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

pubs.acs.org/JPCA
Article

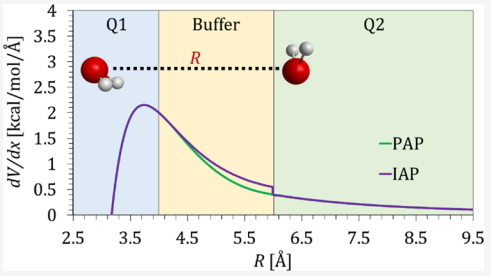
Adaptive-Partitioning Multilayer Dynamics Simulations: 2. Implementations of the Permuted and Interpolated Adaptive-Partitioning Gradients

Anh L. Tran, Emilie B. Guidez,* and Hai Lin*

Cite This: J. Phys. Chem. A 2023, 127, 10320−10333
Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Recently, an adaptive-partitioning multilayer Q1/Q2/MM method was proposed, where Q1 and Q2 denote, respectively, two distinct quantum-mechanical levels of theory and MM, the molecular-mechanical force fields. Such a multilayer model resembles the ONIOM (our own N-layered integrated molecular orbital and molecular mechanics) model by Morokuma and co-workers, but it is distinguished by on-the-fly reclassifying atoms to be Q1, Q2, or MM in dynamics simulations. To smoothly blend the levels of descriptions of the atoms, buffer zones are introduced between adjacent layers, and the energy is smoothly interpolated. In particular, the Q1/Q2 interaction energy was expressed in two different formalisms: permuted and interpolated adaptive-partitioning (PAP and IAP), respectively. While the PAP energy is based on a weighted many-body expansion, the IAP energy is derived via alchemical quantum calculations with interpolated Fock and overlap matrices. In this article, we examine in-depth the irregularities in the IAP energy near the boundary between the buffer and Q2 zones, which were found prominent in some calculations. These irregularities are due to basis-set linear dependencies, which can be effectively suppressed using a cutoff for the weighted atomic orbital coefficients. Furthermore, we derived and implemented the gradients for both PAP and IAP. Test calculations on a series of water cluster models show perfectly smooth gradients in PAP, while a minor discontinuity occurs in IAP gradients at the buffer/Q2 boundary. The energy and gradient discontinuities in IAP become smaller when moving the buffer/Q2 boundary further away from the Q1 center and when increasing the size of the basis sets used. Overall, those discontinuities are controllable, and possible ways to further diminish them are discussed.

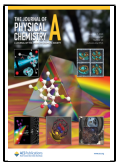


1. INTRODUCTION

Recently, we proposed an adaptive-partitioning multilayer (APML) algorithm,[1] as an extension of the two-layer adaptive-partitioning quantum-mechanics/molecular-mechanics (AP-QM/MM) methods,[2−12] which are one family of algorithms in the more broadly defined adaptive QM/MM for multiscale modeling.[1−41] In APML, an entire model system is divided into multiple layers based on how significant they are to the question being asked. Accordingly, these layers are described at various levels of theory, with the more important layer described at the higher level. The treatment is similar to what is done in the celebrated ONIOM (our own N-layered Integrated molecular orbital and molecular mechanics) models by Morokuma and co-workers.[42,43] The critical difference between ONIOM and APML is the adaptive partitioning in APML. More specifically, during dynamics simulations, the boundaries between these layers remain static in ONIOM, but they are relocated on the fly in APML. As a result, an atom may dynamically be reclassified in APML as high-level QM (denoted Q1), low-level QM (denoted Q2), or MM. In contrast, an atom in ONIOM is always kept at the level of theory that it is assigned to at the beginning of a simulation. [We note that the idea of dynamically relocated interlayer boundaries had also been proposed in the ONIOM-XS (ONIOM with exchanging solvents) algorithm,[14,15] but only two-layer QM/MM descriptions were actually employed.] With its flexibility in adjusting the interlayer boundaries as needed, adaptive QM/MM has been successfully applied to the dynamics simulations of diffusive model systems, such as ions (including protons) migrating in a bulk solution[6,23,41,44−47] or through channel proteins[8,9] as well as solvent molecules exchanged between a protein active site and the bulk solution.[12] The moving boundaries would allow one to characterize the given ion and its various solvation shells with hierarchically varying accuracy and efficiency, with the contents of these solvation shells automatically updated over time.

To avoid abrupt changes in energy and gradients when atoms run across interlayer boundaries, a buffer zone is usually

Received: August 20, 2023
Revised: October 6, 2023
Accepted: October 17, 2023
Published: November 21, 2023



https://doi.org/10.1021/acs.jpca.3c05600
J. Phys. Chem. A 2023, 127, 10320−10333

introduced between two adjacent layers, and the energy and/or gradients are smoothly interpolated as atoms move in the buffer zone. When the energy is interpolated, one issue is the gradients due to the smoothing functions used for interpolations. These gradients are needed to conserve the energy of the model system in dynamics simulations,[2,4,10,24,48] but they may cause artifacts. In fact, this issue is presented not only in the atomistic adaptive QM/MM but also in the mixed-resolution adaptive MM/coarse-grained (CG).[49−52] Many studies have been devoted to addressing this issue.[2,10,24,48,50,53] The treatments can largely be grouped into two categories. In the first category, these gradients are simply ignored.[3,14] This is equivalent to, at each time step, applying external forces to cancel the gradients due to the smoothing functions, effectively equalizing the chemical potentials between the two zones separated by the buffer. The price to pay is that the system is no longer a Hamiltonian system, for which energy is not necessarily conserved. In practical applications, the model system is often coupled to a thermostat. The treatments in the second category are more elaborated. The basic idea is to add a correction term derived from Helmholtz free-energy per particle to the Hamiltonian of the original adaptive algorithm. On average, the gradients of this correction term cancel the gradients due to the smoothing function. First proposed for adaptive MM/CG by Potestio et al.,[53] this idea has also been applied to adaptive QM/MM.[10,24] Overall, algorithms in the second category conserve energy without the artifacts caused by the gradients due to smoothing functions. Because the model remains a Hamiltonian system, these methods are often referred to as Hamiltonian adaptive algorithms, while the schemes in the first category are non-Hamiltonian methods. The drawback of these Hamiltonian algorithms is that they are very expensive because of the extra efforts needed to obtain the correction term from extensive sampling. As a result, they are rarely used, unless the simulations *must* be performed under the microcanonical ensemble. Finally, it should be noted that the relationship between the above Hamiltonian and non-Hamiltonian methods was recognized by Kreis et al.[54] If a Hamiltonian treatment includes a history-dependent friction term and is coupled to a thermostat via the Generalized Langevin Equation, it leads to a non-Hamiltonian method in the first category.

In a three-layer (Q1/Q2/MM) APML model,[1] the adaptive partitioning between Q2 and MM is the same as in the two-layer AP-QM/MM model,[2−12] while that between Q1 and Q2 needs to be developed. In our first APML paper,[1] we proposed two AP methods that both smoothly blend the Q1 and Q2 levels of theory: the permuted-AP (PAP) and the interpolated-AP (IAP) schemes. The PAP algorithm is based on a weighted many-body expansion of the total energy of a model system, whereas the IAP treatment is based on "alchemical" QM calculations with mixed basis sets, Fock matrices, and overlap matrices between Q1 and Q2. Both the weights in PAP and the degrees of mixing in IAP depend on the locations of the atoms in the buffer zone (called buffer atoms), and they change gradually as the atoms vary in position. Although the alchemical-QM treatments in IAP may seem radical, we note that both PAP and IAP are just two different ways to account for the mixed Q1/Q2 nature of the atoms in the buffer zone and that neither PAP nor IAP is fundamentally "more correct" than the other.

Testing on a series of water clusters,[1] we showed that both schemes yielded overall smoothly varying energy when one water molecule is gradually moving across the interlayer boundaries. Despite these encouraging results, small irregularities were recently spotted in some IAP energy curves near the buffer/Q2 boundary. What causes these irregularities? Are they influenced by the location and size of the buffer zone? To what extent do they depend on the selected basis set combination? How will the IAP gradients be impacted? The present paper addresses these questions. We show that these irregularities are caused by the linear dependence of the basis functions for the buffer groups and that a practice ansatz can be employed to largely eliminate them. However, small discontinuities in the IAP energy and gradients at the buffer/Q2 boundary are still observed. Nevertheless, when double-ζ basis sets are used for Q2, these discontinuities are significantly reduced and would likely have minimal impacts on the IAP dynamic simulations.

2. METHODS

2.1. IAP and PAP Treatments of Q1/Q2 Energy. In this section, we provide a succinct introduction of the mechanical-embedding PAP and IAP Hamiltonians for the Q1/Q2 model, while more details can be found in ref 1. Without losing generality, we assume that the first molecule is in the Q1 zone (called the Q1 group), the second to $(N+1)$th molecule in the buffer zone (called buffer groups), and the $(N+2)$th molecule in the Q2 zone (called the Q2 group). For simplicity, we assume that all groups are whole molecules. The Q1 characteristic of the ith group is given by P_i, the value of a smoothing function[14] for the group that varies from 0 at $R_i = R_{max}$ to 1 at $R_i = R_{min}$, where R_i is the distance from the Q1 zone center to a delegate atom (or the centers of mass) of the group (Figure 1). Accordingly, the group's Q2 characteristic is $(1 - P_i)$. The AP potential $V_{12\cdots(N+2)}^{Q1/Q2;\,AP}$ is given by

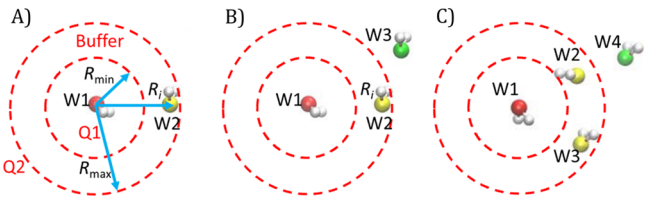


Figure 1. Q1/Q2 model exemplified by a water (A) dimer, (B) trimer, and (C) tetramer consisting of molecules W1 to W4. The Q1 zone is centered at the O atom of W1 (denoted O1). Depending on the distance R_i from O1 to the O atom of Wi (denoted Oi), Wi is in the Q1 zone ($R \le R_{min}$), buffer zone ($R_{min} \le R \le R_{max}$), or Q2 zone ($R_{max} \le R$). H atoms are shown as white spheres, while the O atoms are shown as spheres in red (in the Q1 zone), yellow (in the buffer zone), or green (in the Q2 zone).

$$V_{12\cdots(N+2)}^{Q1/Q2;AP} = V_{12\cdots(N+2)}^{Q2} - V_{12\cdots(N+1)}^{Q2} + V_{12\cdots(N+1)}^{Q1/Q2;AP} \quad (1)$$

where the superscripts indicate the levels of theory at which the groups listed in the subscripts are treated, unless otherwise noted, and AP can be either PAP or IAP. Essentially, eq 1 states that the single-level Q2 description of the Q1-buffer cohort (i.e., the combination of the Q1 and buffer zones), $V_{12\cdots(N+1)}^{Q2}$, is replaced by a mixed Q1/Q2 AP description, $V_{12\cdots(N+1)}^{Q1/Q2;\,AP}$, in the given AP treatment. Note that only the interaction energy is of real interest, which is obtained by

https://doi.org/10.1021/acs.jpca.3c05600
J. Phys. Chem. A 2023, 127, 10320−10333

subtracting the energies of isolated individual groups at the appropriate level of theory.

