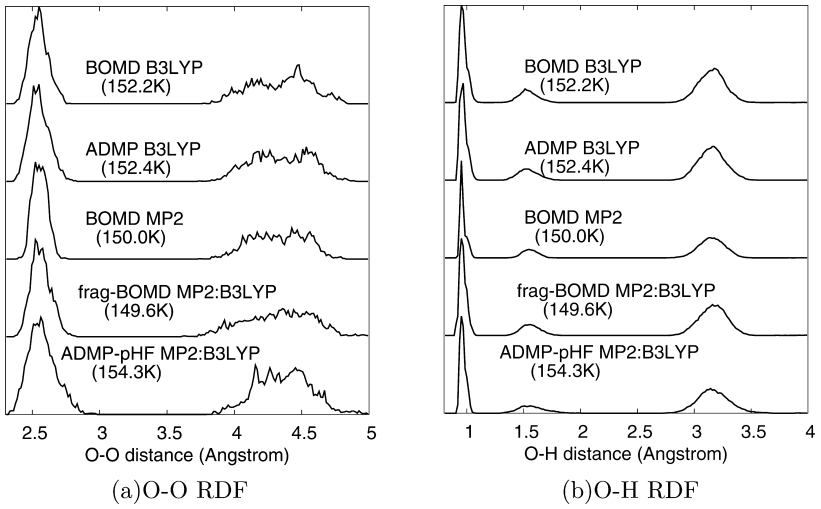


Figure 6. Radial distribution function for O−O and O−H distances from various $H_9O_4^+$ trajectories The statistics are obtained by constructing averages from incremental 2 ps snapshots constructed from the full trajectory. That is, we compute the RDFs using data from 0 to 2 ps, 0.5 to 2.5 ps, and so on, for all trajectories. The resultant distributions are then averaged; standard deviations are presented as error bars in Figure SI-2 of the Supporting Information.

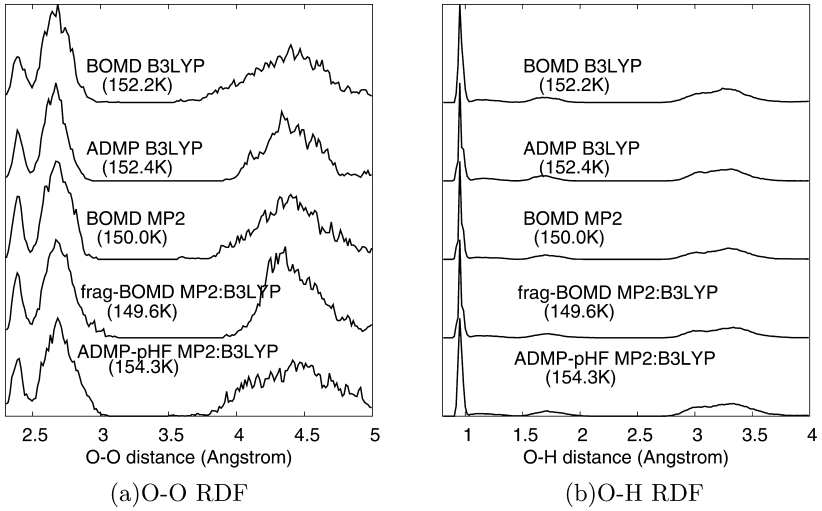


Figure 7. Radial distribution function for O−O and O−H distances from various $H_{13}O_6^+$ trajectories are obtained using the same procedure described in the caption for Figure 6. Error bars can be found in Figure SI-3 of the Supporting Information.

the O−O and O−H distances for the protonated water clusters. From Figures 6 and 7, it is clear that the structural features for the water clusters are captured very well by both ADMP-pHF and frag-BOMD fragment-based dynamics in comparison with the higher level MP2 calculations. In addition, for the protonated water clusters, these calculations are also in agreement with the plain DFT calculations. Specifically, Figures 6(a) and 7(a) clearly show the multiple solvation shells, where the width of the distributions indicate fluctuations that arise from temperature. The inner shells have better ordering in all cases, based on width of the distributions, and this is to expected. The most critical aspect from Figures 6(a) and 7(a) is the fact that the inner shell is tighter in Figure 7(a) with a shorter OO distance. This is due to the strong Zundel-like hydrogen bond present in $H_{13}O_6^+$. Thus, Figure 7(a) shows both Zundel-like and Eigen-like (OO distance ≈2.8 Å) peaks, and this is critically reinforced in all simulations. The peaks beyond 3.7 Å in Figures 6(a) and 7(a) represent the second solvation shell that is generally not expected to be well resolved for these smaller water clusters. Nevertheless, the agreement is reasonable here too but not as good as in the first solvation shell. The OH distributions in Figure 6(b) similarly show an Eigen-like OH distance of approximately 1.5 Å, which informs an asymmetric hydrogen bond; the two shortest OH distributions in Figure 6(b) are consistent with the OO distribution for the first peak in Figure 6(a). Figure 7(b) shows four regions of OH distribution. The covalent OH bond is strongly noted and is followed immediately by a shallow hump at a distance slightly greater than 1 Å. This distance is again the short-strong Zundel-like hydrogen bond feature. Followed by this, we see a peak at a little greater than 1.5 Å, which is consistent with an Eigen geometry. Thus, the $H_{13}O_6^+$ system is critical in that it shows both Zundel and Eigen configurations that are now universally acknowledged as the two central excess proton configurations in water that participate in proton transport.[125] Importantly, all simulations, including ADMP-

DOI: 10.1021/acs.jctc.6b00001
J. Chem. Theory Comput. 2016, 12, 2493−2508

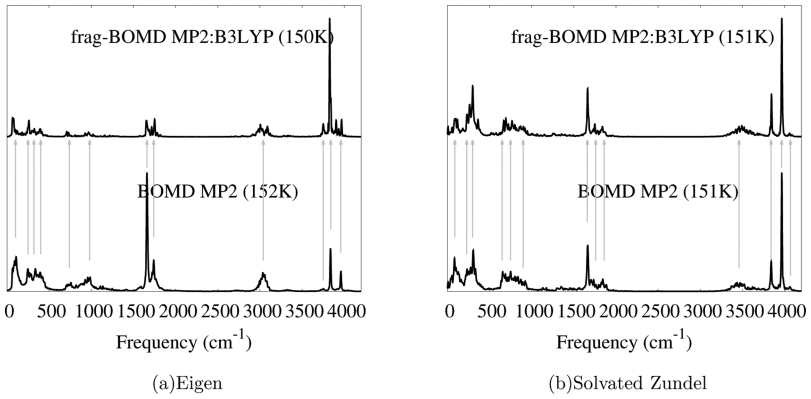


Figure 8. Vibrational density of states calculated from dynamical trajectories at 150 K for (a) $H_9O_4^+$ and (b) $H_{13}O_6^+$. The excellent agreement in frequencies is highlighted using the light colored arrows. In the Supporting Information, we provide details on the statistics and compute error bars.

