


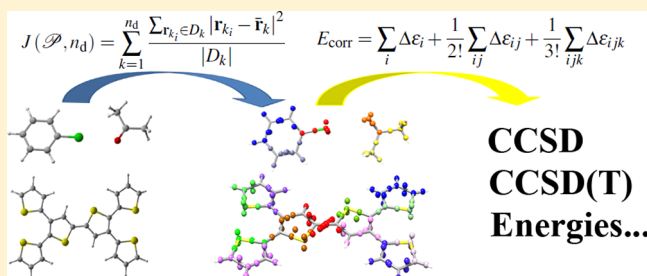
Third-Order Incremental Dual-Basis Set Zero-Buffer Approach: An Accurate and Efficient Way To Obtain CCSD and CCSD(T) Energies

Jun Zhang* and Michael Dolg*

Department of Chemistry, University of Cologne, Cologne, Germany

 Supporting Information

ABSTRACT: An efficient way to obtain accurate CCSD and CCSD(T) energies for large systems, i.e., the third-order incremental dual-basis set zero-buffer approach (inc3-db-B0), has been developed and tested. This approach combines the powerful incremental scheme with the dual-basis set method, and along with the new proposed *K*-means clustering (KM) method and zero-buffer (B0) approximation, can obtain very accurate absolute and relative energies efficiently. We tested the approach for 10 systems of different chemical nature, i.e., intermolecular interactions including hydrogen bonding, dispersion interaction, and halogen bonding; an intramolecular rearrangement reaction; aliphatic and conjugated hydrocarbon chains; three compact covalent molecules; and a water cluster. The results show that the errors for relative energies are <1.94 kJ/mol (or 0.46 kcal/mol), for absolute energies of <0.0026 hartree. By parallelization, our approach can be applied to molecules of more than 30 atoms and more than 100 correlated electrons with high-quality basis set such as cc-pVDZ or cc-pVTZ, saving computational cost by a factor of more than 10–20, compared to traditional implementation. The physical reasons of the success of the inc3-db-B0 approach are also analyzed.



1. INTRODUCTION

Currently, the application of quantum chemistry has gone beyond atoms, small- and medium-size molecules, to more-complex systems, such as clusters, nanoparticles, and biological molecules, as well as surfaces and solids. Besides the larger sizes, the components of these systems are also more diverse. For example, while a water nanoparticle contains only 16 or 17 H₂O molecules, a biological cluster may consist of a highly heterogeneous protein environment surrounding an organic molecule and some transition-metal atoms. Treating all components in a balanced way with sufficient accuracy is a great challenge for quantum chemistry. Density functional theory (DFT) is a valuable tool that can treat large systems with very low computational cost, and thanks to the developments of the last decades, the performance in the fields where traditional DFT is well-known to fail has been significantly improved. However, performance of DFT is still somewhat unpredictable, varying from case to case, and the functional suitable for the system of interest cannot be selected unambiguously without comparison with experiments or benchmark results obtained with highly accurate wave function theory (WFT). Therefore, WFT-based methods are still required for large systems when high accuracy is needed. Coupled-cluster singles and doubles (CCSD)¹ and CCSD with perturbative triple excitations [CCSD(T)]² are the most successful correlated WFT-based methods for single-reference cases. They have an excellent performance when basis sets of reasonable size are applied; in fact, the latter has been recognized as “golden standard” in quantum chemistry. When

combined with correlation-consistent basis sets^{3–7} and various extrapolation techniques,^{8–15} one can systematically approach the complete basis set (CBS) limit, which by now almost became a routine procedure in quantum chemistry.

The largest difficulty for the application of CCSD and CCSD(T) to large systems is the unfavorable scaling of the computational cost, i.e., the formal scaling is $O(n^6)$ and $O(n^7)$, respectively, where n is the number of basis functions. With current computer resources, this makes accurate calculations intractable for systems with more than ~15 atoms. Thus, much effort has been exerted to design new algorithms that decrease the computational cost of CCSD and CCSD(T). The essence of these methods is to exploit the locality of electron correlation. The first successful implementation following these lines was pioneered by Pulay and Sæbø,^{16–20} later, based on this work, Werner and co-workers proposed a series of local correlation methods,^{21–27} including LCCSD²⁶ and LCCSD(T).²⁵ While these methods implement projected atomic orbitals to construct virtual space, some authors have selected alternative ways to do this, e.g., with projected natural orbitals^{28–30} or orbital specific virtuals.^{31–33} Other methods attempting to lower the scaling include e.g., the cluster-in-molecule (CIM) strategy for CCSD and CCSD(T) by Li,^{34–37} CCSD(T) based on fragment molecular orbital (FMO) theory,³⁸ a divide-and-conquer implementation of CCSD(T) by Bartlett,^{39,40} local coupled-cluster singles-doubles theory by

Received: April 6, 2013

Head-Gordon and co-workers,^{41–45} and the divide-expand-consolidate (DEC) coupled-cluster scheme of Jørgensen and co-workers.^{46–48} These methods have attracted considerable attention of the computational chemistry community, and some of them have been applied in, e.g., biological simulations.^{49–51}

Another approach to achieve an efficient implementation of CCSD and CCSD(T) is the incremental scheme. This method was originally proposed by Stoll^{52–57} for the study of periodic systems; however, similar formulas can be traced back to Nesbet's work on the Bethe–Goldstone equations.^{58–60} The incremental scheme has been proven successful in obtaining the correlation energies of periodic solids.^{61,62} Later, it was realized⁶³ that the concept “increment” is strongly related to “cumulant”, an important quantity in solid-state physics⁶⁴ and theoretical chemistry.^{65,66} Incremental expansions can be derived from the Faddeev equation for the many-body problem^{67,68} by a cumulant expansion. Therefore this method gained some applications in theoretical physics, e.g., the analysis of partition functions⁶⁹ and the Hubbard model.⁷⁰ Starting from about 2007, our group began to apply the incremental scheme also for the correlation computation of finite molecules. The accuracy of the molecular incremental scheme has been confirmed for obtaining absolute correlation energies⁷¹ and interaction energies.⁷² It was first applied to CCSD, and later generalized to second-order Møller–Plesset theory (MP2), CCSD(T),^{73,74} and F12 explicit correlation methods.^{75,76} Combined with suitable strategies to further increase the efficiency,^{73,77–79} the incremental scheme for correlation computations has been successfully applied to the energy of water clusters,⁷⁴ property computations such as dipole moments⁸⁰ and optical tensors,⁸¹ and a condensed phase study of hydrogen fluoride.^{82,83} More importantly, since the practical implementation of the incremental scheme is fully automatic and highly parallelized, one can use it like a black box. Moreover, it can be combined with any in-house code or mature quantum chemistry program packages without any need to make changes in the code. This relatively straightforward implementation may be viewed as an advantage of the incremental scheme; however, it is also clear that a higher computational efficiency can be achieved by making modifications to the applied quantum chemistry code, e.g., in order to avoid redundant evaluations of some quantities or to accelerate convergence of the correlation calculations. Therefore, the incremental scheme is a very promising way of obtaining highly accurate energies of large systems efficiently.

In this paper, we present a new implementation of the incremental scheme, i.e., a third-order incremental dual-basis set zero-buffer (inc3-db-B0) correlation approach. This approach combines the incremental scheme for CCSD and CCSD(T) with the dual-basis set method, which means different basis sets are used at the self-consistent field (SCF) and correlation computation stages. By the dual-basis set method (“db”), and especially the powerful zero-buffer approximation (“B0”, see below), the virtual space of the correlation computation can be reduced significantly with little loss in accuracy in both absolute and relative energies for single-reference cases. Moreover, possible ambiguities of the partitioning of the occupied space in domains as well as individual Hartree–Fock (HF) SCF calculations for each increment are avoided. By parallelization, this approach can be applied to the CCSD(T) computations of rather large molecules with large basis sets.

This paper is organized as follows. In section 2, we will first present the basic machinery of the new approach. In section 3, we will test four systems of different chemical nature to show the validity of the new approach, and six other large systems beyond the computational capability of traditional wave function-based correlation methods in order to show the efficiency. In section 4, we conclude our results.

2. THEORY

2.1. Incremental Expansion. In usual correlation computations, the initial SCF step will generate sets of occupied and virtual orbitals. Then, all or a (usually valence) subset of the occupied orbitals will be correlated with all the virtual orbitals via a specific computational scheme, e.g., the coupled-cluster ansatz, to get the correlation energy. In the incremental scheme, the occupied set will be decomposed into several subsets called one-site domains, denoted by i, j, k , etc. The union of two one-site domains i and j is called a two-site domain, denoted by ij , and n -site domains can be defined in an analogous way. For an arbitrary n -site domain X , let ϵ_X be the correlation energy obtained by correlating all occupied orbitals in domain X . Then, we define the first-, second-, and third-order increment as

$$\Delta\epsilon_i := \epsilon_i \quad (1)$$

$$\Delta\epsilon_{ij} := \epsilon_{ij} - \Delta\epsilon_i - \Delta\epsilon_j \quad (2)$$

$$\Delta\epsilon_{ijk} := \epsilon_{ijk} - \Delta\epsilon_{ij} - \Delta\epsilon_{ik} - \Delta\epsilon_{jk} - \Delta\epsilon_i - \Delta\epsilon_j - \Delta\epsilon_k \quad (3)$$

Higher-order increments are defined accordingly. The total correlation energy (E_{corr}) then can be written as

$$E_{\text{corr}} = \sum_i \Delta\epsilon_i + \frac{1}{2!} \sum_{ij} \Delta\epsilon_{ij} + \frac{1}{3!} \sum_{ijk} \Delta\epsilon_{ijk} + \dots \quad (4)$$

Since, in finite systems, the number of domains is finite, this expansion is always ending at order n_d (the number of domains) and exact. We note that expressions such as eqs 2–4 appear in many different fields in both theoretical chemistry and physics, e.g., the energy expression in FMO theory⁸⁴ and the mutual information in complex system studies.⁸⁵ The most obvious analogy possibly is the well-known N -body expansion of a total intermolecular potential:⁸⁶

$$V(\mathbf{r}_1, \dots, \mathbf{r}_N) = \sum_{i < j}^N v_2(\mathbf{r}_i, \mathbf{r}_j) + \sum_{i < j < k}^N v_3(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) + \dots + v_N(\mathbf{r}_1, \dots, \mathbf{r}_N) \quad (5)$$

Thus, while the “locality” of electronic correlation is explored by domain decomposition, the second- and third-order increment can be viewed as “two- and three-body effects”, which take the electronic delocalization nature of the system into account. The higher the order, the less the increments will contribute to the total correlation energies of systems with reasonably well localized occupied orbitals. Therefore, it is possible to truncate the incremental expansion at a certain order without any loss of accuracy. According to our experience^{71,72,74} (also see below), truncating the incremental expansion at third-order is the best choice of keeping high accuracy with low computational cost. We will use a third-order incremental scheme (“inc3”) throughout this paper.

2.2. Domain Decomposition. In order to make the truncated incremental expansion sufficiently accurate, the interactions between domains should be as small as possible. Since the canonical molecular orbitals (CMOs) are rather delocalized due to the orthogonality and symmetry-adaptation requirements, they are not appropriate for a domain decomposition. Thus, CMOs are first localized to get a new set of localized molecular orbitals (LMOs) by the Boys strategy,⁸⁷ which aims to minimize the spatial extent of LMOs; then each LMO ϕ_i is associated to its center of charge $\mathbf{r}_i = \langle \phi_i | \mathbf{r} | \phi_i \rangle$ to get a set of points $\mathcal{P} = \{\mathbf{r}_i | i = 1, 2, \dots, n_{\text{occ}}\}$. Hereafter, we will sometimes use “point” \mathbf{r}_i instead of LMO ϕ_i for the sake of convenience. In order to decompose the occupied space into n_d domains D_1, \dots, D_{n_d} , we minimize, with respect to all possible distributions of orbitals in domains, the function

$$J(\mathcal{P}, n_d) = \sum_{k=1}^{n_d} \frac{\sum_{\mathbf{r}_{k_i} \in D_k} |\mathbf{r}_{k_i} - \bar{\mathbf{r}}_k|^2}{|D_k|} \quad (6)$$

Here, $\bar{\mathbf{r}}_k$ is the geometrical center of the $|D_k|$ point \mathbf{r}_{k_i} values in domain D_k . Obviously, this procedure aims to minimize the distances between the points in a domain. This framework of decomposing a set is called *K*-means clustering (KM)^{88,89} and has been widely used in artificial intelligence, statistics, and chemoinformatics.⁹⁰ In order to perform the optimization of $J(\mathcal{P}, n_d)$, an initial guess (random decomposition) is generated. Then, in each iteration, each point in \mathcal{P} is put into all the domains to determine the domain in which it minimizes $J(\mathcal{P}, n_d)$. The iteration stops when the decomposition does not change. This procedure is very close to the Jacobi sweep algorithm⁹¹ in orbital localization. Note that, in our previous applications, the decomposition was accomplished by the METIS graph partitioning algorithm.⁹² We switch to KM because, first, it is simpler and directly related to our chemical problem, e.g., no connectivity matrix construction step is needed; second, it can be reused in the B0 approximation (see below).

The choice of n_d is critical, since both n_d values that are too small or too large can deteriorate the efficiency and accuracy of the incremental scheme. For the chemical systems with well-defined small “fragments” or “units”, n_d should be the number of those chemical groups. For example, for a cluster of 10 solvent molecules and a small organic molecule, n_d should be set to 11. For systems composed of large “units” or ones with no obvious “fragments”, n_d should be chosen in such a way that, for the highest-order increments in the incremental expansion, the number of orbitals to be correlated is still tractable.

2.3. Dual-Basis Set Implementation. The dual-basis set technique has been widely used in quantum chemistry method development. This terminology was proposed by Jurgens-Lutovsky and Almlöf⁹³ in 1991, but, in fact, King and co-workers had implemented similar strategies a few years earlier.^{94,95} Briefly, this method uses different basis sets at different stages of a quantum chemical computation. Some examples include using dual-basis sets to accelerate MP2^{96,97} and ab initio molecular dynamics.⁹⁸ In the context of the incremental scheme, we use such a strategy: For the basis set \mathcal{B} , at the SCF stage, we will use a smaller basis set \mathcal{B}' , where all the basis functions of angular momenta higher than p are removed from \mathcal{B} ; then for the computation of a domain X , a subset of the functions that was removed is added again to get the correlation energy.

This strategy enables us, in correlation computations of domain X , to use larger basis sets for the region of the system that contributes most to the correlation energy, while, for the rest (or “environment”), \mathcal{B}' is still used. The new basis set applied for the entire domain X is denoted as \mathcal{B}_X . In this way, the virtual space can be significantly reduced compared to the original basis set \mathcal{B} . Note that, in our previous implementation, the reduction of the virtual space is achieved by domain-specific basis sets, using the entire basis set \mathcal{B} at the SCF stage, localization and decomposition, whereas for the computations of increments, the environment was treated with a smaller basis set such as STO-3G⁷³ or SV.⁷⁶

The new dual-basis set method has several advantages. In the previous scheme the difference between the entire basis set and the one used for the increment calculation are significant already at the SCF level. Therefore, for each incremental computation, a SCF and localization procedure has to be performed, and the obtained LMOs must be mapped to the LMOs in basis set \mathcal{B} to identify a domain. This mapping is not unique and thus, in some cases problematic, e.g., for aromatic systems. In the new implementation, since \mathcal{B}' is a subset of \mathcal{B}_X , the orbital relaxation from basis functions of high angular momenta will be small for single-reference cases. Thus a repetition of the SCF and localization are not necessary, and the LMOs in \mathcal{B}' can be exactly projected onto \mathcal{B}_X : the occupied space is preserved and the virtual space is augmented and orthogonalized by the Löwdin procedure.⁹⁹ This generates virtual orbitals which are closest in a least-squares sense to the original virtual orbitals and the added higher angular momentum functions.¹⁰⁰ Therefore, besides saving extra SCF procedures, there is no ambiguous mapping. Also, the environment can be described more accurately in \mathcal{B}' than with basis sets such as STO-3G or SV, which are possibly too small for a good SCF description and also unsuitable for correlation computations. In this way, the dual-basis set incremental scheme can enjoy both high accuracy and efficiency.

Note that as the incremental order increases, the result converges not toward that of a standard CCSD or CCSD(T) calculation, but to the corresponding exact dual-basis set result. To see how the results can differ, we performed a benchmark with 15 small molecules, the details of which can be found in the Supporting Information. The results are listed in Table 1.

Table 1. Benchmark of Dual-Basis Set and Standard CCSD and CCSD(T) Methods

basis set ^a	SCF ^b	CCSD	CCSD(T)
VDZ	0.07853	0.00066	0.00031
VTZ	0.08196	0.00059	0.00023
VQZ	0.08161	0.00058	0.00022
AVDZ	0.08195	0.00077	0.00036
AVTZ	0.08275	0.00061	0.00029
AVQZ	0.07445	0.00053	0.00020

^aVXZ = cc-pVXZ; A = aug. ^bIn the table are the root-mean-square deviations (RMSDs): $\text{RMSD} = (\sum_i^N (E_{\text{ex-db}}(i) - E_{\text{st}}(i))^2 / N)^{1/2}$ (unit: hartree), where $E_{\text{ex-db}}(i)$ and $E_{\text{st}}(i)$ are energies of molecule i computed by dual-basis set and standard correlation methods, respectively. Here, in the dual-basis set computations, the SCF was performed with VXZ(P/S) or AVXZ(P/S) where (P/S) means only basis functions of angular momenta s and p are preserved; the following correlation computation was performed with original VXZ or AVXZ basis sets.

We see that, although energies at the SCF stage differ significantly, the *total* energies have rather small errors, i.e., less than 0.0008 and 0.0004 hartree for CCSD and CCSD(T), respectively. However, the error decreases as the basis set becomes more complete. This can be understood since single excitations in the coupled-cluster ansatz compensate for the omitted orbital relaxation when going from the smaller to the larger basis set. Thus correlation contributions discussed here also contain some orbital relaxation effect. It is noteworthy that, if relative energies are considered, the error becomes even more negligible (see below).

2.4. Zero-Buffer (B0) Approximation. In the previous domain-specific basis incremental scheme, the atoms that contribute significantly to the correlation energy of a domain X were identified as the atoms whose distance from the orbital centers of a domain to be computed is less than a parameter t_{main} .⁷³ In our new implementation, the KM algorithm is used to decompose atoms of the system \mathcal{A} into n_d atomic domains M_1, \dots, M_{n_d} , but the difference is that the \mathbf{r}_{k_i} values in eq 6 are now the coordinates of the atoms \mathbf{x}_{k_i} and $\bar{\mathbf{r}}_k$ is not dynamically determined but fixed to the geometric center of orbital domain D_k . This is the fixed-center KM algorithm, i.e. we try to minimize:

$$J(\mathcal{A}, n_d) = \sum_{k=1}^{n_d} \frac{\sum_{\mathbf{x}_{k_i} \in M_k} |\mathbf{x}_{k_i} - \bar{\mathbf{r}}_k|^2}{|M_k|} \quad (7)$$

In this way, each orbital domain D_k is associated with M_k that shares the same $\bar{\mathbf{r}}_k$. Thus, in the computation of domain X (which is the union of D_k 's), only the atoms from D_k 's associated M_k 's are treated with basis set \mathcal{B} . One may argue that a buffer region “Br” around the M_k values with a radius r should also be treated with \mathcal{B} to ensure the accuracy. As we can see below, this buffer is actually not needed, as long as n_d is reasonable, i.e., a “zero-buffer” or B0 approximation is sufficient for chemical accuracy. In this way, one does not need the manually determined parameter t_{main} . This scheme is shown in Figure 1.

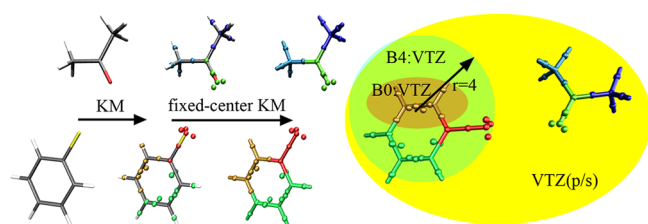


Figure 1. Dual-basis-set zero-buffer (B0) scheme. Left side shows the decomposition of LMOs and atoms into orbital and atomic domains; right side shows the Br approximation.

2.5. An Illustrative Example: Bicyclooctane. Here, we take bicyclooctane as an example to describe the impact of the selected number of domains n_d on the accuracy and efficiency of the inc3-db-B0 approach. Since bicyclooctane has eight carbon atoms, by chemical intuition the largest n_d is 8. We thus performed the inc3-db-B0- and inc3-db-B10-CCSD(T)/VDZ computations of bicyclooctane for $5 \leq n_d \leq 10$. In this case, “B10” is the entire molecule; thus, all increments are evaluated in basis set \mathcal{B} . The results are shown in Figure 2. We find that, for all $n_d \leq 8$, the accuracies of inc3-db-B0 and inc3-db-B10 correlation energies are similar, with an unsigned error of less

than 0.0011 hartree. However, when n_d goes beyond 8, the accuracy of inc3-db-B0 decreases rapidly. This can be understood by looking at Figure 3. For $n_d = 6$ each domain is roughly “one-and-a-half” C–C bond, and the basis functions of other domains have a small contribution. For $n_d = 10$ the molecule is broken into too small pieces, where each small domain contains large energy contributions from the basis functions of other domains. Thus the B0 approximation does not work well and must be extended, e.g. to B10. However, when considering the efficiency, B0 is faster than B10 by about 6 times, revealing that $\text{Br}(r > 0)$ is not worth implementing, since, as long as a reasonable n_d value is selected, inc3-db-B0 can enjoy both high accuracy and efficiency. It is also observed that, while smaller n_d yield a more accurate energy, n_d values that are either too small or too large will decrease the efficiency.

2.6. Distance Screening. The distance of two domains is defined as the closest distance between any two points belonging to different domains. Thus, a truncation distance r_t can be defined. If the distance between D_i and D_j is greater than r_t , the increments comprising the two domains such as $\Delta\epsilon_{ij}$ or $\Delta\epsilon_{ijk}$ can be neglected. This approximation works very efficiently for molecules with a large spatial extent.

2.7. Parallelization. The incremental computation was straightforwardly parallelized: a master node assigns tasks and collects results, whereas several slave nodes compute the increments. The entire procedure is schematically shown in Figure 4.

2.8. Computational Details. In order to test the accuracy and efficiency of our new inc3-db-B0 correlation approach, we selected 10 chemical systems for calibration, i.e., 4 medium-size and 6 larger molecules (Figure 5). The geometries of these systems and the monomers in them were optimized with the three-parameter hybrid Becke exchange and Lee–Yang–Parr correlation functional (B3LYP)¹⁰¹ and SVP basis sets¹⁰² using Gaussian 03,¹⁰³ except that the geometries of 3 and 10 were extracted from the literature.^{104,105} The core electrons were not correlated. CCSD and CCSD(T) computations were performed with Molpro 2010.¹⁰⁶ All calculations were performed on local PC clusters.

In the current work, VDZ and VTZ basis sets were used as \mathcal{B} . At the SCF stage, \mathcal{B}' was set to VXZ(P/S) ($X = \text{D, T}$). The canonical orbitals to be correlated were localized by the Boys strategy.⁸⁷ In the correlation energy computation of domain X , the basis set was augmented to \mathcal{B}_X by the B0 approximation as described above, and the old LMOs were projected onto the new basis set. The number of domains n_d and the screening parameter r_t were set, depending on the systems. The incremental expansion was always truncated at third order. The entire incremental scheme was performed with our in-house code written in C++.

The medium-size systems investigated included the following: a glycine dimer (1), a butane-ethanamide complex (2), a chlorobenzene-acetone complex (3), and educt and product of a $\text{C}_2\text{--C}_6$ enyne allene cyclization reaction (4a, 4b). These systems cover the range from strong hydrogen bonds to weak interactions involving dispersion and halogen bonds, and a rearrangement reaction. The large systems were comprised of hydrocarbon chains $\text{C}_{20}\text{H}_{42}$ (5) and $\text{C}_{20}\text{H}_{22}$ (6), three compact molecules (6aS,11aS)-pterocarpan (7), gibberellin (8), a thiophene dendrimer (9), and the boat-a isomer of a 16-mer water cluster (10). (See Figure 5 and Supporting Information for details.) Note that, in this paper, we are aiming at testing the

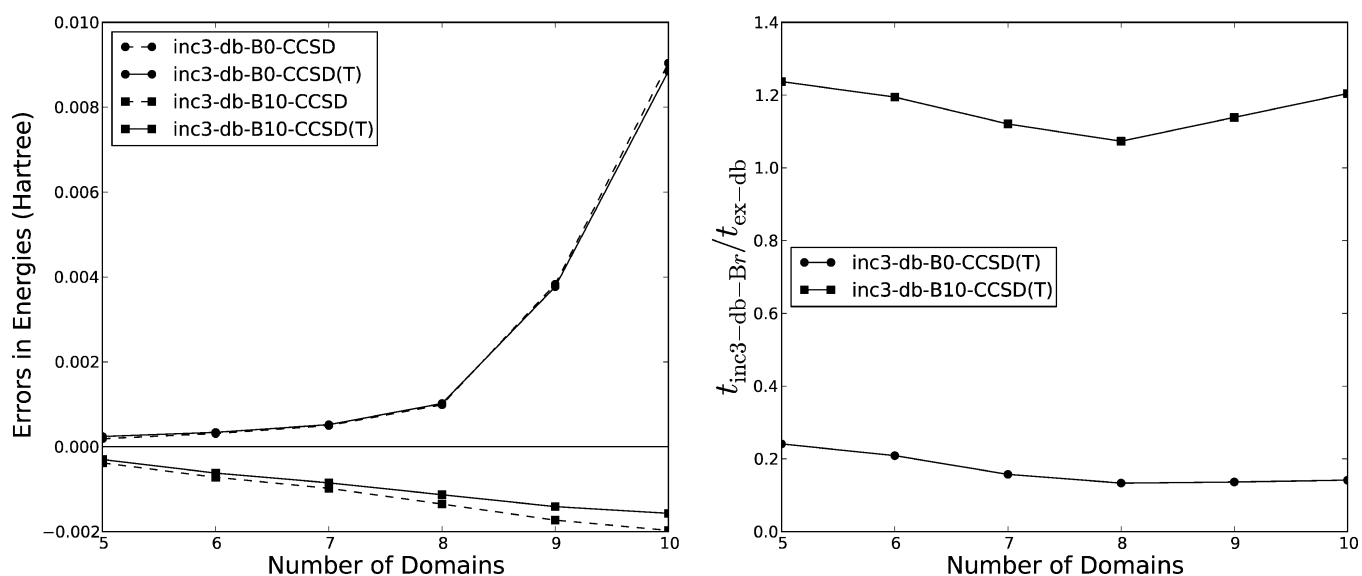


Figure 2. Impact of the number of domains on accuracy and efficiency of the inc3-db-B0 approach. The computations were performed with VDZ basis set, where inc3-db-Br-CCSD(T) ($r = 0, 10$) were parallelized by 10 cores. Here “error” is defined as $E_{\text{ex-db}} - E_{\text{inc3-db-Br}}$; t_X is the wall time of method X.

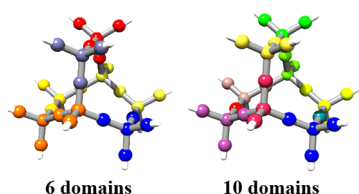


Figure 3. Domain decompositions of bicyclooctane. Licorice denotes bicyclooctane and balls represent centers of charges. The balls of the same color belong to the same domain.

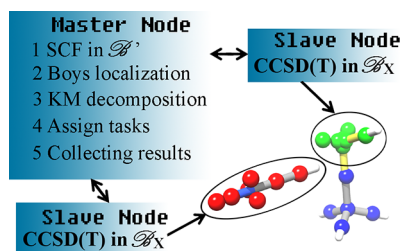


Figure 4. Parallelization of the inc3-db-B0 correlation approach.

accuracy of our methodology rather than exploring a specific system.

3. RESULTS AND DISCUSSION

3.1. Medium-Size Systems. We performed inc3-db-B0 CCSD and CCSD(T) computations on **1**, **2**, **3** and **4a**, **4b** with VDZ and VTZ basis sets. For all complexes and monomers in **1**, **2**, **3**, we also performed exact dual-basis set and standard CCSD and CCSD(T) computations. Since, in actual applications, it is frequently encountered that monomers are within the computational ability but complexes are not, our benchmark here is of practical interest. Moreover, we consider relative energies here in order to check to which extent dual-basis set and standard correlation computations differ. We define the following quantities:

$$\Delta E_{\text{inc3-db-B0}}(\text{AB}) = E_{\text{inc3-db-B0}}(\text{AB}) - E_{\text{ex-db}}(\text{A}) - E_{\text{ex-db}}(\text{B}) \quad (8)$$

$$\Delta E_{\text{ex-db}}(\text{AB}) = E_{\text{ex-db}}(\text{AB}) - E_{\text{ex-db}}(\text{A}) - E_{\text{ex-db}}(\text{B}) \quad (9)$$

$$\Delta E_{\text{st}}(\text{AB}) = E_{\text{st}}(\text{AB}) - E_{\text{st}}(\text{A}) - E_{\text{st}}(\text{B}) \quad (10)$$

Here, $E_{\text{inc3-db-B0}}(\text{X})$, $E_{\text{ex-db}}(\text{X})$, and $E_{\text{st}}(\text{X})$ represent the energies of substance X computed by inc3-db-B0, exact dual-basis set and exact standard correlation methods, respectively.

The domain decomposition of these systems is shown in Figure 6. Obviously, for a given reasonable value of n_d , the KM algorithm can decompose the system in a chemically very meaningful way. For example, the ethanamide in **2** is decomposed into three domains, with each methyl, carbonyl, and amino group comprising a domain; the chlorobenzene in **3** is partitioned into three parts (one C–Cl bond, and two “semi-benzenes”). This implies that our KM algorithm can not only be used in the incremental scheme, but also has potential applications in the identification of molecular 3D structures, beyond 2D.¹⁰⁷ In these computations, we do not use distance screening, since all systems are quite compact, and thus no interactions between domains can be neglected.

The computational results are listed in Tables 2 and 3. Looking at the absolute energies, both for CCSD and CCSD(T), we can see that the inc3-db-B0 approach generates excellent approximations to the exact dual-basis set methods, i.e., even in the worst case, the error is smaller than 0.0004 hartree. Since the good accuracy of the third-order incremental scheme has been confirmed previously,^{71,72,74} our study here further confirms the success of the B0 approximation. This can be explained by the fact that one should use more basis functions of higher angular momenta centered close to the correlated orbitals in order to generate a denser node structure for the computation of the correlation energy, whereas the basis functions which are centered far away have only minor contributions. We already have emphasized for the case of bicyclooctane that the B0 approximation works well only if the decomposition of the molecule into orbital domains and groups of atoms is reasonable.

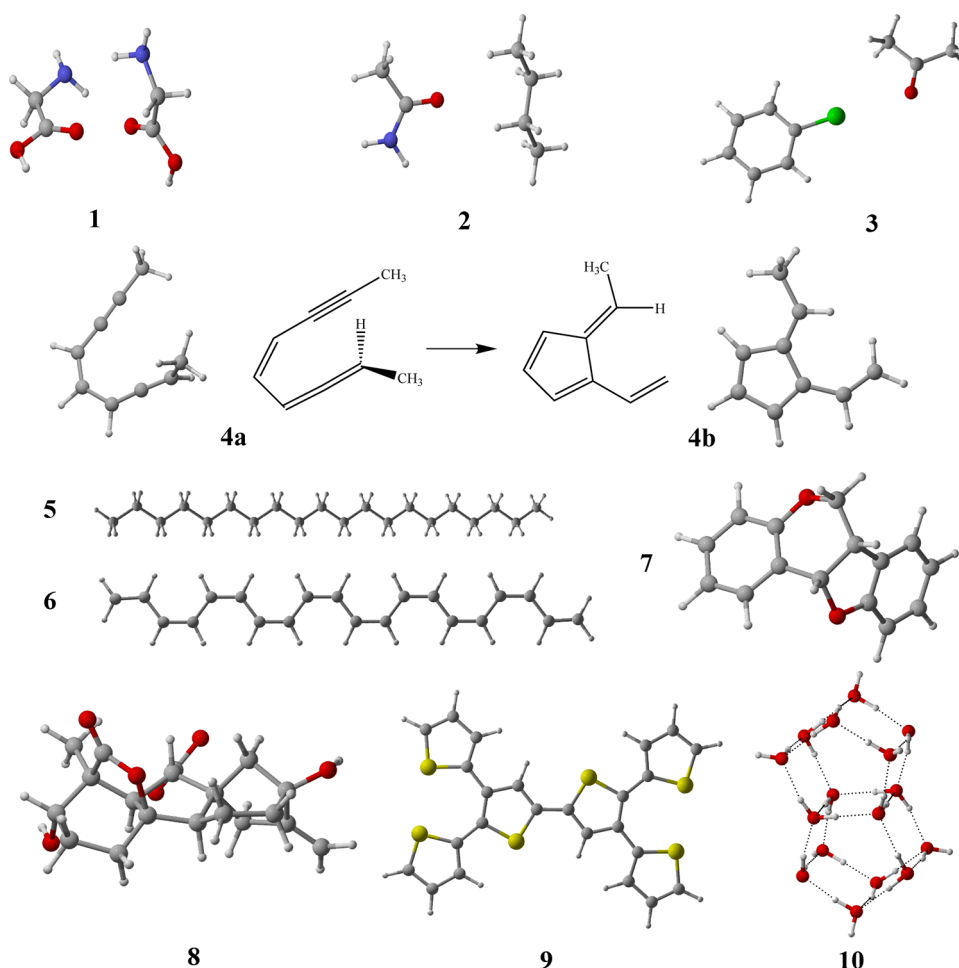


Figure 5. Medium-size and large test systems. Color legend: gray, carbon; white, hydrogen; blue, nitrogen; red, oxygen; green, chlorine; and yellow, sulfur.

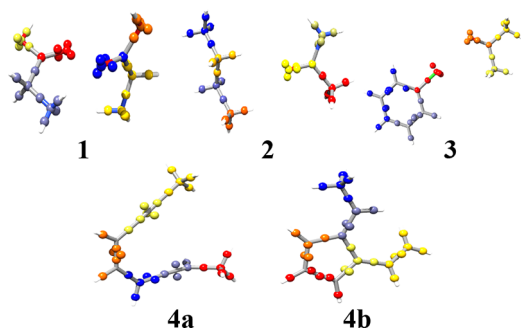


Figure 6. Domain decompositions of the medium-size systems. Balls represent centers of charges. Balls of the same color belong to the same domain.

We find that the inc3-db-B0 and standard correlation methods generate nearly the same relative energies within chemical accuracy (1 kcal/mol or 4.18 kJ/mol), since even in the worst case, the difference is only 1.94 kJ/mol or 0.46 kcal/mol. Hydrogen bonding as in complex 1 is an important interaction mode in chemistry. Our method is highly suitable to generate an accurate energy, since this interaction is localized in the region between the hydrogen bond acceptor and donor, whereas the rest of the system can be decomposed into more regions (of course not too many) to decrease the computational cost without any loss of accuracy. Complex 2 is bound by

dispersion interaction; it particularly represents a weak interaction mode in proteins between the aliphatic and polar side chains of amino acids. Complex 3 is a halogen bonding complex, sharing some similarities with hydrogen bonding but with stronger directionality.¹⁰⁸ Traditional DFT cannot treat these complexes accurately; since the magnitude of these interaction is quite small, accurate correlation methods, especially CCSD(T), must be applied. Our method works very well for these complexes. Interestingly, the performance is better for the more weakly bound complex 3 than for 2. The largest error is 0.40 kJ/mol for inc3-db-B0-CCSD(T)/VTZ of 2. Comparing $\Delta E_{\text{inc3-db-B0}}$ with $\Delta E_{\text{ex-db}}$ suggests that the error is due to the truncated incremental expansion rather than the dual-basis set zero-buffer approximation. For 3, although the highly delocalized benzene ring is decomposed, the accuracy does not decrease. This can be explained that electron delocalization effects can be automatically taken into account in the higher-order (here, second- and third-order) increments. In this case, one 3-site domain is just the entire chlorobenzene. The rearrangement from 4a to 4b is a Schmittel type C_2-C_6 enyne allene cyclization^{109–111} and was carefully studied by theoretical methods.¹¹² This is also a good test case for the incremental scheme, since the geometrical and topological structures of the system change after rearrangement. We see from Figure 6 that KM perfectly decomposed 4a and 4b into chemically meaningful parts. The reaction energies calculated by inc3-db-B0-CCSD and inc3-db-B0-CCSD(T) only have a

Table 2. Accuracy for Medium-Size Systems: inc3-db-B0-CCSD

system	1	2	3	4a	4b
VDZ					
$E_{\text{inc3-db-B0}}$ (Hartree)	−567.40428	−366.61586	−883.21370	−347.85637	−347.78369
$E_{\text{ex-db}}$ (Hartree)	−567.40421	−366.61575	−883.21367	−347.85637	−347.78367
E_{st} (Hartree)	−567.40642	−366.61747	−883.21585	−347.85824	−347.78575
$\Delta E_{\text{inc3-db-B0}}$ (kJ/mol)	−40.25	−12.54	−7.13	−190.82	
$\Delta E_{\text{ex-db}}$ (kJ/mol)	−40.06	−12.27	−7.05	−190.89	
ΔE_{st} (kJ/mol)	−40.11	−12.33	−7.15	−190.32	
error (kJ/mol) ^a	0.14	0.21	−0.02	0.50	
VTZ					
$E_{\text{inc3-db-B0}}$ (Hartree)	−567.95617	−366.98218	−883.67989	−348.17132	−348.10183
$E_{\text{ex-db}}$ (Hartree)	−567.95599	−366.98212	−883.67987	−348.17124	−348.10180
E_{st} (Hartree)	−567.95749	−366.98351	−883.68176	−348.17300	−348.10383
$\Delta E_{\text{inc3-db-B0}}$ (kJ/mol)	−33.14	−8.35	−5.00	−182.45	
$\Delta E_{\text{ex-db}}$ (kJ/mol)	−32.68	−8.18	−4.94	−182.31	
ΔE_{st} (kJ/mol)	−32.77	−8.24	−5.01	−181.60	
error (kJ/mol)	0.37	0.11	−0.01	0.85	

^aError = $\Delta E_{\text{st}} - \Delta E_{\text{inc3-db-B0}}$.

Table 3. Accuracy for Medium-Size Systems: inc3-db-B0-CCSD(T)

system	1	2	3	4a	4b
VDZ					
$E_{\text{inc3-db-B0}}$ (Hartree)	−567.45683	−366.65599	−883.27594	−347.91290	−347.83798
$E_{\text{ex-db}}$ (Hartree)	−567.45677	−366.65592	−883.27591	−347.91262	−347.83796
E_{st} (Hartree)	−567.45738	−366.65647	−883.27661	−347.91309	−347.83868
$\Delta E_{\text{inc3-db-B0}}$ (kJ/mol)	−43.30	−13.90	−8.40	−196.68	
$\Delta E_{\text{ex-db}}$ (kJ/mol)	−43.15	−13.70	−8.32	−196.00	
ΔE_{st} (kJ/mol)	−43.03	−13.69	−8.34	−195.38	
error (kJ/mol) ^a	0.27	0.21	0.06	1.30	
VTZ					
$E_{\text{inc3-db-B0}}$ (Hartree)	−568.04452	−367.04662	−883.77644	−348.25253	−348.18061
$E_{\text{ex-db}}$ (Hartree)	−568.04447	−367.04648	−883.77640	−348.25210	−348.18059
E_{st} (Hartree)	−568.04412	−367.04637	−883.77645	−348.25208	−348.18090
$\Delta E_{\text{inc3-db-B0}}$ (kJ/mol)	−36.56	−10.13	−6.66	−188.81	
$\Delta E_{\text{ex-db}}$ (kJ/mol)	−36.44	−9.76	−6.56	−187.75	
ΔE_{st} (kJ/mol)	−36.35	−9.73	−6.57	−186.87	
error (kJ/mol)	0.21	0.40	0.09	1.94	

^aerror = $\Delta E_{\text{st}} - \Delta E_{\text{inc3-db-B0}}$.

relative error of <1.04%. From Tables 2 and 3, we can see that, for CCSD, the error is mainly due to the dual-basis set method, whereas, with perturbative triples, the error from the truncated incremental expansion and B0 approximation slightly increases and becomes roughly as large as the one due to the dual-basis set method. Summing up, for all cases, we find that the inc3-db-B0 approach can give highly reliable results within chemical accuracy.

In Table 4, we list the total computational times of these calculations, i.e., times including integral evaluation, self-consistent field calculation, orbital localization, partitioning into domains and incremental correlation calculations. We note that, even for the largest systems considered here, the correlation calculations dominate the computational times. Focusing on real times, for VDZ the inc3-db-B0 calculations are longer than the exact dual-basis set computations. However, already for VTZ the inc3-db-B0 approach undoubtedly outperforms the traditional implementation, constituting a savings of 12.5%–36.6% in computational time. We also observed that the larger the system, the more the time savings. Moreover, the inc3-db-B0 approach benefits greatly from

Table 4. Efficiency for Medium-Size Systems

system	basis set	inc3-db-B0-CCSD(T) Time		Exact Dual-Basis Set
		real time	wall time ^a	CCSD(T) time
1	VDZ	6.6 h	0.7 h	4.6 h
1	VTZ	7.4 day	0.9 day	11.3 day
2	VDZ	3.4 h	0.3 h	2.7 h
2	VTZ	3.4 day	5.8 h	5.4 day
3	VDZ	9.7 h	1.2 h	6.4 h
3	VTZ	9.2 day	1.1 day	14.4 day
4a	VDZ	2.3 h	0.2 h	1.7 h
4a	VTZ	2.5 day	4.6 h	2.9 day
4b	VDZ	2.4 h	0.3 h	1.8 h
4b	VTZ	2.5 day	6.9 h	2.9 day

^aAll inc3-db-B0 and exact dual-basis set correlation calculations were performed on nodes equipped with an Intel Core 2 Q6600 2.40 GHz quad-core CPU and 8 GB of RAM. The inc3-db-B0 calculations were parallelized by 16 CPU cores. The time is for a total calculation (i.e., from the integral evaluation, SCF, orbital localization, decomposition into domains, to the correlation stage).

Table 5. Accuracy for Large Systems

system	number of atoms	basis set	CCSD (Hartree)		CCSD(T) (Hartree)
			$E_{\text{inc3-db-B0}}$	$E_{\text{ex-db}}$	$E_{\text{inc3-db-B0}}$
5	62	VDZ	−785.05547	−785.05520	−785.15635
6	42	VDZ	−772.86342	−772.86291	−772.99111
7	29	VTZ	−727.83066	−727.82809	−727.98869
8	49	VDZ	−1186.61284	−1186.61191	−1186.76017
9	44	VDZ	−3306.16564	−3306.16355	−3306.37778
10	48	VDZ	−1220.15527	−1220.15434	−1220.22077

parallelization. Parallelized with only 4 nodes, each equipped with 4 CPU cores (i.e., a total of 16 cores), the computational costs are reduced by a factor of 10–20. In fact, we could accomplish inc3-db-B0 computations in ~ 1 day, while the traditional implementations required more than 14 days. This proves the high efficiency of our new approach. It is also confirmed in computations of large systems (see below), where distance screening leads to additional savings.

3.2. Large Systems. The results and times of these computations are given in Tables 5 and 6, respectively. The

Table 6. Efficiency of Large-Size Systems

system	basis set	ne/nbf ^a	inc3-db-B0-CCSD(T) Time	
			real time	wall time ^b
5	VDZ	122/490	10.5 day	22.3 h
6	VDZ	102/390	11.8 day	18.4 h
7	VTZ	84/678	60.4 day	3.6 day
8	VDZ	136/470	77.6 day	4.9 day
9	VDZ	146/514	112.4 day	4.2 day
10	VDZ	128/384	8.8 day	15 h

^aHere, ne is the number of electrons to be correlated, and nbf is the number of basis set functions. ^bThe inc3-db-B0 correlation calculations of 5, 6, 7, 8, and 10 were parallelized by 20 cores while 9 was parallelized by 30 cores. Each node is equipped with an Intel Core 2 Q6600 2.40 GHz quad-core CPU and 8 GB of RAM. The time is for a total calculation (from the integral evaluation, SCF, orbital localization, decomposition into domains, to the correlation stage).

decompositions of these systems are given in the Supporting Information. The n_d values for 5 and 6 were set to 10. For the alkane chain 5, we set the truncation parameter r_t to 6.0 Å, reducing the number of increments to be computed by 57.1%. No significant loss of accuracy was observed. For the conjugated alkene chain 6 r_t is set to 10.0 Å, since no accurate results were obtained for a value of 6.0 Å due to the strong delocalization of the electrons. The number of increments to be computed is reduced by 12.6%. Interestingly, it is the distance truncation rather than the B0 approximation that has a large impact on the computational accuracy for the conjugated molecule 6. To explain this fact, we again emphasize that, for such systems, the delocalized electronic nature must be taken into account by higher-order increments (explicitly, unions of 1-site domains), which comprise several π -bonds. The errors for the CCSD energies relative to exact dual-basis set CCSD are only 0.0002 and 0.0005 hartree for 5 and 6, respectively. Here, we briefly mention that, in order to select a reasonable r_t , one can perform a fast second-order incremental correlation computation with a small basis set such as STO-3G and plot the relationship between increment $\Delta\epsilon_{ij}$ and the distance between domain i and j . In this way, one can easily see beyond which value of r_t all the increments can be safely neglected.

Clearly, in order to account for, e.g., geometrical differences between molecules or groups in molecules, the value of r_t used in the calculation should be chosen large enough to have some safety margin.

Complexes 7 and 8 are two compact molecules that widely exist in plants. The errors in the CCSD energies are 0.0026 and 0.0009 hartree, respectively. These molecules do not have well-defined chemical units, and the n_d values for both were set to 10. Complex 9 is useful in organic optics and electronics. The value of n_d was set to 12 and the CCSD energy error is 0.0021 hartree. No distance screening was applied for these three systems. The slightly larger errors are due to the too fragile decompositions of these molecules. The water cluster 10 is very suitable for incremental scheme. We use values of $n_d = 16$ and $r_t = 5$ Å and get an error of only 0.0009 hartree for the CCSD energy. Because of the limit of computational ability, the exact dual-basis set CCSD(T) energies could not be computed, but based on the CCSD and CCSD(T) data for medium-size systems and the CCSD results for large systems, we believe that the accuracy is sufficiently high.

For the efficiency, a direct comparison is not available. First, no exact dual-basis set CCSD(T) computations could be performed with our hardware. Second, the dual-basis set CCSD calculations had to be performed with different hardware, since a significantly larger amount of memory than for the incremental calculations was required. However, we can point out that the inc3-db-B0 approach needed only 1–4 days to get both CCSD and CCSD(T) energies, whereas the traditional implementation required 3–14 days for evaluating only the CCSD energies. From Table 6, we can see that the inc3-db-B0 approach can be routinely applied to systems containing more than 100 correlated electrons with basis set of VDZ quality or better. We note here that the size of the large test systems and the basis set size was chosen in such a way that standard CCSD calculations still remain feasible on our current hardware, within a reasonable time frame. We are confident that, by means of the inc3-db-B0 approach, very accurate CCSD and CCSD(T) energies can be obtained routinely also for larger systems and more extended basis sets with the same resources at hand. Corresponding applications are currently performed in our laboratory.

4. CONCLUSION

In this paper, we propose an accurate and efficient implementation of local correlation methods, i.e., the third-order incremental dual-basis set zero-buffer (inc3-db-B0) approach, and apply it to CCSD and CCSD(T). A small basis set is used for the HF computation and then for a domain X, the atoms that contribute most to the correlation energy of X are determined and their basis sets are augmented to perform the correlation computation. We have proven that this scheme yields very accurate absolute and relative energies for a wide

range of systems, including those which are difficult to treat accurately with traditional DFT, e.g., weakly bound complexes or large conjugated systems. The error is <1.94 kJ/mol (or 0.46 kcal/mol). We should note that this accuracy can be realized in the case where some molecules are computed by traditional implementation while others are treated with inc3-db-B0 approach. Thus, our method is very useful in practice. Moreover, since this method reduces the virtual space significantly and is highly suitable for parallelization, it can reduce the computational cost by a factor of 10–20, or more. The larger the system, the more computational effort can be saved.

A new method of domain decomposition, i.e., *K*-means clustering (KM), is proposed. KM is an important technique in inc3-db-B0 approach since it can also be used to determine the atoms that have considerable contribution to the correlation energy of an orbital domain. Once determined, the zero-buffer (B0) approximation can be applied without any loss of accuracy. Thus, this virtual space truncation can be realized without assigning a parameter t_{main} . We also explore how to determine suitable number of domains n_d and truncation distance r_t . The determination of n_d is critical to the success of the B0 approximation. It should be the number of fundamental units constituting the molecule, or a number that does not make the decomposition too fragile.

The physical reasons of the success of the inc3-db-B0 approach root in the fact that, while the B0 approximation successfully explores the locality of electronic correlation, higher-order increments are responsible to treat delocalization effect. Therefore, for delocalized systems, r_t must be determined carefully. In this way, our method works well both for compact and extended, conjugated systems. The systems tractable can now be extended to those containing more than 100 correlated electrons. Our new method is also flexible, fully automatic, and highly parallelized. It can be combined not only with CCSD and CCSD(T), but also with F12 explicit correlation methods. This generalization, together with the application to actual chemical problems, are currently under investigation.

■ ASSOCIATED CONTENT

● Supporting Information

The decomposition of large systems **5**, **6**, **7**, **8**, and **10**, the benchmark of dual-basis set correlation methods with 15 small molecules, and the geometries of **1–10** can be found in the Supporting Information. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*Tel.: ++49 (0)221-470-6893. Fax: ++49 (0)221-470-6896. E-mail: zhangjunqcc@gmail.com (J.Z.), m.dolg@uni-koeln.de (M.D.).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank the Deutsche Forschungsgemeinschaft (DFG) for its support.

■ REFERENCES

- (1) Purvis, G. D.; Bartlett, R. J. A Full Coupled-cluster Singles and Doubles Model: The Inclusion of Disconnected Triples. *J. Chem. Phys.* **1982**, *76*, 1910–1918.
- (2) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A Fifth-Order Perturbation Comparison of Electron Correlation Theories. *Chem. Phys. Lett.* **1989**, *157*, 479–483.
- (3) Thom H. Dunning, J. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- (4) Kendall, R. A.; Thom H. Dunning, J.; Harrison, R. J. Electron Affinities of the First-Row Atoms Revisited. Systematic Basis Sets and Wave Functions. *J. Chem. Phys.* **1992**, *96*, 6796–6806.
- (5) Woon, D. E.; Thom H. Dunning, J. Gaussian Basis Sets for Use in Correlated Molecular Calculations. III. The Atoms Aluminum through Argon. *J. Chem. Phys.* **1993**, *98*, 1358–1371.
- (6) Woon, D. E.; Thom H. Dunning, J. Gaussian Basis Sets for Use in Correlated Molecular Calculations. IV. Calculation of Static Electrical Response Properties. *J. Chem. Phys.* **1994**, *100*, 2975–2988.
- (7) Woon, D. E.; Thom H. Dunning, J. Gaussian Basis Sets for Use in Correlated Molecular Calculations. V. Core-Valence Basis Sets for Boron through Neon. *J. Chem. Phys.* **1995**, *103*, 4572–4585.
- (8) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Basis-set Convergence in Correlated Calculations on Ne, N₂, and H₂O. *Chem. Phys. Lett.* **1998**, *286*, 243–252.
- (9) Fast, P. L.; Sánchez, M. L.; Truhlar, D. G. Infinite Basis Limits in Electronic Structure Theory. *J. Chem. Phys.* **1999**, *111*, 2921–2926.
- (10) Bak, K. L.; Jørgensen, P.; Olsen, J.; Helgaker, T.; Klopper, W. Accuracy of Atomization Energies and Reaction Enthalpies in Standard and Extrapolated Electronic Wave Function/Basis Set Calculations. *J. Chem. Phys.* **2000**, *112*, 9229–9242.
- (11) Varandas, A. J. C. Basis-set Extrapolation of the Correlation Energy. *J. Chem. Phys.* **2000**, *113*, 8880–8887.
- (12) Huh, S. B.; Lee, J. S. Basis Set and Correlation Dependent Extrapolation of Correlation Energy. *J. Chem. Phys.* **2003**, *118*, 3035–3042.
- (13) Schwenke, D. W. The Extrapolation of One-Electron Basis Sets in Electronic Structure Calculations: How It Should Work and How It Can be Made to Work. *J. Chem. Phys.* **2005**, *122*, 014107.
- (14) Bakowies, D. Accurate Extrapolation of Electron Correlation Energies from Small Basis Sets. *J. Chem. Phys.* **2007**, *127*, 164109.
- (15) Feller, D.; Peterson, K. A.; Hill, J. G. On the Effectiveness of CCSD(T) Complete Basis Set Extrapolations for Atomization Energies. *J. Chem. Phys.* **2011**, *135*, 044102.
- (16) Pulay, P. Localizability of Dynamic Electron Correlation. *Chem. Phys. Lett.* **1983**, *100*, 151–154.
- (17) Sæbø, S.; Pulay, P. Local Configuration Interaction: An Efficient Approach for Larger Molecules. *Chem. Phys. Lett.* **1985**, *113*, 13–18.
- (18) Pulay, P.; Sæbø, S. Orbital-invariant Formulation and Second-order Gradient Evaluation in Møller–Plesset Perturbation Theory. *Theor. Chim. Acta.* **1986**, *69*, 357–368.
- (19) Sæbø, S.; Pulay, P. Fourth-Order Møller–Plesset Perturbation Theory in the Local Correlation Treatment. I. Method. *J. Chem. Phys.* **1987**, *86*, 914–922.
- (20) Sæbø, S.; Pulay, P. The Local Correlation Treatment. II. Implementation and Tests. *J. Chem. Phys.* **1988**, *88*, 1884–1890.
- (21) Hampel, C.; Werner, H.-J. Local Treatment of Electron Correlation in Coupled Cluster Theory. *J. Chem. Phys.* **1996**, *104*, 6286–6297.
- (22) Hetzer, G.; Pulay, P.; Werner, H.-J. Multipole Approximation of Distant Pair Energies in Local MP2 Calculations. *Chem. Phys. Lett.* **1998**, *290*, 143–149.
- (23) 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.
- (24) Hetzer, G.; Schütz, M.; Stoll, H.; Werner, H.-J. Low-order Scaling Local Correlation Methods II: Splitting the Coulomb Operator

- in Linear Scaling Local Second-Order Møller–Plesset Perturbation Theory. *J. Chem. Phys.* **2000**, *113*, 9443–9455.
- (25) Schütz, M. Low-order Scaling Local Electron Correlation Methods. III. Linear Scaling Local Perturbative Triples Correction (T). *J. Chem. Phys.* **2000**, *113*, 9986–10001.
- (26) Schütz, M.; Werner, H.-J. Low-order Scaling Local Electron Correlation Methods. IV. Linear Scaling Local Coupled-cluster (LCCSD). *J. Chem. Phys.* **2001**, *114*, 661–681.
- (27) Schütz, M. A New, Fast, Semi-Direct Implementation of Linear Scaling Local Coupled Cluster Theory. *Phys. Chem. Chem. Phys.* **2002**, *4*, 3941–3947.
- (28) Neese, F.; Wennmohs, F.; Hansen, A. Efficient and Accurate Local Approximations to Coupled-electron Pair Approaches: An Attempt to Revive the Pair Natural Orbital Method. *J. Chem. Phys.* **2009**, *130*, 114108.
- (29) Neese, F.; Hansen, A.; Liakos, D. G. Efficient and Accurate Approximations to the Local Coupled Cluster Singles Doubles Method Using a Truncated Pair Natural Orbital Basis. *J. Chem. Phys.* **2009**, *131*, 064103.
- (30) Hansen, A.; Liakos, D. G.; Neese, F. Efficient and Accurate Local Single Reference Correlation Methods for High-spin Open-shell Molecules Using Pair Natural Orbitals. *J. Chem. Phys.* **2011**, *135*, 214102.
- (31) Yang, J.; Kurashige, Y.; Manby, F. R.; Chan, G. K. L. Tensor Factorizations of Local Secondorder Møller–Plesset Theory. *J. Chem. Phys.* **2011**, *134*, 044123.
- (32) Yang, J.; Chan, G. K.-L.; Manby, F. R.; Schütz, M.; Werner, H.-J. The Orbital-specific virtual Local Coupled Cluster Singles and Doubles Method. *J. Chem. Phys.* **2012**, *136*, 144105.
- (33) Krause, C.; Werner, H.-J. Comparison of Explicitly Correlated Local Coupled-cluster Methods with Various Choices of Virtual Orbitals. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7591–7604.
- (34) Li, S.; Ma, J.; Jiang, Y. Linear Scaling Local Correlation Approach for Solving the Coupled Cluster Equations of Large Systems. *J. Comput. Chem.* **2002**, *23*, 237–244.
- (35) Li, S.; Shen, J.; Li, W.; Jiang, Y. An Efficient Implementation of the “Cluster-in-molecule” Approach for Local Electron Correlation Calculations. *J. Chem. Phys.* **2006**, *125*, 074109.
- (36) Li, W.; Piecuch, P.; Gour, J. R.; Li, S. Local Correlation Calculations Using Standard and Renormalized Coupled-Cluster Approaches. *J. Chem. Phys.* **2009**, *131*, 114109.
- (37) Li, W.; Guo, Y.; Li, S. A Refined Cluster-in-Molecule Local Correlation Approach for Predicting the Relative Energies of Large Systems. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7854–7862.
- (38) Mochizuki, Y.; Yamashita, K.; Nakano, T.; Okiyama, Y.; Fukuzawa, K.; Taguchi, N.; Tanaka, S. Higher-order Correlated Calculations Based on Fragment Molecular Orbital Scheme. *Theor. Chem. Acc.* **2011**, *130*, 515–530.
- (39) Flocke, N.; Bartlett, R. J. A Natural Linear Scaling Coupled-Cluster Method. *J. Chem. Phys.* **2004**, *121*, 10935–10944.
- (40) Hughes, T. F.; Flocke, N.; Bartlett, R. J. Natural Linear-Scaled Coupled-Cluster Theory with Local Transferable Triple Excitations: Applications to Peptides. *J. Phys. Chem. A* **2008**, *112*, 5994–6003.
- (41) Subotnik, J. E.; Head-Gordon, M. A Local Correlation Model that Yields Intrinsically Smooth Potential-Energy Surfaces. *J. Chem. Phys.* **2005**, *123*, 064108.
- (42) Subotnik, J. E.; Sodt, A.; Head-Gordon, M. A Near Linear-Scaling Smooth Local Coupled Cluster Algorithm for Electronic Structure. *J. Chem. Phys.* **2006**, *125*, 074116.
- (43) Sodt, A.; Subotnik, J. E.; Head-Gordon, M. Linear Scaling Density Fitting. *J. Chem. Phys.* **2006**, *125*, 194109.
- (44) Subotnik, J. E.; Sodt, A.; Head-Gordon, M. The Limits of Local Correlation Theory: Electronic Delocalization and Chemically Smooth Potential Energy Surfaces. *J. Chem. