JCTC Journal of Chemical Theory and Computation

A Resolution-Of-The-Identity Implementation of the Local Triatomics-In-Molecules Model for Second-Order Møller−Plesset Perturbation Theory with Application to Alanine Tetrapeptide Conformational Energies

Robert A. DiStasio Jr., Yousung Jung, and Martin Head-Gordon*

Department of Chemistry, University of California, Berkeley, California 94720

Received May 5, 2005

Abstract: In this work, we incorporate the resolution-of-the-identity (RI) approximation into the theoretical framework of the local triatomics-in-molecules (TRIM) second-order Møller−Plesset (MP2) perturbation theory model. The resultant model, RI-TRIM MP2, emerges as a robust fourth-order methodology that extends the regime of practical MP2 calculations. With RI-TRIM MP2, correlation energy corrections can easily be obtained for systems that contain more than 125 heavy atoms with a computational timing cost less than those of the prerequisite self-consistent field procedure and popular density functional theory (DFT) alternatives. In this work, the chemical performance of RI-TRIM MP2 is numerically assessed against untruncated RI-MP2 and DFT (B3LYP) in determining the relative energies of 27 different alanine tetrapeptide conformations at the cc-pVXZ (X = D, T, and Q) levels and the results are T → Q extrapolated to the complete basis set limit. As the quality of the basis set employed increases, we report a significant reduction in the error introduced by the RI-TRIM approximation; at the cc-pVDZ level, the root mean-square (RMS) relative error was found as 0.192 kcal/mol and is decreased to an almost negligible 0.040 kcal/mol at the T → Q extrapolated complete basis set limit. Basis set dependence was investigated by computing the RMS (max) deviations from the extrapolated RI-MP2/cc-pV(TQ)Z data set found as 0.377 (0.944) kcal/mol (MP2/cc-pVTZ) and 0.250 (0.591) kcal/mol (TRIM MP2/cc-pVTZ). These deviations are chemically significant when compared against the conformer energy differences, suggesting that to obtain reliably converged relative conformational energies, computations must be done using the cc-pVTZ and cc-pVQZ basis sets followed by extrapolation to the cc-pV(TQ)Z limit. The findings reported herein also provide the first computational evidence demonstrating that the TRIM model approaches exactness as the one-particle basis approaches completeness.

I. Introduction

Over the past two decades, density functional theory (DFT) has emerged as a robust electronic structure method for treating a myriad of chemical problems, especially those that involve large molecular systems.[1−3] Despite its remarkable success as a moderately accurate method, DFT has several significant weaknesses. Current DFT implementations fail to provide a physically correct description of the dispersion forces[4] responsible for base pair stacking in DNA, dimer formation among rare gases, and various other long-range correlation effects.[5] Furthermore, DFT suffers from the well-documented self-interaction problem[6,7] leading to an overall underestimation of activation energies[8] and a general uncertainty in the vicinity of first-order saddle points on potential energy surfaces. Finally, the numerical quadrature techniques employed in DFT (as well as other methodologies that rely on numerical grids such as the class of pseudospectral

* Corresponding author e-mail: mhg@bastille.cchem.berkeley.edu, tel.: 1-510-642-5957.

10.1021/ct050126s CCC: $30.25
Published on Web 08/06/2005

methods) introduce numerical noise and, therefore, further uncertainty into the characterization of the fine details of potential energy surfaces.

During the same time, several attempts have been made at reducing the computational cost of second-order Møller−Plesset (MP2) perturbation theory[9] calculations. MP2 theory is the simplest and most economical wave-function-based improvement to the mean-field Hartree−Fock (HF) approximation that correctly accounts for the physics of dispersion interactions.[10] Because of simplicity in ansatz, MP2 theory is ideal as a starting point for developing methods and techniques that might be extended to calculations employing higher-level correlation theories, that is, those that involve either the coupled-cluster or configuration interaction formalisms or in "ab initio" approaches to improving the functional forms utilized in DFT. Additionally, equilibrium geometries obtained using MP2 theory are typically more reliable than those obtained using the popular DFT alternatives.[11] However, MP2 theory also has several computational limitations worth mentioning. If MP2 theory is formulated in a basis of canonical molecular orbitals (MOs), that is, the orthonormal set of eigenfunctions that diagonalize the Fock matrix, one finds an unphysical fifth-order computational dependence on the system size (M^5).[12] This relatively high computational cost (DFT formally scales as M^3) coupled with the need for large atomic orbital (AO) basis sets for reliable results further limits MP2 from the regime of large molecular systems.[13] The focus of this work is reducing this unnecessary computational cost and extending the current size limitations to practical MP2 calculations. We also note that the performance of MP2 theory when dealing with radicals and systems with small HOMO/LUMO energy gaps[14] has been somewhat disappointing when compared to its treatment of closed-shell systems.[15] However, most large molecular systems, especially those dealt with in a biochemical context, are closed-shell systems and should be unaffected by this drawback.

Several promising routes to decreasing the cost of MP2 computations have surfaced in the literature and can be divided into four distinct categories. The first approach involves a reduction in the computational prefactor but leaves the underlying fifth-order scaling untouched and includes the resolution-of-the-identity (RI) approximation,[16,17] pseudospectral methods,[18] and others.[19] The second class of methods reformulates MP2 theory in an AO basis and exploits the underlying locality with the Laplace transformation leading to linear scaling among one-dimensional systems using small basis sets.[20,21] The third approach directly attacks the aforementioned scaling by making use of localized MOs and, therefore, introduces fundamental changes by ansatz.[22−25] The fourth class of methods includes hybrid methods that have just recently emerged in the literature. One such method combines prefactor reduction (RI), the Laplace transformation, and other approximations such as separate scaling of the same-spin and opposite-spin components of the MP2 energy,[26,27] on the basis of the findings of Grimme.[28] Other hybrid methods are based on the numerous works of Schütz and co-workers;[29−32] of particular note is the incorporation of the density-fitting approach (synonymous in this context with the RI approximation) with the classic local model of Saebø and Pulay,[22] which has produced remarkably low-scaling methods for use in both single-point energy evaluations[33] and geometry optimizations.[34] In doing so, these novel hybrid MP2 methodologies have demonstrated significantly improved computational performance resulting from the combined reduction in the prefactor as well as a fundamental reduction in the underlying scaling.

The purpose of this paper is to report a new hybrid MP2 method and to thoroughly explore its chemical and computational performance. This new method combines prefactor reduction via incorporation of the RI approximation and a reduction in the underlying scaling by utilizing the local triatomics-in-molecules (TRIM) ansatz. Incorporation of the RI approximation into MP2 theory has demonstrated a marked increase in computational efficiency accompanied by tolerable error (~60 μH/atom).[17] Hence, the RI-MP2 methodology itself has extended the regime of molecular systems that can be investigated at the MP2 level of theory. By combining this powerful technique with a formal reduction in the MP2 correlation space via the TRIM ansatz, the resulting methodology, RI-TRIM MP2, emerges as a computationally efficient fourth-order method with the ability to treat even larger molecular systems. One powerful application of RI-TRIM MP2 would be its use in the design and fundamental parameterization of force fields and other semiempirical methods that can be used to more accurately treat large systems of biochemical and nanotechnological interest.

The remainder of this paper is outlined as follows. In Section II, the theory underlying the TRIM model will be briefly reviewed and the incorporation of the RI approximation into the working TRIM framework will be developed. A generalized algorithm for executing an RI-TRIM MP2 calculation will be presented in Section III, accompanied by a detailed analysis of the associated computational requirements. In Section IV, the numerical performance of the RI-TRIM MP2 model will be assessed against the MP2, TRIM MP2, RI-MP2, and DFT (B3LYP) methodologies in the context of determining the relative energies of 27 different alanine tetrapeptide conformations. Further assessment of this new methodology will follow from extrapolation of the untruncated RI-MP2 and RI-TRIM MP2 results to the complete basis set limit. The paper will then finish with some brief conclusions in Section V.

II. Theory

A. Localization of the Occupied and Virtual Subspaces. MP2 theory begins with a delineation of the occupied and virtual single-particle subspaces. When choosing functions to span these subspaces, a balance is sought between redundancy, overcompleteness, and ease of definition. Additionally, the functions must be atom-centered so that atomic truncations can be made that are smooth and globally defined. In this work, we follow the same procedure as that developed in the original TRIM paper[24] and earlier works.[35,36]

Given an AO basis consisting of N functions (O occupied and V virtual), the localized virtual functions will be defined as the set of AOs projected into the virtual subspace, $|\underline{a}_A\rangle =$

$\hat{Q}|\mu_A\rangle = (\hat{1} - \hat{P})|\mu_A\rangle$, following the original work of Saebø and Pulay.[22] In this representation, there is only a small fractional redundancy for calculations employing high-quality basis sets (where $V \approx N$). Since O does not depend on the basis set for a given molecular system, the analogous definition of the localized occupied functions would be tremendously overcomplete. The least overcomplete atomic representation that still exactly spans the occupied subspace is an orthonormal minimal basis of extracted polarized AOs (EPAOs).[37] Since the EPAOs are linear combinations of AOs, the localized occupied functions will be analogously defined as the set of EPAOs projected into the occupied subspace $|\underline{i}_A\rangle = \hat{P}|\alpha_A\rangle = \hat{P}(\sum_\mu \mathbf{B}_{\mu\alpha}|\mu\rangle)$, where $\mathbf{B}$ is the EPAO coefficient matrix.

B. Review of General MP2 Theory. In this section and the remainder of the paper, indices *a*, *b*, *c*, ... will refer to canonical virtual MOs and *i*, *j*, *k*, ... will refer to canonical occupied MOs. We will begin with the following general expression for the MP2 energy in terms of spin−orbitals

$$E_{\mathrm{MP2}} = \frac{1}{2}\sum_{iajb} t_{ij}^{ab}(ia|jb) \tag{1}$$

where t_{ij}^{ab} are the first-order amplitudes (i.e., $|\Psi^{(1)}\rangle = \sum_{iajb} t_{ij}^{ab}|\Psi_{ij}^{ab}\rangle$) given as the solution to the following set of linear equations:

$$\sum_{kcld} \Delta_{iajb;kcld} t_{kl}^{cd} = (ia||jb) \tag{2}$$

where $(ia||jb)$ are the antisymmeterized two-electron repulsion integrals (ERIs) in Mulliken notation, $(ia||jb) = (ia|jb) - (ib|ja)$, and Δ is the somewhat cumbersome energy-difference matrix. In an arbitrary nonorthogonal basis, Δ is an eighth-rank tensor describing the zeroth-order Hamiltonian and is defined as

$$\Delta_{iajb;i'a'j'b'} = F_{ii'}S_{aa'}S_{jj'}S_{bb'} - S_{ii'}F_{aa'}S_{jj'}S_{bb'} + S_{ii'}S_{aa'}F_{jj'}S_{bb'} - S_{ii'}S_{aa'}S_{jj'}F_{bb'} \tag{3}$$

where $F_{aa'}$ and $S_{aa'}$ are elements of the Fock and overlap matrices in the MO basis, respectively. If one is working in a basis of canonical MOs, Δ takes on the following simpler and more familiar form:

$$\Delta_{iajb;i'a'j'b'} = (\epsilon_i - \epsilon_a + \epsilon_j - \epsilon_b)\delta_{ii'}\delta_{aa'}\delta_{jj'}\delta_{bb'} \tag{4}$$

since both the Fock and overlap matrices are diagonal in this representation. Since Δ is diagonal in this representation, the formation of Δ^{-1} is trivial and the first-order amplitudes can easily be obtained via a straightforward evaluation of eq 2.

More generally, the expression for Δ in eq 3 can be simplified using the following matrix-based notation:

$$\Delta = \mathbf{F}_O \otimes \mathbf{S}_V \otimes \mathbf{S}_O \otimes \mathbf{S}_V - \mathbf{S}_O \otimes \mathbf{F}_V \otimes \mathbf{S}_O \otimes \mathbf{S}_V + \mathbf{S}_O \otimes \mathbf{S}_V \otimes \mathbf{F}_O \otimes \mathbf{S}_V - \mathbf{S}_O \otimes \mathbf{S}_V \otimes \mathbf{S}_O \otimes \mathbf{F}_V \tag{5}$$

where $\mathbf{F}_V$ and $\mathbf{F}_O$ ($\mathbf{S}_V$ and $\mathbf{S}_O$) are the MO representations of the Fock (overlap) matrix in the virtual and occupied subspaces, respectively. This expression can be further factorized into a compact direct-product form:

$$\Delta = \tilde{\mathbf{F}}^2 \otimes \tilde{\mathbf{S}}^2 + \tilde{\mathbf{S}}^2 \otimes \tilde{\mathbf{F}}^2 \tag{6}$$

using

$$\tilde{\mathbf{F}}^2 = \mathbf{F}_O\mathbf{S}_V - \mathbf{S}_O\mathbf{F}_V \tag{7}$$

and

$$\tilde{\mathbf{S}}^2 = \mathbf{S}_O\mathbf{S}_V. \tag{8}$$

From eqs 6−8, it is clear that Δ has inherent tensorial structure; its elements can be found as the direct product of two lower-ranked tensors in the mixed canonical occupied and virtual subspaces. Thus, the problem of inverting Δ (and the subsequent evaluation of eq 2 for the amplitudes) can be thought of as being equivalent to the diagonalization of these two lower-rank tensors, namely, $\tilde{\mathbf{F}}^2$ and $\tilde{\mathbf{S}}^2$.

If a local nonorthogonal basis is employed, Δ will not be diagonal and the aforementioned tensorial structure of Δ will be destroyed. Faced with the prohibitively large cost of inverting Δ, the first-order amplitudes must be obtained via an iterative solution of eq 2, as advocated by Saebø and Pulay.[22] However, the iterative solution of eq 2 comes with two fundamental complications. First is the need to store the amplitudes during the iterative procedure. This restriction makes awkward the extension of any local model to higher excitation levels, such as CCSD(T), where there may be an unmanageably large number of local triple amplitudes. Second, the use of redundant nonorthogonal functions followed by local truncations will undoubtedly lead to ill-conditioning in Δ and a myriad of associated convergence issues. Therefore, a viable noniterative solution to eq 2 is sought by retaining the tensorial structure of Δ and developing a methodology that finds a local representation in which it is diagonal.

C. Review of Atom-Centered Truncations. Local approximations to MP2 theory based on atomistic truncations can be characterized in terms of the subset of double substitutions retained in the local ansatz. For reference, the full MP2 correlation space, or the complete set of double substitutions, in an atom-centered basis can be represented in the following symbolic form:

$$\{iajb\}_{\mathrm{FULL}} = \{i\} \otimes \{a\} \otimes \{j\} \otimes \{b\}. \tag{9}$$

Substitutions generally occur from occupied levels centered on two atoms to virtual levels centered on two different atoms. Note that the full set of doubly excited determinants forms a fourth-rank tensor with direct-product structure—a form that enables the indices to be transformed one at a time during the integral transformation from the AO basis to the MO basis, as is done in most conventional MP2 algorithms.

The diatomics-in-molecules (DIM) model retains the following subset of double substitutions:

$$\{iajb\}_{\mathrm{DIM}} = \{(ia)\} \otimes \{(jb)\} \tag{10}$$

where an excitation (ia) occurs from an occupied level to a virtual level on the same atom—and will be referred to as

an atomic single substitution (ASIS). Overall direct-product structure is conserved in the DIM ansatz and is given as the product of the ASIS set with itself.

The TRIM model retains the following subset of double substitutions:

$$\{iajb\}_{\mathrm{TRIM}} = \{(ia)\} \otimes \{j\} \otimes \{b\} \oplus \{i\} \otimes \{a\} \otimes \{(jb)\}. \quad (11)$$

This model is an expansion of the DIM model by inclusion of single-electron transfers (or singly ionic substitutions). Double-electron transfers (or doubly ionic substitutions) are not included in the DIM or TRIM models—as they are often thought of as being the major constituent of basis set superposition error (BSSE).[38,39] Note that direct-product structure in this local ansatz is retained if we consider only half of the substitutions in eq 11. The resultant subset of double substitutions now constitutes the half-TRIM local ansatz.

If one chooses to employ models that retain direct-product structure in the ansatz, namely, the DIM and half-TRIM models described above, the tensorial structure of Δ is conserved and can be exploited during the inversion procedure. The goal now is to develop a formalism that finds a local representation in which Δ is diagonal and also incorporates the powerful RI approximation.

D. Review of the Noniterative Half-TRIM and TRIM Models. Starting with the half-TRIM local correlation ansatz (cf. eq 11)

$$\{iajb\}_{\mathrm{half-TRIM}} = \{(ia)\} \otimes \{j\} \otimes \{b\} \quad (12)$$

the direct-product structure of this model can be exploited using the recanonicalization scheme developed earlier.[24] Indexing the ASIS set by K, the general expression for the energy-difference matrix in eq 3 takes on the following form in the ASIS representation:

$$\Delta^{\mathrm{ASIS}}_{Kjb;K'j'b'} = \mathbf{D}^{\mathrm{ASIS}}_{KK'} S_{jj'} S_{bb'} + \mathbf{S}^{\mathrm{ASIS}}_{KK'} (F_{jj'} S_{bb'} - S_{jj'} F_{bb'}) \quad (13)$$

where $\mathbf{D}^{\mathrm{ASIS}}_{KK'}$ and $\mathbf{S}^{\mathrm{ASIS}}_{KK'}$, the MP2 energy matrix and overlap matrix in the ASIS representation, respectively, are given as

$$\mathbf{D}^{\mathrm{ASIS}}_{KK'} = (\mathbf{FS} - \mathbf{SF})_{KK'} = (\mathbf{B}^{\dagger}\mathbf{SPFPSB})_{ii'}(\mathbf{SQS})_{aa'} - (\mathbf{B}^{\dagger}\mathbf{SPSB})_{ii'}(\mathbf{SQFQS})_{aa'} \quad (14)$$

and

$$\mathbf{S}^{\mathrm{ASIS}}_{KK'} = (\mathbf{B}^{\dagger}\mathbf{SPSB})_{ii'}(\mathbf{SQS})_{aa'} \quad (15)$$

respectively. In eqs 14 and 15, **P**, **Q**, **F**, and **S** are matrix representations in the AO basis and **B** is the matrix defining the transformation from AOs to EPAOs. From here forward, we will use canonical MOs for j and b for computational simplicity. With this simplification, the energy-difference matrix in the ASIS representation in eq 13 takes on the following simpler form:

$$\Delta^{\mathrm{ASIS}}_{Kjb;K'j'b'} = \mathbf{D}^{\mathrm{ASIS}}_{KK'} \delta_{jj'} \delta_{bb'} + \mathbf{S}^{\mathrm{ASIS}}_{KK'} (\epsilon_j \delta_{bb'} - \delta_{jj'} \epsilon_b) \quad (16)$$

since both the Fock and overlap matrices are diagonal in the canonical representation. At this point, we have Δ as a sum of two separable terms, with a side length equal to the total number of ASIS pairs (N_K) given by

$$N_K = \sum_A N_A M_A \quad (17)$$

where N_A and M_A represent the number of local virtuals (atomic basis functions) and local occupieds (minimal basis functions or EPAOs) on atom A, respectively.

To further simplify this expression for Δ, we need to find a local representation in which both of the separable terms in eq 16 are diagonal. To do so, we proceed now in a manner analogous to the derivation of the canonical Hartree−Fock equations.[40] Symmetric orthogonalization (using $\mathbf{S}^{-1/2}_{\bar{K}\bar{K}}$) of the energy-difference matrix in the ASIS representation brings us into the orthogonalized ASIS, or OASIS, representation where Δ takes on the following form:

$$\Delta^{\mathrm{OASIS}}_{\bar{K}jb;\bar{K}'j'b'} = \mathbf{D}^{\mathrm{OASIS}}_{\bar{K}\bar{K}'} \delta_{jj'} \delta_{bb'} + \delta_{\bar{K}\bar{K}'} (\epsilon_j \delta_{bb'} - \delta_{jj'} \epsilon_b). \quad (18)$$

In eq 18, the second separable term (that involving the ASIS overlap matrix) has been brought into diagonal form and $\bar{K}$ has been introduced as the index of substitutions in the OASIS representation. Next, Δ is brought into the desired "canonical" OASIS representation (COASIS) via diagonalization of the MP2 energy matrix in the OASIS representation

$$\mathbf{D}^{\mathrm{OASIS}}_{\bar{K}\bar{K}'} = \sum_{\bar{P}} \mathbf{U}_{\bar{K}\bar{P}} \epsilon_{\bar{P}} \mathbf{U}_{\bar{P}\bar{K}'}. \quad (19)$$

This allows us to write Δ in its final form, in the mixed COASIS and canonical orbital representation

$$\Delta^{\mathrm{COASIS}}_{\bar{P}jb;\bar{P}'j'b'} = (\epsilon_{\bar{P}} + \epsilon_j - \epsilon_b) \delta_{\bar{P}\bar{P}'} \delta_{jj'} \delta_{bb'} \quad (20)$$

where $\bar{P}$ has been introduced as the index of substitutions in the COASIS representation. Now that Δ is diagonal, finding its inverse is trivial, allowing for a noniterative solution to eq 2 for the amplitudes. Knowledge of the amplitudes allows us to write down the following energy expression for the half-TRIM model (cf. eqs 1, 2, and 20):

$$E_{\mathrm{half-TRIM}} = -\frac{1}{2} \sum_{\bar{P}jb} \frac{\bar{V}_{\bar{P}jb} \bar{\bar{V}}_{\bar{P}jb}}{\epsilon_{\bar{P}} + \epsilon_j - \epsilon_b} \quad (21)$$

where $\bar{V}_{\bar{P}jb}$ and $\bar{\bar{V}}_{\bar{P}jb}$ are the appropriately transformed symmeterized and antisymmeterized ERIs in the COASIS representation [i.e., $\bar{V}_{\bar{P}jb} = (\bar{P}|jb)_C$ and $\bar{\bar{V}}_{\bar{P}jb} = (\bar{P}|jb)_C - (\bar{P}|jb)_E$] given as

$$V_{\bar{P}jb} = \sum_{\bar{K}K} \mathbf{U}_{\bar{P}\bar{K}} \mathbf{S}^{-1/2}_{\bar{K}K} V_{Kjb} = \sum_K \mathbf{T}_{\bar{P}K} V_{Kjb} \quad (22)$$

and **T** is a transformation matrix that brings quantities in the ASIS representation into the COASIS representation. (The subscripts C and E, above, refer to Coulomb and exchange integrals, respectively.)

Since the energy expression for the half-TRIM model includes a factor of $^1/_2$ for double-counting energy contributions and the half-TRIM ansatz retains only half of the TRIM subset of double substitutions, the TRIM energy should be double that of the half-TRIM energy. However, this overcounts the DIM contribution to the energy, so we have the

following expression as the working approximation to the final TRIM energy:

$$E_{\mathrm{TRIM}} = 2E_{\mathrm{half-TRIM}} - E_{\mathrm{DIM}}. \tag{23}$$

For completeness, the expression for the DIM energy in the COASIS representation is given as

$$E_{\mathrm{DIM}} = -\frac{1}{2}\sum_{\bar{P}\bar{Q}}\frac{\bar{V}_{\bar{P}\bar{Q}}\bar{\bar{V}}_{\bar{P}\bar{Q}}}{\epsilon_{\bar{P}} + \epsilon_{\bar{Q}}}. \tag{24}$$

At this point, we have developed a formalism that finds a local representation in which Δ is diagonal allowing for a noniterative solution of the MP2 energy; the only task that remains is the formal incorporation of the RI approximation into the theoretical framework of our local MP2 model.

E. Introduction of the RI Approximation. The RI approximation[17] involves the insertion of an approximate resolution-of-the-identity into the Hilbert space of two interacting charge densities, that is,

$$(ia|jb) = (ia|\hat{1}|jb) \approx \sum_{MN}(ia|M)(M|N)^{-1}(N|jb) \tag{25}$$

where the sum is over all members of a fitted auxiliary basis set (typically 2−4 times larger than its complementary AO basis). In other words, the use of the RI approximation involves replacement of four-centered ERIs with two- and three-centered ERIs. Under this approximation, the complete set of four-centered ERIs can now be constructed via a single matrix multiplication step:

$$(ia|jb) = \sum_{N}\mathbf{B}^{N}_{ia}\,\mathbf{B}^{N}_{jb} \tag{26}$$

where

$$\mathbf{B}^{N}_{ia} = \sum_{M}(ia|M)(M|N)^{-1/2} \tag{27}$$

is a modified coefficient matrix that includes the two- and three-centered ERIs in the auxiliary and mixed MO/auxiliary bases, respectively.

To incorporate the RI approximation into the TRIM model, all four-centered ERIs in the localized (DIM) or mixed canonical/localized (half-TRIM) bases must be expressed in terms of **B** matrices. In the expressions for $E_{\mathrm{half-TRIM}}$ (eq 21) and E_{DIM} (eq 24), one can identify four different types of four-centered ERIs, namely, $(\bar{P}|jb)_C$ and $(\bar{P}|jb)_E$ from the half-TRIM model and $(\bar{P}|\bar{Q})_C$ and $(\bar{P}|\bar{Q})_E$ from the DIM model. In the ASIS representation, these ERIs are given in more explicit detail as

$$(K|jb)_C = (\underline{i'a'}|jb), \tag{28}$$

$$(K|jb)_E = (\underline{i'}b|j\underline{a'}), \tag{29}$$

$$(K|L)_C = (\underline{i'a'}|\underline{j''b''}), \tag{30}$$

and

$$(K|L)_E = (\underline{i'b''}|\underline{j''a'}), \tag{31}$$

where underlined orbitals with the same number of primes are localized orbitals centered on the same atom (orbitals without primes and underlines are canonical MOs). To form these integrals within the RI approximation, the prescriptions given in eqs 26 and 27 can be used on eqs 28−31 to determine the corresponding set of **B** matrices that must first be constructed:

$$(K|jb)_C = (\underline{i'a'}|jb) = \sum_{N}\mathbf{B}^{N}_{\underline{i'a'}}\mathbf{B}^{N}_{jb} = \sum_{N}\mathbf{B}^{N}_{K}\mathbf{B}^{N}_{jb}, \tag{32}$$

$$(K|jb)_E = (\underline{i'}b|j\underline{a'}) = \sum_{N}\mathbf{B}^{N}_{\underline{i'}b}\mathbf{B}^{N}_{j\underline{a'}}, \tag{33}$$

$$(K|L)_C = (\underline{i'a'}|\underline{j''b''}) = \sum_{N}\mathbf{B}^{N}_{\underline{i'a'}}\mathbf{B}^{N}_{\underline{j''b''}} = \sum_{N}\mathbf{B}^{N}_{K}\mathbf{B}^{N}_{L}, \tag{34}$$

and

$$(K|L)_E = (\underline{i'b''}|\underline{j''a'}) = \sum_{N}\mathbf{B}^{N}_{\underline{i'b''}}\mathbf{B}^{N}_{\underline{j''a'}}. \tag{35}$$

Using eq 22, the integrals given in eqs 32−35 can be transformed into the COASIS representation, and the energy can then be directly evaluated using eqs 21, 23, and 24. Therefore, the central theme in the evaluation of the RI-TRIM MP2 energy is the computationally efficient formation and subsequent manipulation of the unique set of **B** matrices given by eqs 32−35. Now that we have formally incorporated the RI approximation into the theoretical framework of our local MP2 model, we now describe in some detail the computational algorithm that constructs and manipulates this set of **B** matrices necessary to evaluate E_{TRIM} in the next section.

III. Algorithm

An overview of the RI-TRIM MP2 algorithm is given in Chart 1. In each of the following subsections, the individual functions in Chart 1 will be discussed in greater detail.

A. Function 00: Self-Consistent Field Procedure and EPAO Algorithm. In Function 00, the standard self-consistent field (SCF) procedure is executed, generating the following set of matrices necessary during the post Hartree−Fock computation: the overlap (**S**), Fock (**F**), and density (**P**) matrices in the AO basis and the coefficient (**C**) and orbital energy (ϵ) matrices in the MO basis. Additionally, a separate algorithm that generates the EPAO coefficient (**B**) matrix is called immediately following the generation of a SCF. The computational cost associated with the SCF procedure formally scales cubically with system size, and the associated memory and storage requirements are minimal in the context of our RI-TRIM MP2 algorithm.

B. Functions 01−02: RI Overhead Routines. Function 01 is essentially one of two mandatory overhead routines for computations involving the RI approximation. In Function 01, the two-centered ERIs, $(M|N)$ are assembled using the Coulomb metric in the auxiliary basis. A singular value decomposition is then performed on $(M|N)$ to obtain the desired inverse square-root matrix necessary for constructing the various **B** matrices described earlier. The computational cost of forming $(M|N)^{-1/2}$ is cubic with respect to system size ($\sim X^3$ where X is the number of auxiliary basis functions) and is attributed entirely to the diagonalization of $(M|N)$.

Chart 1. A General Flowchart of the RI-TRIM MP2 Algorithm

Function 00 SCF Procedure; Formation of Overlap (S), MO Coefficient (C), Fock (F), Density (P), Orbital Energy (ε), and EPAO Coefficient (B) Matrices

Function 01 RI- Overhead (Part I): Formation of the $(M\,|\,N)^{-1/2}$ Matrix

Function 02 RI- Overhead (Part II): Formation of the $(ia\,|\,M)$ Matrix

Function 03 TRIM Overhead: Formation of V, O, and T Transformation Matrices and COASIS Energy $(\varepsilon_{\bar{P}})$ Matrix

Function 04 Formation of B Matrices (Part I)

Function 05 Formation of B Matrices (Part II)

Function 06 Construction of the DIM and TRIM Exchange Integrals and the B_K^N Matrix

Function 07 Canonicalization of the DIM Exchange Integrals and the B_K^N Matrix
Construction of the DIM Coulomb Integrals
Evaluation of the DIM Energy

Function 08 Canonicalization of the TRIM Exchange Integrals
Construction of the TRIM Coulomb Integrals
Evaluation of the TRIM Energy

The memory requirements associated with the diagonalization of $(M|N)$, namely, $3X^2$, also sets the minimum overall memory requirements of any algorithm that employs the RI approximation. In our RI-TRIM MP2 algorithm, we have decided on $5X^2$ as the overall memory requirement, which allows for treatment of system sizes on the order of a C_{90} alkane chain at the cc-pVDZ level with 2 GB of available memory. For future use, the $(M|N)^{-1/2}$ matrix is written to disk with quadratic storage requirements (X^2).

In Function 02, the remainder of the overhead associated with the RI approximation is completed as the three-centered ERIs, $(ia|M)$, are constructed in the mixed MO/auxiliary basis. This set of three-centered ERIs is formed in a two-step transformation procedure from the preassembled set of three-centered ERIs in the mixed AO/auxiliary basis $(\mu\nu|M)$. In the first transformation step, $\mu \rightarrow i$ via

$$(i\nu|M) = \sum_{\mu} C_{\mu i}(\mu\nu|M) \tag{36}$$

with an associated quartic computational cost of $\sim O(nb2)X$, where $(nb2)$ is the number of significant AO basis set pairs (determined by the chosen integral threshold value). In the second transformation step, $\nu \rightarrow a$ via

$$(ia|M) = \sum_{\nu} C_{\nu a}(i\nu|M) \tag{37}$$

with an associated quartic computational cost of $\sim NOVX$. The set of $(ia|M)$ is also written to disk with cubic storage requirements (OVX) for use in later functions.

C. Function 03: TRIM Overhead Routine. The use of localized MOs in MP2 theory via utilization of the TRIM ansatz also comes with an associated overhead, all of which is managed in Function 03. In this function, the MP2 energy $(\mathbf{D}_{KK'}^{\mathrm{ASIS}})$ and overlap $(\mathbf{S}_{KK'}^{\mathrm{ASIS}})$ matrices in the ASIS representation are constructed according to the prescriptions given by eqs 14 and 15, respectively. A singular value decomposition is then performed on $\mathbf{S}_{KK'}^{\mathrm{ASIS}}$ to obtain the desired inverse square-root matrix necessary for symmetric orthogonalization of the ASIS representation. The computational cost of forming $\mathbf{S}_{\bar{K}\bar{K}}^{-1/2}$ is cubic with respect to system size $(\sim N_K^3)$, where N_K is the number of ASIS pairs given by eq 17, a quantity that grows linearly with system size. In practice, $N_K \approx X$, so the computational cost of diagonalizing $\mathbf{S}_{KK'}^{\mathrm{ASIS}}$ can directly be compared to that of diagonalizing $(M|N)$ in Function 01. After $\mathbf{D}_{KK'}^{\mathrm{ASIS}}$ is brought into the OASIS representation $(\mathbf{D}_{\bar{K}\bar{K}'}^{\mathrm{OASIS}})$ by both left and right matrix multiplication with $\mathbf{S}_{\bar{K}\bar{K}}^{-1/2}$, it is brought into diagonal form by another diagonalization step (total computational cost of $3N_K^3$ steps). The eigenvector matrix $\mathbf{U}_{\bar{P}\bar{K}}$ which diagonalizes $\mathbf{D}_{\bar{K}\bar{K}'}^{\mathrm{OASIS}}$ is multiplied on the right by $\mathbf{S}_{\bar{K}\bar{K}}^{-1/2}$ to form the desired $\mathbf{T}$ transformation matrix, which brings quantities in the ASIS representation into the COASIS representation, with a cubic computational cost of N_K^3. The corresponding eigenvalue matrix $\epsilon_{\bar{P}}$ provides us with the energies associated with each substitution in the COASIS representation and is used directly in the energy evaluation. Both of these quantities, $\mathbf{T}$ and $\epsilon_{\bar{P}}$, are written to disk for future use with an associated quadratic storage requirement $(\sim N_K^2)$. In addition, two other transformation matrices are necessary for computations involving localized MOs: the $\mathbf{V}$ and $\mathbf{O}$ matrices, which transform canonical MOs to localized virtual and occupied orbitals, respectively. From our definitions of localized virtual and occupied functions given in Part A of Section II, it follows that

$$\mathbf{V} = C_V^{\dagger}\mathbf{S} \tag{38}$$

and

$$\mathbf{O} = C_O^{\dagger}\mathbf{SB} \tag{39}$$

Chart 2. A More Detailed Look at the Algorithm in Function 04. Array and Disk (HD) Ordering Is Given with Index Speeds Decreasing from Left to Right within Square Brackets

Load $(M|N)^{-1/2}$, V, and O matrices from disk

Loop over occupied batches, ob

Loop over $i \in ob$

Load $(ia\|M)\ \forall\ a, M, given\ i$ from disk	$[a, M]$
Make $B_{ia}^{N} = \sum_M (ia\|M)(M\|N)^{-1/2}\ \forall\ a, N, given\ i$	$[a, N]$
Make $\tilde{B}_{ia}^{N} = (B_{ia}^{N})^{\dagger}\ \forall\ a, N, given\ i$	$[N, a]$
Accumulate $B_{ia}^{N}\ \forall\ a, N, i \in ob$	$[a, N, i \in ob]$
Accumulate $\tilde{B}_{ia}^{N}\ \forall\ a, N, i \in ob$	$[N, a, i \in ob]$

End Loop over $i \in ob$

Make $B_{i\underline{a}}^{N} = \sum_a B_{ia}^{N} V_{a\underline{a}}\ \forall\ \underline{a}, N, i \in ob$	$[N, i \in ob, \underline{a}]$
Write $B_{ia}^{N}\ \forall\ a, N, i \in ob$ to disk	HD: $[a, N, i]$

Loop over atoms, A

Extract $B_{i\underline{a}}^{N}\ \forall\ \underline{a} \in A, N, i \in ob$	$[N, i \in ob, \underline{a} \in A]$
Write $B_{i\underline{a}}^{N}\ \forall\ \underline{a} \in A, N, i \in ob$ to disk	HD: $[N, i \in ob, \underline{a} \in A, ob, A]$

End Loop over atoms, A

Loop over virtual batches, vb

Extract $\tilde{B}_{ia}^{N}\ \forall\ a \in vb, N, i \in ob$	$[N, a \in vb, i \in ob]$
Write $\tilde{B}_{ia}^{N}\ \forall\ a \in vb, N, i \in ob$ to disk	HD: $[N, a \in vb, i, vb]$

End Loop over virtual batches, vb

End Loop over occupied batches, ob

where C_V and C_O are the virtual and occupied blocks of the MO coefficient matrix, **S** is the overlap matrix in the AO basis, and **B** is the EPAO coefficient matrix. These quantities are constructed with an associated cubic computational cost ($\sim N^2V$ for **V** and $\sim N^2O + MN^2$ for **O**) and are written to disk with quadratic storage requirements for later use in the construction of the modified RI coefficient matrices.

D. Functions 04−06: Construction of B Matrices and Exchange Integrals. Function 04 is the first of three functions that assembles and stores the various **B** matrices necessary for constructing the four-centered ERIs. The manner in which we proceed uses a fairly involved batching scheme to carefully balance CPU and I/O timings under the working memory constraint of $5X^2$, cubic I/O and disk requirements, and quartic computational cost. Without this batching scheme, our algorithm was susceptible to dominance by I/O time stemming from an unmanageable number of hard drive seeks. Given that the $(ia|M)$ matrix is written to disk with the overall order $[a, M, i]$, that is, a is the fastest index and i is the slowest index, we suggest the algorithm depicted in Chart 2.

The batch sizes involved in the loops over occupied and virtual batches are determined by the overall memory constraint of $5X^2$—a quantity unique to the molecular system being considered. There are two matrix multiplication steps in Function 04: the formation of the canonical $\mathbf{B}_{ia}^{N}$ matrix and the mixed canonical occupied/local virtual $\mathbf{B}_{i\underline{a}}^{N}$ matrix with associated computational costs of $\sim OVX^2$ and $\sim OVNX$, respectively. As shown in eqs 32 and 33, these matrices will be directly used in the construction of the TRIM Coulomb and TRIM exchange integrals. In the next function, the transposed canonical **B** matrix, $\tilde{\mathbf{B}}_{ia}^{N}$, now in the proper order on disk, will be transformed into the mixed local occupied/canonical virtual $\mathbf{B}_{i\underline{a}}^{N}$ matrix via matrix multiplication with **O**. In Function 04, there are associated cubic storage requirements ($2OVX$ and ONX) for writing both canonical **B** matrices and the mixed canonical occupied/local virtual **B** matrix to disk.

In Function 05, another **B** matrix is constructed, namely, the mixed local occupied/canonical virtual $\mathbf{B}_{i\underline{a}}^{N}$ matrix, from the transposed canonical $\tilde{\mathbf{B}}_{ia}^{N}$ matrix formed and written to disk earlier in Function 04 (Chart 3). This transformation step involves a single matrix multiplication with an associated quartic computational cost of $\sim OVMX$. As was done previously in Function 04 with $\mathbf{B}_{ia}^{N}$, the localized index in $\mathbf{B}_{i\underline{a}}^{N}$ is also grouped by atom and written selectively to disk with cubic storage requirements (MVX). This atomistic grouping will allow us to easily construct the four-centered ERIs given in eqs 28−31, where atom-based selection is crucial. From eq 33, it is clear that the $\mathbf{B}_{i\underline{a}}^{N}$ matrix will be directly used in the construction of the TRIM exchange integrals.

Function 06 is the final routine that deals with assembling the unique set of **B** matrices necessary to evaluate E_{TRIM}. In this function (Chart 4), our previous array reordering and selective disk writing is exploited as we construct two

Chart 3. A More Detailed Look at the Algorithm in Function 05

Loop over virtual batches, vb	
Load $\widetilde{B}_{ia}^{N}\ \forall\ a \in vb, N, i$ from disk	$[N, a \in vb, i]$
Loop over $a \in vb$	
Extract $\widetilde{B}_{ia}^{N}\ \forall\ N, i, given\ a$	$[N, i]$
Make $B_{i\underline{a}}^{N} = \sum_i \widetilde{B}_{ia}^{N} O_{i\underline{i}}\ \forall\ \underline{i}, N, given\ a$	$[N, \underline{i}]$
Accumulate $B_{i\underline{a}}^{N}\ \forall\ N, \underline{i}, a \in vb$	$[N, \underline{i}, a \in vb]$
End Loop over $a \in vb$	
Loop over atoms, A	
Extract $B_{i\underline{a}}^{N}\ \forall\ N, \underline{i} \in A, a \in vb$	$[N, \underline{i} \in A, a \in vb]$
Write $B_{i\underline{a}}^{N}\ \forall\ N, \underline{i} \in A, a \in vb$ to disk	HD: $[N, \underline{i} \in A, a, A]$
End Loop over atoms, A	
End Loop over virtual batches, vb	

differently ordered local $\mathbf{B}_{ia}^{N}$ matrices in a fairly complex loop over atoms. These $\mathbf{B}_{i\underline{a}}^{N}$ matrices are not only used to directly assemble the DIM exchange integrals via eq 35 but also contain the $\mathbf{B}_{K}^{N}$ matrix necessary for the construction of the TRIM and DIM Coulomb integrals (cf. eqs 32 and 34). This routine contains four matrix multiplication steps and their associated computational costs: the formation of $\mathbf{B}_{i\underline{a}}^{N}$ from both $\mathbf{B}_{i\underline{a}}^{N}$ ($\sim MNVX$) and $\mathbf{B}_{ia}^{N}$ ($\sim MNOX$) as well as the construction of both the TRIM ($\sim N_K OVX$) and DIM ($\sim N_K^2 X$) exchange integrals. For later retrieval, the $\mathbf{B}_{K}^{N}$ matrix and the set of TRIM exchange integrals, $(K|jb)_E$, are written to disk with associated $N_K X$ and $N_K OV$ storage requirements, respectively. Note that the recently assembled set of DIM exchange integrals is held in memory as it will be used next in Function 07 in the evaluation of E_{DIM}.

E. Functions 07−08: Construction of Coulomb Integrals, Canonicalization, and Energy Evaluation. In Function 07, the transformation matrix, **T**, is read from the disk and used to bring both the DIM exchange integrals and the $\mathbf{B}_{K}^{N}$ matrix from the ASIS representation into the COASIS representation via the prescription given in eq 22. The computational cost associated with bringing the DIM exchange integrals into the COASIS representation is cubic ($2N_K^3$) and is attributed to the left- and right-hand matrix multiplication of $(K|L)_E$ with **T**. We can circumvent multiple *ASIS* $|\rightarrow$ *COASIS* integral transformations by direct canonicalization of the $\mathbf{B}_{K}^{N}$ matrix, which can be directly used in the assembly of the DIM and TRIM Coulomb integrals in the COASIS representation. This transformation (i.e., $\mathbf{B}_{K}^{N}$ $|\rightarrow \mathbf{B}_{\bar{P}}^{N}$) has a cubic computational cost of $\sim N_K^2 X$, and $\mathbf{B}_{P}^{N}$ is easily written to disk with a minimal quadratic storage requirement ($N_K X$). The $\mathbf{B}_{P}^{N}$ matrix is also used to construct the DIM Coulomb integrals via eq 34 with cubic computational effort ($\sim N_K^2 X$). With both the DIM Coulomb and exchange integrals in memory, E_{DIM} is evaluated with negligible computational cost using eq 24.

The RI-TRIM MP2 algorithm is then completed in Function 08, as $E_{\mathrm{half-TRIM}}$ is incremented in a loop over occupied batches in which the TRIM exchange integrals are read from disk and brought into the mixed COASIS and canonical representation as the TRIM Coulomb integrals are assembled via matrix multiplication of $\mathbf{B}_{\bar{P}}^{N}$ and $\mathbf{B}_{ia}^{N}$. The computational costs associated with these steps are both quartic, namely, $\sim N_K^2 OV$ for the canonicalization of the TRIM exchange integrals and $\sim N_K OVX$ for the construction of the TRIM Coulomb integrals.

F. Summary of Computational Requirements. As mentioned earlier, the RI-TRIM MP2 algorithm was written under the following constraints: quadratic memory not to exceed $5X^2$, cubic I/O and storage requirements, and quartic computational cost. Within the eight functions described above, there are nine steps with quartic-associated computational cost and six different quantities are written to disk with cubic storage requirements. The dominant computational steps are attributed to the formation of the canonical $\mathbf{B}_{ia}^{N}$ matrix ($\sim OVX^2$) in Function 04, and the formation of the TRIM exchange and TRIM Coulomb integrals [both ($\sim N_K OVX$)] in Function 06 and Function 08, respectively. Overall savings with respect to RI-MP2, which scales asymptotically with the fifth-power of the system size [the quintic computational cost is $\sim O^2V^2X$ attributed to the construction of the full set of $(ia|jb)$], should be proportional to the number of atoms in the systems under investigation. By introducing the local TRIM ansatz into RI-MP2 theory, we have essentially replaced the aforementioned quintic computational step with six quartic computational steps necessary to build the set of **B** matrices and integrals necessary for the evaluation of the RI-TRIM MP2 energy.

IV. Performance of the Model

One interesting challenge that remains in modern theoretical chemistry is the accurate prediction of the relative energies of a series of polypeptide conformations. More often than not, relative energies among polypeptide conformations are on the order of 1 kcal/mol (or less) and are highly dependent on correlation energy corrections arising from post-HF calculations.[41] In this section, the numerical performance and computational timings of the RI-TRIM MP2 methodology will be assessed against MP2, TRIM MP2, RI-MP2, and DFT (B3LYP) in the prediction of the relative orderings of 27 different alanine tetrapeptide conformations. What one finds

Chart 4. A More Detailed Look at the Algorithm in Function 06

Loop over atoms, A	
Load $B_{\underline{i}b}^{N}\ \forall\ N,\underline{i}\in A,b$ from disk	$[N,\underline{i}\in A,b]$
Loop over $\underline{i}\in A$	
Extract $B_{\underline{i}b}^{N}\ \forall\ N,b,given\ \underline{i}$	$[N,b]$
Make $B_{\underline{i}\underline{b}}^{N}=\sum_b B_{\underline{i}b}^{N}V_{b\underline{b}}\ \forall\ N,\underline{b},\underline{i}\in A$	$[N,\underline{b}]$
Accumulate $B_{\underline{i}b}^{N}\ \forall\ N,b,\underline{i}\in A$	$[N,b,\underline{i}\in A]$
Accumulate $B_{\underline{i}\underline{b}}^{N}\ \forall\ N,\underline{b},\underline{i}\in A$	$[N,\underline{b},\underline{i}\in A]$
End Loop over $\underline{i}\in A$	
Load $B_{j\underline{a}}^{N}\ \forall\ N,\underline{a}\in A,j$ from disk	$[N,j\in ob,\underline{a}\in A,ob]$
Loop over $\underline{a}\in A$	
Extract $B_{j\underline{a}}^{N}\ \forall\ N,j,given\ \underline{a}$	$[N,j]$
Make $B_{\underline{j}\underline{a}}^{N}=\sum_a B_{j\underline{a}}^{N}O_{j\underline{j}}\ \forall\ N,\underline{j},\underline{a}\in A$	$[N,\underline{j}]$
Accumulate $B_{j\underline{a}}^{N}\ \forall\ N,j,\underline{a}\in A$	$[N,j,\underline{a}\in A]$
Accumulate $B_{\underline{j}\underline{a}}^{N}\ \forall\ N,\underline{j},\underline{a}\in A$	$[N,\underline{j},\underline{a}\in A]$
End Loop over $\underline{a}\in A$	
Extract $B_{K}^{N}\ \forall\ N,K\in A$	$[N,P\in A]$
Write $B_{K}^{N}\ \forall\ N,K\in A$ to disk	HD: $[N,P]$
Make $(K\mid jb)_E=(\underline{i'}b\mid j\underline{a'})=\sum_N B_{\underline{i'}b}^{N}B_{j\underline{a'}}^{N}\ \forall\ \underline{i}\in A,\underline{a}\in A,b,j$	$[b,\underline{i}\in A,j,\underline{a}\in A]$
Write $(K\mid jb)_E\ \forall\ K\in A,b,j$ to disk in occupied batches	HD: $[b,P\in A,j\in ob,A,ob]$
Loop over atoms, B	
Extract $B_{\underline{j}\underline{a}}^{N}\ \forall\ N,\underline{j}\in B,\underline{a}\in A$	$[N,\underline{j}\in B,\underline{a}\in A]$
Extract $B_{\underline{i}\underline{b}}^{N}\ \forall\ N,\underline{b}\in B,\underline{i}\in A$	$[N,\underline{b}\in B,\underline{i}\in A]$
Make $(K\mid L)_E=(\underline{i'b''}\mid \underline{j''a'})=\sum_N B_{\underline{i'b''}}^{N}B_{\underline{j''a'}}^{N}\ \forall\ \underline{i}\,\&\,\underline{a}\in A,\underline{j}\,\&\,\underline{b}\in B$	$[\underline{j}\in B,\underline{a}\in A,\underline{b}\in B,\underline{i}\in A]$
End Loop over atoms, B	
End Loop over atoms, A	

is that the aforementioned challenge is directly addressed by RI-TRIM MP2, a method that not only accurately provides the relative energies of these conformations but does so with a computational timing cost that is negligible with respect to the iterative procedures necessary in both the SCF and DFT formalisms.

All calculations discussed in this work were performed on a 2 GHz Apple XServe with 2 GB of available memory and a triply striped 7200 rpm hard drive with 100 GB of available disk space. The RI-TRIM MP2 code was implemented into a development version of Q-Chem 2.0[42] and was used for all calculations. The geometries of the alanine tetrapeptide conformations were optimized at the HF/6-31G** level of theory[41] and are provided in the Supporting Information. In all post-HF computations performed, the frozen-core approximation was introduced to reduce computational timings and resources. In all cases, self-consistency was determined by DIIS errors below 10^{-8} au and the numerical threshold utilized in all integral evaluations was set at 10^{-12} au. All calculations were performed utilizing the correlation consistent basis sets (cc-pVXZ where X = D, T, and Q) of Dunning[43] and their respective complementary auxiliary basis sets provided by Weigend et al.[17,44]

A. Chemical Tests. For a series of 27 alanine tetrapeptide conformations, single-point energy calculations were first performed using the MP2, TRIM MP2, RI-MP2, RI-TRIM MP2, and DFT methodologies and the cc-pVDZ and cc-pVTZ basis sets. From this set of calculations, relative conformational energies were computed and are presented in Table 1.

A fairly simple statistical analysis of this set of data (Table 2) yields several quantities worth further discussion. A direct comparison of RI-MP2 with MP2 yields the relative error introduced by the RI approximation. The root-mean-square (RMS) error using both the cc-pVDZ and cc-pVTZ basis sets was computed as 0.003 kcal/mol and is completely negligible with respect to the relative conformational energies given in Table 1. By comparing RI-TRIM MP2 to TRIM

Table 1. Relative Energies (in kcal/mol) for a Series of 27 Different Alanine Tetrapeptide Conformations Using the cc-pVDZ (first column) and cc-pVTZ (second column) Basis Sets (and Their Respective Complementary Auxiliary Basis Sets where Applicable)[a]

conformation	MP2		RI-MP2		TRIM MP2		RI-TRIM MP2		DFT	
1	6.635	5.075	6.637	5.080	6.203	4.722	6.202	4.723	3.175	0.994
2	6.045	4.927	6.047	4.931	5.684	4.614	5.684	4.615	2.904	1.230
3	0.293	0.511	0.294	0.512	0.235	0.449	0.237	0.449	0.000	0.681
4	7.311	6.345	7.313	6.349	6.953	6.025	6.952	6.026	5.126	3.716
5	7.543	5.989	7.543	5.993	7.215	5.717	7.211	5.718	5.441	3.398
6	2.994	2.960	2.994	2.962	2.859	2.839	2.858	2.839	2.681	2.756
7	6.575	6.680	6.573	6.682	6.471	6.570	6.469	6.569	7.445	7.490
8	4.802	4.466	4.798	4.467	4.742	4.461	4.739	4.461	5.852	5.935
9	9.282	8.153	9.288	8.156	9.058	8.040	9.063	8.041	6.756	6.119
10	7.860	7.781	7.859	7.781	7.721	7.726	7.721	7.725	7.716	8.061
11	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.016	1.626
12	1.158	0.467	1.155	0.466	1.188	0.508	1.186	0.507	2.878	2.788
13	4.706	4.122	4.710	4.125	4.441	3.880	4.442	3.881	2.249	1.210
14	5.983	5.116	5.980	5.118	5.766	4.934	5.761	4.933	4.962	3.716
15	1.708	1.802	1.701	1.800	1.738	1.879	1.733	1.877	4.075	4.641
16	4.701	3.669	4.699	3.670	4.590	3.589	4.588	3.589	4.566	3.695
17	2.801	3.236	2.800	3.236	2.753	3.185	2.753	3.185	2.728	4.007
18	2.182	2.115	2.187	2.118	1.995	1.948	1.998	1.949	0.053	0.000
19	4.630	4.084	4.636	4.087	4.428	3.933	4.432	3.935	2.482	1.985
20	2.785	2.118	2.784	2.120	2.639	2.003	2.637	2.004	2.011	1.188
21	2.724	2.824	2.722	2.827	2.608	2.758	2.605	2.759	2.771	3.041
22	5.535	5.791	5.532	5.791	5.391	5.696	5.389	5.695	5.396	5.893
23	5.972	5.867	5.972	5.868	5.840	5.775	5.839	5.774	5.823	6.006
24	4.012	3.952	4.011	3.953	3.895	3.863	3.894	3.862	4.088	4.148
25	3.578	2.894	3.579	2.897	3.435	2.779	3.434	2.780	3.032	2.431
26	2.046	1.103	2.044	1.105	1.988	1.064	1.984	1.064	2.592	1.779
27	5.621	4.416	5.622	4.419	5.477	4.336	5.476	4.336	5.228	4.079

[a] The number of AO (auxiliary) basis functions employed in the cc-pVDZ and cc-pVTZ calculations was 390 (1428) and 908 (2260), respectively.

Table 2. RMS (Maximum) Errors among Various Different Methodologies (in kcal/mol) Using the cc-pVDZ and cc-pVTZ Basis Sets

	cc-pVDZ	cc-pVTZ
RI-MP2 vs MP2	0.003 (0.007)	0.003 (0.005)
RI-TRIM vs TRIM MP2	0.003 (0.006)	0.001 (0.002)
TRIM MP2 vs MP2	0.192 (0.432)	0.157 (0.353)
RI-TRIM MP2 vs RI-MP2	0.193 (0.435)	0.159 (0.357)
DFT vs RI-MP2	1.574 (3.463)	1.816 (4.086)
DFT vs RI-TRIM MP2	1.420 (3.027)	1.688 (3.729)

MP2, we are now in a position to numerically assess the magnitude of the error introduced by the RI approximation within the TRIM framework. From Table 2, it is clear that RI-TRIM MP2 does so with exceptional consistency with computed RMS errors of 0.003 kcal/mol (cc-pVDZ) and 0.001 kcal/mol (cc-pVTZ). On the basis of the relative magnitude of this error, it is safe to conclude that, within the context of determining the relative energies of these polypeptides, the data obtained from calculations utilizing the RI approximation are indistinguishable from those obtained without using the RI approximation. This allows us the powerful option of pursuing the relative energies of this set of polypeptides at the complete basis set limit by utilizing the higher-quality cc-pVQZ basis set. At the cc-pVQZ level, TRIM MP2 and MP2 calculations on systems of this size are prohibitively expensive and can now be replaced by their computationally efficient RI analogues with no associated loss in accuracy.

Comparing the TRIM MP2 and MP2 data sets allows for quantification of the local error, that is, the error attributed to truncation of the MP2 correlation space introduced by the TRIM model. The RMS local error was computed as 0.192 and 0.157 kcal/mol at the cc-pVDZ and cc-pVTZ levels, respectively. The analogous comparison of RI-TRIM MP2 to RI-MP2 yields a consistent measure of the local error, namely, 0.193 kcal/mol (cc-pVDZ) and 0.159 kcal/mol (cc-pVTZ). Although the magnitude of the local error decreased as the quality of the basis set was increased, the local error is significant enough to introduce relative ordering shifts among close-lying conformers when the cc-pVDZ and cc-pVTZ basis sets are employed. This finding further justifies the need to explore the relative energetics of this system of conformers at the now possible cc-pVQZ level of theory. We would also like to point out that the incorporation of the RI approximation into the theoretical framework of TRIM MP2 does not introduce any additional error; that is, there are no additive or compound errors arising from the simultaneous use of these approximations.

Another interesting quantity of note is the assessment of the popular B3LYP[45,46] functional against the MP2 methods examined in this work. The role of dispersion interactions in the relative energies of polypeptide conformations has been found to be significant in a recent work by Yu et al.[47] In

fact, the electron correlation effects obtained at the MP2 level were found to be critical in determining the binding energies of dipeptide conformations; in treating alanine dipeptide, for example, DFT is unable to predict the formation of the very stable peptide bond.[48] In the case of alanine and glycine dipeptides, it was also shown that the inclusion of higher-order correlation effects [i.e., from CCSD(T)] was not significant with respect to the MP2 results.[48] On the basis of these findings, we feel justified in using the present MP2 results as the benchmark for comparison in this work. We, therefore, conclude that the RMS error computed by comparing DFT to RI-MP2 at the cc-pVDZ level, 1.574 kcal/mol, is completely unacceptable for accurately predicting the relative conformational energies within a biochemical context. In fact, this error increases to 1.816 kcal/mol using the higher-quality cc-pVTZ basis set! This trend is consistently reproduced in the comparison of DFT to RI-TRIM MP2, where the RMS errors were found as 1.420 and 1.688 kcal/mol, at the cc-pVDZ and cc-pVTZ levels, respectively. To put this error analysis into a more global perspective, our findings clearly show that the RMS DFT error is at least 7 times larger than the error introduced by using localized MOs within the MP2 framework.

At the cc-pVQZ level, single-point energy calculations were performed using the RI-MP2, RI-TRIM MP2, and DFT methodologies for the same series of 27 alanine tetrapeptide conformations. From this set of calculations, relative conformational energies were computed and are presented in Table 3. The RMS (max) local error at this level was computed as 0.090 (0.195) kcal/mol; again we observe a substantial decrease in the magnitude of this error with the use of a higher-quality basis set. Although the magnitude of this error is still large enough to occasionally introduce discrepancies in relative orderings (at the cc-pVQZ level, RI-MP2 and RI-TRIM MP2 only disagree once with respect to relative ordering among these 27 different conformations), we emphasize the achievement of error that is less than 0.1 kcal/mol. The RMS (max) DFT errors with respect to RI-MP2 and RI-TRIM MP2 were found as 1.733 (3.212) and 1.674 (3.168) kcal/mol, respectively. Even at the cc-pVQZ level, DFT fails to perform at an acceptable level for accurately treating the relative energetics within this system of polypeptides conformers and will no longer be discussed. These findings provide another instance where DFT fails to converge as rapidly as expected with the use of high-quality correlation consistent basis sets, as was also found when investigating heats of formation for the hydrogenation of N_2[49] and in the determination of atomization energies.[50]

Since all calculations were performed utilizing the correlation consistent basis sets (cc-pVXZ where X = D, T, and Q) of Dunning,[43] we now have the option of utilizing the following prescription:[13]

$$E_{XY} = E_{\mathrm{SCF},Y} + \frac{X^3 E_{\mathrm{CORR},X} - Y^3 E_{\mathrm{CORR},Y}}{X^3 - Y^3} \qquad Y > X \quad (40)$$

to yield the cc-pV(DT)Z and cc-pV(TQ)Z extrapolated data sets (provided in Table 4). At the cc-pV(DT)Z level, the RMS (max) local error introduced by RI-TRIM MP2 has been reduced to 0.146 (0.324) kcal/mol—a modest (~10%) improvement from the RMS error at the cc-pVTZ level, 0.159 kcal/mol. As the complete basis set limit is approached at the cc-pV(TQ)Z level,[13] the RMS (max) error is significantly reduced to 0.040 (0.082) kcal/mol. It should be noted that the magnitude of this error, just slightly below *kT* at room temperature (~0.5 kcal/mol), is not likely to be significant within a biochemical context. Just by using eq 40, we were able to reduce the error introduced by RI-TRIM MP2 by a factor of 2.5 over the error at the cc-pVQZ level. At this level, there are no discrepancies in the RI-TRIM MP2 and RI-MP2 predictions of the relative energy orderings within this series—the first computational evidence that the TRIM model approaches exactness as the complete basis set limit is approached.

At this point, we are also in a position to numerically address the claim that local MP2 methods, which do not include the doubly ionic excitations partly responsible for BSSE, are more accurate than full MP2 when smaller basis sets are employed.[38,39] Taking the RI-MP2/cc-pV(TQ)Z data set as our "best numbers", a direct comparison with MP2/cc-pVDZ and TRIM MP2/cc-pVDZ yields RMS (max) deviations of 1.096 (2.504) and 0.939 (2.072) kcal/mol, respectively. At the cc-pVTZ level, the MP2 and TRIM MP2 methods yield RMS (max) deviations of 0.377 (0.944) and 0.250 (0.591) kcal/mol, respectively. On the basis of these

Table 3. Relative Energies (in kcal/mol) for a Series of 27 Different Alanine Tetrapeptide Conformations Using the cc-pVQZ Basis Set (and the Complementary Auxiliary Basis Set where Applicable)[a]

conformation	RI-MP2	RI-TRIM MP2	DFT
1	4.236	4.042	0.742
2	4.231	4.053	1.056
3	0.534	0.498	1.251
4	5.729	5.545	3.350
5	5.323	5.170	3.263
6	2.879	2.813	3.228
7	6.590	6.522	7.730
8	4.538	4.528	6.333
9	7.859	7.802	6.264
10	7.725	7.707	8.367
11	0.000	0.000	2.184
12	0.340	0.372	3.266
13	3.618	3.474	1.225
14	4.686	4.583	3.862
15	2.080	2.124	5.291
16	3.511	3.462	3.925
17	3.366	3.336	4.638
18	1.811	1.704	0.000
19	3.739	3.648	2.033
20	1.730	1.663	1.270
21	2.802	2.754	3.336
22	5.756	5.708	6.318
23	5.762	5.714	6.335
24	3.899	3.849	4.521
25	2.493	2.430	2.442
26	0.706	0.689	1.806
27	4.014	3.972	4.155

[a] The number of AO (auxiliary) basis functions employed in the cc-pVQZ calculations was 1760 (3850).

Table 4. Relative Energies (in kcal/mol) for a Series of 27 Different Alanine Tetrapeptide Conformations Based on the Extrapolated cc-pV(DT)Z (First Column) and cc-pV(TQ)Z (Second Column) Data Sets

conformation	RI-MP2		RI-TRIM MP2	
1	5.553	4.130	5.229	4.055
2	5.390	4.190	5.095	4.113
3	0.693	0.571	0.628	0.554
4	6.884	5.730	6.577	5.648
5	6.353	5.261	6.101	5.197
6	3.148	2.895	3.031	2.871
7	6.931	6.665	6.816	6.628
8	4.477	4.640	4.492	4.626
9	8.197	7.923	8.129	7.907
10	7.766	7.791	7.745	7.800
11	0.000	0.000	0.000	0.000
12	0.346	0.290	0.391	0.315
13	4.538	3.655	4.304	3.584
14	5.395	4.679	5.224	4.637
15	1.707	2.194	1.803	2.213
16	3.784	3.549	3.715	3.523
17	3.414	3.418	3.360	3.404
18	2.468	1.913	2.308	1.852
19	4.319	3.816	4.188	3.770
20	2.335	1.755	2.231	1.724
21	3.009	2.918	2.963	2.883
22	5.899	5.824	5.823	5.812
23	5.979	5.815	5.901	5.800
24	4.115	3.976	4.036	3.955
25	3.118	2.495	3.012	2.471
26	1.223	0.674	1.189	0.675
27	4.513	4.020	4.456	4.008

deviations, we can report that the local MP2 is only slightly more accurate (∼15%) than full MP2 at the cc-pVDZ level but substantially more accurate (∼34%) at the cc-pVTZ level; any further comment on the relationship between this improvement and BSSE without performing a series of counterpoise-corrected calculations on these conformers would be speculation. For comparison, we present the same error analysis using the LMP2/cc-pVTZ(-f) results of Friesner et al.,[41] which provides a RMS (max) error of 0.865 (1.690) kcal/mol [relative conformational energies at the LMP2/cc-pVTZ(-f) level are provided in Table S4 of the Supporting Information].

We find the basis set dependence of the relative conformational energies at the RI-MP2 and RI-TRIM MP2 levels surprising. Relative conformational energies are usually assumed to be fairly independent of the basis set because of the likelihood of error cancellation. We had performed the cc-pVQZ level calculations and the associated T → Q extrapolation largely for completeness—to demonstrate convergence in our new methodology. However, the RMS (max) deviations from RI-MP2/cc-pV(TQ)Z found as 0.377 (0.944) kcal/mol (MP2/cc-pVTZ) and 0.250 (0.591) kcal/mol (TRIM MP2/cc-pVTZ) are chemically significant when compared against the conformer energy differences. It therefore appears that this level of effort, that is, computations performed systematically at the cc-pVTZ and cc-pVQZ levels followed by extrapolation to cc-pV(TQ)Z, is necessary to obtain reliably converged relative conformational energies.

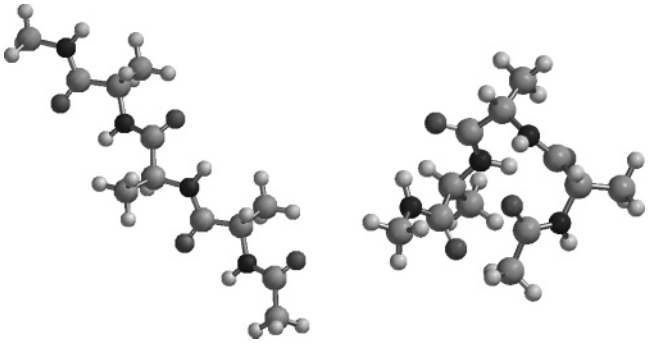

Figure 1. Graphical representations of the linear (1-D, conformation 1) conformation (left) and the globular (3-D, conformation 3) conformation (right) of the alanine tetrapeptide. These figures were created using the Spartan '04 package.[51]

The biophysical implications of our most reliable relative conformational energies are also quite interesting. In particular, among the low energy conformers, conformations 1 and 2 are extended (linear) while conformation 3 retains a compact hairpin geometry (globular). The energy difference between these geometries represents the energy gap between extended and folded backbone conformations—a fundamental quantity that propagates throughout a force field. We find this energy gap to be approximately 3.5 kcal/mol [RI-TRIM MP2/cc-pV(TQ)Z]. This number can be directly contrasted with the small basis canonical MP2 results that are closer to 5.5−6.0 kcal/mol, perhaps associated with intramolecular basis set superposition error. Additionally, previous local MP2 calculations using small basis sets yielded a gap of about 2.5 kcal/mol (Table S4 of the Supporting Information), while B3LYP, even at the cc-pVQZ level, predicts this energy difference as virtually zero. These findings reemphasize that one noteworthy application of the RI-TRIM MP2 methodology would be its use in the fundamental parameterization of force fields that treat systems of biochemical interest.

B. Computational Timings. In this section, we will explore the relationship between computational timings and system size for the following methodologies: SCF, MP2, TRIM MP2, RI-MP2, and RI-TRIM MP2. For clarity, we chose to omit DFT from the plots as the CPU time for DFT is almost always slightly larger (but on the same order of magnitude) than SCF. In doing so, we have considered linear (1-D) and globular (3-D) alanine tetrapeptide (see Figure 1), octapeptide, and hexadecapeptide conformations (the Cartesian coordinates for each of these model systems is provided in the Supporting Information). Using the linear and globular alanine tetrapeptide conformations, the computational scaling of methodology with angular momentum will also be evaluated by a direct comparison of CPU timings and basis set size for the SCF, RI-MP2, and RI-TRIM MP2 methods.

In Figure 2, a plot of log CPU time versus system size is presented for the aforementioned methodologies. This plot clearly shows that the CPU timings associated with the SCF, MP2, and TRIM MP2 methodologies strongly depend on

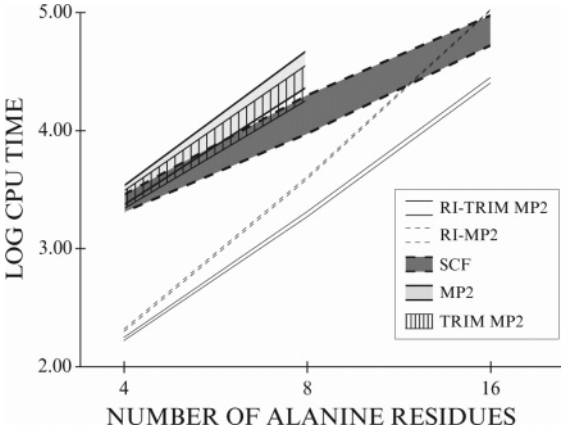

Figure 2. Plot of log CPU time vs system size for the SCF, MP2, TRIM MP2, RI-MP2, and RI-TRIM MP2 methodologies using the cc-pVDZ basis set. Each methodology has two associated lines: the bottom line represents data from linear (1-D) conformations, and the top line represents data from globular (3-D) conformations.

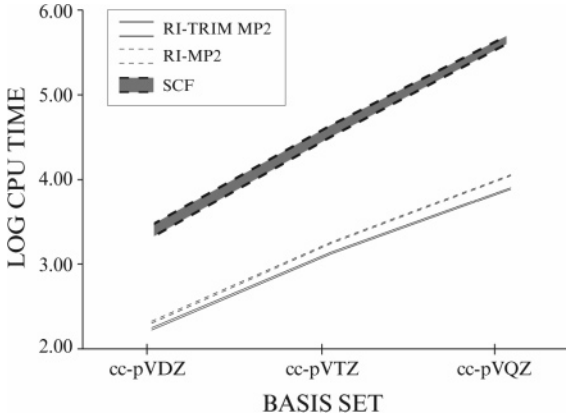

Figure 3. Plot of log CPU time vs basis set size for the SCF, RI-MP2, and RI-TRIM MP2 methodologies on linear (1-D, bottom line) and globular (3-D, top line) alanine tetrapeptide conformations.

system dimensionality, whereas the methods employing the RI approximation do not. From this plot, it is also evident that the iterative SCF procedure is the limiting time factor for calculations involving the RI approximation within a certain regime of system sizes. This trend relies strongly on the fact that we are using relatively tight DIIS convergence and integral screening threshold values; all estimates made in this section are based on these parameters and will need to be adjusted if one is using the default (looser) SCF parameters in most computational software packages. Using linear regression analysis, the crossover point between SCF and RI-MP2 was computed at approximately 75 heavy atoms, just shy of the globular (3-D) alanine hexadecapeptide data point [the linear (1-D) crossover occurs slightly earlier at approximately 64 heavy atoms]. The RI-TRIM MP2 methodology clearly extends the regime of practical MP2 calculations even further, as it is still less expensive than SCF at 125 heavy atoms—a system size corresponding to 25 alanine residues (for the 1-D case, the crossover occurs at approximately 104 heavy atoms, corresponding to a linear chain of 21 alanine residues). Put another way, one has to wait approximately 26 h for the MOs necessary to perform an RI-TRIM MP2 calculation, which takes less than 8 h (alanine hexadecapeptide at the cc-pVDZ level). From these data, it is now evident that some future research efforts should be focused on reducing the computational cost of the SCF procedure.

To explore the relationship between these methodologies and angular momentum in the AO basis set utilized, a plot of log CPU time versus basis set size is presented for SCF, RI-MP2, and RI-TRIM MP2 in Figure 3. This data set clearly shows that methods incorporating the RI approximation are scaling, on average, one power of L (highest angular momentum in the basis set) less than the SCF procedure and are, once again, independent of the system dimensionality. Using linear regression analysis, these scaling factors were found as 2.3 (RI-MP2) and 2.2 (RI-TRIM MP2) after normalization to 3.0 for the SCF methodology. This finding strongly supports the aforementioned point that SCF is the time limiting factor in RI-MP2 and RI-TRIM MP2 calculations. For alanine tetrapeptide at the cc-pVQZ level, a SCF was reached in approximately 6 days; this should be compared to the 2 h necessary to perform an RI-TRIM MP2 calculation. Hence, at the level where correlation energy corrections are most crucial, the RI-TRIM MP2 methodology has emerged with the ability to not only provide them with an acceptable level of accuracy but with very small relative timing costs.

V. Conclusions

In this work, the RI approximation has been incorporated into the theoretical framework of the local TRIM MP2 model to form a novel local fourth-order methodology, RI-TRIM MP2. On the basis of a series of stringent chemical and computational tests, RI-TRIM MP2 has demonstrated the ability to deliver accurate predictions of the relative energies of a series of 27 different alanine tetrapeptide conformations with RMS relative errors found as 0.192 kcal/mol at the cc-pVDZ level and an almost negligible 0.040 kcal/mol at the T → Q extrapolated complete basis set limit. The accurate prediction of the relative energies of this series of 27 alanine tetrapeptide conformations demonstrated a significant dependence on the basis set with RMS (max) deviations from the extrapolated RI-MP2/cc-pV(TQ)Z data set computed as 0.377 (0.944) kcal/mol (MP2/cc-pVTZ) and 0.250 (0.591) kcal/mol (TRIM MP2/cc-pVTZ). These findings strongly suggest that computations must be done using the cc-pVTZ and cc-pVQZ basis sets followed by extrapolation to the cc-pV(TQ)Z limit in order to obtain reliably converged relative conformational energies.

Correlation energy corrections using RI-TRIM MP2 can easily be obtained for systems that contain more than 125 heavy atoms with a computational timing cost less than that of the prerequisite SCF procedure and popular DFT alternatives. The RI-TRIM MP2 methodology reduces the current size limitations on systems that can be treated at the MP2 level and should be used to extend the regime of practical MP2 calculations when RI-MP2 becomes the rate-determining computational step (for systems containing more than 75 heavy atoms). With these findings, some future research

efforts will be directed toward reducing the computational cost of the iterative SCF procedure by utilizing linear scaling algorithms,[52−54] density fitting,[55] or a dual-basis approach.[56] The combination of an appropriately optimized SCF algorithm and the RI-TRIM MP2 gradient would then allow us to perform geometry optimizations on systems containing more than 125 heavy atoms at the MP2 level; this would be highly desirable as small biochemical and nanotechnological systems of interest could then be routinely explored in a number of chemical applications.

Acknowledgment. Funding for this work comes from a subcontract of Q-Chem Inc. as part of an NIH SBIR grant. We gratefully acknowledge useful discussions with Prof. Jay Ponder and Prof. Pengyu Ren, who suggested we re-examine the conformational energies of the alanine tetrapeptides and forwarded the optimized structures to us. We also thank Judy Hicks of Plush Design for her assistance in creating some of the figures in this work and Laura Rutherford for general research assistance. M.H.-G. is a part-owner of Q-Chem Inc.

Supporting Information Available: Absolute energies for the SCF and DFT methods at the cc-pVDZ, cc-pVTZ, and cc-pVQZ levels; absolute correlation energies for the MP2 and TRIM MP2 methods at the cc-pVDZ and cc-pVTZ levels; absolute correlation energies for the RI-MP2 and RI-TRIM MP2 methods at the cc-pVDZ, cc-pVTZ, and cc-pVQZ levels; relative conformational energies at the LMP2/cc-pVTZ(-f) level; and cartesian coordinates for all systems investigated. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1989.

(2) Kohn, W.; Becke, A. D.; Parr, R. G. *J. Phys. Chem.* **1996**, *100*, 12974−12980.

(3) Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*; Wiley-VCH: New York, 2000.

(4) Kristyan, S.; Pulay, P. *Chem. Phys. Lett.* **1994**, *229*, 175−180.

(5) Jung, Y.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2004**, *6*, 2008−2011.

(6) Johnson, B. G.; Gonzales, C. A.; Gill, P. M. W.; Pople, J. A. *Chem. Phys. Lett.* **1994**, *221*, 100−108.

(7) Perdew, J. P.; Zunger, A. *Phys. Rev. B: Condens. Matter* **1981**, *23*, 5048−5079.

(8) Lynch, B. J.; Fast, P. L.; Harris, M.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, *104*, 4811−4815.

(9) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618−622.

(10) Pople, J. A.; Binkley, J. S.; Seeger, R. *Int. J. Quantum Chem.* **1976**, *S10*, 1−19.

(11) For example, see Tables 8−4 of ref 3 and references therein.

(12) Head-Gordon, M. *J. Phys. Chem.* **1996**, *100*, 13213−13225.

(13) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. *J. Chem. Phys.* **1997**, *106*, 9639−9646.

(14) Byrd, E. F. C.; Sherrill, C. D.; Head-Gordon, M. *J. Phys. Chem. A* **2001**, *105*, 9736−9747.

(15) Helgaker, T.; Gauss, J.; Jorgensen, P.; Olsen, J. *J. Chem. Phys.* **1997**, *106*, 6430−6440.

(16) Feyereisen, M.; Fitzgerald, G.; Komornicki, A. *Chem. Phys. Lett.* **1993**, *208*, 359−363.

(17) Weigend, F.; Haser, M.; Patzelt, H.; Ahlrichs, R. *Chem. Phys. Lett.* **1998**, *294*, 143−152.

(18) Friesner, R. A.; Murphy, R. B.; Beachy, M. D.; Ringnalda, M. N.; Pollard, W. T.; Dunietz, B. D.; Cao, Y. X. *J. Phys. Chem. A* **1999**, *103*, 1913−1928.

(19) Pulay, P.; Saebø, S.; Wolinski, K. *Chem. Phys. Lett.* **2001**, *344*, 543−552.

(20) Ayala, P. Y.; Scuseria, G. E. *J. Chem. Phys.* **1999**, *110*, 3660−3671.

(21) Ayala, P. Y.; Kudin, K. N.; Scuseria, G. E. *J. Chem. Phys.* **2001**, *115*, 9698−9707.

(22) Saebø, S.; Pulay, P. *Annu. Rev. Phys. Chem.* **1993**, *44*, 213−236.

(23) Schütz, M.; Hetzer, G.; Werner, H. J. *J. Chem. Phys.* **1999**, *111*, 5691−5705.

(24) Lee, M. S.; Maslen, P. E.; Head-Gordon, M. *J. Chem. Phys.* **2000**, *112*, 3592−3601.

(25) Subotnik, J. E.; Head-Gordon, M. *J. Chem. Phys.* **2005**, *122*, 034109-1−034109-9.

(26) Jung, Y.; Lochan, R. C.; Dutoi, A. D.; Head-Gordon, M. *J. Chem. Phys.* **2004**, *121*, 9793−9802.

(27) Lochan, R. C.; Jung, Y.; Head-Gordon, M. *J. Phys. Chem. A* **2005**, in press.

(28) Grimme, S. *J. Chem. Phys.* **2003**, *118*, 9095−9102.

(29) Schütz, M.; Hetzer, G.; Werner, H. J. *J. Chem. Phys.* **1999**, *111*, 5691−5705.

(30) Hetzer, G.; Schütz, M.; Stoll, H.; Werner, H. J. *J. Chem. Phys.* **2000**, *113*, 9443−9455.

(31) Schütz, M. *J. Chem. Phys.* **2000**, *113*, 9986−10001.

(32) Schütz, M.; Werner, H. J. *J. Chem. Phys.* **2001**, *114*, 661−681.

(33) Werner, H. J.; Manby, F. R.; Knowles, P. J. *J. Chem. Phys.* **2003**, *118*, 8149−8160.

(34) Schütz, M.; Werner, H. J.; Lindh, R.; Manby, F. R. *J. Chem. Phys.* **2004**, *121*, 737−750.

(35) Maslen, P. E.; Head-Gordon, M. *Chem. Phys. Lett.* **1998**, *283*, 102−108.

(36) Maslen, P. E.; Head-Gordon, M. *J. Chem. Phys.* **1998**, *109*, 7093−7099.

(37) Lee, M. S.; Head-Gordon, M. *Int. J. Quantum Chem.* **2000**, *76*, 169−184.

(38) Saebø, S.; Tong, W.; Pulay, P. *J. Chem. Phys.* **1993**, *98*, 2170−2175.

(39) Schütz, M.; Rauhut, G.; Werner, H. J. *J. Phys. Chem. A* **1998**, *102*, 5997−6003.

(40) For a detailed derivation of the canonical Hartree−Fock equations, see: Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: An Introduction to Advanced Electronic Structure Theory*; Dover: New York, 1989.

(41) Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Friesner, R. A. *J. Am. Chem. Soc.* **1997**, *119*, 5908−5920.

(42) Kong, J.; White, C. A.; Krylov, A. I.; Sherrill, C. D.; Adamson, R. D.; Furlani, T. R.; Lee, M. S.; Lee, A. M.; Gwaltney, S. R.; Adams, T. R.; Ochsenfeld, C.; Gilbert, A. T. B.; Kedziora, G. S.; Rassolov, V. A.; Maurice, D. R.; Nair, N.; Shao, Y.; Besley, N. A.; Maslen, P. E.; Dombroski, J. P.; Dachsel, H.; Zhang, W. M.; Korambath, P. P.; Baker, J.; Byrd, E. F. C.; Van Voorhis, T.; Oumi, M.; Hirata, S.; Hsu, C. P.; Ishikawa, N.; Florian, J.; Warshel, A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M.; Pople, J. A. *J. Comput. Chem.* **2000**, *21*, 1532−1548.

(43) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007−1023.

(44) Weigend, F.; Kohn, A.; Hattig, C. *J. Chem. Phys.* **2002**, *116*, 3175−3183.

(45) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(46) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter* **1998**, *37*, 785−789.

(47) Yu, C.-H.; Norman, M. A.; Schäfer, L.; Ramek, M.; Peeters, A.; van Alsenoy, C. *J. Mol. Struct.* **2001**, *567*−*568*, 361−374.

(48) Chaudhuri, P.; Canuto, S. *THEOCHEM* **2002**, *577*, 267−279.

(49) Sekusak, S.; Frenking, G. *THEOCHEM* **2001**, *541*, 17−29.

(50) Wang, N. X.; Wilson, A. K. *J. Chem. Phys.* **2004**, *121*, 7632−7646.

(51) *Spartan '04 for Windows*; Wavefunction, Inc.: Irvine, CA.

(52) Challacombe, M. *J. Chem. Phys.* **1999**, *110*, 2332−2342.

(53) Schwegler, E.; Challacombe, M.; Head-Gordon, M. *J. Chem. Phys.* **1997**, *106*, 9708−9717.

(54) Challacombe, M.; Schwegler, E. *J. Chem. Phys.* **1997**, *106*, 5526−5536.

(55) Polly, R.; Werner, H. J.; Manby, F. R.; Knowles, P. J. *Mol. Phys.* **2004**, *102*, 2311−2321.

(56) Liang, W.; Head-Gordon, M. *J. Phys. Chem. A* **2004**, *108*, 3206−3210.

CT050126S