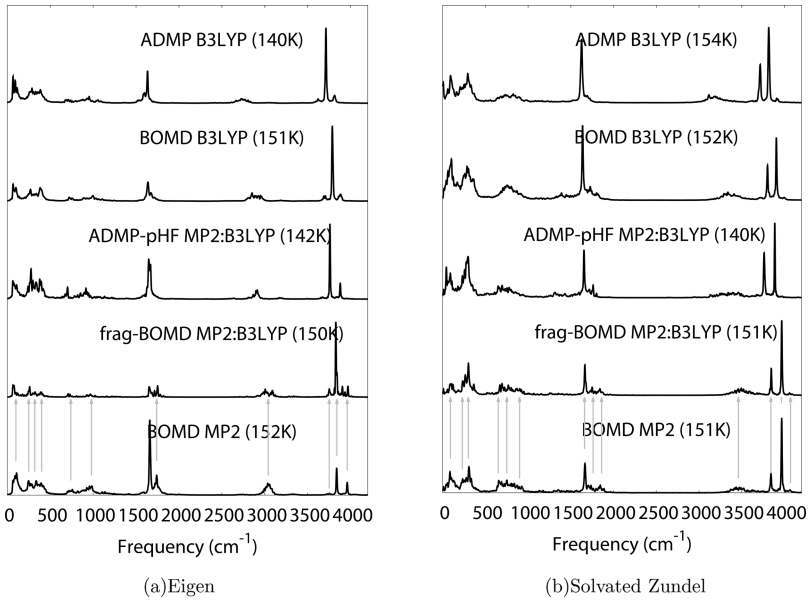


Figure 9. Vibrational density of states calculated from dynamical trajectories at 150 K for (a) $H_9O_4^+$ and (b) $H_{13}O_6^+$. The extended Lagrangian formalism yields a slight redshift in the higher frequency peaks greater than 2500 cm^{-1} but is in reasonably good agreement in the lower frequency region. The redshift is reduced as compared to ADMP B3LYP, which does not include the MP2 treatment. In the Supporting Information, we provide details on the statistics and compute error bars.

pHF, reproduce the structural features present in the MP2 BO dynamics.

This brings in the critical question of computational efficiency of fragment-based AIMD methods. We discuss these aspects in Section VI.

V. DYNAMICAL PROPERTIES FROM CORRELATION FUNCTIONS

Vibrational density of states (VDOS) are calculated for each trajectory from the Fourier transform of the velocity autocorrelation function. As we know from previous studies,[61−63,126] the temperature can play a significant role in VDOS spectra since this controls the extent to which anharmonicity is sampled during these classical nuclear trajectory calculations. Hence, we tune the initial conditions such that all trajectories have an average temperature of about 150 K. This is also roughly the temperature of argon-tagged action spectroscopy[127] as seen in previous calculations.[62,65] The VDOS spectra are presented in Figures 8 and 9. To obtain these figures, we compute multiple spectra using 3 ps segments from dynamics that are each shifted by 500 fs. That is, we compute the vibrational density of states using data from 0 to 3 ps, 0.5 to 3.5 ps, and so on, for all trajectories. These spectra are then used to compute averages and standard deviations. The averages are presented in Figures 8 and 9, and the standard deviations are shown using error bars in Figures SI-4 through SI-8 of the Supporting Information. Specifically, in Figure 8 of the main text and Figures SI-4 through SI-6, we compute the vibrational density of states obtained from full MP2 and frag-BOMD MP2:B3LYP, and the fact that *the frequencies are in excellent agreement all through the spectral range is shown using arrows to guide the eye.* In Figure 9 of the main text and Figures SI-7 and SI-8, we provide the same for the full breadth of simulations. (Also note that to check convergence, we also computed the same using 2 ps snapshots. That is, we also computed the vibrational density of states using data from 0 to 2 ps, 0.5 to 2.5 ps, and so on, for all trajectories. No perceivable

DOI: 10.1021/acs.jctc.6b00001
J. Chem. Theory Comput. 2016, 12, 2493−2508

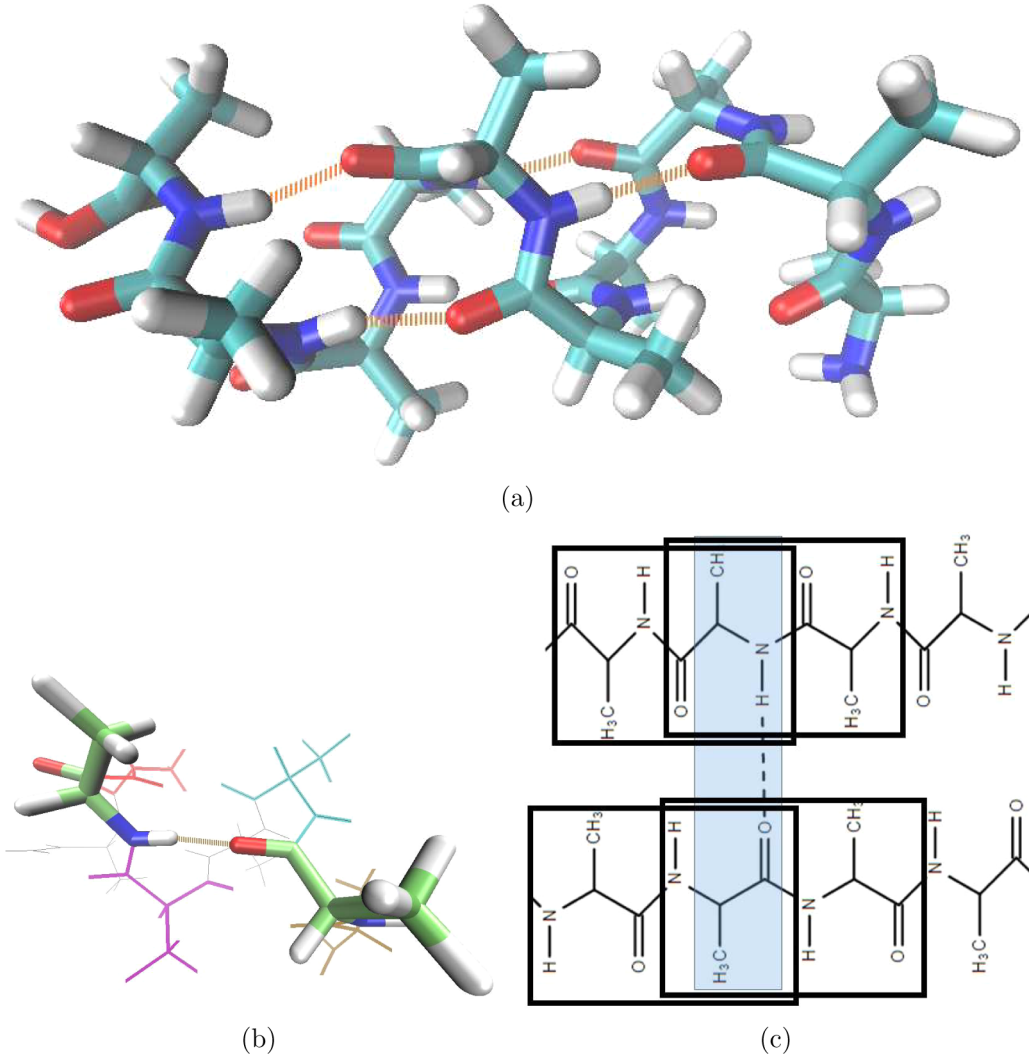


Figure 10. All-trans alanine decamer. (a) Helical view with hydrogen bonds shown. (b) and (c) Portion of the polypeptide with fragments labeled. For example the primary fragment shown in licorice in (b) includes the CO−HN hydrogen bond, and it has an overlap with the purple, red, cyan, and tan fragments. This is also shown in (c). (The secondary overlapping fragments are automatically generated using a bit-manipulation algorithm presented in ref 32.) In this manner, critical hydrogen bonds responsible for the helix formation are captured as part of the fragmentation mechanism.

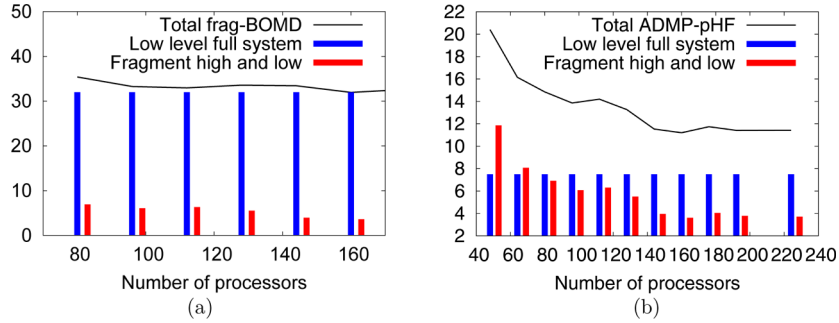


Figure 11. All calculations are performed on Intel Xeon E5-2650 v2 processors. The vertical axis is in units of CPU minutes. See text for details.

difference was noted from the spectra presented here.) It is clear from these figures that the DFT results are redshifted with respect to those from MP2. This behavior is drastically improved upon in Figure 8 when the MP2 results are compared with those from MP2:B3LYP fragment-based BOMD. This is perhaps the most significant message here: that the fragment-based approaches are able to provide results in quantitative agreement with higher level electronic structural results. The ADMP-pHF result is redshifted with respect to the MP2 BOMD and MP2:B3LYP fragment-based BOMD, but this redshift is substantially reduced as compared to the respective ADMP B3LYP redshifts seen in Figure 9. Furthermore, it appears that these redshifts are more localized to the higher frequencies as note in Figures SI-7 and SI-8.

DOI: 10.1021/acs.jctc.6b00001
J. Chem. Theory Comput. 2016, 12, 2493−2508

This is to be expected since one of the central ideas behind ADMP and similar extended Lagrangian formalisms is that the electronic dynamics is mapped onto a set of fictitious variables that obey classical-like dynamics. The mapping is enforced by including an inertia tenor as outlined above, and the choice of this inertia tensor does have a role on the higher frequency correlation functions. This aspect has also been noted in previous publications.[50,55] At the end, it appears the structural properties from ADMP-pHF are quantitatively accurate, and correlation functions are also accurate in the lower end of the spectrum. Future work will deal with improvement of these methods in the higher frequency regimes. Fragment-based BOMD, on the contrary, has accuracy that is very much comparable to standard full post-Hartree−Fock BOMD (with lesser computational overhead) and hence appears to be well suited for production level calculations.

VI. COMPUTATIONAL EFFICIENCY OF THE ADMP-pHF SCHEME

The system studied to demonstrate computational efficiency is the all-trans alanine decamer shown in Figure 10. The full system MP2 timings are prohibitive for this case.

As noted in the histogram plots presented in Figure 11, the frag-BOMD and ADMP-pHF simulations contain three separate parts: (i) the DFT energy and gradient calculation for the full system, (ii) the DFT energy and gradient calculations for the fragments, and (iii) MP2 energy and gradient calculations for fragments. The parallelism for the fragment-based frag-BOMD simulations is organized such that the full system low level calculation is performed on a separate node, comprising 16 shared memory processors, in parallel with the fragment high and low level pairs of jobs that are spawned to other similar nodes. The parallelism for the fragment-based ADMP-pHF is organized in a slightly different manner. Here, the full system low level job is performed on a separate node (the parent node), and the fragment high and low level pairs of jobs are spawned to other similar nodes *after* completion of the low level calculation. This leads to a slight inefficiency in the current ADMP-pHF implementation, which will be corrected in future versions of the code, but as is already clear from the discussion below, the ADMP-pHF calculation is quite fast.

In Figure 11(a) and (b), we present the total time per dynamics step (black curve), the CPU time required for the full system (low level) DFT calculation (blue histograms), and the average CPU time required for both low and high level calculations for each fragment (red histograms). In both figures, the CPU time required for the full system low level (blue) is flat because this calculation is performed on a single node with 16 processors for both ADMP-pHF and frag-BOMD. Further reduction in the total time step can be obtained for both ADMP-pHF and frag-BOMD by increasing the number of processors for full system low level, but since this is independent of the fragmentation algorithm (and is dependent on the restrictions on implementation of the underlying electronic structure software), we only use a fixed number of processors here. Thus, all electronic structure energy and gradient calculations are performed on 16 processors in this benchmark. In general, one would expect a similar scaling behavior for any number of processors (assuming the electronic structure method and program used parallelizes appropriately).

The next critical observation in Figure 11(a) is that the full system calculation essentially forces the total CPU time to flatten out at 80 processors and beyond, and the resources required for the full system calculation are nearly a factor of 4 greater than that for the fragment calculations. This is a major bottleneck for fragment-based frag-BOMD when applied to large systems. The ADMP-pHF method, as shown from the blue histogram in Figure 11(b), reduces the time required for the low level calculation by nearly a factor of 4, and now the full system and fragment calculations require comparable resources. This allows for better load balance and higher code efficiency for a larger number of processors and directly translates toward the greater efficiency for the ADMP-pHF scheme. In addition, one notices that the total CPU time (black curve) never approaches the blue histogram in Figure 11(b) unlike the case in Figure 11(a). This is partially due to the code inefficiency mentioned above, where the low level calculation is performed before the fragment calculations are spawned. Despite this inefficency, the scaling is better for ADMP-pHF, and this scaling is expected to further improve in future versions of the computer program. Furthermore, at larger numbers of processors (greater than 200), the low level calculation again uses resources that are nearly a factor of 4 greater than the fragment calculation. This will require that the lower level calculations be performed over a larger number of processors (that is, repartitioning the node allocation), which should further improve the scaling for number of processors greater than 200.

Given that structural features are well preserved for strongly hydrogen bonded systems and for peptides, the ADMP-pHF implementation is expected to have an advantage in large scale simulations.

VII. CONCLUDING REMARKS

In this publication, we introduce a hybrid extended Lagrangian, post-Hartree−Fock Born−Oppenheimer dynamics approach, where the two schemes are combined through a recently developed[32] fragment-based electronic structure formalism[32] that employs the set-theoretic principle of inclusion−exclusion[42] to generalize the well-known ONIOM[5] approach. Here, a large system is first treated using DFT and then recursively partitioned into overlapping subsections where the electronic description is improved through higher levels of electronic structure treatment; the intersection regions between the overlapping subsections are appropriately corrected as allowed by the principle of inclusion−exclusion in conjunction with ONIOM. The hybrid dynamical method described here propagates the electronic structure for the full system by an extended Lagrangian scheme, whereas the fragments are treated using converged post-HF methods. This significantly reduces the cost for electronic structure calculations on the full system and will be beneficial for ab initio dynamics treatment of large systems where repeated evaluation of energy and gradients of the entire system may be computationally prohibitive.

This approach is benchmarked on a set of well-studied protonated water clusters, $H_9O_4^+$ and $H_{13}O_6^+$ and polypeptide fragments: proline tripeptide and alanine polypeptide. All dynamics trajectories show good energy conservation in the subkcal/mol range. The structural features of the protonated water clusters show that the fragment-based ADMP-PHF approach provides results in good agreement with BO dynamics at MP2 level of theory. The analysis of vibrational

DOI: 10.1021/acs.jctc.6b00001
J. Chem. Theory Comput. 2016, 12, 2493−2508

density of states also indicates large improvements in ADMP-pHF over pure extended Lagrangian results at the lower level of theory; furthermore, ADMP-pHF provides better agreement with the spectra calculated from higher level Born–Oppenheimer dynamics. In all cases the computational gain in frag-BOMD is orders of magnitude greater than the full MP2 calculations, and ADMP-pHF improves further on this performance. The algorithm is written in parallel and potentially allows AIMD treatment of large systems. Two challenges remain in fragment-based ab initio molecular dynamics: (a) The topology of fragmentation could change during dynamical simulations; thus, an adaptive fragmentation and energy extrapolation scheme needs to be developed in order to study processes involving structural rearrangements, such as proton transfer reactions and conformational changes in polypeptides. (b) The calculation of forces introduced by electronic embedding could be expensive,[43,44] and computational feasibility studies are required. These aspects will be probed in future publications.

APPENDIX A: NUCLEAR AND DENSITY MATRIX GRADIENTS ASSOCIATED WITH EQS 4 AND 5

Since $\mathbf{P}_{\mathbf{low},0}$ is propagated according to eq 5, the gradient term $\frac{\partial E_{\mathrm{low}}(0)}{\partial \mathbf{R}}\Big|_{\mathbf{P}_{\mathbf{low},0}}$ in eq 6 contains forces from the unconverged density matrix in addition to the standard Born–Oppenheimer gradients, and these aspects have been well described in refs 48 and 51. We provide the resulting gradient expressions here, starting from eq 35 in ref 51. Note that this expression is identical to eq 14 in ref 48. (Also see the discussion following eqs 14 and 15 in ref 48.)

$$\frac{\partial E_{\mathrm{low}}(0)}{\partial \mathbf{R}_{i,j}}\Big|_{\mathbf{P}_{\mathbf{low},0}} = \frac{\partial E^{\mathrm{BO}}_{\mathrm{low}}(0)}{\partial \mathbf{R}_{i,j}}\Big|_{\mathbf{P}_{\mathbf{low},0}} + \mathrm{Tr}\left[[\tilde{\mathbf{P}}_{\mathbf{low},0}, \mathbf{F}_{\mathbf{low},0}]\left(\tilde{\mathbf{Q}}_{\mathbf{low},0}\frac{d\mathbf{U}_{\mathbf{low},0}}{d\mathbf{R}_{i,j}}\mathbf{U}^{-1}_{\mathbf{low},0} - \tilde{\mathbf{P}}_{\mathbf{low},0}\mathbf{U}^{-\mathrm{T}}_{\mathbf{low},0}\frac{d\mathbf{U}^{\mathrm{T}}_{\mathbf{low},0}}{d\mathbf{R}_{i,j}}\right)\right] \tag{A1}$$

where in the first equation $\tilde{\mathbf{Q}}_{\mathbf{low},0} = \mathbf{I} - \tilde{\mathbf{P}}_{\mathbf{low},0}$ and $\tilde{\mathbf{P}}_{\mathbf{low},0} = 3\mathbf{P}^2_{\mathbf{low},0} - 2\mathbf{P}^3_{\mathbf{low},0}$ is the McWeeny purified density matrix.[128] The quantity, $\mathbf{U}_{\mathbf{low},0}$ transforms the density matrix from the nonorthogonal basis to an orthogonal basis. That is, the nonorthogonal Gaussian basis overlap matrix, $\mathbf{S}'_{\mathbf{low},0}$ is factorized as $\mathbf{S}'_{\mathbf{low},0} \equiv \mathbf{U}^{\mathrm{T}}_{\mathbf{low},0}\mathbf{U}_{\mathbf{low},0}$. For Löwdin symmetric orthonormalization, $\mathbf{U}_{\mathbf{low},0} \equiv (\mathbf{S}'_{\mathbf{low},0})^{1/2}$, and for Cholesky decomposition $\mathbf{U}_{\mathbf{low},0}$ is upper triangular. The quantity $\mathbf{U}_{\mathbf{low},0}$ may also represent a singular value decomposition of $\mathbf{S}'_{\mathbf{low},0}$, and in this case, the square root of the singular values ($\mathbf{S}'_{\mathbf{low},0}$ is positive semidefinite) is to be absorbed in to the left and right singular vectors of $\mathbf{S}'_{\mathbf{low},0}$ to define $\mathbf{U}_{\mathbf{low},0}$. The quantity, $\mathbf{F}_{\mathbf{low},0}$, is the Fock matrix for the whole system at the low level of theory in an ortho-normal basis and includes the exchange correlation function for density functional implementation. In the nonorthogonal basis, it has the form

$$\mathbf{F}'^{\nu,\sigma}_{\mathbf{low},0} \equiv \mathbf{h}'_{\nu,\sigma} + \mathbf{G}'(\tilde{\mathbf{P}}'_{\mathbf{low},0})_{\nu,\sigma} + \frac{\partial E_{\mathrm{xc}}}{\partial \mathbf{P}'^{\nu,\sigma}_{\mathbf{low},0}} \tag{A2}$$

where the matrix elements ν,σ, are listed as subscripts and superscripts to allow clarity. Here, the one-electron terms are $\mathbf{h}'_{\nu,\sigma}$, two-electron terms are $\mathbf{G}'(\tilde{\mathbf{P}}'_{\mathrm{low},0})^{\nu},\sigma$, and E_{xc} is the exchange correlation function. In addition, $\mathbf{P}'_{\mathbf{low},0}$ is the density matrix in the nonorthogonal basis. That is, $\mathbf{P}_{\mathbf{low},0} \equiv \mathbf{U}_{\mathbf{low},0}\ \mathbf{P}'_{\mathbf{low},0}\ \mathbf{U}^{\mathrm{T}}_{\mathbf{low},0}$, $\mathbf{h} = \mathbf{U}^{-\mathrm{T}}_{\mathbf{low},0}\ \mathbf{h}'\ \mathbf{U}^{-1}_{\mathbf{low},0}$, etc. The term $\frac{\partial E^{\mathrm{BO}}_{\mathrm{low}}(0)}{\partial \mathbf{R}_{i,j}}\Big|_{\mathbf{P}_{\mathbf{low},0}}$ in eq A1 represents the standard Born–Oppenheimer forces from ref 129

$$\frac{\partial E^{\mathrm{BO}}_{\mathrm{low}}(0)}{\partial \mathbf{R}_{i,j}}\Big|_{\mathbf{P}_{\mathbf{low},0}} = \mathrm{Tr}\left[\frac{d\mathbf{h}'}{d\mathbf{R}_{i,j}}\tilde{\mathbf{P}}'_{\mathbf{low},0} + \frac{1}{2}\frac{\partial \mathbf{G}'(\tilde{\mathbf{P}}'_{\mathbf{low},0})}{\partial \mathbf{R}_{i,j}}\Big|_{\mathbf{P}'_{\mathbf{low},0}}\tilde{\mathbf{P}}'_{\mathbf{low},0}\right] - \mathrm{Tr}\left[\mathbf{F}'_{\mathbf{low},0}\tilde{\mathbf{P}}'_{\mathbf{low},0}\frac{d\mathbf{S}'_{\mathbf{low},0}}{d\mathbf{R}_{i,j}}\tilde{\mathbf{P}}'_{\mathbf{low},0}\right] + \frac{\partial E_{\mathrm{xc}}}{\partial \mathbf{R}_{i,j}}\Big|_{\mathbf{P}_{\mathbf{low},0}} + \frac{\partial V_{NN}}{\partial \mathbf{R}_{i,j}} \tag{A3}$$

Equation A1 has been carefully derived and discussed in refs 48 and 51, but it is clear that additional terms exist in the forces when the density matrix is not fully converged. We may further simplify eq A1 as follows

$$\begin{aligned}\frac{\partial E_{\mathrm{low}}(0)}{\partial \mathbf{R}_{i,j}}\Big|_{\mathbf{P}_{\mathbf{low},0}} - \frac{\partial E^{\mathrm{BO}}_{\mathrm{low}}(0)}{\partial \mathbf{R}_{i,j}}\Big|_{\mathbf{P}_{\mathbf{low},0}} &= \mathrm{Tr}\left[[\tilde{\mathbf{P}}_{\mathbf{low},0}, \mathbf{F}_{\mathbf{low},0}]\left((1-\tilde{\mathbf{P}}_{\mathbf{low},0})\frac{d\mathbf{U}_{\mathbf{low},0}}{d\mathbf{R}_{i,j}}\mathbf{U}^{-1}_{\mathbf{low},0} - \tilde{\mathbf{P}}_{\mathbf{low},0}\mathbf{U}^{-\mathrm{T}}_{\mathbf{low},0}\frac{d\mathbf{U}^{\mathrm{T}}_{\mathbf{low},0}}{d\mathbf{R}_{i,j}}\right)\right] \\ &= \mathrm{Tr}\left[[\tilde{\mathbf{P}}_{\mathbf{low},0}, \mathbf{F}_{\mathbf{low},0}]\left(\frac{d\mathbf{U}_{\mathbf{low},0}}{d\mathbf{R}_{i,j}}\mathbf{U}^{-1}_{\mathbf{low},0} - \tilde{\mathbf{P}}_{\mathbf{low},0}\mathbf{U}^{-\mathrm{T}}_{\mathbf{low},0}\frac{d\mathbf{S}'_{\mathbf{low},0}}{d\mathbf{R}_{i,j}}\mathbf{U}^{-1}_{\mathbf{low},0}\right)\right]\end{aligned} \tag{A4}$$

It is seen from eq A1 that the difference between forces used as a result of the extended Lagrangian propagation and the standard BOMD forces (that were used in ref 32) is proportional to the commutator $[\tilde{\mathbf{P}}_{\mathbf{low},0}, \mathbf{F}_{\mathbf{low},0}]$, which is a measure of the convergence of the electronic wavefunction. This is a critical term that arises here due to the nonorthogonality of the basis and is absent in standard implementations of the Car–Parrinello scheme. This commutator has been shown to be related to the norm of the fictitious inertia tensor of the propagated density matrix,[51] represented as $\mu_{\mathrm{low},0}$ in eq 3. For the cases when the density matrix is converged, the commutator tends to zero and could be neglected, and in that case, the ADMP gradient reduces to the standard BOMD gradient.

