# Adaptive, Geometric Networks for Efficient Coarse-Grained *Ab Initio* Molecular Dynamics with Post-Hartree−Fock Accuracy

Timothy C. Ricard, Cody Haycraft, and Srinivasan S. Iyengar*

Department of Chemistry and Department of Physics, Indiana University, 800 East Kirkwood Avenue, Bloomington, Indiana 47405, United States

S Supporting Information

ABSTRACT: We introduce a new coarse-graining technique for *ab initio* molecular dynamics that is based on the adaptive generation of connected geometric networks or graphs specific to a given molecular geometry. The coarse-grained nodes depict a local chemical environment and are *networked* to create edges, triangles, tetrahedrons, and higher order simplexes based on (a) a Delaunay triangulation procedure and (b) a method that is based on molecular, bonded and nonbonded, local interactions. The geometric subentities thus created, that is nodes, edges, triangles, and tetrahedrons, each represent an energetic measure for a specific portion of the molecular system, capturing a specific set of interactions. The energetic measure is constructed in a manner consistent with ONIOM and allows assembling an overall molecular energy that is purely based on the geometric network derived from the molecular conformation. We use this approach to obtain accurate MP2 energies for polypeptide chains containing up to 12 amino-acid monomers (123 atoms) and DFT energies up to 26 amino-acid monomers (263 atoms). The energetic measures are obtained at much reduced computational costs; the approach currently yields MP2 energies at DFT cost and DFT energies at PM6 cost. Thus, in essence the method performs an efficient "coarse-graining" of the molecular system to accurately reproduce the electronic structure properties. The method is comparable in principle to several fragmentation procedures recently introduced in the literature, including previous procedures introduced by two of the authors here, but critically differs by overcoming the computational bottleneck associated with adaptive fragment creation without spatial cutoffs. The method is used to derive a new, efficient, ab initio molecular dynamics formalism (both Born−Oppenheimer and Car−Parrinello-style extended Lagrangian schemes are presented) a critical hallmark of which is that, at each dynamics time-step, multiple electronic structure packages can be simultaneously invoked to assemble the energy and forces for the full system. Indeed, in this paper, as an illustration, we use both Psi4 and Gaussian09 simultaneously at every time-step to perform AIMD simulations and also the energetic benchmarks. The approach works in parallel (currently over 100 processors), and the computational implementation is object oriented in C++. MP2 and DFT based on-the-fly dynamics results are recovered to good accuracy from the coarse-grained AIMD methods introduced here at reduced costs as highlighted above.
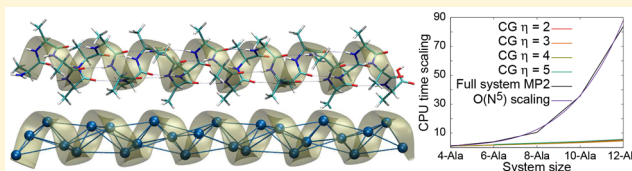
I. INTRODUCTION

Ab initio molecular dynamics (AIMD)[1−7] has had critical impact on the study of fundamental chemical problems[8−19] and in complex processes in biological,[20−26] atmospheric,[27−29] and materials[30−32] chemistry, but the straightforward application of AIMD to challenging problems is limited by the need for accurate electronic structure at every instant in time during a classical trajectory. This makes the routine application of post-Hartree−Fock theory in AIMD for large systems less affordable. There have been several algorithmic improvements[33−40] including the advent of reduced-scaling post-Hartree−Fock methods,[41−45] but DFT-type treatment, despite its many challenges,[46−48] is the only practically reliable option available for modeling large systems. In this regard there have been several new fragment-based electronic structure methods that have become recently available,[49−53] some of which have been applied to AIMD[54−64] studies. In a few of these cases classical trajectories are obtained in agreement with MP2 and CCSD levels of theory.[60−62,64] Both Born−Oppenheimer[60] and

extended-Lagrangian (Car−Parrinello type) options are available for the methods in refs 61 and 62, and in these cases, the molecular fragments are determined efficiently, "on-the-fly," using a bit-manipulation algorithm.[60] These developments may in the future push the rigorous description of complex chemical, biological, and materials processes.

Influenced by developments in molecular fragmentation[49,50,53−55,60−62,65−81] and facilitated by work on many-body theory applied toward computing potential surfaces for quantum nuclear dynamics,[82,83] here we introduce a geometric network theory to efficiently compute the instantaneous electronic structure for AIMD simulations in large systems. One goal that is achieved here is to make MP2-based AIMD studies efficient for medium-to-large sized polypeptide systems. Toward this we "coarse-grain" large systems, and the associated coarse-grained nodes (or *CG-nodes*) are connected to form a
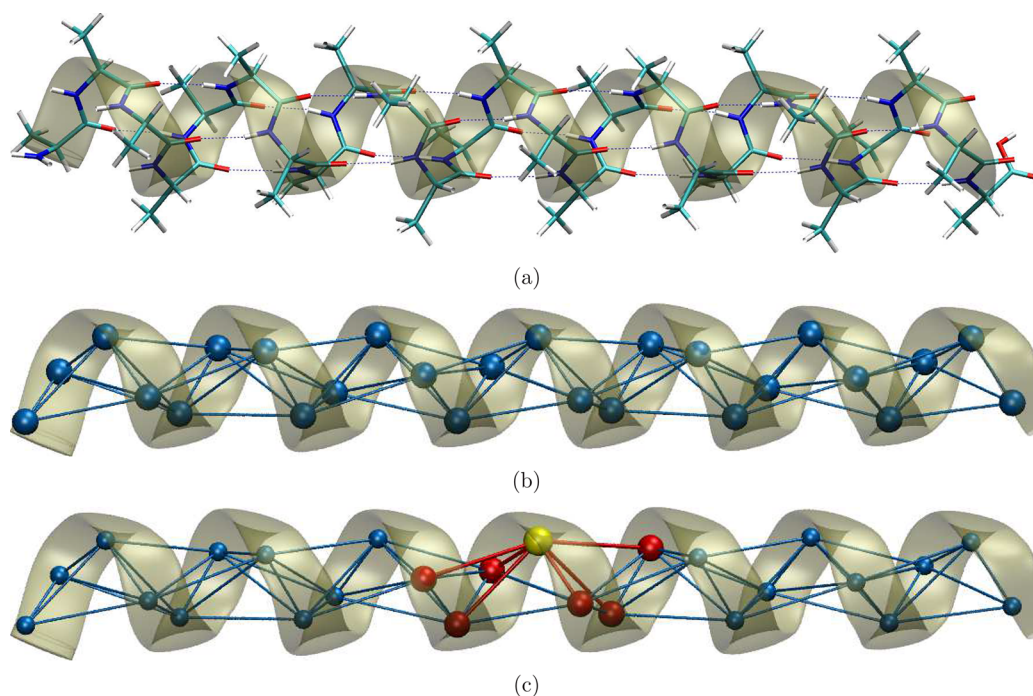
Figure 1. Part (b) shows the filtered Delaunay triangulation of a helical polypeptide chain shown in part (a). The nodes represent individual coarse-grained sites. Edges connecting these nodes represent critical interactions that include both nearest neighbor interactions and critical nonbonded interactions. The interactions captured by such geometric decompositions are to be computed using electronic structure theory. Part (c) selectively displays the connectivity neighborhood (red) of one node (yellow) for the special Delaunay triangulation shown in part (b).
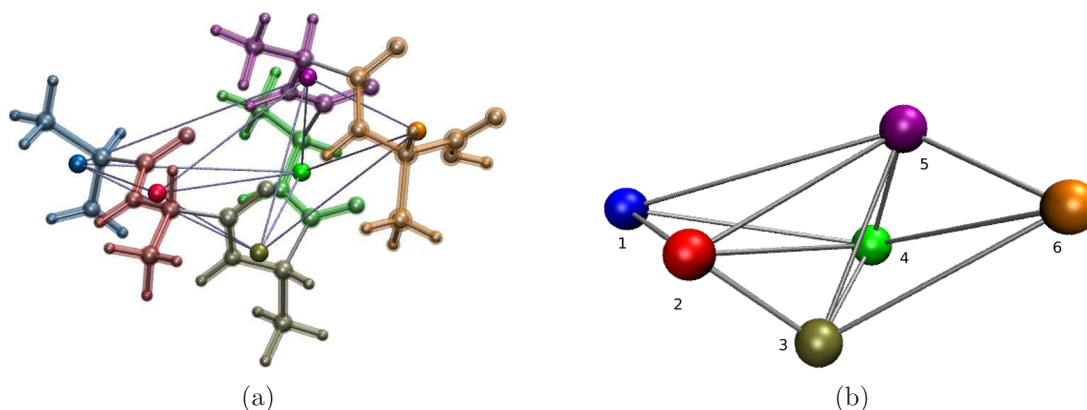

Figure 2. Part (b) highlights features of the Delaunay triangulation scheme for the molecular representation in part (a).

geometric network or a connected graph. The nature of the connectivity in this network is designed to capture all critical bonded and nonbonded interactions in the system. To achieve this in a general fashion for arbitrary systems, here, the level of connectivity in the coarse-grained network is defined based on (a) a filtered (local-envelope-based) Delaunay triangulation[84−87] and (b) a chemical connectivity-based procedure. This allows us to include interactions beyond those from near neighbors and may be adaptively determined during dynamics. Edges connecting nodes (*CG-edges*) are obtained from the aforementioned geometric network and used to provide a decomposition of the full system energy. We gauge the extent to which such an energy may efficiently capture the full system MP2 or DFT energy and allow efficient AIMD treatment.

The paper is organized as follows: In Section II we discuss the coarse-grained network-based energy decomposition procedure. In Section III we provide benchmarks for accuracy of MP2 conformational energies; multiple polyalanine, Ala$_n$,

conformers ranging in size from $n = 4 \cdots 26$ are studied. We find that the network decomposition provides an accurate description of conformational stabilization and the computational cost is reduced to that of DFT when MP2 accuracy is necessary and to that of PM6 when DFT accuracy is desired. Thus, in essence the method performs an efficient "coarse-graining" of the molecular system to accurately reproduce the electronic structure properties. In Section IV we present AIMD trajectories that conserve the total energy to inside the sub-kcal/mol range. The approximate trajectories are compared to the full system DFT and MP2 calculations and found to be in good agreement. A variety of measures are compared. Most notably we gauge the Fourier transform of the velocity autocorrelation function to probe the differences in spectral signatures obtained from the trajectories. Conclusions are given in Section V. The Supporting Information (SI) is provided with additional AIMD data and additional computational benchmarks for accuracy.

II. COARSE-GRAINED *AB INITIO* MOLECULAR DYNAMICS USING GEOMETRIC NETWORKS TO ALLOW DFT AND POST-HARTREE FOCK ACCURACY FOR LARGE SYSTEMS

II.A. Nodal Architectures for Molecular Systems through Simplex Decomposition. We begin by introducing a coarse-grained (CG) version of molecular systems as represented in Figure 1. Here, for illustration, a polypeptide chain is presented, where each amino acid is represented as a coarse-grained unit or a "*CG-node*". These nodes are system dependent and may capture a chemically significant portion of the system without eliminating critical interactions, while maintaining a suitable size for the desired optimization of computation. The *CG-nodes* are connected to form a network, or connected graph, through an edge generation scheme. Two such schemes are employed here: (a) a filtered Delaunay triangulation[84,85] procedure and (b) a connectivity-displacement-based interaction procedure, both of which are discussed below.