In PAP, the mixed potential of the Q1-buffer cohort is a sum of weighted many-body interaction terms,

$$V_{12\cdots(N+1)}^{\mathrm{Q1/Q2;PAP}} = V_{12\cdots(N+1)}^{\mathrm{Q1/Q2;1-body}} + \Delta V_{12\cdots(N+1)}^{\mathrm{Q1/Q2;2-body}} + \Delta V_{12\cdots(N+1)}^{\mathrm{Q1/Q2;3-body}} + \cdots \quad (2)$$

The weighted many-body interaction terms contain both Q1- and Q2-level contributions:

$$V_{12\cdots(N+1)}^{\mathrm{Q1/Q2;1-body}} = \sum_{i=1}^{N+1} [P_i V_i^{\mathrm{Q1}} + (1 - P_i) V_i^{\mathrm{Q2}}] \quad (3)$$

$$\Delta V_{12\cdots(N+1)}^{\mathrm{Q1/Q2;2-body}} = \sum_{i=1}^{N} \sum_{j=i+1}^{N+1} [P_i P_j \Delta V_{ij}^{\mathrm{Q1}} + (1 - P_i P_j) \Delta V_{ij}^{\mathrm{Q2}}] \quad (4)$$

$$\Delta V_{12\cdots(N+1)}^{\mathrm{Q1/Q2;3-body}} = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \sum_{k=j+1}^{N+1} [P_i P_j P_k \Delta V_{ijk}^{\mathrm{Q1}} + (1 - P_i P_j P_k) \Delta V_{ijk}^{\mathrm{Q2}}] \quad (5)$$

where $\Delta V_{ij} = V_{ij} - (V_i + V_j)$ and $\Delta V_{ijk} = V_{ijk} - (V_{ij} + V_{ik} + V_{jk}) + (V_i + V_j + V_k)$ are the two- and three-body contributions in a regular many-body expansion, respectively. The zero of potential energy in the one-body term for a given group is arbitrary and often conveniently set to the energy at its equilibrium geometry at the given Q1 or Q2 level.

In contrast, IAP computes alchemically the mixed potential of the Q1-buffer cohort $V_{12\cdots(N+1)}^{\mathrm{Q1/Q2;\,IAP}}$, employing an interpolated basis set for the buffer groups:

$$\phi_i^{\mathrm{Q1/Q2}} = P_i \phi_i^{\mathrm{Q1}} + (1 - P_i) \phi_i^{\mathrm{Q2}} \quad (6)$$

where the respective basis sets for the Q1 and Q2 groups, $\{\phi^{\mathrm{Q1}}\}$ and $\{\phi^{\mathrm{Q2}}\}$, are combined in a similar way to a contracted basis set. An atomic orbital (AO) μ centered at an atom belonging to group i inherits the smoothing function value of group i, i.e., $P_\mu = P_i$, and we combine the respective shells from each basis set before the self-consistent field (SCF) iterations. The resulting alchemical Fock matrix $f_{\mu\nu}^{\mathrm{Q1/Q2}}$ is a hybrid of the weighted Fock matrices at the Q1 ($f_{\mu\nu}^{\mathrm{Q1}}$) and Q2 ($f_{\mu\nu}^{\mathrm{Q2}}$) levels and a new Fock matrix ($f_{\mu\nu}^{\mathrm{Q1+Q2}}$) constructed using the interpolated basis set:

$$f_{\mu\nu}^{\mathrm{Q1/Q2}} = P_\mu P_\nu f_{\mu\nu}^{\mathrm{Q1}} + (1 - P_\mu)(1 - P_\nu) f_{\mu\nu}^{\mathrm{Q2}} + f_{\mu\nu}^{\mathrm{Q1+Q2}} \quad (7)$$

$$f_{\mu\nu}^{\mathrm{Q1}} = h_{\mu\nu}^{\mathrm{Q1}} + g_{\mu\nu}^{\mathrm{Q1}} = h_{\mu\nu}^{\mathrm{Q1}} + \sum_{\lambda\sigma} P_\lambda P_\sigma D_{\lambda\sigma} \left([\mu\nu|\lambda\sigma] - \frac{1}{2}[\mu\sigma|\lambda\nu] \right) \quad (8)$$

$$f_{\mu\nu}^{\mathrm{Q2}} = h_{\mu\nu}^{\mathrm{Q2}} + g_{\mu\nu}^{\mathrm{Q2}} = h_{\mu\nu}^{\mathrm{Q2}} + \sum_{\lambda\sigma} (1 - P_\lambda)(1 - P_\sigma) D_{\lambda\sigma} \left([\mu\nu|\lambda\sigma] - \frac{1}{2}[\mu\sigma|\lambda\nu] \right) \quad (9)$$

where $h_{\mu\nu}$ and $g_{\mu\nu}$ are the one- and two-electron terms, respectively, $D_{\lambda\sigma}$ is the charge density matrix element for atomic orbitals λ and σ, and appropriate weights are applied. Note that $f_{\mu\nu}^{\mathrm{Q1}}$ and $f_{\mu\nu}^{\mathrm{Q2}}$ are Fock matrices constructed only with the Q1 and Q2 basis sets, respectively, whereas $f_{\mu\nu}^{\mathrm{Q1+Q2}}$ is constructed with the interpolated basis set. Similarly, the alchemical overlap matrix $S_{\mu\nu}^{\mathrm{Q1/Q2}}$ is given by

$$S_{\mu\nu}^{\mathrm{Q1/Q2}} = P_\mu P_\nu S_{\mu\nu}^{\mathrm{Q1}} + (1 - P_\mu)(1 - P_\nu) S_{\mu\nu}^{\mathrm{Q2}} + S_{\mu\nu}^{\mathrm{Q1+Q2}} \quad (10)$$

where the overlap matrix $S_{\mu\nu}^{\mathrm{Q1+Q2}}$ is constructed with the interpolated basis set. With the alchemical Fock and overlap matrixes, the SCF calculations yield the alchemical energy for the Q1-buffer cohort. To get the needed interaction energy, one must subtract from this energy the energies of isolated individual groups from the total energy at the corresponding Q1 (if in the Q1 zone) or mixed Q1/Q2 (if in the buffer zone) level of theory.

In the current implementation, the smoothing function P_i is applied to all AOs in all integrals. Therefore, if four orbitals are located on the same atom, the smoothing function is applied 4 times. As long as P_i varies smoothly, the scaling changes smoothly. One may argue that such a treatment is "non-physical." However, the overall IAP (and PAP) model is nonphysical with scaling, regardless of how the scaling is applied. Thus, our focus is to obtain smoothly varying energies and gradients unless the scaling leads to significantly erroneous electronic structures. We have so far not found any significant problems in this regard in our calculations.

2.2. Smoothing Function. The smoothing function used for group i in both AP-QM/MM[2−12] and APML methods[1] is a fifth-order polynomial, first introduced by Kercharoen and Morokuma:[14]

$$P_i = -6\alpha_i^5 + 15\alpha_i^4 - 10\alpha_i^3 + 1 \quad (11)$$

where the value of α_i is given by

$$\alpha_i = \frac{R_i - R_{\min}}{R_{\max} - R_{\min}} \quad (12)$$

The function P_i varies smoothly between 1 at $R_i = R_{\min}$ and 0 at $R_i = R_{\max}$, leading to overall smooth potential energy curves for model systems in our previous investigation.[1] However, as will be shown in Section 4.1 of the results, some irregularities may occur near the buffer/Q2 boundary. These irregularities are due to linear dependencies induced when P_i becomes very small.

To address this issue, a cutoff value (*coef-cut*) for the weighted coefficients (eq 13) is introduced. Weighted coefficients are given by

$$C_{\mathrm{weighted}} = P_\mu \times c_\mu \quad (13)$$

In eq 13, P_μ is the value of the smoothing function for the atomic orbital ϕ_μ and c_μ is the atomic orbital coefficient. For instance, if *coef-cut* = 0.01, when a weighted coefficient (C_{weighted}) falls below this value, C_{weighted} is set to 0.01.

2.3. IAP and PAP Gradients. Following the definition of the Hamiltonians, gradient calculations are straightforward, but care must be taken to ensure that the gradients associated with individual groups are properly accounted for in the subtractions. In our current implementation, the weights of individual groups and therefore the weighting coefficients in the Hamiltonian are treated as constants for a given geometry in each Q1/Q2 setup. The gradients of the smoothing functions are thus not included, as discussed in Section 1.

https://doi.org/10.1021/acs.jpca.3c05600
J. Phys. Chem. A 2023, 127, 10320−10333

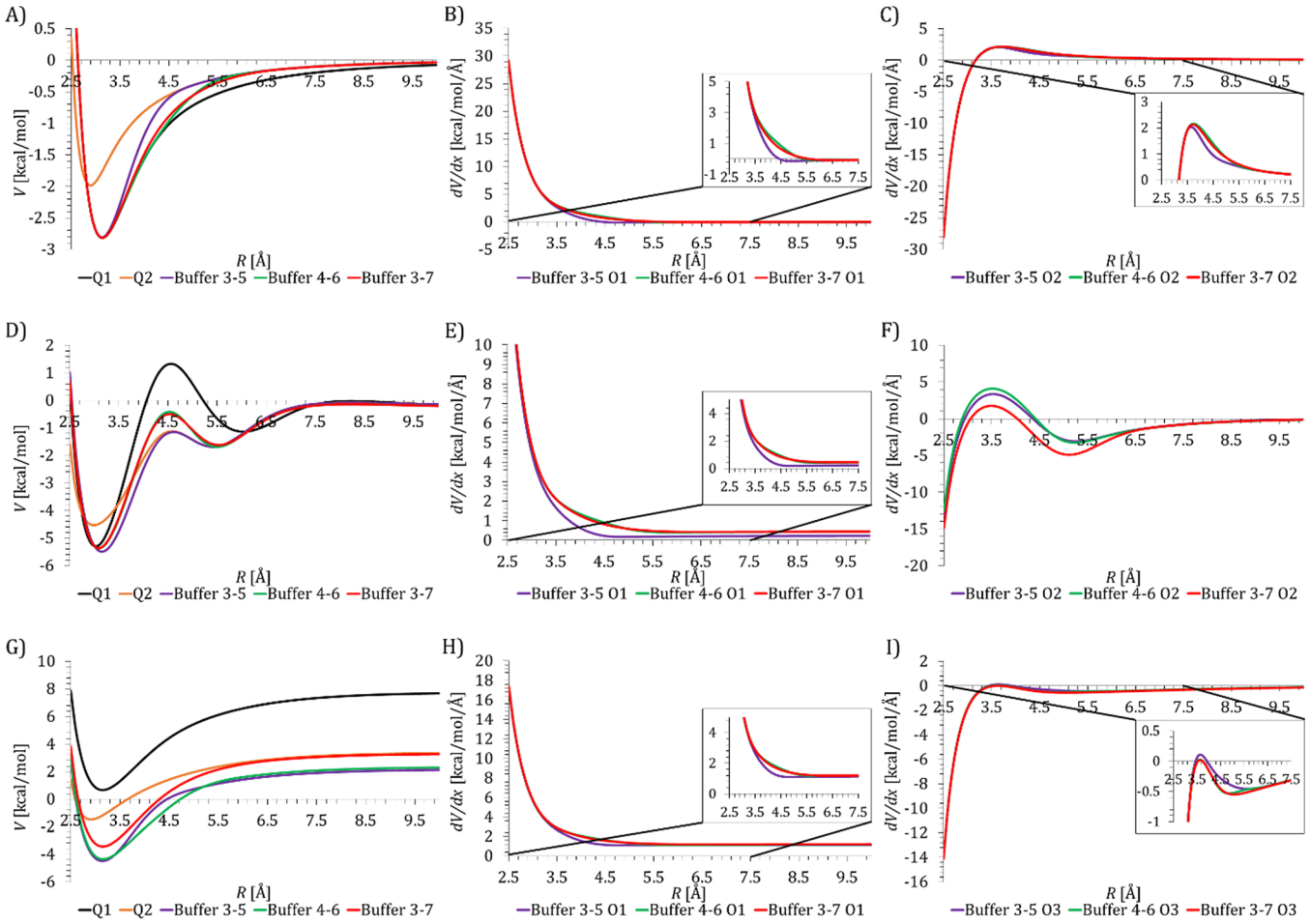


Figure 2. (A) Q1, Q2, and PAP energy profiles for the water dimer model, where R is the distance between the oxygen atoms of the fixed W1 and moving W2 (denoted O1 and O2, respectively). The basis set combination Q1 = 6-31G(d)/Q2 = STO-3G is employed. (B) Similar to (A), but for the x component of the PAP gradient (dV/dx) at O1. (C) Similar to (B), but for the x components of the PAP gradient (dV/dx) at O2. (D−F) Similar to (A−C), but for the water trimer model. (G, H) is similar to (A, B), but for the tetramer model, and R is the distance between the fixed O1 and moving O3 (oxygen of W3). (I) Similar to (H), but at O3. Only PAP curves are shown for the gradients.

Let us consider the derivative of the AP energy with respect to nuclear coordinate X_A, where atom A belongs to group i:

$$\frac{\partial V^{\mathrm{Q1/Q2;AP}}_{1,2,\ldots,N+2}}{\partial X_{A\in i}} = \frac{\partial V^{\mathrm{Q2}}_{1,2,\ldots,N+2}}{\partial X_{A\in i}} - \frac{\partial V^{\mathrm{Q2}}_{1,2,\ldots,N+1}}{\partial X_{A\in i}} + \frac{\partial V^{\mathrm{Q1/Q2;AP}}_{1,2,\ldots,N+1}}{\partial X_{A\in i}} \tag{14}$$

The first two terms on the right-hand side are the Q2 gradients with respect to the nuclear coordinate X_A for the entire system and the Q1-buffer cohort, respectively. The third term is the AP gradient for the Q1-buffer cohort. If group i is in the Q2 zone, the last two terms vanish, and the gradient of the AP energy is simply the gradient of the Q2 energy.

For PAP, the gradients of the last term in the right-hand side of eq 14 (truncated after the 2-body interaction) are given by

$$\begin{aligned}\frac{\partial V^{\mathrm{Q1/Q2;PAP}}_{12\ldots(N+1)}}{\partial X_{A\in i}} &= P_i\frac{\partial V^{\mathrm{Q1}}_{i}}{\partial X_{A\in i}} + (1-P_i)\frac{\partial V^{\mathrm{Q2}}_{i}}{\partial X_{A\in i}} \\ &+ \sum_{ij}\left(P_iP_j\left(\frac{\partial V^{\mathrm{Q1}}_{ij}}{\partial X_{A\in i}} - \frac{\partial V^{\mathrm{Q1}}_{i}}{\partial X_{A\in i}}\right)\right. \\ &\left. + (1-P_iP_j)\left(\frac{\partial V^{\mathrm{Q2}}_{ij}}{\partial X_{A\in i}} - \frac{\partial V^{\mathrm{Q2}}_{i}}{\partial X_{A\in i}}\right)\right)\end{aligned} \tag{15}$$

For IAP, the gradients $\frac{\partial V^{\mathrm{Q1/Q2;IAP}}_{1,2,\ldots,N+1}}{\partial X_{A\in i}}$ are computed using common procedures[55,56] except that the orbital coefficients are weighted by the appropriate smoothing function (eq 13). In addition, care must be taken to subtract the energy derivatives $\frac{\partial V^{\mathrm{Q1/Q2;IAP}}_{i}}{\partial X_{A\in i}}$ for the individual group i.

3. COMPUTATIONS

The energy and gradient of both PAP and IAP are implemented in the software GAMESS.[57−59] For simplicity in data analysis and effectiveness in pinpointing the problems, we chose the Hartree−Fock level of theory for all test calculations. For the IAP formalism, the interaction energies and gradients were computed using several basis set combinations (Q1/Q2): 6-31G(d)[60−62]/STO-3G,[63] 6-31G(d)/3-21G,[64] 6-311G(2df,2pd)[65,66]/6-31G(d,p),[60−62] Dunning's cc-pVTZ/cc-pVDZ,[67] and the Sapporo TZP/DZP.[68,69] The mixed basis sets are detailed in Table S1 of the Supporting Information. Unless otherwise specified, we will present the calculations with the combination Q1 = 6-31G(d)/Q2 = STO-3G, which shows a drastic difference between the Q1 and Q2 basis sets. Such large differences between basis sets make the combination a critical test for our algorithms,

https://doi.org/10.1021/acs.jpca.3c05600
J. Phys. Chem. A 2023, 127, 10320−10333

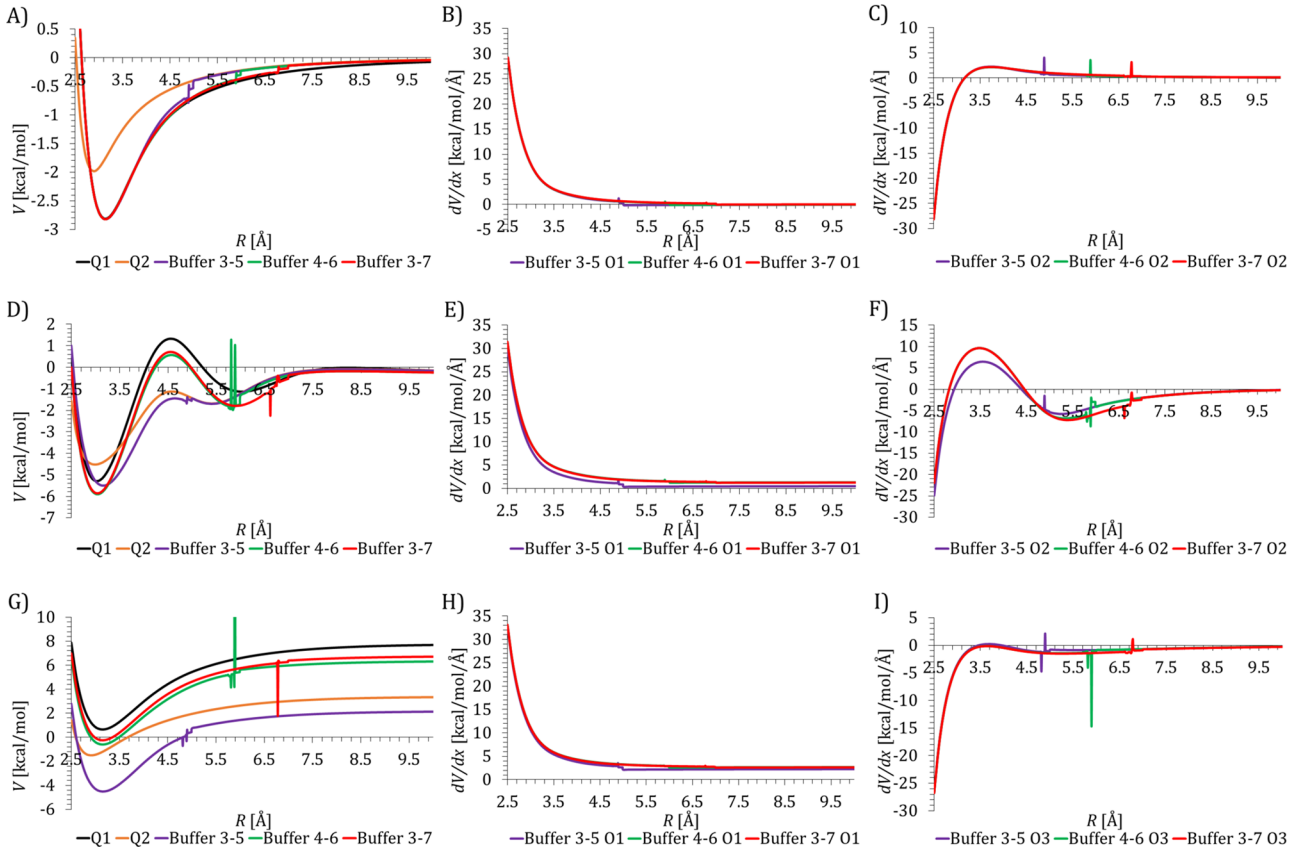


Figure 3. (A) Q1, Q2, and IAP energy profiles with no cutoff applied to the smoothing function (*coef-cut* = 0) for the water dimer model, where R is the distance between O1 and O2, respectively. The basis set combination Q1 = 6-31G(d)/Q2 = STO-3G is employed. (B) Similar to (A), but for the x components of the IAP gradient (dV/dx) at O1. (C) Similar to (B), but at O2. (D−F) Similar to (A−C), but for the water trimer model. (G, H) is similar to (A, B), but for the tetramer model, and R is the distance between O1 and O3. (I) Similar to (H), but at O3. Only IAP curves are shown for the gradients.

although the minimum basis set STO-3G is admittedly very unlikely to be employed in practical applications aiming at meaningful accuracy. The coefficients and exponents of all tested basis sets were obtained from the basis set exchange repository.[70,71]

A set of water-cluster models (Figure 1, see also Table S2 in the Supporting Information) was employed for testing the PAP and IAP algorithms. The oxygen atoms serve as the delegated atoms in the measurements of the distance R. The Q1 zone was centered at O1, the oxygen atom of the first water molecule (W1). For convenience, a buffer zone is denoted by R_{min}−R_{max}; for example, a buffer of 3−5 Å features R_{min} = 3 Å and R_{max} = 5 Å. Three different setups of buffer zone were tested: 3−5, 4−6, and 3−7 Å. The 3−5 Å buffer gives the most critical test because it is not only narrow in thickness but also close to the QM zone center, although it leads to a QM-buffer cohort that is probably too small for actual applications. In the dimer (Figure 1A) and trimer (Figure 1B) models, the second water molecule (W2) moved across the buffer zone from the Q1 to Q2 zones, while the traveling molecule is the third water molecule (W3) in the tetramer (Figure 1C) model. All the other water molecules were fixed at their respective positions and were classified as Q1, buffer, or Q2 according to the setup of the buffer zone (see Table S3 in the Supporting Information). All water molecules were kept rigid during the calculations.

During the scan of an energy profile, the mobile water molecule was displaced along the x direction, initially at a large step of ΔR_L = 0.1 Å but switched to a much smaller step of ΔR_S = 0.01 Å once its oxygen atom moved near (within 15% of the buffer thickness) the buffer/Q2 boundary. With such a fine step size, it is unlikely to miss the irregularities in the energy curve. The discontinuity in a potential energy curve at the buffer/Q2 boundary, $\Delta V(R_{max})$, was estimated as the absolute difference between the extrapolated energy $V_{ex}(R_{max})$ and the actually computed energy $V(R_{max})$ at the boundary:

$$\Delta V(R_{max}) = |V_{ex}(R_{max}) - V(R_{max})| \quad (16)$$

The extrapolated energy was based on the data of the last two steps in the scan within the buffer zone, assuming that the energy changes are the same from $R_{max} - 2\Delta R_S$ to $R = R_{max} - \Delta R_S$ and from $R = R_{max} - \Delta R_S$ to $R = R_{max}$. In other words, the gradient $\frac{dV}{dR}$ was assumed to be the same in the last two steps, which is a reasonable approximation given the small step size of 0.01 Å. This led to the following equation:

$$\Delta V(R_{max}) = |V(R_{max}) + V(R_{max} - 2\Delta R_S) - 2V(R_{max} - \Delta R_S)| \quad (17)$$

A similar procedure was used to estimate the discontinuities in the gradient components.

4. RESULTS

4.1. PAP Energies and Gradients. The PAP energy profiles are plotted against R, the distance between O1 fixed at

https://doi.org/10.1021/acs.jpca.3c05600
J. Phys. Chem. A 2023, 127, 10320−10333

Table 1. Number of MOs Removed by the GAMESS Code due to Linear Dependencies for the Water Dimer Model with the Basis Set Combination Q1 = 6-31G(d)/Q2 = STO-3G

num. MOs removed	buffer 3−5 Å		buffer 4−6 Å		buffer 3−7 Å	
	R_i (Å)	P_i ($\times 10^{-3}$)	R_i (Å)	P_i ($\times 10^{-3}$)	R_i (Å)	P_i ($\times 10^{-3}$)
0	<4.82	>7.40	<5.82	>7.40	<6.63	>7.40
1	4.82−4.85	3.76−6.34	5.82−5.85	3.76−6.34	6.63−6.71	3.41−6.86
2	4.86−4.87	2.49−3.08	5.86−5.87	2.49−3.08	6.72−6.74	2.49−3.08
5	4.88	1.97	5.88	1.97	6.76−6.77	1.74−1.97
6	4.89	1.53	5.89	1.53	6.78	1.53
9	4.90	1.16	5.90	1.16	6.80	1.16
12	4.91−4.99	0.00124−0.851	5.91−5.99	0.00124−0.851	6.81−6.99	0.000156−1.00

the QM zone center and the moving water's oxygen O2 (or O3), in Figure 2A,D,G for the water dimer, trimer, and tetramer, respectively, where the single-level Q1 and Q2 energy profiles are also plotted for comparison. The x components of the gradients (dV/dx) of O1 are plotted in Figure 2B,E,H for the dimer, trimer, and tetramer, respectively. The counterparts of the moving water's oxygen are shown in Figure 2C,F,I, respectively. Additionally, dV/dx of the other oxygen atoms is given in Figure S1 of the Supporting Information. The gradients dV/dy and dV/dz of all oxygen atoms are displayed in Figures S2 and S3, respectively. It can be seen that all energy and gradient curves are smooth throughout the Q1, buffer, and Q2 regions for all three buffer zones (3−5, 4−6, and 3−7 Å).

First, let us look at the energy profiles. The Q1 and Q2 curves are quite distinct from each other due to significant differences between the 6-31G(d) and STO-3G basis sets. The PAP energy curves generally lie between them, showing subtle variances depending on the specific model and the choice of buffer zone. (i) For the water dimer model, all PAP curves change gradually from Q1 to Q2 in the buffer region as R increases. A steeper change is observed for a narrower buffer, where the smoothing function decreases more rapidly as R increases. (ii) For the water trimer model, all energy curves display two minima: the first minimum near $R \sim 3.0$ Å at both Q1 and Q2, and the second minima near $R \sim 6.0$ Å at Q1 and ∼5.5 Å at Q2, respectively. All three buffers slightly (by ∼0.2 Å) shift the location of the first minima toward larger R. As to the second minimum, the 3−5 Å buffer reproduces the Q2 results, while the 4−6 and 3−7 Å buffers shift the location by ∼0.3 Å. The forward barrier from the first to the second minima is the highest at Q1 (∼7 kcal/mol) and lowest at Q2 (∼3 kcal/mol), with the PAP ones (4 to 5 kcal/mol) in between. (iii) For the tetramer, because of the higher complexity of the model, the PAP energy curves do not necessarily match either the Q1 or Q2 curve even when W3 moves outside the buffer zone.

Next, we examine the gradients. Overall, the gradient curves in a given panel, which are obtained with different buffer setups, have shapes similar to each other. The 3−5 Å buffer possesses a small QM zone and a narrow buffer zone, often leading to results that are more distinct from those of the other two buffers. This is especially true for the water oxygen atoms that remain in the Q1 or Q2 zone during the energy scans. For example, O1 stays in Q1 all the time, and the largest difference in its gradient components is found to be in dV/dx for the trimer model, which is ∼2 kcal/mol/Å between the 3−5 Å buffer and the other two buffers. (Please note that the vertical axes in the plots of the Supporting Information are generally of much smaller scales.) For the moving water's oxygen, the gradients show more noticeable discrepancies between buffer setups. For example, for O2 in the trimer model, dV/dx with the 3−5 Å buffer differs by up to ∼3 kcal/mol/Å from those with the other two buffers. However, no obvious pattern can be noted. Inspections of the dV/dy and dV/dz curves reach similar conclusions. These results imply that in general the impacts by the PAP interpolations may be insignificant on the dynamics of the Q1 and Q2 atoms (unless they are close to the buffer zone), while the buffer atoms could be more substantially affected. A wide buffer may alleviate the impacts.

4.2. IAP Energies and Gradients: Basis Set Linear Dependency in Alchemical-QM Calculations. The potential energy curves and gradients with respect to the x coordinate (dV/dx) obtained for the water models by IAP without any special treatment (i.e., no cutoff is applied) are shown in Figure 3. While the IAP energy profiles resemble those of the PAP, they are not mandated to match the latter. The prominent issue is, however, the irregularities that manifest as "jumps" or "spikes" near the buffer/Q2 boundaries in the plots. Taking the dimer model as an example, for the 3−5 Å buffer, irregularities appear at the same locations ($R \sim 4.8$ Å) in the energy and gradient plots. Increasing the size of the buffer zone (3−7 Å) or moving the buffer zone further away from the Q1 center (4−6 Å) pushes the irregularities further away, alleviating the impacts on the Q1 zone, but does not eliminate them.

We found that the origin of this issue is the basis set linear dependencies in the alchemical-QM computations in the IAP model. Specifically, as the buffer group moves close to the buffer/Q2 boundary, the value of P_i becomes very small, causing numerical difficulties in the variational calculations. The GAMESS code thus removes the linearly dependent molecular orbitals (MOs) from the variation space to avoid an ill-defined $\mathbf{S}^{-1/2}$ matrix, leading to numerical noise when diagonalizing the Fock matrix.

We will use the dimer model as an example. There are in total 38 atomic orbitals, 19 from 6-31G(d) for W1, and 19 from mixed 6-31G(d)/STO-3G for W2. Table 1 shows the number of linearly dependent MOs removed from the variation space. For all three buffers, more MOs are dropped as W2 approaches the buffer/Q2 boundary. By the time W2 is 0.01 Å from the buffer/Q2 boundary, nearly a third of the MOs are removed from the variation space. We note that the number of linearly dependent MOs is the same for the alchemical-QM calculations of both the dimer and monomer, so the percentage of MO removed will be even more prominent (over half) for the monomer. When W2 is in the buffer zone, the interaction energy between the two molecules is given by

https://doi.org/10.1021/acs.jpca.3c05600
J. Phys. Chem. A 2023, 127, 10320−10333

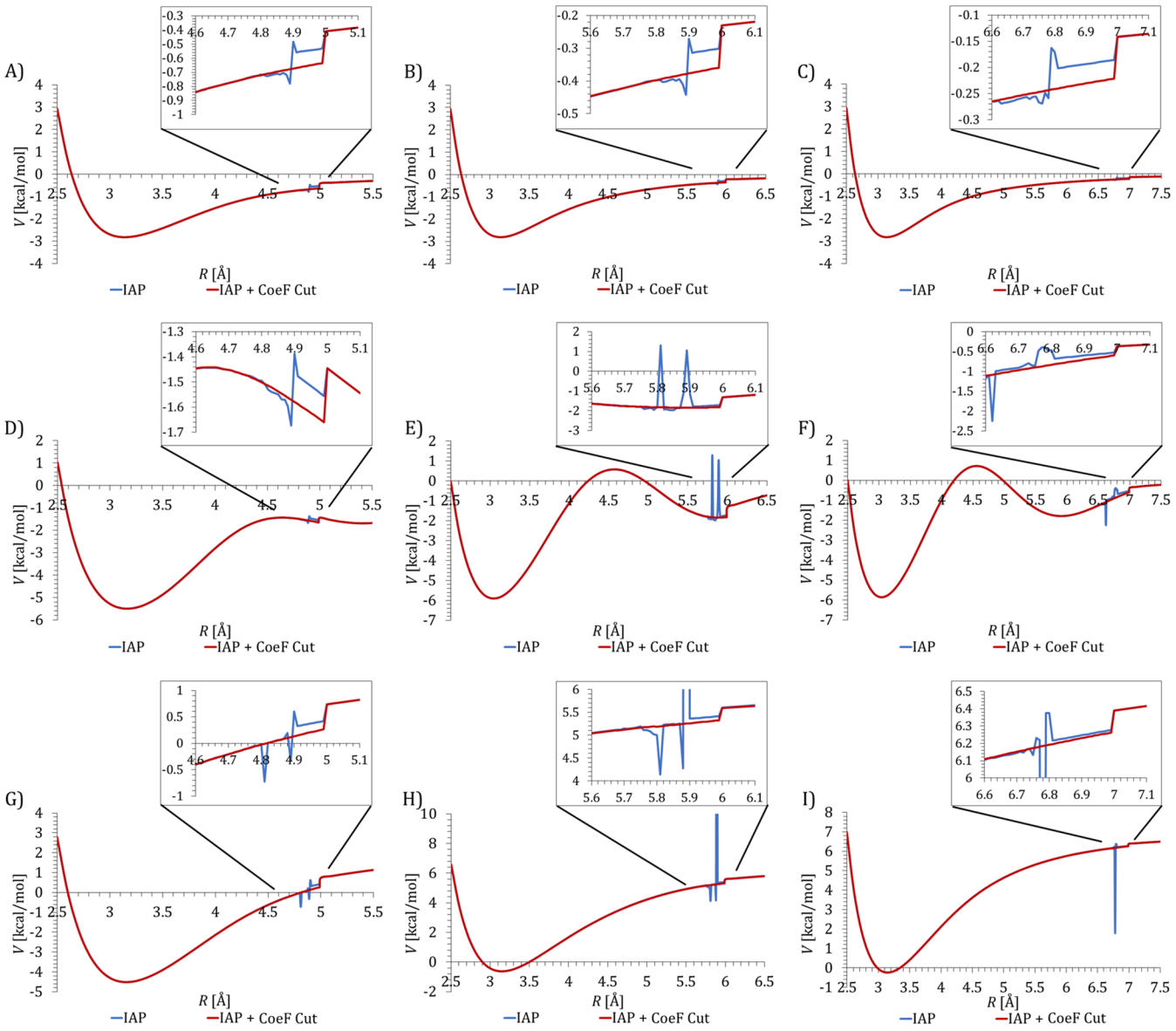


Figure 4. IAP potential energy profiles (blue: no cutoff; red: with *coef-cut* = 0.01) for the water dimer model with the (A) 3−5 Å, (B) 4−6 Å, and (C) 3−7 Å buffer zones, respectively. The basis set combination Q1 = 6-31G(d)/Q2 = STO-3G is employed. The energy profiles for the water trimer model are shown in (D−F), and those for the tetramer model are shown in (G−I).

$$\Delta V = V_{12}^{\mathrm{Q1/Q2;IAP}} - V_{1}^{\mathrm{Q1}} - V_{2}^{\mathrm{Q1/Q2;IAP}} \tag{18}$$

Errors in the energy are therefore introduced in both IAP terms (the dimer energy $V_{12}^{\mathrm{Q1/Q2;\,IAP}}$ and the monomer energy $V_{2}^{\mathrm{Q1/Q2;\,IAP}}$), but these errors do not exactly cancel each other out. It is evident in Table 1 that the number of MOs dropped correlates with the value of P_i. For a 2 Å wide buffer, the linear dependencies occur within 0.2 Å of the buffer/Q2 boundary, regardless of the position of the buffer zone. For a 4 Å wide buffer zone, they occur within 0.4 Å of the buffer/Q2 boundary.

Another issue that we ran into was that the SCF did not converge at several geometries during the potential energy curves (near the buffer/Q2 boundary) with the default cutoff for electron integrals (10^{-9}) in GAMESS. To achieve convergence, we tightened the cutoff to 10^{-11}, resulting in more integrals being included in the SCF procedure.

Clearly, numerical noises caused by removing linearly dependent functions are added to the IAP energies and gradients when the smoothing function P_i becomes very small. The locations and magnitudes of the irregularities depend on the specific model systems and/or different basis set combinations. One may attempt to avoid these linear dependencies by tightening the linear dependence threshold. However, difficulties in SCF convergence will likely arise.

4.3. IAP Energies and Gradients: Cutoff (*Coef-Cut*) for the Weighted Coefficients. Several ways to alleviate linear dependency issues have been suggested, but no universally applicable solution has been found to the best of our knowledge.[72,73] It is *a priori* difficult to predict exactly when these issues will occur during a simulation and how they may affect the results. As shown in Table 1, linear dependencies begin to occur when P_i values are smaller than ∼0.01 for the water dimer. We thus introduce a cutoff value *coef-cut* = 0.01 for the weighted MO coefficients (eq 13). We note that a different cutoff value may be more suitable for other model systems or other Q1/Q2 basis set combinations. As can easily

https://doi.org/10.1021/acs.jpca.3c05600
J. Phys. Chem. A 2023, 127, 10320−10333

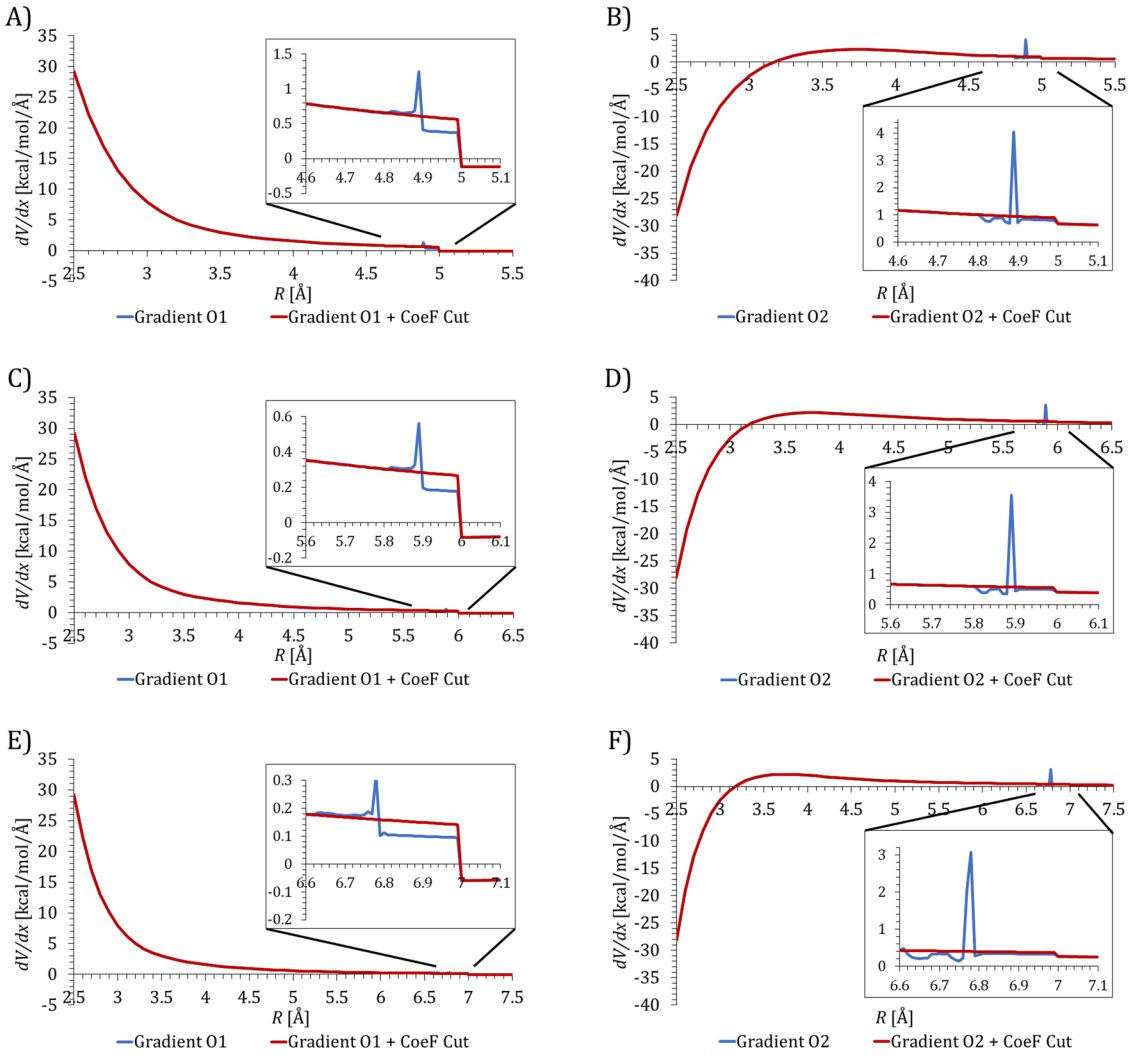


Figure 5. IAP gradient components dV/dx (blue: no cutoff; red: with *coef-cut* = 0.01) for the oxygen atoms (A) O1 and (B) O2 of the water dimer model using the 3−5 Å buffer and the basis set combination Q1 = 6-31G(d)/Q2 = STO-3G. R is the distance between O1 and O2. (C, D): the same as (A, B), but with 4−6 Å buffer. (E, F): the same as (A, B), but with the 3−7 Å buffer.

be seen in Figure 4, using the cutoff *coef-cut* essentially eliminates the irregularities in the energy profiles near the buffer/Q2 boundary for all systems and buffer zones, which is very encouraging. As expected, the irregularities in the gradients also disappear, as exemplified by the gradient components for the dimer model: dV/dx in Figure 5 as well as dV/dy and dV/dz in Figures S4 and S5 of the Supporting Information. For brevity, we do not elaborate the results for the trimer and tetramer, but we point out that they are also essentially free of these irregularities, as shown in Figures S6−S11 of the Supporting Information. Some tiny fluctuations smaller than 0.0005 kcal/mol/Å can be seen in the gradient curves, which are likely due to random numerical noises. In addition, no SCF convergence issue occurred when a *coef-cut* was applied with the default integral cutoff of 10^{-9}.

One intriguing question is how the electron integrals are affected by the interpolations and *coef-cut*. We found that, when no *coef-cut* value is used, large fluctuations occur in the one-electron and two-electron energy terms. When a *coef-cut* value is used, the one-electron and two-electron energy terms vary smoothly. This is exemplified by a single water molecule in the 3−5 Å buffer zone in Figure S12 of the Supporting Information.

Unfortunately, a discontinuity at the buffer/Q2 boundary remains due to two factors. The first factor is the "intrinsic" discontinuity of the current IAP Q1/Q2 energy definition (eq 1) at the buffer/Q2 boundary, as exemplified by the dimer model. When W2 is in the Q2 zone, infinitesimally close to the buffer/Q2 boundary, the interactions between W1 and W2 are determined at the Q2 level. However, when W2 is in the buffer zone infinitesimal to the buffer/Q2 boundary, the interactions are computed with W1 at the Q1 level and W2 at the Q2 level. The two computed interactions can show considerable discrepancies with each other if Q1 and Q2 differ significantly, as in the case of 6-31G(d) and STO-3G here. For larger Q1 and Q2 basis sets, the two computed interactions both move toward the complete basis set limit, yielding better agreements with each other (Section 4.4). A modified definition of IAP energy may help resolve this issue, as will be discussed in the following Section 5.

The second factor that may contribute to the remaining discontinuity is the large change in variational space between when the mobile water molecule is in the buffer and Q2 zones. A water molecule in the buffer zone possesses both Q1 and Q2 basis functions, even though the weighted coefficients for the Q1 functions are very small. However, it only has Q2 basis

https://doi.org/10.1021/acs.jpca.3c05600
J. Phys. Chem. A 2023, 127, 10320−10333

functions as soon as it enters the Q2 zone. The availability of more (Q1) basis functions to the water in the buffer zone lowers the energy.

Table 2 shows the estimated energy discontinuity at the buffer-Q2 boundary, which varies between 0.08 and 0.5 kcal/mol. The discontinuity in energy also leads to discontinuity in gradients. Table 3 lists the average and maximum changes in gradient components, and the statistics are taken on all three buffers. These changes are usually smaller than 1 kcal/mol/Å and tend to decrease as the buffer/Q2 boundary is pushed further away from the Q1 center. However, sometimes, the maximum changes can be much larger, reaching the order of several kcal/mol/Å.

Table 2. Energy Discontinuity (in kcal/mol) at the Buffer/Q2 Boundary with Q1 = 6-31G(d)/Q2 = STO-3G and *Coef-Cut* = 0.01

	buffer 3−5 Å	buffer 4−6 Å	buffer 3−7 Å
dimer	0.223	0.127	0.079
trimer	0.223	0.512	0.220
tetramer	0.457	0.266	0.125

We also tested another idea, which uses a cutoff in the smoothing function value P_i itself (denoted P_i-*cut*) instead of a cutoff in the weighted coefficients. The results obtained with P_i-*cut* = 0.01 are essentially the same as those obtained with *coef-cut* = 0.01 (see Tables S4 and S5 and Figures S13 and S14 in the Supporting Information). However, we found that the use of a cutoff to the P_i value sometimes leads to difficulties converging the SCF at certain geometries where P_i is small. At these geometries, a tighter integral cutoff value still had to be used to achieve convergence, which increased the computational cost. Therefore, P_i-*cut* is deemed not to be as robust as *coef-cut*.

4.4. IAP Energies and Gradients: Basis Set Combinations. One important factor that notably affects the discontinuities in the IAP energy and gradients is the basis set combination. To gain further insights, additional sets of basis set combinations were tested: 6-31G(d)/3-21G, cc-pVTZ/cc-pVDZ, 6-311G(2df,2pd)/6-31G(d,p), and Sapporo TZP/DZP. The 3−5 Å buffer with a *coef-cut* value of 0.01 was used in these calculations, except for the 6-311G(2df,2pd)/6-31G(d,p) combination, where a *coef-cut* of 0.05 is used. The energy profiles are illustrated in Figure 6, while the gradient curves are shown in Figures S15−S27 of the Supporting Information. The energies and gradients overall vary smoothly up to the buffer/Q2 boundary with the remaining discontinuity at the Q2 boundary.

The discontinuities in the energy and dV/dx gradient profiles are analyzed in Table 4 and Figure 7, respectively. The discontinuities in the dV/dy and dV/dz gradient profiles are summarized in Figures S28 and S29 of the Supporting Information. In general, a larger Q2 basis set tends to yield a smaller discontinuity in both the energy and gradients. In fact, when a polarized double-ζ (instead of the minimum STO-3G) is used for Q2, which is usually required to provide reasonable accuracy in practical applications, the discontinuities become much smaller for both the energies (∼0.2 kcal/mol or less) and gradients (∼0.5 kcal/mol/Å or less). If both Q1 and Q2 approach the complete basis set limit, then the discontinuities would be expected to become vanishingly small.

5. DISCUSSION

In this work, the energies and gradients of the PAP and IAP methods were computed and analyzed for the three model water systems. For PAP, the energies and gradients vary smoothly as a water molecule travels across the Q1, buffer, and Q2 zones. For IAP, irregularities in the energies and gradients were observed in the proximity of the buffer/Q2 boundary. They are caused by numerical noises due to the linear dependencies introduced in the basis functions as the scaled atomic orbital coefficients approach zero in the alchemical-QM calculations. The GAMESS code handles the linearly dependent MOs by removing them from the variational space in SCF, causing numerical noises in the IAP energy and gradients. Encouragingly, these numerical noise-induced irregularities were effectively removed by introducing a cutoff value for the

Table 3. Average and Maximum Changes in the Gradient Components dV/dx, dV/dy, and dV/dz (in kcal/mol/Å) at the Buffer/Q2 Boundary with Q1 = 6-31G(d)/Q2 = STO-3G and *Coef-Cut* = 0.01, Computed over All Three Buffers

			O1	O2	O3	O4
dimer	dV/dx	Ave	0.412	0.156	n/a	n/a
		Max	0.684	0.217	n/a	n/a
	dV/dy	Ave	0.279	0.275	n/a	n/a
		Max	0.431	0.447	n/a	n/a
	dV/dz	Ave	0.000221	0.000459	n/a	n/a
		Max	0.000281	0.000808	n/a	n/a
trimer	dV/dx	Ave	0.388	0.945	0.609	n/a
		Max	0.679	1.44	0.939	n/a
	dV/dy	Ave	0.303	1.75	0.204	n/a
		Max	0.427	3.93	0.592	n/a
	dV/dz	Ave	0.000356	0.000569	0.000662	n/a
		Max	0.000521	0.000803	0.00171	n/a
tetramer	dV/dx	Ave	0.398	0.397	0.423	0.103
		Max	0.664	0.808	0.470	0.170
	dV/dy	Ave	0.327	0.441	1.21	0.0737
		Max	0.505	0.711	1.70	0.160
	dV/dz	Ave	0.000412	0.000883	0.00178	0.0000988
		Max	0.000656	0.00144	0.00269	0.000231

https://doi.org/10.1021/acs.jpca.3c05600
J. Phys. Chem. A 2023, 127, 10320−10333

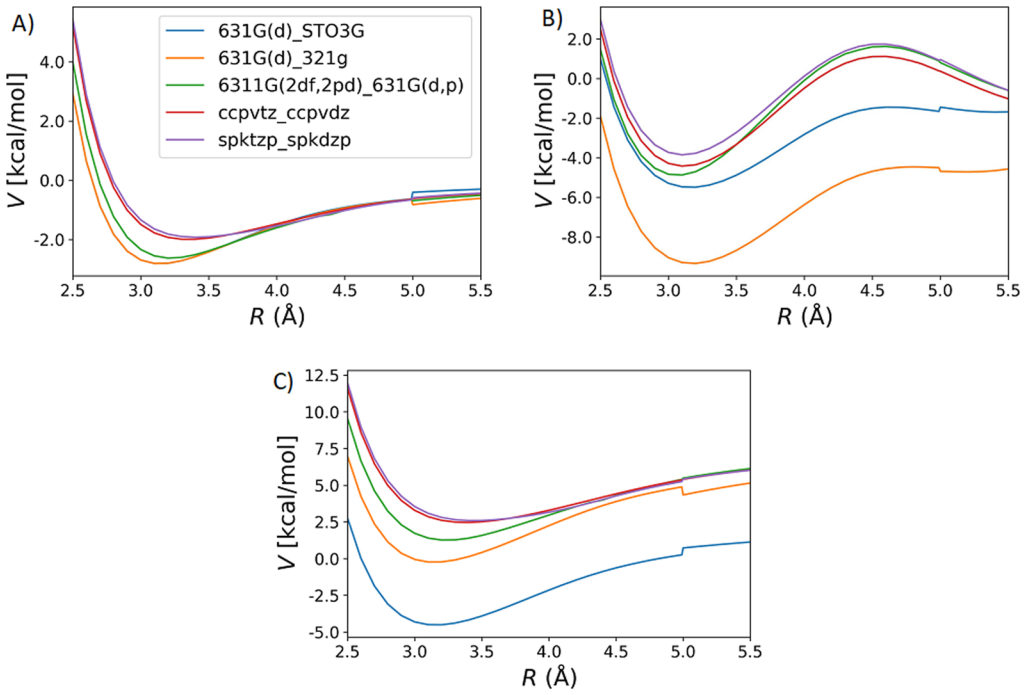


Figure 6. IAP energy profiles of the (A) dimer, (B) trimer, and (C) tetramer models with the 3−5 Å buffer, employing various basis set combinations. A *coef-cut* value of 0.01 is used for all except the 6-311G(2df,2pd)/6-31G(d,p) combination where a *coef-cut* value of 0.05 is used. *R* is the distance between O1 and the oxygen of the moving water molecule.

Table 4. Energy Discontinuity (in kcal/mol) at the Buffer/Q2 Boundary with Different Basis Set Combinations and *Coef-Cut* = 0.01

	dimer	trimer	tetramer
6-31G(d)/STO-3G	0.223	0.223	0.457
6-31G(d)/3-21G	0.177	0.177	0.561
6-311G(2df,2pd)/6-31G(d,p)[a]	0.0552	0.0554	0.0668
cc-pVTZ/cc-pVDZ	0.0194	0.0111	0.0242
Sapporo TZP/DZP	0.0655	0.0652	0.160

[a] *coef-cut* = 0.05.

scaled atomic orbital coefficients, preventing them from getting too small.

However, discontinuities in the energy and gradient remained at the buffer/Q2 boundary, caused by the intrinsic discontinuity in the current IAP energy definition as well as by the changes in variation space between when the moving water molecule is in the Q2 and buffer zones. These discontinuities are mostly small (typically less than 0.5 kcal/mol for the energy and 1 kcal/mol/Å for the gradients) but can go higher. In our tests, the gradient reached 4 kcal/mol/Å for the trimer computed with the 6-31G(d)/STO-3G basis set combination. These discontinuities tend to decrease as the buffer zone moves further away from the Q1 center. They also become less prominent with an increasing basis set size. In particular, when a polarized double-ζ basis set is used for Q2, the discontinuities are so small (~0.2 kcal/mol or less in energy and ~0.5 kcal/mol/Å or less in gradients) that they are unlikely to have significant influences on the dynamic simulations.

As an extra check, we ran an additional test on the Eigen cation ($H_9O_4^+$) with the 3−5 Å buffer and *coef-cut* = 0.01 using the cc-pVTZ/cc-pVDZ basis set combination. Eigen cation is one of the common forms of a hydrated proton, which is often considered to be delocalized over the hydronium ion and its solvation shell. The diffusion of a hydrated proton is dominated by the Grotthuss mechanism through reorganizing the network of hydrogen and covalent bonds of the water molecules.[74,75] Thus, the study of hydrate protons is very important and has been the theme of a number of adaptive QM/MM simulations.[5,6,8,9,12,23,41,45−47] Here, similar to what we did in the water clusters, we gradually pulled one water molecule away from the solvation shell of the hydronium ion while keeping the other atoms fixed. The model and results are detailed in Figures S30−S32 in the Supporting Information. The discontinuities are again found to be very small: ~0.16 kcal/mol in energy and less than 0.24 kcal/mol/Å in the gradient components. These results are well in line with the above assessments of the discontinuities (~0.2 kcal/mol or less in energy and ~0.5 kcal/mol/Å or less in gradient components) on the water clusters, even though the energy curve of the positively charged Eigen cation has a deeper well. Although this is just one very simple case, the agreement is encouraging.

One possible solution to alleviate the remaining bumps could be a different definition of the IAP energy from the current one. The IAP energy has been defined through eq 1 so that it resembles the IMOMO (integrated MO plus MO)[76] energy as much as possible, but we pay a price for doing so by having an intrinsic discontinuity at the buffer/Q2 boundary. On the other hand, although being similar to IMOMO is something nice to have, it is not a prerequisite. Furthermore, the current IAP definition does not take full advantage of the power of alchemical-QM calculations, which can put Q2 groups together with the Q1-buffer cohort in one calculation, i.e., a calculation with all atoms in the Q1, buffer, and Q2 zones where the atoms in different zones are described differently in the Fock and overlap matrices. Inclusion of the Q2 atoms will increase the computational costs of the alchemical calculation,

https://doi.org/10.1021/acs.jpca.3c05600
J. Phys. Chem. A 2023, 127, 10320−10333

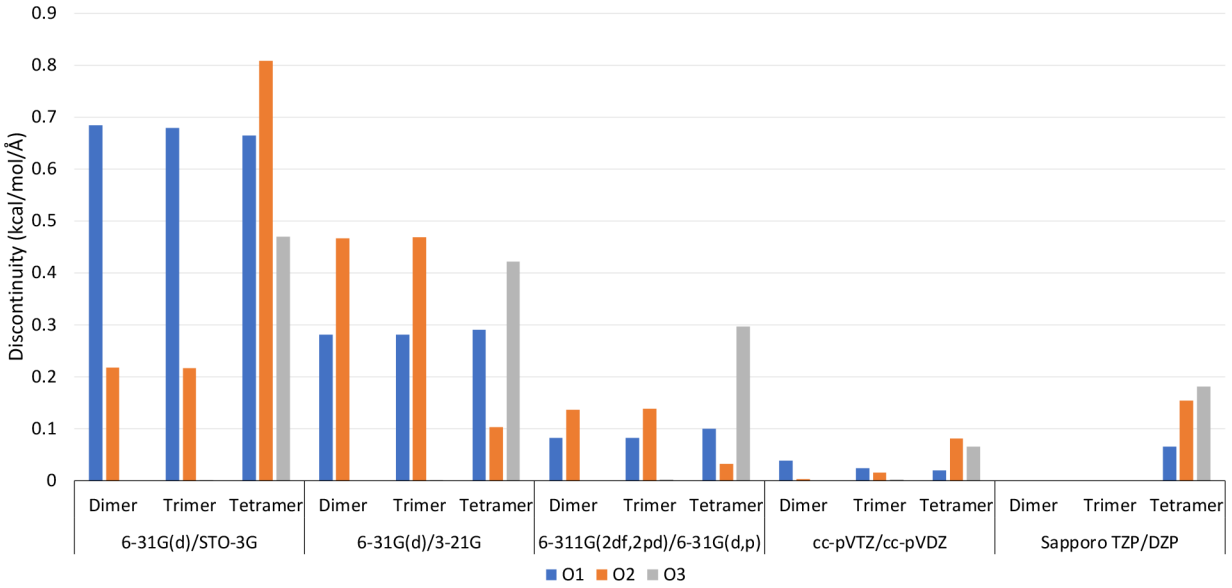


Figure 7. Discontinuities in the gradient components dV/dx (kcal/mol/Å) at the buffer/Q2 boundary for the water dimer, trimer, and tetramer models with the 3−5 Å buffer for various basis set combinations. A *coef-cut* value of 0.01 is used, except for the 6-311G(2df,2pd)/6-31G(d,p) combination where *coef-cut* = 0.05. The discontinuities in the dimer and trimer for Sapporo TZP/SZP are too small to be visible. Because the discontinuities for O4 are too small to be visible for all basis set combinations, we omitted O4 from the plot.

and the trade-off is fewer calculations because we would no longer need to do the two single-level Q2 calculations in eq 1. This option will be explored in the future.

Another ansatz could be used to scale the Fock and overlap matrix elements for the alchemical-QM calculations without mixing the basis sets, such as

$$\begin{cases} F_{\mu\upsilon}^{\mathrm{IAP}} = F_{\mu\upsilon}^{\mathrm{Q1}} \text{ and } S_{\mu\upsilon}^{\mathrm{IAP}} = S_{\mu\upsilon}^{\mathrm{Q1}}, \text{ if } \mu, \nu \in \mathrm{Q1} \\ F_{\mu\upsilon}^{\mathrm{IAP}} = P_\nu F_{\mu\upsilon}^{\mathrm{Q1}} + (1 - P_\nu) F_{\mu\upsilon}^{\mathrm{Q2}} \text{ and } S_{\mu\upsilon}^{\mathrm{IAP}} \\ \quad = P_\nu S_{\mu\upsilon}^{\mathrm{Q1}} + (1 - P_\nu) S_{\mu\upsilon}^{\mathrm{Q2}}, \text{ if } \mu \in \mathrm{Q1}, \nu \in \text{buffer} \\ F_{\mu\upsilon}^{\mathrm{IAP}} = P_\mu P_\nu F_{\mu\upsilon}^{\mathrm{Q1}} + (1 - P_\mu P_\nu) F_{\mu\upsilon}^{\mathrm{Q2}} \text{ and } S_{\mu\upsilon}^{\mathrm{IAP}} \\ \quad = P_\mu P_\nu S_{\mu\upsilon}^{\mathrm{Q1}} + (1 - P_\mu P_\nu) S_{\mu\upsilon}^{\mathrm{Q2}}, \text{ if } \mu, \nu \in \text{buffer} \end{cases} \tag{19}$$

In this approach, the Fock and overlap matrices are constructed with the original Q1 ($F_{\mu\upsilon}^{\mathrm{Q1}}$, $S_{\mu\upsilon}^{\mathrm{Q1}}$) and Q2 ($F_{\mu\upsilon}^{\mathrm{Q2}}$, $S_{\mu\upsilon}^{\mathrm{Q2}}$) basis sets, and the smoothing coefficients are subsequently applied to each element based on the group where the basis functions μ and ν are located. This work will be the subject of a subsequent publication.

6. CONCLUSIONS

In conclusion, the PAP and IAP energy and gradient profiles were analyzed for the Q1/Q2 systems. While the PAP energy and gradients vary smoothly, irregularities due to linear dependencies were observed near the buffer/Q2 boundary for IAP. These irregularities were fixed by introducing a cutoff in the weighted atomic orbital coefficients. However, discontinuities at the Q2 buffer boundary remain. These discontinuities are due to both the current definition of the IAP energy and the change in variational space when a molecule moves from the buffer to the Q2 zone. It was shown that these discontinuities can be made smaller by moving the buffer/Q2 boundary further away from the Q1 center and increasing the basis set sizes. Dynamics simulations and new definitions of IAP energy will be explored in future work.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpca.3c05600.

Oxygen and hydrogen mixed basis sets; Cartesian coordinates of the water cluster models; classifications of water molecules in these models; remaining discontinuities in IAP energy and gradients with P_i-*cut* = 0.01 for the water cluster models; discontinuities in the IAP gradient components with the different basis set combinations for the water cluster models; additional PAP gradient components for the water cluster models; additional IAP gradient components for water dimer with *coef-cut* = 0.01; gradients of all atoms for the water trimer and tetramer; one-electron and two-electron IAP energies of the water molecule in the 3−5 buffer zone with *coef-cut* = 0.01 and *coef-cut* = 0.00 (i.e., no cutoff); IAP energy profiles with P_i-*cut* = 0.01; IAP gradients dV/dx with P_i-*cut* = 0.01 for water dimer; IAP gradient components for the water dimer, trimer and tetramer and various basis set combinations; summaries of IAP gradient component discontinuities for dV/dy and dV/dz with various basis set combinations for the water clusters; model of the Eigen cation; IAP energy profile of Eigen cation; and IAP gradient components of the Eigen cation (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Emilie B. Guidez − *Department of Chemistry, University of Colorado Denver, Denver, Colorado 80217, United States;* orcid.org/0000-0003-1961-0469; Email: emilie.guidez@ucdenver.edu

https://doi.org/10.1021/acs.jpca.3c05600
J. Phys. Chem. A 2023, 127, 10320−10333

Hai Lin − *Department of Chemistry, University of Colorado Denver, Denver, Colorado 80217, United States;* orcid.org/0000-0002-3525-9122; Email: hai.lin@ucdenver.edu

Author

Anh L. Tran − *Department of Chemistry, University of Colorado Denver, Denver, Colorado 80217, United States;* orcid.org/0000-0001-5367-1682

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpca.3c05600

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work is supported by the National Science Foundation (CHE-2153441) and National Institute of General Medical Sciences (1R15GM141728-01). This work is also supported by the University of Colorado Denver. Calculations were run on the CU Denver Alderaan cluster, supported by NSF (2019089)

■ REFERENCES

(1) Mato, J.; Duster, A. W.; Guidez, E. B.; Lin, H. Adaptive-partitioning multilayer dynamics simulations: 1. On-the-fly switch between two quantum levels of theory. *J. Chem. Theory Comput.* **2021**, *17*, 5456−5465.

(2) Heyden, A.; Lin, H.; Truhlar, D. G. Adaptive partitioning in combined quantum mechanical and molecular mechanical calculations of potential energy functions for multiscale simulations. *J. Phys. Chem. B* **2007**, *111* (9), 2231−2241.

(3) Pezeshki, S.; Davis, C.; Heyden, A.; Lin, H. Adaptive-partitioning QM/MM dynamics simulations: 3. Solvent molecules entering and leaving protein binding sites. *J. Chem. Theory Comput.* **2014**, *10* (11), 4765−4776.

(4) Pezeshki, S.; Lin, H. Adaptive-partitioning redistributed charge and dipole schemes for QM/MM dynamics simulations: On-the-fly relocation of boundaries that pass through covalent bonds. *J. Chem. Theory Comput.* **2011**, *7* (11), 3625−3634.

(5) Pezeshki, S.; Lin, H. Recent developments in adaptive QM/MM. In *Quantum Modeling of Complex Molecular Systems*; Rivail, J.-L.; Ruiz-Lopez, M.; Assfeld, X., Eds.; Challenges and Advances in Computational Chemistry and Physics, Springer, 2015; pp 93−113.

(6) Pezeshki, S.; Lin, H. Adaptive-partitioning QM/MM for molecular dynamics simulations: 4. Proton hopping in bulk water. *J. Chem. Theory Comput.* **2015**, *11* (6), 2398−2411.

(7) Pezeshki, S.; Lin, H. Recent developments in QM/MM methods towards open-boundary multi-scale simulations. *Mol. Simul.* **2015**, *41* (1−3), 168−189. (acccessed 2014/05/12)

(8) Duster, A. W.; Lin, H. Tracking proton transfer through titratable amino acid side chains in adaptive QM/MM simulations. *J. Chem. Theory Comput.* **2019**, *15* (11), 5794−5809.

(9) Duster, A. W.; Garza, C. M.; Aydintug, B. O.; Negussie, M. B.; Lin, H. Adaptive partitioning QM/MM for molecular dynamics simulations: 6. Proton transport through a biological channel. *J. Chem. Theory Comput.* **2019**, *15*, 892−905.

(10) Duster, A.; Wang, C.-H.; Lin, H. Adaptive QM/MM for molecular dynamics simulations: 5. On the energy-conserved permuted adaptive-partitioning schemes. *Molecules* **2018**, *23* (9), 2170.

(11) Duster, A. W.; Wang, C.-H.; Garza, C. M.; Miller, D. E.; Lin, H. Adaptive quantum/molecular mechanics: What have we learned, where are we, and where do we go from here? *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2017**, *7* (5), 1310. e1310-n/a

(12) Duster, A.; Garza, C.; Lin, H. Adaptive partitioning QM/MM dynamics simulations for substrate uptake, product release, and solvent exchange. In *Computational Approaches for Studying Enzyme Mechanism*, Voth, G. A. Ed.; Methods in Enzymology, Elsevier, 2016; pp 342−358.

(13) Kerdcharoen, T.; Liedl, K. R.; Rode, B. M. A QM/MM simulation method applied to the solution of Li^+ in liquid ammonia. *Chem. Phys.* **1996**, *211*, 313−323.

(14) Kerdcharoen, T.; Morokuma, K. ONIOM-XS: an extension of the ONIOM method for molecular simulation in condensed phase. *Chem. Phys. Lett.* **2002**, *355* (3,4), 257−262.

(15) Kerdcharoen, T.; Morokuma, K. Combined quantum mechanics and molecular mechanics simulation of Ca^{2+}/ammonia solution based on the ONIOM-XS method: Octahedral coordination and implication to biology. *J. Chem. Phys.* **2003**, *118* (19), 8856−8862.

(16) Bulo, R. E.; Ensing, B.; Sikkema, J.; Visscher, L. Toward a practical method for adaptive QM/MM simulations. *J. Chem. Theory Comput.* **2009**, *5* (9), 2212−2221.

(17) Nielsen, S. O.; Bulo, R. E.; Moore, P. B.; Ensing, B. Recent progress in adaptive multiscale molecular dynamics simulations of soft matter. *Phys. Chem. Chem. Phys.* **2010**, *12* (39), 12401−12414.

(18) Takenaka, N.; Kitamura, Y.; Koyano, Y.; Nagaoka, M. The number-adaptive multiscale QM/MM molecular dynamics simulation: Application to liquid water. *Chem. Phys. Lett.* **2012**, *524*, 56−61.

(19) Bulo, R. E.; Michel, C.; Fleurat-Lessard, P.; Sautet, P. Multiscale modeling of chemistry in water: Are we there yet? *J. Chem. Theory Comput.* **2013**, *9* (12), 5567−5577. (acccessed 2013/11/19)

(20) Waller, M. P.; Kumbhar, S.; Yang, J. A density-based adaptive quantum mechanical/molecular mechanical method. *ChemPhysChem* **2014**, *15* (15), 3218−3225.

(21) Watanabe, H. C.; Kubař, T.; Elstner, M. Size-consistent multipartitioning QM/MM: A stable and efficient adaptive QM/MM method. *J. Chem. Theory Comput.* **2014**, *10* (10), 4242−4252. (acccessed 2014/10/02)

(22) Böckmann, M.; Doltsinis, N. L.; Marx, D. Adaptive switching of interaction potentials in the time domain: An extended lagrangian approach tailored to transmute force field to QM/MM simulations and back. *J. Chem. Theory Comput.* **2015**, *11* (6), 2429−2439.

(23) Jiang, T.; Boereboom, J. M.; Michel, C.; Fleurat-Lessard, P.; Bulo, R. E. Proton transfer in aqueous solution: Exploring the boundaries of adaptive QM/MM. In *Quantum Modeling of Complex Molecular Systems*, Rivail, J.-L.; Ruiz-Lopez, M.; Assfeld, X.. Eds.; Springer International Publishing, 2015; pp 51−91.

(24) Boereboom, J. M.; Potestio, R.; Donadio, D.; Bulo, R. E. Toward Hamiltonian adaptive QM/MM: Accurate solvent structures using many-body potentials. *J. Chem. Theory Comput.* **2016**, *12* (8), 3441−3448.

(25) Zheng, M.; Waller, M. P. Adaptive quantum mechanics/molecular mechanics methods. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2016**, *6* (4), 369−385.

(26) Dohm, S.; Spohr, E.; Korth, M. Developing adaptive QM/MM computer simulations for electrochemistry. *J. Comput. Chem.* **2017**, *38* (1), 51−58.

(27) Field, M. J. An algorithm for adaptive QC/MM simulations. *J. Chem. Theory Comput.* **2017**, *13* (5), 2342−2351.

(28) Zheng, M.; Kuriappan, J. A.; Waller, M. P. Toward more efficient density-based adaptive QM/MM methods. *Int. J. Quantum Chem.* **2017**, *117* (6), No. e25336.

(29) Boereboom, J. M.; Fleurat-Lessard, P.; Bulo, R. E. Explicit solvation matters: Performance of QM/MM solvation models in nucleophilic addition. *J. Chem. Theory Comput.* **2018**, *14*, 1841−1852.

(30) Hofer, T. S.; Hünenberger, P. H. Absolute proton hydration free energy, surface potential of water, and redox potential of the hydrogen electrode from first principles: QM/MM MD free-energy simulations of sodium and potassium hydration. *J. Chem. Phys.* **2018**, *148* (22), No. 00799. 222814/222811−222828

(31) Watanabe, H. Improvement of performance, stability and continuity by modified size-consistent multipartitioning quantum mechanical/molecular mechanical method. *Molecules* **2018**, *23* (8), 1882.
(32) Zheng, M.; Waller, M. P. Yoink: An interaction-based partitioning API. *J. Comput. Chem.* **2018**, *39* (13), 799−806.
(33) Chen, H.; Liao, M.; Wang, H.; Wang, Y.; Zhang, L. Adaptive QM/MM coupling for crystalline defects. *Comput. Methods Appl. Mech. Eng.* **2019**, *354*, 351−368.
(34) Watanabe, H. C.; Cui, Q. Quantitative analysis of QM/MM boundary artifacts and correction in adaptive QM/MM simulations. *J. Chem. Theory Comput.* **2019**, *15* (7), 3917−3928.
(35) Zhang, B.; Altarawy, D.; Barnes, T.; Turney, J. M.; Schaefer, H. F. Janus: An extensible open-source software package for adaptive QM/MM methods. *J. Chem. Theory Comput.* **2019**, *15* (8), 4362−4373.
(36) Yang, Z.-H Extending scaled-interaction adaptive-partitioning QM/MM to covalently bonded systems. *Phys. Chem. Chem. Phys.* **2020**, *22* (32), 17987−17998.
(37) Yang, Z.-H On-the-fly determination of active region centers in adaptive-partitioning QM/MM. *Phys. Chem. Chem. Phys.* **2020**, *22* (34), 19307−19317.
(38) Yang, Z.-H Speed-dependent adaptive partitioning in QM/MM MD simulations of displacement damage in solid-state systems. *Phys. Chem. Chem. Phys.* **2021**, *23* (5), 3417−3428.
(39) Bachmann, J.; Doltsinis, N. L. Adaptive partitioning molecular dynamics using an extended Hamiltonian approach. *J. Chem. Phys.* **2021**, *155* (14), 144104.
(40) Wang, J.-N.; Liu, W.; Li, P.; Mo, Y.; Hu, W.; Zheng, J.; Pan, X.; Shao, Y.; Mei, Y. Accelerated computation of free energy profile at ab initio quantum mechanical/molecular mechanics accuracy via a semiempirical reference potential. 4. Adaptive QM/MM. *J. Chem. Theory Comput.* **2021**, *17* (3), 1318−1325.
(41) Watanabe, H. C.; Yamada, M.; Suzuki, Y. Proton transfer in bulk water using the full adaptive QM/MM method: integration of solute- and solvent-adaptive approaches. *Phys. Chem. Chem. Phys.* **2021**, *23* (14), 8344−8360.
(42) Svensson, M.; Humbel, S.; Froese, R. D. J.; 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 $Pt(P(t\text{-}Bu)_3)_2 + H_2$ oxidative addition. *J. Phys. Chem.* **1996**, *100* (50), 19357−19363.
(43) 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.; et al. The ONIOM method and its applications. *Chem. Rev.* **2015**, *115* (12), 5678−5796.
(44) Rode, B. M.; Hofer, T. S.; Randolf, B. R.; Schwenk, C. F.; Xenides, D.; Vchirawongkwin, V. Ab initio quantum mechanical charge field (QMCF) molecular dynamics: a QM/MM − MD procedure for accurate simulations of ions and complexes. *Theor. Chem. Acc.* **2006**, *115* (2), 77−85.
(45) Hofer, T. S. Probing proton transfer reactions in molecular dynamics—A crucial prerequisite for QM/MM simulations using dissociative models. In *Quantum Modeling of Complex Molecular Systems*, Rivail, J.-L.; Ruiz-Lopez, M.; Assfeld, X.. Eds.; Springer International Publishing, 2015; pp 115−134.
(46) Talachutla, S.; Bhat, S.; Duster, A. W.; Lin, H. Improved indicator algorithms for tracking a hydrated proton as a local structural defect in Grotthuss diffusion in aqueous solutions. *Chem. Phys. Lett.* **2021**, *784*, No. 139121.
(47) Yan, S.; Wang, B.; Lin, H. Tracking the delocalized proton in concerted proton transfer in bulk water. *J. Chem. Theory Comput.* **2023**, *19* (2), 448−459.
(48) Heyden, A.; Truhlar, D. G. Conservative algorithm for an adaptive change of resolution in mixed atomistic/coarse-grained multiscale simulations. *J. Chem. Theory Comput.* **2008**, *4* (2), 217−221.
(49) Delle Site, L.; Praprotnik, M. Molecular systems with open boundaries: Theory and simulation. *Phys. Rep.* **2017**, *693*, 1−56.
(50) Ensing, B.; Nielsen, S. O.; Moore, P. B.; Klein, M. L.; Parrinello, M. Energy conservation in adaptive hybrid atomistic/coarse-grain molecular dynamics. *J. Chem. Theory Comput.* **2007**, *3* (3), 1100−1105. (acccessed 2013/11/29)
(51) Delle Site, L. Some fundamental problems for an energy-conserving adaptive-resolution molecular dynamics scheme. *Phys. Rev. E* **2007**, *76* (4), No. 047701.
(52) Praprotnik, M.; Delle Site, L.; Kremer, K. Adaptive resolution molecular-dynamics simulation: Changing the degrees of freedom on the fly. *J. Chem. Phys.* **2005**, *123* (22), 32286. 224106/224101−224114
(53) Potestio, R.; Fritsch, S.; Español, P.; Delgado-Buscalioni, R.; Kremer, K.; Everaers, R.; Donadio, D. Hamiltonian adaptive resolution simulation for molecular liquids. *Phys. Rev. Lett.* **2013**, *110* (10), No. 108301.
(54) Kreis, K.; Donadio, D.; Kremer, K.; Potestio, R. A unified framework for force-based and energy-based adaptive resolution simulations. *Europhys. Lett.* **2014**, *108* (3), 30007.
(55) Szabo, A.; Ostlund, N. S. *Modern quantum chemistry: Introduction to advanced electronic structure theory*; McGraw-Hill, 1989.
(56) Komornicki, A.; Fitzgerald, G. Molecular gradients and hessians implemented in density functional theory. *J. Chem. Phys.* **1993**, *98* (2), 1398−1421.
(57) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; et al. General atomic and molecular electronic structure system. *J. Comput. Chem.* **1993**, *14* (11), 1347−1363.
(58) Gordon, M. S.; Schmidt, M. W. Chapter 41 - Advances in electronic structure theory: GAMESS a decade later. In *Theory and Applications of Computational Chemistry*, Dykstra, C. E.; Frenking, G.; Kim, K. S.; Scuseria, G. E. Eds.; Elsevier, 2005; pp 1167−1189.
(59) Barca, G. M. J.; Bertoni, C.; Carrington, L.; Datta, D.; Silva, N. D.; Deustua, J. E.; Fedorov, D. G.; Gour, J. R.; Gunina, A. O.; Guidez, E.; et al. Recent developments in the general atomic and molecular electronic structure system. *J. Chem. Phys.* **2020**, *152* (15), 154102.
(60) Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-consistent molecular-orbital methods. IX. An extended Gaussian-type basis for molecular-orbital studies of organic molecules. *J. Chem. Phys.* **1971**, *54* (2), 724−728.
(61) Hariharan, P. C.; Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theor. Chim. Acta* **1973**, *28* (3), 213−222.
(62) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self-consistent molecular orbital methods. XII. Further extensions of Gaussian-type basis sets for use in molecular orbital studies of organic molecules. *J. Chem. Phys.* **1972**, *56* (5), 2257−2261.
(63) Hehre, W. J.; Stewart, R. F.; Pople, J. A. Self-consistent molecular-orbital methods. I. Use of Gaussian expansions of Slater-type atomic orbitals. *J. Chem. Phys.* **1969**, *51* (6), 2657−2664.
(64) Binkley, J. S.; Pople, J. A.; Hehre, W. J. Self-consistent molecular orbital methods. 21. Small split-valence basis sets for first-row elements. *J. Am. Chem. Soc.* **1980**, *102* (3), 939−947.
(65) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *J. Chem. Phys.* **1980**, *72* (1), 650−654.
(66) Frisch, M. J.; Pople, J. A.; Binkley, J. S. Quadratic configuration interaction. A general technique for determining electron correlation energies. *J. Chem. Phys.* **1984**, *80* (7), 3265−3269.
(67) Dunning, T. H., Jr. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90* (2), 1007−1023.
(68) Noro, T.; Sekiya, M.; Koga, T. Correlating basis sets for the H atom and the alkali-metal atoms from Li to Rb. *Theor. Chem. Acc.* **2003**, *109* (2), 85−90.
(69) Noro, T.; Sekiya, M.; Koga, T. Contracted polarization functions for the atoms helium through neon. *Theor. Chem. Acc.* **1997**, *98* (1), 25−32.
(70) Schuchardt, K. L.; Didier, B. T.; Elsethagen, T.; Sun, L.; Gurumoorthi, V.; Chase, J.; Li, J.; Windus, T. L. Basis set exchange: A

https://doi.org/10.1021/acs.jpca.3c05600
J. Phys. Chem. A 2023, 127, 10320−10333

community database for computational sciences. *J. Chem. Inf. Model.* **2007**, *47* (3), 1045−1052.
(71) Pritchard, B. P.; Altarawy, D.; Didier, B.; Gibson, T. D.; Windus, T. L. New basis set exchange: an open, up-to-date resource for the molecular sciences community. *J. Chem. Inf. Model.* **2019**, *59* (11), 4814−4820.
(72) Jakubikova, E.; Rappé, A. K.; Bernstein, E. R. Exploration of basis set issues for calculation of intermolecular interactions. *J. Phys. Chem. A* **2006**, *110* (31), 9529−9541.
(73) Høyvik, I.-M. The spectrum of the atomic orbital overlap matrix and the locality of the virtual electronic density matrix. *Mol. Phys.* **2020**, *118* (21−22), No. e1765034.
(74) Agmon, N. The Grotthuss mechanism. *Chem. Phys. Lett.* **1995**, *244* (5−6), 456−462.
(75) Marx, D. Proton transfer 200 years after von Grotthuss: Insights from ab initio simulations. *ChemPhysChem* **2006**, *7* (9), 1848−1870.
(76) Svensson, M.; Humbel, S.; Morokuma, K. Energetics using the single point IMOMO (integrated molecular orbital plus molecular orbital) calculations: Choices of computational levels and model system. *J. Chem. Phys.* **1996**, *105* (9), 3654−3661.

https://doi.org/10.1021/acs.jpca.3c05600
J. Phys. Chem. A 2023, 127, 10320−10333