The density gradient expression in eqs 5 and 7 has also been extensively discussed in refs 48 and 49, and the corresponding expression for idempotent density matrices is concisely written as

DOI: 10.1021/acs.jctc.6b00001
J. Chem. Theory Comput. 2016, 12, 2493–2508

$$\left.\frac{\partial E^{\mathrm{PIE-ONIOM}}(\mathbf{R}, \mathbf{P}_{\mathrm{low},\mathbf{0}})}{\partial \mathbf{P}_{\mathrm{low},\mathbf{0}}}\right|_{\mathbf{R}} \equiv \left.\frac{\partial E_{\mathrm{low}}(0)}{\partial \mathbf{P}_{\mathrm{low},\mathbf{0}}}\right|_{\mathbf{R}} = [[\mathbf{F}_{\mathrm{low},\mathbf{0}}, \tilde{\mathbf{P}}_{\mathrm{low},\mathbf{0}}], \tilde{\mathbf{P}}_{\mathrm{low},\mathbf{0}}] \quad \text{(A5)}$$

and again, this term tends to zero for the converged (BO) cases.

For all simulations in this study, the fictitious inertia tensor for valence electrons are fixed at the scalar value $\mu_{\mathrm{valence}} = 180$ au, and core orbitals are weighted based on μ_{valence} and corresponding diagonal elements of the Fock matrix that determines $E_{\mathrm{low}}(0)$.[48,49] This setup of fictitious mass has been widely used for many applications[32,67−69,130,131] and shown to be accurate in reproducing structural and vibrational features.

APPENDIX B: CONJUGATE HAMILTONIAN FOR THE LAGRANGIAN IN EQ 3 AND ASSOCIATED CONSERVATION PROPERTIES

The conjugate Hamiltonian for the Lagrangian presented in eq 3 is given by the Legendre transform[132]

$$\mathcal{H} = \mathrm{Tr}(\mathcal{W}_{\mathrm{low},\mathbf{0}}\mathbf{W}_{\mathrm{low},\mathbf{0}}) + \mathrm{Tr}(\mathcal{V}^{\mathbf{T}}\mathbf{V}) - \mathcal{L} \quad \text{(B1)}$$

where $\mathcal{W}_{\mathrm{low},\mathbf{0}}$ and $\mathcal{V}$ are the conjugate momenta for $\mathbf{P}_{\mathrm{low},\mathbf{0}}$ and $\mathbf{R}$, respectively, and are given by

$$\mathcal{W}_{\mathrm{low},\mathbf{0}} = \frac{\partial \mathcal{L}}{\partial \mathbf{W}_{\mathrm{low},\mathbf{0}}} = \boldsymbol{\mu}_{\mathrm{low},0}^{1/2}\mathbf{W}_{\mathrm{low},\mathbf{0}}\boldsymbol{\mu}_{\mathrm{low},0}^{1/2} \quad \text{(B2)}$$

and

$$\mathcal{V} = \frac{\partial \mathcal{L}}{\partial \mathbf{V}} = \mathrm{MV} \quad \text{(B3)}$$

Using eqs B2 and B3 in eq B1, one obtains the conjugate Hamiltonian as

$$\mathcal{H} = \frac{1}{2}\mathrm{Tr}[\mathcal{V}^{\mathbf{T}}\mathbf{M}^{-1}\mathcal{V}] + \frac{1}{2}\mathrm{Tr}[(\boldsymbol{\mu}_{\mathrm{low},0}^{-1/4}\mathcal{W}_{\mathrm{low},\mathbf{0}}\boldsymbol{\mu}_{\mathrm{low},0}^{-1/4})^2] + E^{\mathrm{PIE-ONIOM}}(\mathbf{R}, \mathbf{P}_{\mathrm{low},\mathbf{0}}) + \mathrm{Tr}[\boldsymbol{\Lambda}_{\mathrm{low},\mathbf{0}}(\mathbf{P}_{\mathrm{low},\mathbf{0}}^{\mathbf{2}} - \mathbf{P}_{\mathrm{low},\mathbf{0}})] \quad \text{(B4)}$$

It is the conservation properties of this Hamiltonian that are described in Table 1. The fact that such a quantity must be conserved is seen from the total derivative of $\mathcal{H}$ with respect to t, i.e.,

$$\frac{d\mathcal{H}}{dt} = \mathrm{Tr}\left[\frac{\partial \mathcal{H}}{\partial \mathbf{P}_{\mathrm{low},\mathbf{0}}}\frac{d\mathbf{P}_{\mathrm{low},\mathbf{0}}}{dt} + \frac{\partial \mathcal{H}}{\partial \mathcal{W}_{\mathrm{low},\mathbf{0}}}\frac{d\mathcal{W}_{\mathrm{low},\mathbf{0}}}{dt} + \frac{\partial \mathcal{H}}{\partial \mathbf{R}}\frac{d\mathbf{R}}{dt} + \frac{\partial \mathcal{H}}{\partial \mathcal{V}}\frac{d\mathcal{V}}{dt}\right]$$
$$= \mathrm{Tr}\left[\left(\left.\frac{\partial E^{\mathrm{PIE-ONIOM}}(\mathbf{R}, \mathbf{P}_{\mathrm{low},\mathbf{0}})}{\partial \mathbf{P}_{\mathrm{low},\mathbf{0}}}\right|_{\mathbf{R}} + \boldsymbol{\Lambda}_{\mathrm{low},\mathbf{0}}\mathbf{P}_{\mathrm{low},\mathbf{0}} + \mathbf{P}_{\mathrm{low},\mathbf{0}}\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_{\mathrm{low},\mathbf{0}}\right)\mathbf{W}_{\mathrm{low},\mathbf{0}}\right]$$
$$+ \mathrm{Tr}\left[\mathcal{W}_{\mathrm{low},\mathbf{0}}\frac{d^2\mathbf{P}_{\mathrm{low},\mathbf{0}}}{dt^2}\right] + Tr\left[\frac{\partial E^{\mathrm{PIE-ONIOM}}(\mathbf{R}, \mathbf{P}_{\mathrm{low},\mathbf{0}})}{\partial \mathbf{R}}\mathbf{V}\right] + \mathrm{Tr}\left[\mathrm{VM}\frac{d^2\mathbf{R}}{dt^2}\right] \quad \text{(B5)}$$

where the definitions in eqs B2 and B3 have been used, and it is also assumed that $\mu_{\mathrm{low},0}$ is time independent. Now, using the Lagrange equations of motion, i.e., eqs 5 and 4, it follows that

$$\frac{d\mathcal{H}}{dt} = 0 \quad \text{(B6)}$$

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.6b00001.

Additional figures describing the evolution of total energy during dynamics for all simulations and figures pertaining to the statistics in the radial distribution and vibrational density of states. (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: iyengar@indiana.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research is supported by the National Science Foundation Grant NSF CHE-1058949 to S.S.I. In addition funding from the Indiana University's Faculty Research Support Program is also acknowledged.

REFERENCES

(1) Field, C.; Bash, P. A.; Karplus, M. A Combined Quantum Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations. *J. Comput. Chem.* **1990**, *11*, 700.

(2) Leforestier, C. Classical Trajectories Using the Full Ab Initio Potential Energy Surface $H^- + CH_4 \rightarrow CH_4 + H^-$. *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) Yang, W. Direct Calculation of Electron Density in Density-Functional Theory. *Phys. Rev. Lett.* **1991**, *66*, 1438.

(5) Maseras, F.; Morokuma, K. A New "Ab Initio + Molecular Mechanics" Geometry Optimization Scheme of Equilibrium Structures and Transition States. *J. Comput. Chem.* **1995**, *16*, 1170.

(6) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. The Effective Fragment Potential Method: A QM-Based MM Approach to Modeling Environmental Effects in Chemistry. *J. Phys. Chem. A* **2001**, *105*, 293.

(7) 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.

(8) 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.

(9) 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.

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

(11) 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.

(12) Wang, L.-W.; Zhao, Z.; Meza, J. Linear-Scaling Three-Dimensional Fragment method for Large-scale Electronic Structure Calculations. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2008**, *77*, 165113.

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

(14) 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*, 104109.

DOI: 10.1021/acs.jctc.6b00001
J. Chem. Theory Comput. 2016, 12, 2493−2508

(15) 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*, 203−208.
(16) 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.
(17) 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.
(18) 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*, 094118.
(19) 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*, 064113.
(20) Hirata, S. Fast Electron-Correlation Methods for Molecular Crystals: an Application to the α, $\beta(1)$, and $\beta(2)$ Modifications of Solid Formic Acid. *J. Chem. Phys.* **2008**, *129*, 204104.
(21) Kamiya, M.; Hirata, S.; Valiev, M. Fast Electron-Correlation Methods for Molecular Crystals Without Basis Set Superposition Errors. *J. Chem. Phys.* **2008**, *128*, 074103.
(22) Brorsen, K. R.; Minezawa, N.; Xu, F.; Windus, T. L.; Gordon, M. S. Fragment Molecular Orbital Molecular Dynamics with the Fully Analytic Energy Gradient. *J. Chem. Theory Comput.* **2012**, *8*, 5008.
(23) Brorsen, K. R.; Zahariev, F.; Nakata, H.; Fedorov, D. G.; Gordon, M. S. Analytic Gradient for Density Functional Theory Based on the Fragment Molecular Orbital Method. *J. Chem. Theory Comput.* **2014**, *10*, 5297.
(24) 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.
(25) 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*, 054503.
(26) Chung, L. W.; Hirao, H.; Li, X.; Morokuma, K. The ONIOM Method: Its Foundation and Applications to Metalloenzymes and Photobiology. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012**, *2*, 327.
(27) Raghavachari, K.; Saha, A. Accurate Composite and Fragment-Based Quantum Chemical Models for Large Molecules. *Chem. Rev.* **2015**, *115*, 5643.
(28) 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.
(29) Collins, M. A.; Bettens, R. P. A. Energy-Based Molecular Fragmentation Methods. *Chem. Rev.* **2015**, *115*, 5607.
(30) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. Fragmentation Methods: A Route to Accurate Calculations on Large Systems. *Chem. Rev.* **2012**, *112*, 632.
(31) Sumner, I.; Iyengar, S. S. Combining Quantum Wavepacket *Ab Initio* Molecular Dynamics (QWAIMD) with QM/MM and QM/QM Techniques: Implementation Blending ONIOM and Empirical Valence Bond Theory. *J. Chem. Phys.* **2008**, *129*, 054109.
(32) 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*, 3978−3991.
(33) Humbel, S.; Sieber, S.; Morokuma, K. The IMOMO Method: Integration of Different Levels of Molecular Orbital Approximations for Geometry Optimization of Large Systems: Test for N-Butane Conformation and SN_2 Reaction: RCl+Cl$^-$. *J. Chem. Phys.* **1996**, *105*, 1959.
(34) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. NIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. a Test for Diels-Alder Reactions and $Pt(P(T\text{-}Bu)_3)_2$ + H_2 Oxidative Addition. *J. Phys. Chem.* **1996**, *100*, 19357.
(35) Svensson, M.; Humbel, S.; Morokuma, K. Energetics Using the Single Point IMOMO (Integrated Molecular Orbital+molecular Orbital) Calculations: Choices of Computational Levels and Model System. *J. Chem. Phys.* **1996**, *105*, 3654.
(36) Dapprich, S.; Komaromi, 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**, *461-462*, 1.
(37) Maseras, F. The IMOMM Method Opens the Way for the Accurate Calculation of "real" Transition Metal Complexes. *Chem. Commun.* **2000**, *19*, 1821.
(38) Vreven, T.; Morokuma, K. On the Application of the IMOMO (Integrated Molecular Orbital + Molecular Orbital) Method. *J. Comput. Chem.* **2000**, *21*, 1419.
(39) Vreven, T.; Morokuma, K.; Farkas, O.; Schlegel, H. B.; Frisch, M. J. Geometry Optimization with QM/MM, ONIOM, and Other Combined Methods. I. MicroIterations and Constraints. *J. Comput. Chem.* **2003**, *24*, 760.
(40) Vreven, T.; Morokuma, K. Investigation of the $S_0 \rightarrow S_1$ Excitation in Bacteriorhodopsin with the ONIOM(MO:MM) Hybrid Method. *Theor. Chem. Acc.* **2003**, *109*, 125.
(41) Vreven, T.; Byun, K. S.; Komaromi, I.; Dapprich, S.; Montgomery, J. A.; Morokuma, K.; Frisch, M. J. Combining Quantum Mechanics with Molecular Mechanics Methods in ONIOM. *J. Chem. Theory Comput.* **2006**, *2*, 815.
(42) Björklund, A.; Husfeldt, T.; Koivisto, M. Set Partitioning via Inclusion Exclusion. *SIAM J. Comput.* **2009**, *39*, 546.
(43) Hratchian, H. P.; Parandekar, P. V.; Raghavachari, K.; Frisch, M. J.; Vreven, T. QM:QM Electronic Embedding Using Mulliken Atomic Charges: Energies and Analytic Gradients in an ONIOM Framework. *J. Chem. Phys.* **2008**, *128*, 034107.
(44) Mayhall, N. J.; Raghavachari, K.; Hratchian, H. P. ONIOM-Based QM:QM Electronic Embedding Method Using Löwdin Atomic Charges: Energies and Analytic Gradients. *J. Chem. Phys.* **2010**, *132*, 114107.
(45) Parrinello, M.; Rahman, A. Crystal Structure and Pair Potentials: A Molecular-Dynamics Study. *Phys. Rev. Lett.* **1980**, *45*, 1196.
(46) Remler, D. K.; Madden, P. A. Molecular Dynamics Without Effective Potentials Via the Car-Parrinello Approach. *Mol. Phys.* **1990**, *70*, 921.
(47) Marx, D.; Hutter, J. Ab Initio Molecular Dynamics: Theory And Implementation. In *Modern Methods and Algorithms of Quantum Chemistry*; Grotendorst, J., Ed.; John Vonneumann Institute For Computing: Julich, 2000; Vol. *1*; p 301.
(48) 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.
(49) Iyengar, S. S.; Schlegel, H. B.; Millam, J. M.; Voth, G. A.; Scuseria, G. E.; Frisch, M. J. Ab Initio Molecular Dynamics: Propagating the Density Matrix with Gaussian Orbitals. II. Generalizations Based on Mass-Weighting, Idempotency, Energy Conservation and Choice of Initial Conditions. *J. Chem. Phys.* **2001**, *115*, 10291.
(50) Schlegel, H. B.; Iyengar, S. S.; Li, X.; Millam, J. M.; Voth, G. A.; Scuseria, G. E.; Frisch, M. J. Ab Initio Molecular Dynamics: Propagating the Density Matrix with Gaussian Orbitals. III. Comparison with Born-Oppenheimer Dynamics. *J. Chem. Phys.* **2002**, *117*, 8694.
(51) Iyengar, S. S.; Schlegel, H. B.; Voth, G. A.; Millam, J. M.; Scuseria, G. E.; Frisch, M. J. Ab Initio Molecular Dynamics: Propagating the Density Matrix with Gaussian Orbitals. IV. Formal Analysis of the Deviations from Born-Oppenheimer Dynamics. *Isr. J. Chem.* **2002**, *42*, 191.

DOI: 10.1021/acs.jctc.6b00001
J. Chem. Theory Comput. 2016, 12, 2493−2508

(52) Iyengar, S. S.; Frisch, M. J. Effect of Time-Dependent Basis Functions and Their Superposition Error on Atom-Centered Density Matrix Propagation (ADMP): Connections to Wavelet Theory of Multi-Resolution Analysis. *J. Chem. Phys.* **2004**, *121*, 5061.
(53) Iyengar, S. S.; Schlegel, H. B.; Voth, G. A. Atom-Centered Density Matrix Propagation: Generalizations Using Bohmian Mechanics. *J. Phys. Chem. A* **2003**, *107*, 7269.
(54) 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.
(55) Herbert, J. M.; Head-Gordon, M. Curvy-Steps Approach to Constraint-Free Extended-Lagrangian Ab Initio Molecular Dynamics, Using Atom-Centered Basis Functions: Convergence Toward Born Oppenheimer Trajectories. *J. Chem. Phys.* **2004**, *121*, 11542.
(56) Iyengar, S. S.; Schlegel, H. B.; Scuseria, G. E.; Frisch, M. J.; Millam, J. M. Comment On "Curvy-Steps Approach to Constraint-Free Extended-Lagrangian Ab Initio Molecular Dynamics, Using Atom-Centered Basis Functions: Convergence Toward Born Oppenheimer Trajectories". *J. Chem. Phys.* **2005**, *123*, 027101.
(57) Herbert, J. M.; Head-Gordon, M. Response to "Comment on 'Curvy-steps approach to constraint-free extended-Lagrangian ab initio molecular dynamics, using atom-centered basis functions: Convergence toward Born−Oppenheimer trajectories'". *J. Chem. Phys.* **2005**, *123* (123), 027102.
(58) Cukier, R. I.; Deutch, J. M. Microscopic Theory of Brownian Motion: The Multiple-Time-Scale Point of View. *Phys. Rev.* **1969**, *177*, 240.
(59) Andersen, H. C. Molecular Dynamics Simulations at Constant Pressure And/or Temperature. *J. Chem. Phys.* **1980**, *72*, 2384.
(60) Parrinello, M.; Rahman, A. Study of an F Center in Molten KCl. *J. Chem. Phys.* **1984**, *80*, 860.
(61) 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*, 084309.
(62) Iyengar, S. S. Further Analysis of the Dynamically Averaged Vibrational Spectrum for the "magic" Protonated 21-Water Cluster. *J. Chem. Phys.* **2007**, *126*, 216101.
(63) 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.
(64) 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.
(65) 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*, 244301.
(66) Iyengar, S. S.; Li, X.; Sumner, I. The Study of Dynamically Averaged Vibrational Spectroscopy of Atmospherically Relevant Clusters Using Ab Initio Molecular Dynamics in Conjunction with Quantum Wavepackets. *Adv. Quantum Chem.* **2008**, *55*, 333.
(67) Vimal, D.; Pacheco, A. B.; Iyengar, S. S.; Stevens, P. S. Experimental and Ab Initio Dynamical Investigations of the Kinetics and Intramolecular Energy Transfer Mechanisms for the OH+1,3-Butadiene Reaction Between 263 and 423 K at Low Pressure. *J. Phys. Chem. A* **2008**, *112*, 7227.
(68) Pacheco, A. B.; Dietrick, S. M.; Stevens, P. S.; Iyengar, S. S. Pump-Probe" Atom-Centered Density Matrix Propagation Studies to Gauge Anharmonicity and Energy Repartitioning in Atmospheric Reactive Adducts: Case Study of the OH + Isoprene and OH + Butadiene Reaction Intermediates. *J. Phys. Chem. A* **2012**, *116*, 4108.
(69) 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 OH + Isoprene Reaction Intermediate Using *Ab-Initio* Molecular Dynamics. *J. Phys. Chem. A* **2012**, *116*, 399.
(70) Petersen, M. K.; Iyengar, S. S.; Day, T. J. F.; Voth, G. A. The Hydrated Proton at Water Liquid/Vapour Interfaces. *J. Phys. Chem. B* **2004**, *108*, 14804.
(71) 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.
(72) Petersen, P. B.; Saykally, R. J. Evidence for an Enhanced Hydronium Concentration at the Liquid Water Surface. *J. Phys. Chem. B* **2005**, *109*, 7976.
(73) Petersen, P. B.; Saykally, R. J. On the Nature of Ions at the Liquid Water Surface. *Annu. Rev. Phys. Chem.* **2006**, *57*, 333−364.
(74) Levering, L. M.; Sierra-Hernandez, M. R.; Allen, H. C. Observation of Hydronium Ions at the Air - Aqueous Acid Interface: Vibrational Spectroscopic Studies of Aqueous HCI, HBr, and HI. *J. Phys. Chem. C* **2007**, *111*, 8814.
(75) Mucha, M.; Frigato, T.; Levering, L. M.; Allen, H. C.; Tobias, D. J.; Dang, L. X.; Jungwirth, P. Unified Molecular Picture of the Surfaces of Aqueous Acid, Base, and Salt Solutions. *J. Phys. Chem. B* **2005**, *109*, 7617.
(76) Kudin, K. N.; Car, R. Why Are Water-Hydrophobic Interfaces Charged? *J. Am. Chem. Soc.* **2008**, *130*, 3915.
(77) Gibson, D. A.; Ionova, I. V.; Carter, E. A. A Comparison of Car-Parrinello and Born-Oppenheimer Generalized Valence-Bond Molecular-Dynamics. *Chem. Phys. Lett.* **1995**, *240*, 261.
(78) Hartke, B.; Carter, E. A. Ab Initio Molecular-Dynamics with Correlated Molecular Wavefunctions - Generalized Valence Bond Molecular-Dynamics and Simulated Annealing. *J. Chem. Phys.* **1992**, *97*, 6569.
(79) Hartke, B.; Carter, E. A. Spin Eigenstate Dependent Hartree-Fock Molecular Dynamics. *Chem. Phys. Lett.* **1992**, *189*, 358.
(80) Martyna, G.; Cheng, C.; Klein, M. L. *J. Chem. Phys.* **1991**, *95*, 1318.
(81) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Parandekar, P. V.; Mayhall, N. J.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian*, Development Version, Revision H.37+; Gaussian, Inc., Wallingford, CT, 2010.
(82) Niklasson, A. M. N.; Cawkwell, M. J. Generalized extended Lagrangian Born-Oppenheimer molecular dynamics. *J. Chem. Phys.* **2014**, *141*, 164123.
(83) 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.
(84) Steele, R. P. Multiple-timestep ab initio molecular dynamics using an atomic basis set partitioning. *J. Phys. Chem. A* **2015**, *119*, 12119−12130.
(85) Pomes, R.; Roux, B. Structure and Dynamics of a Proton Wire: A Theoretical Study of H+ Translocation Along the Single-File Water Chain in the Gramicidin a Channel. *Biophys. J.* **1996**, *71*, 19.
(86) Pomes, R.; Roux, B. Theoretical Study of H^+ Translocation Along a Model Proton Wire. *J. Phys. Chem.* **1996**, *100*, 2519.

DOI: 10.1021/acs.jctc.6b00001
J. Chem. Theory Comput. 2016, 12, 2493−2508

(87) Decornez, H.; Drukker, K.; Hammes-Schiffer, S. Solvation and Hydrogen-Bonding Effects on Proton Wires. *J. Phys. Chem. A* **1999**, *103*, 2891.
(88) Brewer, M. L.; Schmitt, U. W.; Voth, G. A. The Formation and Dynamics of Proton Wires in Channel Environments. *Biophys. J.* **2001**, *80*, 1691.
(89) 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.
(90) Neidle, S.; Berman, H. M.; Shieh, H. S. Highly Structured Water Network in Crystals of a Deoxydinucleoside-Drug Complex. *Nature* **1980**, *288*, 129.
(91) 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.
(92) 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.
(93) McEwan, M. J.; Phillips, L. F. *Chemistry of the Atmosphere*; Eward Arnold: London, 1975.
(94) Wayne, R. P. *Chemistry of the Atmosphere*; Clarendon Press: Oxford, 1994.
(95) Iyengar, S. S. Dynamical Effects on Vibrational and Electronic Spectra of Hydroperoxyl Radical Water Clusters. *J. Chem. Phys.* **2005**, *123*, 084310.
(96) Okumura, M.; Yeh, L. I.; Myers, J. D.; Lee, Y. T. Infrared Spectra of the Solvated Hydronium Ion: Vibrational Predissociation Spectroscopy of Mass-Selected H_3O^+ $(H_2O)_nH_2$. *J. Phys. Chem.* **1990**, *94*, 3416.
(97) 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 $H^+(H_2O)_n$ (N = 6 to 27) Clusters. *Science* **2004**, *304*, 1137.
(98) Miyazaki, M.; Fujii, A.; Ebata, T.; Mikami, N. Infrared Spectroscopic Evidence for Protonated Water Clusters Forming Nanoscale Cages. *Science* **2004**, *304*, 1134.
(99) 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.
(100) 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.
(101) Hammer, N. I.; Diken, E. G.; Roscioli, J. R.; Johnson, M. A.; Myshakin, E. M.; Jordan, K. D.; McCoy, A. B.; Huang, X.; Bowman, J. M.; Carter, S. The Vibrational Predissociation Spectra of the $H_5O_2^+$ $\cdot RG_n$(RG = Ar,Ne) clusters: Correlation of the solvent perturbations in the free OH and shared proton transitions of the Zundel ion. *J. Chem. Phys.* **2005**, *122*, 244301.
(102) Diken, E. G.; Headrick, J. M.; Roscioli, J. R.; Bopp, J. C.; Johnson, M. A.; McCoy, A. B. Fundamental Excitations of the Shared Proton in the $H_3O_2^-$ and $H_5O_2^+$ Complexes. *J. Phys. Chem. A* **2005**, *109*, 1487.
(103) Asmis, K. R.; Pivonka, N. L.; Santambrogio, G.; Brümmer, M.; Kaposta, C.; Neumark, D. M.; Wöste, L. Gas-Phase Infrared Spectrum of the Protonated Water Dimer. *Science* **2003**, *299*, 1375.
(104) Fridgen, T. D.; MacAleese, P. L.; Maitre; McMahon, T. B.; Boissel, P.; Lemaire, J. Infrared Spectra of Homogeneous and Heterogeneous Proton-Bound Dimers in the Gas Phase. *Phys. Chem. Chem. Phys.* **2005**, *7*, 2747.
(105) Vendrell, O.; Gatti, F.; Meyer, H.-D. Dynamics and Infrared Spectroscopy of the Protonated Water Dimer. *Angew. Chem., Int. Ed.* **2007**, *46*, 6918.
(106) Xie, Y.; Remington, R. B.; Schaefer, H. F., III The Protonated Water Dimer: Extensive Theoretical Studies of $H_5O_2^+$. *J. Chem. Phys.* **1994**, *101*, 4878.
(107) Sadhukhan, S.; Munoz, D.; Adamo, C.; Scuseria, G. E. Predicting Proton Transfer Barriers with Density Functional Methods. *Chem. Phys. Lett.* **1999**, *306*, 83.
(108) Tsai, C. J.; Jordan, K. D. Theoretical Study of the $(H_2O)_6$ Cluster. *Chem. Phys. Lett.* **1993**, *213*, 181.
(109) 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.
(110) 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.
(111) Kryachko, E. S. Ab Initio Studies of the Conformations of Water Hexamer: Modelling the Penta-Coordinated Hydrogen-Bonded Pattern in Liquid Water. *Chem. Phys. Lett.* **1999**, *314*, 353.
(112) Hodges, M. P.; Wales, D. J. Global Minima of Protonated Water Clusters. *Chem. Phys. Lett.* **2000**, *324*, 279.
(113) Cheng, H.-P. Water Clusters: Fascinating Hydrogen-Bonding Networks, Solvation Shell Structures, and Proton Motion. *J. Phys. Chem. A* **1998**, *102*, 6201.
(114) Sadeghi, R. R.; Cheng, H.-P. The Dynamics of Proton Transfer in a Water Chain. *J. Chem. Phys.* **1999**, *111*, 2086.
(115) Cheng, H.-P.; Krause, J. L. The Dynamics of Proton Transfer in $H_5O_2^+$. *J. Chem. Phys.* **1997**, *107*, 8461.
(116) 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: $H^+(H_2O)_{21}$ and $H^+(H_2O)_{28}$. *J. Chem. Phys.* **2005**, *122*, 074315.
(117) 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*, 234504.
(118) 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.
(119) Wei, D.; Salahub, D. R. Hydrated Proton Clusters: Ab Initio Molecular Dynamics Simulation and Simulated Annealing. *J. Chem. Phys.* **1997**, *106*, 6086.
(120) Kaledin, M.; Kaledin, A. L.; Bowman, J. M. Vibrational Analysis of the $H_5O_2^+$ Infrared Spectrum Using Molecular and Driven Molecular Dynamics. *J. Phys. Chem. A* **2006**, *110*, 2933.
(121) Shi, L.; Holliday, A. E.; Khanal, N.; Russell, D. H.; Clemmer, D. E. Configurationally-Coupled Protonation of Polyproline-7. *J. Am. Chem. Soc.* **2015**, *137*, 8680.
(122) Servage, K. A.; Fort, K. L.; Silveira, J. A.; Shi, L.; Clemmer, D. E.; Russell, D. H. Unfolding of Hydrated Alkyl Diammonium Cations Revealed by Cryogenic Ion Mobility-Mass Spectrometry. *J. Am. Chem. Soc.* **2015**, *137*, 8916−8919.
(123) Shi, L.; Holliday, A. E.; Shi, H.; Zhu, F.; Ewing, M. A.; Russell, D. H.; Clemmer, D. E. Characterizing Intermediates Along the Transition from Polyproline I to Polyproline II Using Ion Mobility Spectrometry-Mass Spectrometry. *J. Am. Chem. Soc.* **2014**, *136*, 12702.
(124) Bakowies, D.; Thiel, W. Hybrid Models for Combined Quantum Mechanical and Molecular Mechanical Approaches. *J. Phys. Chem.* **1996**, *100*, 10580.
(125) Agmon, N. The Grotthuss Mechanism. *Chem. Phys. Lett.* **1995**, *244*, 456.
(126) 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*, 184308.
(127) Roscioli, J. R.; McCunn, L. R.; Johnson, M. A. Quantum Structure of the Intermolecular Proton Bond. *Science* **2007**, *316*, 249.
(128) McWeeny, R. Some Recent Advances in Density Matrix Theory. *Rev. Mod. Phys.* **1960**, *32*, 335.
(129) Pulay, P. *Ab Initio* Calculation of Force Constants and Equilibrium Geometries in Polyatomic Molecules. *Mol. Phys.* **1969**, *17*, 197.

DOI: 10.1021/acs.jctc.6b00001
J. Chem. Theory Comput. 2016, 12, 2493−2508

(130) 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.
(131) Phatak, P.; Venderley, J.; Debrota, 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.
(132) Goldstein, H.; Poole, C.; Safko, J. *Classical Mechanics*; Addison Wesley: San Francisco, 2002.

DOI: 10.1021/acs.jctc.6b00001
J. Chem. Theory Comput. 2016, 12, 2493−2508