II.A.1. Delaunay Triangulation of Molecular Systems. In the Delaunay triangulation procedure, a three-dimensional network, or a connected graph, is created with *CG-nodes* as vertices; *CG-edges* connect *CG-nodes* when these obey the so-called Delaunay triangulation condition.[84,85] In three dimensions this condition requires that no nodes may appear within the circumscribing sphere for each tetrahedral arrangement of nodes. For this reason, in Figure 2, there is no edge that connects nodes 2 and 6. Such an edge would create the tetrahedron with vertices, 2-3-5-6, which would contain node 4 in the interior of the circumscribing sphere and violate the Delaunay condition. Delaunay triangulation allows an orthogonal real-space decomposition of the molecular framework into tetrahedral (or 3-simplex) domains known as convex hulls. [Simplexes are *n*-dimensional generalizations of triangles, with the tetrahedron being a three-dimensional example.] For example, Figure 2 contains four convex hulls with vertices, 1-2-3-4, 1-2-4-5, 2-3-4-5, and 3-4-5-6 which do not overlap. There are a variety of commercial and open-source software packages which can provide Delaunay triangulations for objects in *n*-dimensions. Here, we make use of the Computational Geometry Algorithms Library (CGAL),[88] which allows incremental construction of the triangulation and scales as $O(\mathrm{N}\log(N))$, N being the number of vertices in the geometric network. A similar method has been employed for potential surface sampling,[89] quantum-dynamical simulations,[90] and in computational geophysics applications.[91] It may also be noted that objects obtained from Delaunay triangulation are *dual-space* analogues of Voronoi diagrams.[84,85]

Once the decomposition is carried out, a *CG-edge* in this wire-frame geometric network is interpreted as a set-theoretic union of the molecular fragments depicted as part of the two *CG-nodes* that it connects. Our choice of a Delaunay triangulation to define nodal connectivity ensures that critical local interactions are captured in the network topology. Furthermore, interactions beyond a chosen spatial cutoff are ignored, and thus we include a local envelope of simplexes within our Delaunay triangulation, where the size of the local spatial envelope may depend on the strength of the local interactions present in the system. For example, in Figure 1, the local envelope that encapsulates each *CG-edge* is large enough to include all connected amino-acid monomers and four additional pairs of amino-acid monomers that interact through

space to capture nonbonded interactions within a specific distance cutoff. For the 3_{10}-helices that we discuss later in this publication, these nonlocal interactions capture hydrogen-bond interactions that are a critical part of protein secondary structure. An additional edge is required to be included to capture the same for the α-helix conformation.

II.A.2. Nodal Architectures Based on Molecular Connectivity. The connectivity-displacement-based network-generation procedure considers bonding and nonbonding (weak) interactions contained in the system. In this paper, we utilize the covalent bonding network to determine this framework, but for other systems hydrogen bonding network connectivities may also be included within the definition of connectivity for the determination of the connected graph. This scheme begins with *CG-edge* formation between adjacent,"neighboring", nodes based on molecular connectivity. Following this, for *CG-edges* that share a *CG-node*, the peripheral nodes may be connected to form a new *CG-edge* that captures first degree, nonbonded interactions. Consequently the connectivity-displacement parameter, η, is the degree of separation allowed for "interaction" through *CG-edges*, where $\eta = 2$ implies edge formation between neighboring *CG-nodes* as described above. Furthermore, $\eta = 3$ creates additional *CG-edges* between *CG-nodes* connected through a chain of three *CG-nodes*. In all cases, as in the Delaunay scheme, the *CG-edges* are a set-theoretic union of the elements in both *CG-nodes*. This connectivity-displacement interaction scheme generates a *static* network invariant to evolution of the system as compared to the Delaunay scheme, which is inherently *dynamical and adaptive*. Furthermore, the convex hulls contained within the simplexes defined using the connectivity-based measure are *not* spatially orthogonal, in contrast with those from Delaunay triangulation. The impact of this feature may be gauged from the flexibility to include the 1−3 edge interaction in Figure 2 here, which cannot be included within the Delaunay procedure as a result of the Delaunay condition noted above.

II.B. Energy Estimates from Systematic and Recursive Network Decomposition of Molecular Systems. Once a molecular system is "coarse-grained" to obtain a geometric network as discussed in the previous subsection, we use it to construct an energetic measure for the whole system. Specifically, here, we define an energetic measure for the whole system as a linear combination of energies obtained from electronic structure calculations performed using molecular fragments belonging to *CG-nodes* and *CG-edges* as

$$E_{\mathrm{network}} = E^{\mathrm{level},0} + \sum_{\alpha} \Delta E_{\alpha}^{\mathrm{cor.},1} - \sum_{I} \Delta E_{I}^{\mathrm{cor.},1} \left[\sum_{l=2}^{p_I} (-1)^l m_l \right] \quad (1)$$

Here, $E^{\mathrm{level},0}$ represents a reference electronic energy for the whole system, which could be obtained from a low level electronic structure calculation on the full system. The quantity "α" in eq 1 represents a *CG-edge* and I represents a *CG-node*. The quantity $\Delta E_{\alpha}^{\mathrm{cor.},1}$ is a correction to $E^{\mathrm{level},0}$ arising from the subsystem that is depicted by the edge α and is obtained as

$$\Delta E_{\alpha}^{\mathrm{cor.},1} = E_{\alpha}^{\mathrm{level},1} - E_{\alpha}^{\mathrm{level},0} \quad (2)$$

in a fashion consistent with the well-known ONIOM[60,92] method. Here, the terms "level, 0" and "level, 1" represent increasing levels of electronic structure theory. The energy correction associated with *CG-node*, *I*, is also written in a similar fashion as in eq 2, and hence

$$E_{\text{network}} = E^{\text{level},0} + \sum_\alpha (E_\alpha^{\text{level},1} - E_\alpha^{\text{level},0})$$
$$- \sum_I (E_I^{\text{level},1} - E_I^{\text{level},0}) \left[\sum_{l=2}^{p_I} (-1)^l m_l \right] \quad (3)$$

The square-bracketed term in eqs 1 and 3 incorporates an overcounting correction for CG-node, I, and p_I represents the number of CG-edges that contain the CG-node, I. For example, in Figure 1(c), the CG-node highlighted in yellow belongs to six edges (two that are bonded and four that interact through space). As a result, the sum over the index l in eq 1 essentially represents a loop over the intersections of CG-edges, and, for example, $l = 2$ represents the two-edge intersections on node I, $l = 3$ represents the three-edge intersections on node I, and so on. The phase $(-1)^l$ is selected such that $(l - 1)$-edge contributions are removed from (l)-edge contributions to account for overcounting of CG-nodal contributions. The quantity m_l represents the multiplicity of l in p_I, that is

$$m_l \equiv \binom{p_I}{l} = \frac{p_I!}{(p_I - l)! \cdot l!} \quad (4)$$

which is the number of times that l-edges appear inside the local neighborhood network that defines the neighbors of CG-node, I. However, also notice that the bracketed term, "[···]" in eq 3, is a partial sum of the binomial expansion for $(x-1)^{p_I}$ for $x = 1$, with terms $l = 0$ and $l = 1$ omitted. Hence,

$$(x - 1)^{p_I}|_{x=1} = \sum_{n=0}^{p_I} \binom{p_I}{n} x^{p_I - n} (-1)^n \Big|_{x=1} = \sum_{n=0}^{p_I} \binom{p_I}{n} (-1)^n = 0 \quad (5)$$

As a result,

$$\sum_{n=2}^{p_I} (-1)^n m_l = \binom{p_I}{1} - \binom{p_I}{0} = p_I - 1 \quad (6)$$

and eq 1 may be simplified to

$$E_{\text{network}} = E^{\text{level},0} + \sum_\alpha \Delta E_\alpha^{\text{cor.},1} - \sum_I \Delta E_I^{\text{cor.},1}(p_I - 1) \quad (7)$$

This simplification is a key step in the formalism presented here and greatly reduces the computational cost in determining contributions from CG-edge overlaps.

The procedure developed above can be easily generalized to include all embedded CG-simplexes within the constructed wireframe geometric network. Although, we do not benchmark this generalization here, we present the general expression in Appendix A. The expression in Appendix A, when combined with either Delaunay triangulation or connectivity-displacement parameter η, may be used to derive a systematic approach for fragmentation and molecular coarse-graining to achieve energy and forces in agreement with high level electronic structure theory in a computationally efficient manner. For example, there are essentially two dimensions where the CG-technique introduced here can be systematically improved.

(a) For any given truncation of the generalized version of eq A1 (for example, truncated in this paper to CG-edges as given in eq 7), the parameter η discussed in Section II.A is an expansion coefficient that allows us to "tune-in" many-body interactions between coarse-grained units of the specific size depicted by the truncation order. In

Section III, we have exhaustively probed all possible fragmentation protocols as allowed within the truncated description presented in eq 7.

(b) The other dimension where the CG-description can be systematically improved is by introducing energy contributions from faces and all embedded simplexes as stated above, but this generalization is not numerically benchmarked in this paper. (See Appendix A.)

II.C. Computational Efficiency of the Nodal Decomposition Scheme. Fragmentation methods[51,60,71,73−76,93] often consider overlapping subsystems to introduce many-body interactions, while maintaining smaller fragment sizes to optimize computational complexity. The associated total energy of the system is obtained by assembling the energies of all fragments with suitable removal of contributions from common subfragments as allowed by the inclusion−exclusion principle.[94] There are several fragment-based methods that benefit from such an approach.[50,51,53,60−62,71−74]

The connections between fragment-based electronic structure methods and the approach developed here can be understood through inspection of Figure 3. For example, in
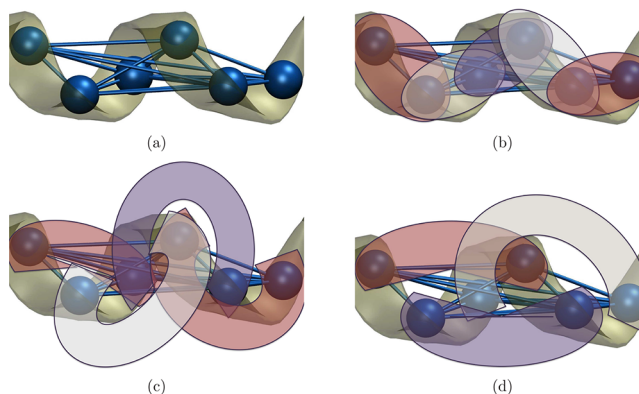

(a)   (b)

(c)   (d)

Figure 3. Connections between the CG-approach and molecular fragmentation are illustrated here for a 3_{10} polypeptide helical fragment. The CG-network in part (a) comprises (i) chemically bonded (nearest neighbor) fragments (illustrated using Venn diagrams in part (b)), that is connectivity-displacement parameter, $\eta = 2$, (ii) next nearest neighbors, that is, parts (b) and (c) together represent $\eta = 3$, and finally (iii) parts (b), (c), and (d) together represent all interactions obtained in part (a) from connectivity-displacement parameter, $\eta = 4$. Importantly, the CG-network in part (a) includes all interactions depicted in parts (b), (c), and (d). Computing the overlapping fragments and their respective weights from all such interactions are appreciably reduced by the CG-algorithm introduced in this paper.

Figure 3(b), chemically bonded neighbor interactions are depicted by portraying CG-edges and corresponding fragments that would be used in the above listed fragmentation protocols. In Figures 3(c) and 3(d) longer-range interactions are depicted using both the corresponding CG-edges and molecular fragments depicted as Venn diagrams. While computing the overlapping regions between these sets may become challenging, all possible many-body interactions are included in such a decomposition which may be critical for modeling complex systems. The computational challenge appears in two parts: (i) the "derivative" fragments formed from overlaps would need to be identified from all of the predefined "primitive" fragments, and (ii) the overcounting coefficient for each of these fragments would be needed. The first problem needs determination of all

possible overlaps for every fragment with all other fragments, and this procedure would scale as $O(N^2)$ in the worst case and $O(N)$ when some local spatial cutoff condition is enforced. Next, the procedure needs to be iterated to compose all overlapping fragments between any three primitive fragments; the number of such regions would potentially grow up to $O(N^3)$ without local truncation. In general, high order overlaps will need to be considered, and for an adaptive scheme, essential for dynamics, the algorithm must cycle until reaching a point where all overlapping regions have been exhausted, requiring a potentially exponential scaling number of overlap calculations

$$\sum_{n=1} \binom{N_{\text{frag}}}{n} \tag{8}$$

where N_{frag} is the total number of fragments created from the algorithm, and n is the order of overlaps, that is $n = 2$ checks for number overlaps between pairs of fragments and so on. The complexity depicted in eq 8 grows exponentially but tails off as n increases for fixed N_{frag}. Once all such overlapping derivative fragments are found, the appropriate coefficients (that correct for overcounting) would be needed for the overall energy expression obtained from the principle of inclusion-exclusion.[94] The summation of all of these coefficients grants the weights of the derivative fragments, and as before the calculation of these coefficients through direct determination of all possible combinations would require an exponential number of operations.

The approach proposed here, arising from eq 7, circumvents the exponential scaling bottleneck described above. Specifically, the scaling of the algorithm depends on the number of edges obtained from Delaunay triangulation. Since the number of nodes are related to the number of edges through η, the CG-algorithm would require roughly

$$O\left(\frac{N_{\text{frag}}}{2\eta} \log\left(\frac{N_{\text{frag}}}{2\eta}\right)\right) \tag{9}$$

evaluations. We further note that even for moderately sized systems our algorithm shows orders of magnitude speed up over the exponential scaling number of comparisons required from eq 8. In practice we see great reduction in computational costs. For example, a 16-amino-acid fragment with $\eta = 4$ yields 42 edges in the CG-network and hence 42 primary fragments in a fragmentation protocol. These 42 primary fragments will require $\sum_{n=1}^{2*(\eta-1)} \binom{42}{n}$ checks to discover all possible fragments obtained from intersections involving the primary fragments. This is effectively done in the current CG-algorithm in an extremely efficient fashion and requires only 9 ms of CPU time (on an 8-core Intel i7, at 3.4 GHz and 8GB RAM). Thus, we observe significant cost savings, allowing more two-body interactions to be efficiently included. (As an illustration note that the two-body interactions captured within an $\eta = 4$ calculation involve all interactions represented in Figures 3(b)–3(d).) The consideration of higher many-body interaction would follow similar scaling to eq 9. For instance three-body interactions (*CG-triangles*) would be linearly dependent on number of monomers (*CG-nodes*) but quadratically dependent on the distance cutoff (η) leading to an expression similar to eq 9.

II.D. Further Computational Enhancements through Extended Lagrangian Treatment of $E^{\text{level},0}$.

The theoretical framework is benchmarked in Section III for accuracy in computing isomer stabilization energies for polypeptide chains and is demonstrated, in Section IV, through *ab initio* molecular dynamics simulations. When DFT accuracy is desired, $E^{\text{level},0}$ is computed using PM6, and when MP2 accuracy is desired, $E^{\text{level},0}$ is obtained from DFT. Thus, for MP2 accuracy, the costs toward computing electronic energies for N number of *CG-nodes* would scale as $O(N*(M^{3.5} + (M)^5))$ where M is related to the number of electronic basis functions (or size) within each *CG-node*. Similarly, the costs for the corresponding *CG-edge* calculations would be $O((\eta*N)*((2M)^{3.5} + (2M)^5))$. The full low level calculation would scale as $O((NM)^{3.5})$. Therefore, in the large system limit the calculation of $E^{\text{level},0}$ would quickly dominate the total CPU time, reducing the overall scaling to order of $N^{3.5}$. These formal results are further substantiated through computational scaling curves provided in Section III.

Thus, in the large system limit, the efficiency of the algorithm is constrained by the calculation of $E^{\text{level},0}$. Since the greatest impact of the $E^{\text{level},0}$-expense is on AIMD simulations, here, in addition to studying the full Born–Oppenheimer molecular dynamics implementation of eq 7, this energy functional is also embedded within an extended Lagrangian[95,96] framework, and the electronic parameters that determine $E^{\text{level},0}$ are propagated with the nuclear degrees of freedom through a simple adjustment of the relative electron–nuclear time-scales. This is based on the previously demonstrated Atom-centered Density Matrix Propagation with post-Hartree–Fock accuracy (ADMP-pHF)[61] and is reminiscent of the Car–Parrinello method[5,97] but provides post-Hartree–Fock electronic structure accuracy, "on-the-fly". Thus, we introduce the extended Lagrangian

$$\mathcal{L} = \frac{1}{2}\text{Tr}[\mathbf{V^T M V}] + \frac{1}{2}\text{Tr}\left[(\boldsymbol{\mu}_{\text{level},0}^{1/4} \mathbf{W}_{\text{level},0} \boldsymbol{\mu}_{\text{level},0}^{1/4})^2\right]$$
$$- E_{\text{network}}(\mathbf{R}, \mathbf{P}_{\text{level},0}) - \text{Tr}[\boldsymbol{\Lambda}_{\text{level},0}(\mathbf{P}_{\text{level},0}^2 - \mathbf{P}_{\text{level},0})] \tag{10}$$

Here $\mathbf{M}$, $\mathbf{R}$, and $\mathbf{V}$ are the nuclear masses, positions, and velocities. The single particle density matrix $\mathbf{P}_{\text{level},0}$ determines $E^{\text{level},0}$ and has velocity ($\mathbf{W}_{\text{level},0}$) with fictitious inertia tensor $\boldsymbol{\mu}_{\text{level},0}$. Thus, the dynamical parameters depicted in eq 10 are $\{\mathbf{R}, \mathbf{V}; \mathbf{P}_{\text{level},0}, \mathbf{W}_{\text{level},0}\}$, which each evolve in time through velocity Verlet[98] integration. The values of the fictitious inertia tensor are chosen as discussed in refs 61 and 62. Specifically, the single particle orbitals represented within $\mathbf{P}_{\text{level},0}$ are mass-weighted as per the respective Fock matrix elements as outlined in ref 62 (See eqs 5 and 6 in ref 62.) in such a way that the core-orbitals are weighted with a greater inertia as compared to the valence. This allows an adjustment of time-scales for the electronic degrees of freedom represented by $\mathbf{P}_{\text{level},0}$ leading to simultaneous propagation with nuclear degrees of freedom through eq 10. This process is very much akin to the well-known Car–Parrinello scheme but differs here by allowing "on-the-fly" post-Hartree–Fock accuracy as well as larger time-steps as compared to those traditionally used in Car–Parrinello dynamics, as a result of $\boldsymbol{\mu}_{\text{level},0}$. However, as noted in ref 62, the frequency of nuclear motion is perturbed quadratically (red-shifted), by the choice of $\boldsymbol{\mu}_{\text{level},0}$, and hence a scaling factor is required for the frequencies. The scaling factor is system independent, and we use the same scaling factor as in ref 62 for

all our studies. (Precise values are mentioned in the results section.)

To preserve the idempotency[7,99,100] of $P_{level,0}$, a Lagrange multiplier matrix, $\Lambda_{level,0}$, is used. The Lagrangian is integrated using velocity Verlet,[98] while the Lagrangian multiplier matrices maintain the idempotency of $P_{low,0}$ through an iterative process at each step, as described in refs 61, 62, and 100. It is critical to note that only the low level electronic structure, $P_{level,0}$, is propagated in time, and the remaining fragment energies are obtained through convergence as done in ADMP-pHF in ref 61. The nuclear and density matrix gradients have similar forms as those discussed in ref 61.

III. CONFORMER STABILIZATION STUDIES ON POLYALANINE SYSTEMS WITH POST-HARTREE-FOCK ACCURACY

Here we probe the accuracy and efficiency in computing electronic energies and gradients in agreement with post-Hartree–Fock theories using the CG-methods described above. Toward this, we consider polyalanine chains, Ala_n, of lengths in the range of 4 through 26. In Section III.A, we first discuss our coarse-graining strategies for polypeptide chains following which isomer stabilization energies are benchmarked in Section III.B. The choice of polyalanine chains is based on their significant helical propensities[101−103] due to hydrogen bonding.[102,104−108]

In Section III.B, we consider the conformational stability of two sets of linear and helical polyalanine structures. Examples of these structures for Ala_{10} are found in Figure 4. The first set of structures was obtained through geometry optimization at
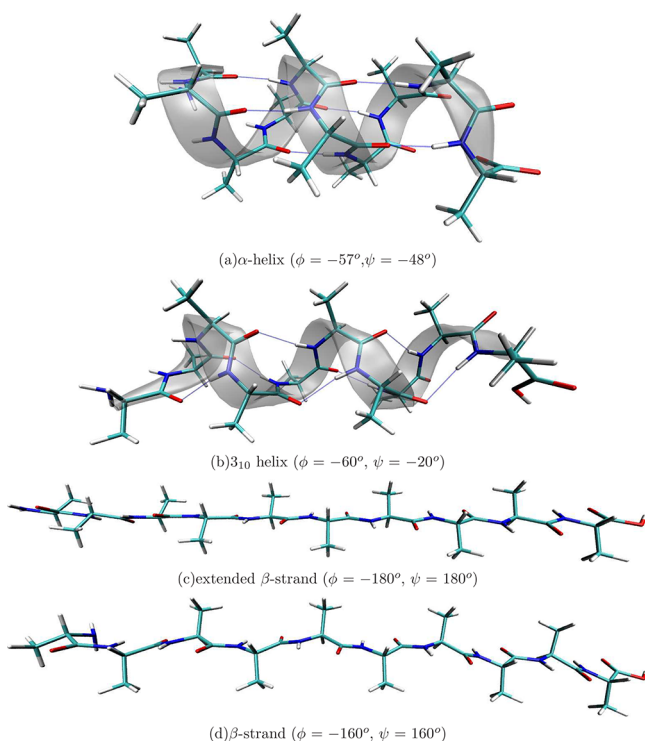

(a)α-helix ($\phi = -57^o, \psi = -48^o$)

(b)3_{10} helix ($\phi = -60^o, \psi = -20^o$)

(c)extended β-strand ($\phi = -180^o, \psi = 180^o$)

(d)β-strand ($\phi = -160^o, \psi = 160^o$)

Figure 4. Four different structural motifs of Ala_{10} are shown here along with the associated backbone dihedral angles α-helix (a) and extended β-strand (c) are nonoptimized structures, while 3_{10} helix (b) and β-strand (d) were optimized from the respective structures at B3LYP/6-31++G(d,p). Additional structural characteristics are discussed in Section III.B.

the B3LYP/6-31++G(d,p) level of electronic structure theory to obtain the β-strand and 3_{10} helical conformers in gas phase. Helical starting structures (Figure 4(a)) converged to the expected[106,107] 3_{10} form (Figure 4(b)) that is characterized by ($i \rightarrow i + 3$) hydrogen bonding with peptide-backbone dihedral angles[109] at approximately $\phi = -60^\circ$ and $\psi = -20^\circ$. The linear conformation (Figure 4(c)) converged to the β-strand secondary structure (Figure 4(d)), with dihedral angles at approximately $\phi = -160^\circ$ and $\psi = 160^\circ$. We have also used α-helix (Figure 3(a)), with ($i \rightarrow i + 4$) hydrogen bonding and dihedral angles at about $\phi = -57^\circ$ and $\psi = -48^\circ$, and the completely extended straight chain (extended β-strand) (Figure 3(c)) with dihedral angles at $\phi = -180^\circ$ and $\psi = 180^\circ$ in our benchmarks here since these have a critical role in the dynamics of the short peptides[104,107] and as parts of larger biological systems.[110,111] Furthermore, these nonequilibrium structures may also be sampled during the dynamics studies such as those constructed later in this publication. For all cases the fragmentation protocols are as described in Section III.A.

III.A. Effective Coarse-Graining Strategies for Polypeptide Chains. Since the peptide bond, NH−CO, is considered to have a partial double bond character,[109] our CG-nodes are defined by breaking the backbone bond between the α-carbon and carbonyl carbon forming CHR-NH-CO CG-node fragments; but, when chemical bonds are broken to create these coarse-grained units, dangling valencies are created. Hence, in a fashion consistent with the ONIOM methodology,[60,61] the dangling valencies are saturated by the use of link atoms; hydrogen atoms were used as link atoms for this work. Furthermore, as discussed in refs 24, 112, and 113, the forces computed on the link atoms during the CG-node and CG-edge electronic structure calculations are correctly transformed back to obtain corrections to forces on the real system atoms through appropriate Jacobians.

To now determine the CG-edges in an automated and adaptive fashion, we utilize the two schemes discussed in Section II.A, and the associated edges are shown in Figure 5 for a Ala_6 conformer. Our first scheme to find these CG-edges considers the bond connectivity information between nodes. That is the CG-nodal unit displacement along the bonding network of the system is used here rather than any Cartesian criteria. For polypeptide systems, sequential displacement between the C_α along the peptide backbone was used for this system. The linear bond topology of the polypeptides leads the quantity p_I to be upper bounded by $2*\eta$, when η is chosen as the connectivity cutoff to be included in forming the network (and is defined in Section II.B). Figure 5 shows a sample set of edges from 6-alanine with increasing η. As the value of η increases, the expectation is that the full system energy would be better represented with increasing longer range interactions being considered. If η is taken as the full residue length of the system, then it would consider all possible dimers (CG-edges) within the system. While increasing η would require more calculations for additional edges, this is offset here through parallelism on a distributed computing environment.

Delaunay triangulation, with the application of a filter based on Cartesian distance, was the second method explored to determine CG-edges. This method uses a simplex decomposition of a field of vertices that are chosen as the centers of mass of each CG-node. The result is the nonoverlapping partitioning, or tessellation, of the three-dimensional Cartesian space enclosed by our molecule, into space-filling polyhedra. The edges formed in this fashion are our CG-edges after passing
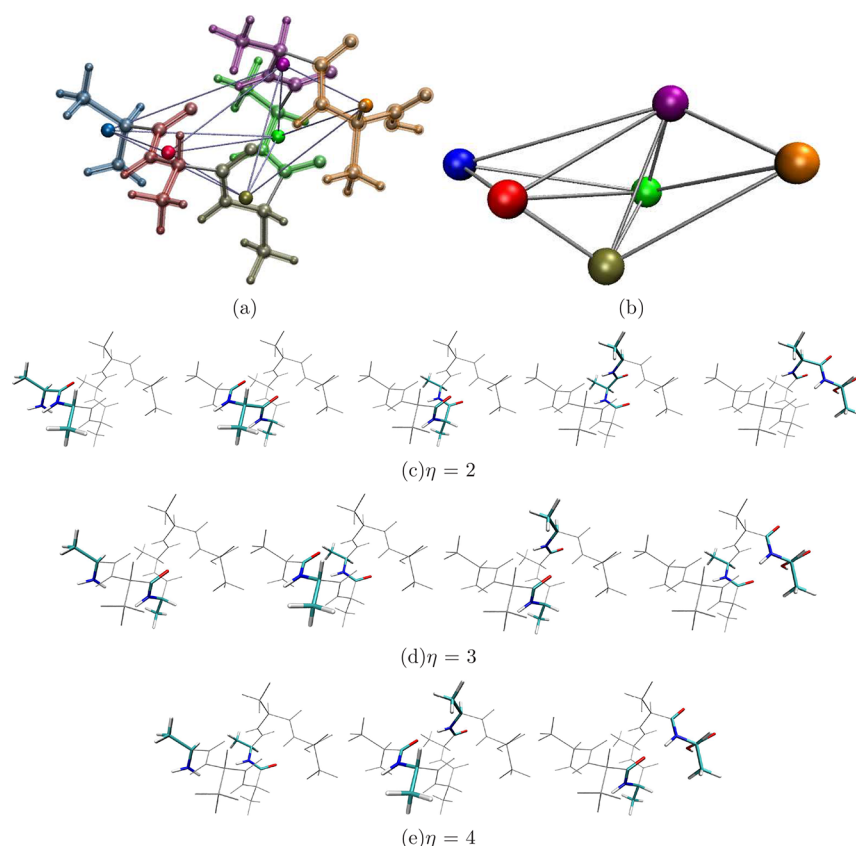
Figure 5. α-Helical Ala$_6$ structure (a) is subject to Delaunay triangulation of the *CG-node* center of masses, and associated simplex decomposition is shown in part (b). Edge generation through the connectivity scheme is also presented for increasing values of η (parts (c)−(e)). These figures also show an important difference between the connectivity-based and Delaunay triangulation schemes. The edge generation between amino acids 1 (blue) and 3 (tan) is shown on the left figure in part (d). However, such an edge is absent in the Delaunay interpretation in part (b), since the associated tetrahedron would have another node inside the circumcircle thus violating the Delaunay condition as discussed earlier in Section II.A. Thus, the Delaunay scheme may not include certain nonbonded interactions which would be part of the connectivity-displacement-based scheme.
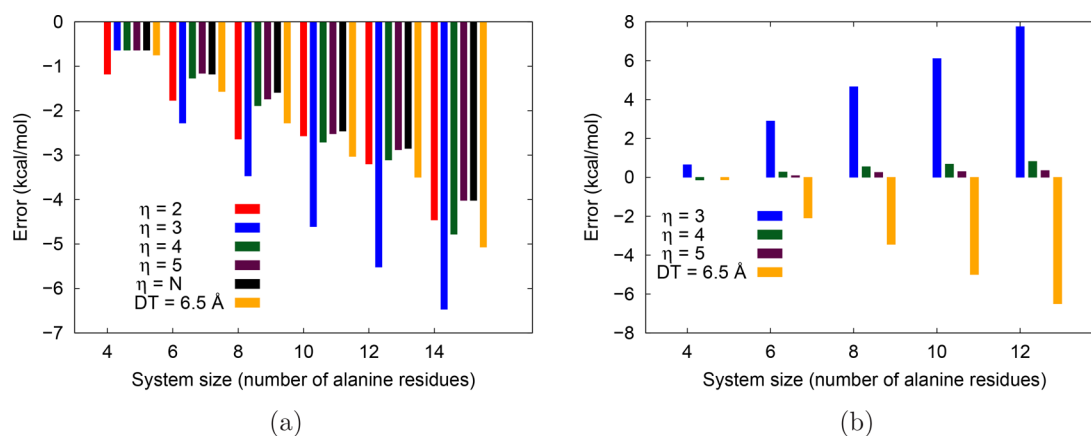

Figure 6. Errors in the conformational energy for the gas phase optimized structures, with increasing η. This figure also includes a Delaunay triangulation of the center of mass corresponding to the *CG-nodes* with a 6.5 Å filter. B3LYP:PM6 calculations (a) used 6-31++G(d,p) basis and MP2:B3LYP calculations (b) used 6-31+G(d,p) basis for both electronic structure methods. Note the results of $\eta = 2$ in (b) were excluded as the errors were too large and obscured the rest of the data. Due to the extensive nature of the error, the corresponding per residue errors are shown in SI, Figure SI-1, and the absolute energy errors are shown Figures SI-3 and SI-4. The per-residue errors are generally less than 1 kcal/mol for all systems. The legend key "η=N" in part (a) represents the case where *all amino-acid dimers* were connected with an edge and used in the calculation. Indeed, the approach in Section II.B allows efficient computation of many-body effects to all orders as described in Section II.C and in Figure 7.

through a Cartesian distance filter to remove edges that represent insignificant interactions. For our system we set the filter at 6.5 Å, which would represent the distance between hydrogen-bonded amino-acid pairs. This scheme has the advantage of capturing interactions based on locality rather than based on bond connectivity, providing an adaptive definition of the network, and adding and removing edges based on the relevance of interactions. Thus, the filtered
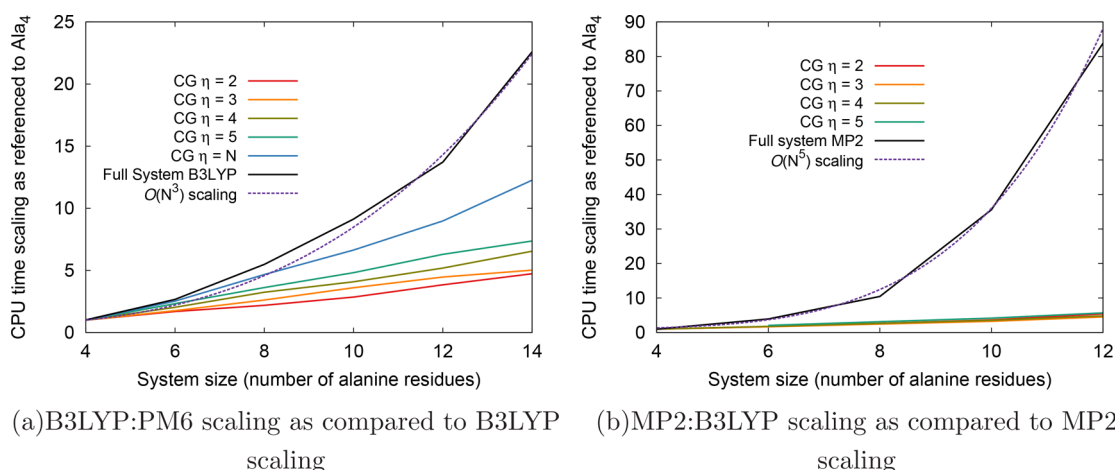
Figure 7. Computational efficiency of CG-AIMD is portrayed here. In essence MP2 accuracy is achieved at much-reduced (B3LYP) cost as seen in part (b). Furthermore, the algorithm discussed in Sections II.B and II.C allows efficient treatment of *all* many-body interactions at the chosen geometric *CG-edge* truncation noted in eq 7 (also see Appendix A, eq A1). As seen in part (a) though, the choice of η particularly affects the scaling of DFT accuracy, since the number of required fragment-DFT calculations increases quadratically with increasing η for the implementation in eq 7. For the range of system sizes considered in part (b), a similar quadratic scaling has little effect in comparison with the inherent $O(N^5)$ scaling of MP2, which dominates the figure.

Delaunay triangulation scheme produces a dynamically defined network, unlike our connectivity-based scheme, and would evolve within an AIMD calculation to capture interactions between components of protein secondary structure better. Furthermore, for larger scale production calculations, it is foreseeable that one will need to combine both connectivity-based and Delaunay-based simplex decomposition techniques, given that critical nonbonded interactions may fluctuate during dynamics.

III.B. Isomer Stabilization Energies. All four conformers depicted in Figure 4 were considered at a variety of residue lengths. At both B3LYP/6-31++G(d,p) and MP2/6-31+G(d,p) levels, it was found that the 3_{10} helical conformation was most stable for all Ala$_n$ conformations considered here. The conformational stability of the 3_{10} helix was found to be greater with MP2. [Here conformational stability is defined as the difference in absolute energy between the helical structure and the linear conformer.] These stabilization energies act as benchmarks for our coarse-graining methods. [Absolute energies are also benchmarked, and these results are described in the SI.] While the optimized 3_{10}-helices and β-strands were studied using polypeptide lengths in the range 4 through 14, the nonoptimized α-helix and extended extended β-strand were considered using polypeptide lengths in the range 4 through 26.

III.B.1. Error in Stabilization of Minimum Energy Structures: 3_{10}-Helices and β-Strands. The chosen systems are coarse-grained as discussed in the previous subsection. Both filtered Delaunay triangulation and connectivity-displacement-based ideas are explored. Figure 6 shows the accuracy of our scheme in obtaining conformer stabilization energies for the optimized structures with increasing residue length and increasing η. The error in conformer stabilization energy is here defined as

$$\Delta E_{\text{conform.}} = \left(E_{\beta\text{-strand,high}} - E_{3_{10}\text{-helix,high}} \right)$$
$$- \left(E_{\beta\text{-strand,CG}} - E_{3_{10}\text{-helix,CG}} \right) \quad (11)$$

where the conformer stabilization is the difference in absolute energies of the 3_{10} helix and the β-strand. The error in conformer stabilization energy is the difference between the coarse-grained technique as compared to that from conformer stabilization computed using the higher level of theory used in the corresponding coarse grain calculation. As noted in Figure 6(b) MP2:B3LYP extrapolation shows monotonic corrections to the stabilization energy with the increase in connectivity-displacement factor η. The network with only chemically bonded neighbor edges ($\eta = 2$) was excluded from Figure 6(b) since the error was significantly larger as compared to those from other values of η. The expansion to the first degree nonbonded neighboring edges ($\eta = 3$) shows a manageable but increasingly large error as the system size grows; but with further increase in η, the error in conformational stabilization energy is significantly reduced to far less than 1 kcal/mol error at $\eta = 4$ and with further monotonic refinement for $\eta = 5$. These η values allow the inclusion of ($i \rightarrow i + 3$) interactions, thereby properly accounting for the hydrogen bonding interactions that are critical for the 3_{10} helical conformational stability. Yet, the errors are extensive with respect to the size of the system, and hence we also present, per residue, errors in stabilization and in absolute energies in the SI. Furthermore, the complexity of the CG-calculations grows with the complexity of the full system low level calculation as seen in Figure 7. In Figure 7(b), for smaller systems such as Ala$_4$, the computational effort is dominated by the high level fragment calculations, but for larger systems, the full system low level calculation dominates the calculation.

It may also be useful to consider the behavior of Hartree–Fock as low level, while requiring MP2 at the desired higher level. This may be attractive since the fragment SCF calculations do not need to be performed again for MP2 given their availability during the Hartree–Fock fragment calculations. However, as seen in Figure 7(b), for larger systems, the computational scaling is determined by the full system low level calculation. As a result of this, the reuse of SCF results for the fragment MP2 calculations yields no significant gain in computational overhead, but this result does provide another rationalization toward the extended Lagrangian formalism in Section II.D. Thus, given the very good accuracy of the B3LYP:MP2 calculations at lower values of η, we proceed with this choice for the remaining portion of the paper.
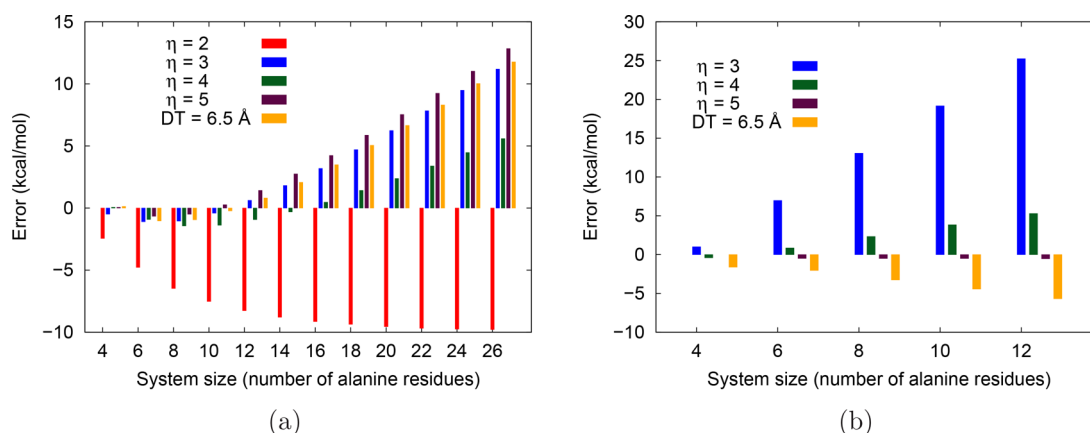
Figure 8. Errors in conformational stabilization energy between the α-helix and fully extended β-strand. B3LYP:PM6 calculations (part (a)) used 6-31++G(d,p) basis, and MP2:B3LYP calculations (part (b)) used 6-31+G(d,p) basis with the Psi4 package used for the higher level calculation. As in Figure 6 the results of $\eta = 2$ in (b) were excluded since the errors were too large and obscured the rest of the data. Due to the extensive nature of the error, the corresponding per residue errors are shown in SI, Figure SI-2, and the absolute energy errors are shown Figures SI-3 and SI-4.

It must also be noted here, that two different electronic structure packages were used here to compute the MP2 and B3LYP energies. For MP2, the Psi4 package[114] was used, whereas for DFT the Gaussian09[115] suite of electronic structure programs was used. While, this does not affect the accuracy of the results (to show this, in SI, Figure SI-5, we also present results where both MP2 and DFT fragments were computed using Gaussian09), it does provide us with additional flexibility given the wide range of options available in various individual electronic structure packages. In fact this is one of the critical hallmarks of our approach where the energies and gradients from *CG-edge*, *CG-node*, etc. are sufficiently decoupled such that completely different electronic structure environments may be used within one single calculation. This kind of partition is also done here for the AIMD calculations reported in Section IV, with further discussions on these computational details reserved for a future publication.

In contrast to the MP2:B3LYP calculations, the B3LYP:PM6 corrections (Figure 6(a)) start at a moderate accuracy with neighboring edges ($\eta = 2$) but converge to roughly 0.25 kcal/mol per residue stabilization error. This systematic error, even with increase in η, may be due to the semiempirical PM6 method, that would need to be improved through dispersion corrections.[116,117]

The use of Delaunay triangulation with a distance filter of 6.5 Å for the MP2:B3LYP demonstrates approximately the same magnitude of error as connectivity-displacement factor $\eta = 3$, but it overstabilizes the helical structure as compared to the linear structure. This is due to the lower level of refinement in the treatment of the linear structures, which with a Delaunay filter distance of 6.5 Å yields a similar simplex network as in the case of the connectivity-displacement calculation with $\eta = 2$; the helical structures' networks on the contrary remain close to $\eta = 4$. Although the helical structures have a greater gain with a larger network, the linear structures also benefit from a more expanded network. These networks from the filtered Delaunay when treated at B3LYP:PM6 display a similar error as $\eta = 4$ since the accuracy for the linear structure for this level of theory is relatively invariant to choice of η as compared to the 3_{10} helical structure.

III.B.2. Error in Stabilization of Structures That Are Not Stationary Points: α-Helix and Extended β-Strand. As in Figure 6, Figure 8 shows the accuracy of our scheme but with

the conformational energy calculated as the difference in absolute energy between idealized α-helix (Figure 4(a)) and the fully extended β-strand (Figure 4(c)). These structures allow us to investigate nonoptimized, but relevant, structures that may appear during dynamics. The α-helix is a very common peptide motif,[101−103] whereas an extended β-strand is often considered the least stable conformation of short chain polypeptides.[107] The MP2:B3LYP extrapolation for these stability calculations showed the same trend as for the optimized structures discussed above, that is monotonic convergence to subkcal/mol accuracy with increasing η, with convergence at $\eta = 5$ which is expected since the α-helical conformation has hydrogen bonding between ($i \rightarrow i + 4$) residues. However, this conformational energy dependence on η implies that as the system evolves during dynamics the peptide may deform from the 3_{10}- to α-helix requiring a different coarse-grained network to properly capture the interactions. Setting $\eta = 5$ may resolve that particular deformation, but clearly other structural changes may also occur during dynamics. The simulations conducted in this paper, however, use a fixed CG-scheme since we do not see these transformations during the dynamics calculations presented here. Future publications will evaluate aspects regarding change in network connectivity.

Turning to the B3LYP:PM6 treatment of these structures, the smaller polypeptide chains (Ala$_4$ through Ala$_{12}$) show good accuracy for the structures when the choice of η is greater than 2, but as with the systems in Figure 6, further refinement shows minimal gains in accuracy. The absolute energies for fully extended β-strand conformers are not significantly affected by choice of η, while the choice of η for the α-helix leads to the oscillatory behavior in Figure 6; but as the system size is increased, the overall error in the fully extended β-strand structure increases but remains fairly invariant to the choice of η, but the α-helix error decreases and then oscillates with the choice of η. These systematic errors are again indicative of the lack of dispersion correction in PM6 and B3LYP.

For the case of Delaunay triangulation, the MP2:B3LYP treatment produced very similar errors as it did for the optimized set of structures above. Although, due to the required change of η to capture the hydrogen bond, the magnitude of its error is now similar to $\eta = 4$ but yet again overstabilizing the helical conformer over the linear. For the B3LYP:PM6 treatment the Delaunay scheme errors closely resemble the
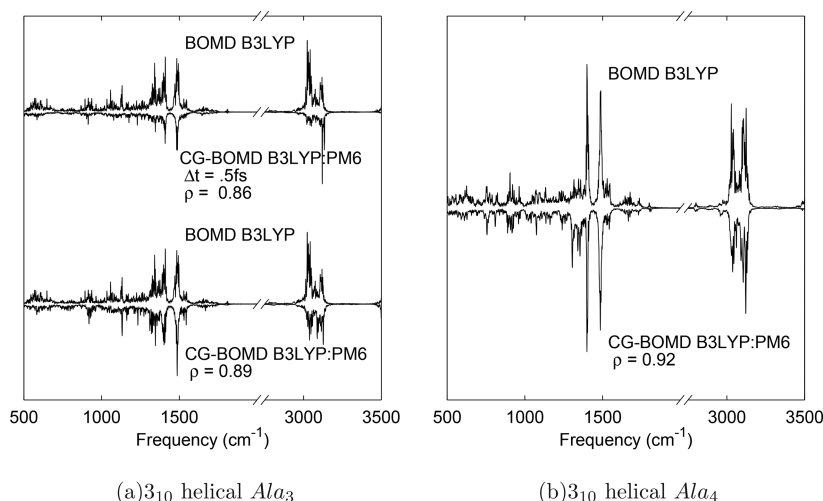
Figure 9. Vibrational density of states from B3LYP level dynamical trajectories for helical initial conformations of Ala_3 (a) and Ala_4 (b). We present the full system with a positive amplitude and each of the coarse-grained dynamics (CG) with a negative amplitude for ease of comparison.

errors for $\eta = 3$. This occurs as the CG-network for α-helix closely resembles $\eta = 5$, while the linear CG-network is effectively $\eta = 2$.

In summary, for both optimized as well as nonequilibrium structures, the MP2:B3LYP studies provide monotonic improvement with increasing η stabilizing at the stage where hydrogen bonds preserving secondary structure are correctly captured. The B3LYP:PM6 study, on the contrary, yields a systematic error with increasing size which is perhaps due to lack of dispersion correction.

IV. BENCHMARKING THE COARSE-GRAINED *AB INITIO* MOLECULAR DYNAMICS TRAJECTORIES

In this section, we test the CG-AIMD scheme for utility in efficiently computing classical trajectories for medium-sized polypeptide fragments in the gas phase. There are essentially two flavors of classical trajectory methods that we utilize here. The Born−Oppenheimer molecular dynamics[1,2,4] is used here with electronic energy defined as per eq 1. The associated nuclear gradients are used in conjunction with the velocity Verlet scheme[98] to propagate the molecular framework. These gradients appear in parts and include the full system gradients (gradients of $E^{level,0}$ with respect to the nuclear variables) at the chosen lower level of theory and *CG-nodes* and *CG-edges* (gradients of $\Delta E_\alpha^{cor,1}$ and $\Delta E_I^{cor,1}$ in eq 1 with respect to the nuclear variables). The latter include contributions from multiple levels of theory as indicated by eq 2. Furthermore, in our case these energies and gradients are obtained using two different electronic structure packages, Gaussian09[115] and Psi4,[114] and these calculations are done in parallel using an MPI protocol. The final gradients required for BOMD are assembled exactly as prescribed by eq 1 and are closely related to the equivalent discussion refs 60 and 61. Furthermore, we utilize the appropriate Jacobians to back-transform the link atom gradients on to the Cartesian framework of the molecular system.

It is already clear that the full system energy and gradients could become a bottleneck for large systems. To overcome this, we propagate the electronic structure parameters that describe the lower level full system (for example DFT) using a Car−Parrinello-like extended Lagrangian treatment[5] as described in eq 10. Our method differs from the traditional Car−Parrinello

approach in that we choose to propagate the single particle density matrix, and N-representability is assured through the idempotency condition and conservation of particle number using the Lagrangian in eq 10. Further details can be found in refs 61, 62, and 100. Furthermore, for production simulations here we chose a fictitious inertia tensor where the valence orbitals have an inertia of 180 au (=0.1 amu·bohr²), and the core orbitals are weighted as per their respective Fock matrix value as discussed in refs 61 and 62. This choice is known to perform well for hydrogen bonded systems.[28,61,118−121] The numerical value of the inertia-tensor determines the maximum time-step for CG-ADMP-pHF dynamics with larger values allowing larger time-steps.

Both CG-BOMD and CG-ADMP-pHF trajectories were computed by integrating the equations of motion using velocity Verlet[98] integration with a variety of step sizes with simulation details presented in tables in the Supporting Information. No thermostats were employed, hence the total energy would be expected to be well-conserved with acceptable drifts in total energy. An effective temperature for each simulation is obtained using average kinetic energy with associated fluctuations. We gauge the effectiveness of a dynamics trajectories by recording the energy conservation drift in total energy during dynamics. Energy conservation is reliant upon the smoothness of the potential energy surface sampled by the system, which implies a conservative Hamiltonian since these simulations are in the microcanonical regime. The conservation of total energy is also critical as the time-correlation functions are computed to obtain vibrational density of states, and these involve nuclear velocities that require a conservative Hamiltonian dynamics. All dynamics trajectories discussed here use CG-networks where the *CG-nodes* are defined as discussed in Section III.A, and *CG-edges* are a union of *CG-nodes*. Furthermore, the measure η is chosen to include a range of values including all possible amino-acid dimer units, since the conformational energy was noted to improve with η (Sections III.B.1 and III.B.2). We studied the dynamics of Ala_3 and Ala_4 in helical and β-strand initial conformations under treatment by both CG and standard electronic structure methods; the simulation parameters and energy conservation data are summarized and tabulated in the SI. As noted in the tables in SI, all simulations conserve the total energy within a few of a tenths of a kcal/mol.
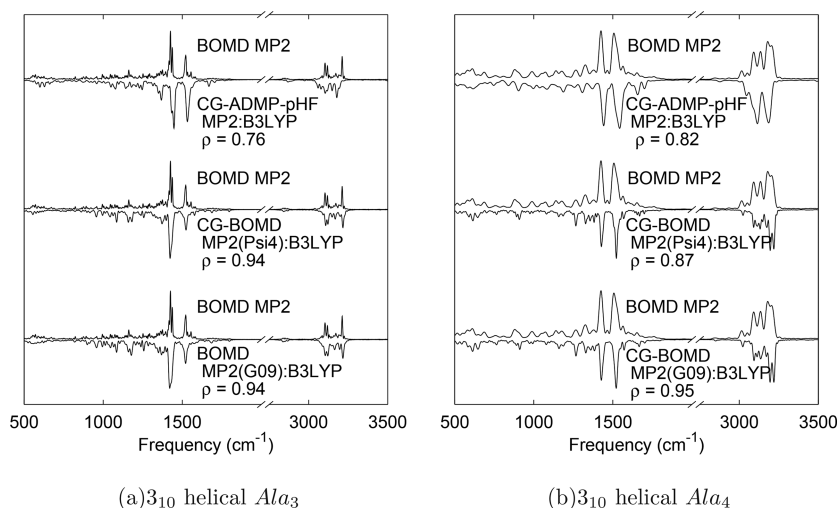
(a) 3_{10} helical Ala_3

(b) 3_{10} helical Ala_4

Figure 10. Vibrational density of states calculated from MP2 level dynamical trajectories for helical initial conformations of Ala_3 (a) and Ala_4 (b). We present the full system with a positive amplitude and each of the coarse-grained dynamics (CG) with a negative amplitude for ease of comparison. The CG-ADMP-pHF trajectories frequencies were shifted by a scaling factor of 1.021 as in previous work.[62] See the discussion in Section IIII.D.

We next inspect the vibrational density of states to gauge the level of agreement between the coarse-grained trajectories and higher level full system calculations. As in previous work,[9,10,18,19,122] we compute the absolute value of Fourier transform of the nuclear velocities ($\tilde{\mathbf{V}}(\omega)$) to provide the spectral activity involved in each trajectory at a given frequency.

$$
I_V(\omega) = \lim_{T \to \infty} \int_{t=0}^{t=T} dt \, \exp(-\iota\omega t)\langle \mathbf{V}(0)\cdot\mathbf{V}(t)\rangle
$$
$$
= \tilde{\mathbf{V}}(\omega)\cdot\tilde{\mathbf{V}}(\omega) = |\tilde{\mathbf{V}}(\omega)|^2 \quad (12)
$$

This expression, based on the convolution theorem,[123] obtains the vibrational density of states ($I(\omega)$) from the nuclear velocities. Here we utilize this to gauge agreement between trajectories by probing the level to which spectral activity is reproduced. In addition, as stated in Section IIII.D and examined in more detail in ref 62, a uniform scaling factor of 1.021 was applied to the frequencies in CG-ADMP-pHF trajectories.

In order to quantify this agreement, we calculate the correlation coefficient between the vibrational density of states for the CG-BOMD and CG-ADMP-pHF trajectories (represented as $\mathbf{I}_{V,CG}$) and the BOMD trajectories where the full system energy and forces are computed at the target level of theory at every step (represented as $\mathbf{I}_{V,BOMD}$) as

$$
\rho = \sqrt{\frac{\mathbf{I}_{V,BOMD}\cdot\mathbf{I}_{V,CG}}{\|\mathbf{I}_{V,BOMD}\|\,\|\mathbf{I}_{V,CG}\|}} \quad (13)
$$

which is the Pearson correlation coefficient[123] of the two. The more similar the two spectra, the closer ρ would be to identity.

We present selected spectra for B3LYP (Figure 9) and MP2 (Figure 10); the spectra for the other dynamics trajectories can be found in the SI. These spectra show quantitative agreement between the coarse-grained and full system dynamics. The essential features from the full benchmarks appear in the CG-AIMD vibrational states as well. The most pronounced peaks in the lower frequency end of the spectra, here presented, represent the Amide I and Amide II modes. These modes are often used to aid in the characterization of polypeptide conformations,[124,125] and our CG-method is able to correctly replicate these important bands. Across the spectra, the CG-

AIMD trajectories were able to reproduce the methyl and α-carbon−hydrogen stretch regime (about 3100 cm^{-1} to 3200 cm^{-1}). The DFT results are red-shifted with respect to the MP2 derived density of states. The given coarse-grained dynamics matches this relationship quite well and shows qualitative agreement to the full system. The CG-BOMD for MP2:B3LYP density of states had correlation coefficients greater than 0.9 with the full BOMD MP2 trajectories, and CG-BOMD for B3LYP:PM6 had coefficients greater than 0.8. This demonstrates good agreement between these trajectories.

V. CONCLUSIONS

In this publication, we introduce a geometric network formalism to coarse grain large molecular systems to ease the computation of electronic structure gradients for use in *ab initio* molecular dynamics (AIMD) trajectories. These geometric networks were constructed through the partitioning of the system into nonoverlapping nodes which are connected by edges to form a connected graph. We employed two schemes to determine edge formation: a filtered Delaunay triangulation[84,85] and a connectivity-displacement procedure. An energetic measure for the full system is constructed using this geometric network. This energetic measure is shown to be analogous to previously developed fragmentation methods,[51,60,71,73−76,93] but it circumvents the exponential scaling costs associated with determining fragment overlaps and weights consistent with the Principle of Inclusion and Exclusion.[94] Furthermore, the current approach also provides an efficient computational framework to evaluate higher order many-body effects in a completely adaptive fashion.

The scaling of post-Hartree−Fock methods limits the utilization of these electronic structure methods for larger systems and makes *ab initio* molecular dynamics simulations cost prohibitive for any system larger than a few dozen atoms. The goal for this work is to provide alternatives that can bridge this scaling problem. This is specifically achieved here by replacing the scaling costs for full system electronic structure treatment with a cheaper method. The higher levels of treatment are reserved for the *CG-edges* and *CG-nodes* that are significantly smaller and system size independent, thus contributing only a constant computational scaling, restricting

the computational cost to the lower level treatment on the full system. Furthermore, the individual *CG-edge* and *CG-node* calculations at all levels of treatment are completely decoupled and hence can be spawned on multiple nodes on a distributed computer-architecture. This implies that multiple electronic structure packages may be used interchangeably; in this work we used the Gaussian09[115] and Psi4[114] packages.

Through conformational benchmarks with varying lengths of polyalanine chains, it was shown that the aforementioned energetic measure can achieve MP2 level accuracy with DFT cost. Next it was demonstrated that the use of this geometric network to determine the gradients in use in AIMD, within both Born–Oppenheimer and extended Lagrangian formalisms, conserved the total energy to within the sub-kcal/mol range. These coarse-grained dynamics trajectories were able to reproduce the spectral features found in the full system benchmark comparison trajectories.

APPENDIX A: HIGH-ORDER GENERALIZATIONS TO EQ 7 PROVIDE A SYSTEMATIC ROUTE TO INCREASING ACCURACY

The procedure developed in Section IIII.B is generalized to include all embedded *CG-simplexes*. The essence of this idea in captured in the expression below

$$E_{\text{network}}^{N\text{-}gon} = E^{\text{level},0} + \sum_{\alpha} \Delta E_{\alpha,N}^{\text{cor.},1}$$
$$- \sum_{n=(N-1)}^{1} (-1)^{N-n} \left\{ \sum_{\alpha} \Delta E_{\alpha,n}^{\text{cor.},1} \left[\sum_{m=n+1}^{N} (-1)^{m-n} (p_{\alpha}^{n,m} - 1) \right] \right\} \tag{A1}$$

where summation in the third term on the right side is presented in reverse to include corrections from the largest subclusters (*n-gons*) and then consider corrections arising from each set of progressively smaller subsystems. The term $\Delta E_{\alpha,N}^{\text{cor.},1}$ is a generalization to the corresponding term in eq 7 but now includes energy corrections from embedded *N-gons*. Furthermore, the square-bracketed term incorporates an overcounting correction for the α-th *n-gon*, with energy correction, $\Delta E_{\alpha,n}^{\text{cor.},1}$. The term $p_{\alpha}^{n,m}$ is essentially a generalization of p_I, in eqs 1 and 3, and represents the number of times the α-th *n-gon* appears in all *m-gons*.

In addition, multiple levels of theory can be added by simply expanding $E^{\text{level},0}$ in terms of additional layers. Alternately, specific *CG-nodes*, *CG-edges*, or higher order *CG-simplices* may be further decomposed using a similar approach.

ASSOCIATED CONTENT

ⓢ Supporting Information
The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.8b00186.

> Additional AIMD data and additional computational benchmarks for accuracy and tables of simulation details (PDF)

AUTHOR INFORMATION

Corresponding Author
*E-mail: iyengar@indiana.edu.

ORCID ⓘ
Srinivasan S. Iyengar: 0000-0001-6526-2907

REFERENCES

(1) Wang, I. S. Y.; Karplus, M. Dynamics of Organic Reactions. *J. Am. Chem. Soc.* **1973**, *95*, 8160.

(2) Leforestier, C. Classical Trajectories Using the Full Ab Initio Potential Energy Surface H⁻+CH₄ → CH₄+H⁻. *J. Chem. Phys.* **1978**, *68*, 4406.

(3) Helgaker, T.; Uggerud, E.; Jensen, H. J. A. Integration of the Classical Equations of Motion on Ab Initio Molecular Potential Energy Surfaces Using Gradients and Hessians: Application to Translational Energy Release upon Fragmentation. *Chem. Phys. Lett.* **1990**, *173*, 145.

(4) Bolton, K.; Hase, W. L.; Peslherbe, G. H. In *Modern Methods for Multidimensional Dynamics Computation in Chemistry*; Thompson, D. L., Ed.; World Scientific, Singapore, 1998; Chapter Direct Dynamics of Reactive Systems, p 143.

(5) Car, R.; Parrinello, M. Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.* **1985**, *55*, 2471.

(6) Remler, D. K.; Madden, P. A. Molecular Dynamics Without Effective Potentials Via the Car-Parrinello Approach. *Mol. Phys.* **1990**, *70*, 921.

(7) 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.

(8) Alhambra, C.; Byun, K.; Gao, J. The Geometry of Water in Liquid Water from Hybrid Ab Initio-Monte Carlo and Density Functional-Molecular Dynamics Simulations. *Combined Quantum Mechanical and Molecular Mechanics Methods*; Gao, J., Thompson, M., Eds.; ACS Symposium Series 712; 1998; p 35.

(9) 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.

(10) 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 (H₅O₂⁺) Cation. *J. Chem. Theory Comput.* **2012**, *8*, 4876.

(11) 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.

(12) Hayes, R. L.; Fattal, E.; Govind, N.; Carter, E. A. Long Live Vinylidene! A New View of the H₂CC: → HC≡CH Rearrangement from ab Initio Molecular Dynamics. *J. Am. Chem. Soc.* **2001**, *123*, 641–657.

(13) Li, J.; Pacheco, A. B.; Raghavachari, K.; Iyengar, S. S. A Grotthuss-like proton shuttle in the anomalous C₂H₃⁺ Carbocation: Energetic and vibrational properties for isotopologues. *Phys. Chem. Chem. Phys.* **2016**, *18*, 29395.

(14) Sager, L. M.; Iyengar, S. S. Proton relays dictate stability and spectroscopy in anomalous carbocations: Case studies on C₂H₅⁺ and C₃H₃⁺. *Phys. Chem. Chem. Phys.* **2017**, *19*, 27801.

(15) Sumner, I.; Iyengar, S. S. Quantum Wavepacket *Ab Initio* Molecular Dynamics: An Approach for Computing Dynamically Averaged Vibrational Spectra Including Critical Nuclear Quantum Effects. *J. Phys. Chem. A* **2007**, *111*, 10313.

(16) Li, J.; Li, X.; Iyengar, S. S. Vibrational Properties of Hydrogen Bonded Systems Using the Multi-Reference Generalization to the "On-The-Fly" Electronic Structure Within Quantum Wavepacket *Ab Initio* Molecular Dynamics (QWAIMD). *J. Chem. Theory Comput.* **2014**, *10*, 2265.

(17) Pylaeva, S. A.; Elgabarty, H.; Sebastiani, D.; Tolstoy, P. M. Symmetry and dynamics of FHF⁻ anion in vacuum, in CD₂Cl₂ and in

CCl_4. Ab initio MD study of fluctuating solvent-solute hydrogen and halogen bonds. *Phys. Chem. Chem. Phys.* **2017**, *19*, 26107−26120.

(18) Li, X.; Moore, D. T.; Iyengar, S. S. Insights from First Principles Molecular Dynamics Studies Towards Infra-Red Multiple-Photon and Single-Photon Action Spectroscopy: Case Study of the Proton-Bound Di-Methyl Ether Dimer. *J. Chem. Phys.* **2008**, *128*, 184308.

(19) Li, X.; Oomens, J.; Eyler, J. R.; Moore, D. T.; Iyengar, S. S. Isotope Dependent, Temperature Regulated, Energy Repartitioning in a Low-Barrier, Short-Strong Hydrogen Bonded Cluster. *J. Chem. Phys.* **2010**, *132*, 244301.

(20) 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.

(21) Wong, K.-Y.; Gao, J. Insight into Phosphodiesterase Mechanism from Combined QM/MM Molecular Dynamics Simulations. *FEBS J.* **2011**, *278*, 2579.

(22) Harris, D. L. Oxidation and Electronic State Dependence of Proton Transfer in the Enzymatic Cycle of Cytochrome P450eryF. *J. Inorg. Biochem.* **2002**, *91*, 568.

(23) Lin, Y.-l.; Gao, J. Kinetic Isotope Effects of L-Dopa Decarboxylase. *J. Am. Chem. Soc.* **2011**, *133*, 4398.

(24) 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.

(25) Iyengar, S. S.; Sumner, I.; Jakowski, J. Hydrogen Tunneling in an Enzyme Active Site: A Quantum Wavepacket Dynamical Perspective. *J. Phys. Chem. B* **2008**, *112*, 7601.

(26) 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.

(27) Gerber, R. B.; Sebek, J. Dynamics simulations of atmospherically relevant molecular reactions. *Int. Rev. Phys. Chem.* **2009**, *28*, 207−222.

(28) 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.

(29) Hammerich, A. D.; Finlayson-Pitts, B. J.; Gerber, R. B. NOx Reactions on Aqueous Surfaces with Gaseous HCl: Formation of a Potential Precursor to Atmospheric Cl Atoms. *J. Phys. Chem. Lett.* **2012**, *3*, 3405−3410.

(30) Iannuzzi, M.; Parrinello, M. Proton Transfer in Heterocycle Crystals. *Phys. Rev. Lett.* **2004**, *93*, 025901.

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

(32) Lin, I.-H.; Lu, Y.-H.; Chen, H.-T. Nitrogen-doped C_{60} as a robust catalyst for CO oxidation. *J. Comput. Chem.* **2017**, *38*, 2041−2046.

(33) Iannuzzi, M.; Laio, A.; Parrinello, M. Efficient Exploration of Reactive Potential Energy Surfaces Using Car-Parrinello Molecular Dynamics. *Phys. Rev. Lett.* **2003**, *90*, 238302.

(34) Elber, R.; Ghosh, A.; Cardenas, A. Long Time Dynamics of Complex Systems. *Acc. Chem. Res.* **2002**, *35*, 396.

(35) Bylaska, E. J.; Weare, J. Q.; Weare, J. H. Extending molecular simulation time scales: Parallel in time integrations for high-level quantum chemistry and complex force representations. *J. Chem. Phys.* **2013**, *139*, 074114.

(36) Tuckerman, M. E.; Berne, B. J.; Rossi, A. Molecular Dynamics Algorithm for Multiple Time Scales: Systems with Disparate Masses. *J. Chem. Phys.* **1991**, *94*, 1465.

(37) Steele, R. P. Multiple-timestep ab initio molecular dynamics using an atomic basis set partitioning. *J. Phys. Chem. A* **2015**, *119*, 12119−12130.

(38) Chen, M.; Yu, T.-Q.; Tuckerman, M. E. Locating landmarks on high-dimensional free energy surfaces. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 3235−3240.

(39) Yu, T.-Q.; Chen, P.-Y.; Chen, M.; Samanta, A.; Vanden-Eijnden, E.; Tuckerman, M. Order-parameter-aided temperature-accelerated sampling for the exploration of crystal polymorphism and solid-liquid phase transitions. *J. Chem. Phys.* **2014**, *140*, 214109.

(40) Zhang, X.; Lu, G.; Curtin, W. A. Multiscale quantum/atomistic coupling using constrained density functional theory. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2013**, *87*, 054113.

(41) Ayala, P. Y.; Scuseria, G. E. Linear scaling second-order Møller-Plesset theory in the atomic orbital basis for large molecular systems. *J. Chem. Phys.* **1999**, *110*, 3660−3671.

(42) Schütz, M.; Hetzer, G.; Werner, H.-J. Low-order scaling local electron correlation methods. I. Linear scaling local MP2. *J. Chem. Phys.* **1999**, *111*, 5691−5705.

(43) Distasio, R. A.; Steele, R. P.; Rhee, Y. M.; Shao, Y.; Head-Gordon, M. An improved algorithm for analytical gradient evaluation in resolution-of-the-identity second-order Møller-Plesset perturbation theory: Application to alanine tetrapeptide conformational analysis. *J. Comput. Chem.* **2007**, *28*, 839−856.

(44) Pavošević, F.; Pinski, P.; Riplinger, C.; Neese, F.; Valeev, E. F. Sparse Maps-A systematic infrastructure for reduced-scaling electronic structure methods. IV. Linear-scaling second-order explicitly correlated energy with pair natural orbitals. *J. Chem. Phys.* **2016**, *144*, 144109.

(45) Sode, O.; Hirata, S. Second-order many-body perturbation study of solid hydrogen fluoride under pressure. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7765−7779.

(46) Klimes, J.; Michaelides, A. Perspective: Advances and challenges in treating van der Waals dispersion forces in density functional theory. *J. Chem. Phys.* **2012**, *137*, 120901.

(47) Mori-Sanchez, P.; Cohen, A. J.; Yang, W. T. Many-electron self-interaction error in approximate density functionals. *J. Chem. Phys.* **2006**, *125*, 201102.

(48) Cohen, A. J.; Mori-Sanchez, P.; Yang, W. T. Challenges for Density Functional Theory. *Chem. Rev.* **2012**, *112*, 289.

(49) 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.

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

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

(52) 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.

(53) 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.

(54) 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.

(55) 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.

(56) Lange, A. W.; Voth, G. A. Multi-State Approach to Chemical Reactivity in Fragment Based Quantum Chemistry Calculations. *J. Chem. Theory Comput.* **2013**, *9*, 4018.

(57) Collins, M. A. Can Systematic Molecular Fragmentation Be Applied to Direct Ab Initio Molecular Dynamics? *J. Phys. Chem. A* **2016**, *120*, 9281−9291.

(58) Liu, J.; Zhu, T.; Wang, X.; He, X.; Zhang, J. Z. H. Quantum Fragment Based ab Initio Molecular Dynamics for Proteins. *J. Chem. Theory Comput.* **2015**, *11*, 5897.

(59) Willow, S. Y.; Salim, M. A.; Kim, K. S.; Hirata, S. Ab initio molecular dynamics of liquid water using embedded fragment second-order many-body perturbation theory towards its accurate property prediction. *Sci. Rep.* **2015**, *5*, 14358.

(60) 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.

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

(62) Haycraft, C.; Li, J.; Iyengar, S. S. On-the-fly" Ab initio molecular dynamics with coupled cluster accuracy. *J. Chem. Theory Comput.* **2017**, *13*, 1887.

(63) Liu, J.; Zhu, T.; Wang, X.; He, X.; Zhang, J. Z. H. Quantum Fragment Based ab Initio Molecular Dynamics for Proteins. *J. Chem. Theory Comput.* **2015**, *11*, 5897−5905.

(64) Liu, J.; He, X.; Zhang, J. Z. H.; Qi, L.-W. Hydrogen-bond structure dynamics in bulk water: insights from ab initio simulations with coupled cluster theory. *Chem. Sci.* **2018**, *9*, 2065.

(65) Yang, W. Direct Calculation of Electron Density in Density-Functional Theory. *Phys. Rev. Lett.* **1991**, *66*, 1438.

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

(67) 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.

(68) 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.

(69) 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.

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

(71) 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.

(72) 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.

(73) 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.

(74) 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.

(75) 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.

(76) 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.

(77) 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.

(78) 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.

(79) Kamiya, M.; Hirata, S.; Valiev, M. Fast Electron-Correlation Methods for Molecular Crystals Without Basis Set Superposition Errors. *J. Chem. Phys.* **2008**, *128*, 074103.

(80) 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.

(81) 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.

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

(83) Braams, B. J.; Bowman, J. M. Permutationally invariant potential energy surfaces in high dimensionality. *Int. Rev. Phys. Chem.* **2009**, *28*, 577.

(84) Bowyer, A. Computing Dirichlet tessellations. *Comput. J.* **1981**, *24*, 162−166.

(85) Watson, D. Computing the n-dimensional Delaunay tessellation with applications to Voronoi polytopes. *Comput. J.* **1981**, *24*, 167−172.

(86) Aurenhammer, F. Voronoi Diagrams − A survey of a fundamental geometric data structure. *ACM Comput. Survey* **1991**, *23*, 345.

(87) Okabe, A.; Boots, B.; Sugihara, K.; Chiu, S. N. *Spatial Tessellations − Concepts and applications of Voronoi diagrams*; John Wiley and Sons: 2000.

(88) Hert, S.; Seel, M. dD Convex Hulls and Delaunay Triangulations. In *CGAL User and Reference Manual*, 4.10 ed.; CGAL Ed. Board: 2017.

(89) DeGregorio, N.; Iyengar, S. S. Efficient and adaptive methods for computing accurate potential surfaces for quantum nuclear effects: Applications to hydrogen transfer reactions. *J. Chem. Theory Comput.* **2018**, *14*, 30−47.

(90) Coffey, T. M.; Wyatt, R. E.; Schieve, W. C. Reconstruction of the Time-Dependent Wave Function Exclusively from Position Data. *Phys. Rev. Lett.* **2011**, *107*, 230403.

(91) Sun, L.; Yeh, G.-T.; Ma, X.; Lin, F.; Zhao, G. Engineering applications of 2D and 3D finite element mesh generation in hydrogeology and water resources. *Comput. Geosci* **2017**, *21*, 733−758.

(92) 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.

(93) Deev, V.; Collins, M. A. Approximate ab initio energies by systematic molecular fragmentation. *J. Chem. Phys.* **2005**, *122*, 154102.

(94) Björklund, A.; Husfeldt, T.; Koivisto, M. Set Partitioning via Inclusion Exclusion. *SIAM J. Comput.* **2009**, *39*, 546.

(95) Andersen, H. C. Molecular Dynamics Simulations at Constant Pressure And/or Temperature. *J. Chem. Phys.* **1980**, *72*, 2384.

(96) Parrinello, M.; Rahman, A. Crystal Structure and Pair Potentials: A Molecular-Dynamics Study. *Phys. Rev. Lett.* **1980**, *45*, 1196.

(97) Hutter, J. Car-Parrinello Molecular Dynamics. *WIREs-Comp. Mol. Sci.* **2012**, *2*, 604−612.

(98) 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.

(99) McWeeny, R. Some Recent Advances in Density Matrix Theory. *Rev. Mod. Phys.* **1960**, *32*, 335.

(100) 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.

(101) Padmanabhan, S.; Marqusee, S.; Ridgeway, T.; Laue, T. M.; Baldwin, R. L. Relative Helix-Forming Tendencies of Nonpolar Amino Acids. *Nature* **1990**, *344*, 268−270.

(102) Chakrabartty, A.; Kortemme, T.; Baldwin, R. L. Helix propsensities of the amino acids measured in alanine-based peptides without helix-stablizing side-chain interactions. *Protein Sci.* **1994**, *3*, 843−852.

(103) Bagińska, K.; Makowska, J.; Wiczk, W.; Kasprzykowski, F.; Chmurzyński, L. Conformational studies of alanine-rich peptide using CD and FTIR spectroscopy. *J. Pept. Sci.* **2008**, *14*, 283−289.

(104) Rossi, M.; Scheffler, M.; Blum, V. Impact of Vibrational Entropy on the Stability of Unsolvated Peptide Helices with Increasing Length. *J. Phys. Chem. B* **2013**, *117*, 5574−5584.

(105) Levy, Y.; Jortner, J.; Becker, O. M. Solvent effects on the energy landscapes and folding kinetics of polyalanine. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 2188−2193.

(106) Jagielska, A.; Skolnick, J. Origin of intrinsic 3_{10}-helix versus strand stability in homopolypeptides and its implications for the accuracy of the Amber force field. *J. Comput. Chem.* **2007**, *28*, 1648−1657.

(107) Wieczorek, R.; Dannenberg, J. J. Comparison of Fully Optimized α- and 3_{10}−Helices with Extended ß-Strands. An ONIOM Density Functional Theory Study. *J. Am. Chem. Soc.* **2004**, *126*, 14198−14205.

(108) Li, J.; Wang, Y.; Chen, J.; Liu, Z.; Bax, A.; Yao, L. Observation of α−Helical Hydrogen-Bond Cooperativity in an Intact Protein. *J. Am. Chem. Soc.* **2016**, *138*, 1824−1827.

(109) Nelson, D. L.; Cox, M. M. In *Lehninger Principles of Biochemistry*, 4th edition ed.; Freeman: 2004.

(110) Tonlolo, C.; Benedetti, E. The polypeptide 3_{10}-helix. *Trends Biochem. Sci.* **1991**, *16*, 350−353.

(111) Loughlin, W. A.; Tyndall, J. D. A.; Glenn, M. P.; Fairlie, D. P. Beta-Strand Mimetics. *Chem. Rev.* **2004**, *104*, 6085−6118.

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

(113) Dapprich, S.; Komáromi, I.; Byun, K.; 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*, 1−21.

(114) Parrish, R. M.; Burns, L. A.; Smith, D. G. A.; Simmonett, A. C.; DePrince, A. E.; Hohenstein, E. G.; Bozkaya, U.; Sokolov, A. Y.; Di Remigio, R.; Richard, R. M.; Gonthier, J. F.; James, A. M.; McAlexander, H. R.; Kumar, A.; Saitow, M.; Wang, X.; Pritchard, B. P.; Verma, P.; Schaefer, H. F.; Patkowski, K.; King, R. A.; Valeev, E. F.; Evangelista, F. A.; Turney, J. M.; Crawford, T. D.; Sherrill, C. D. Psi4 1.1: An Open-Source Electronic Structure Program Emphasizing Automation, Advanced Libraries, and Interoperability. *J. Chem. Theory Comput.* **2017**, *13*, 3185−3197.

(115) 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.

(116) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

(117) Rezac, J.; Fanfrlik, J.; Salahub, D.; Hobza, P. Semiempirical Quantum Chemical PM6Method Augmented by Dispersion and H-Bonding Correction Terms Reliably Describes Various Types of Noncovalent Complexes. *J. Chem. Theory Comput.* **2009**, *5*, 1749−1760.

(118) 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.

(119) 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.

(120) 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.

(121) 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.

(122) Iyengar, S. S. Further Analysis of the Dynamically Averaged Vibrational Spectrum for the "Magic" Protonated 21-Water Cluster. *J. Chem. Phys.* **2007**, *126*, 216101.

(123) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in C*; Cambridge University Press: New York, 1992.

(124) Krishnan, M.; Gupta, V. Vibration Spectra of Alpha-helix of Poly-alanine. *Chem. Phys. Lett.* **1970**, *6*, 231−234.

(125) Dousseau, F.; Pézolet, M. Determination of the Secondary Structure Content of Proteins in Aqueous Solutions from Their Amide I and Amide II Infrared Bands. Comparison between Classical and Partial Least-Squares Methods. *Biochem.* **1990**, *29*, 8771−8779.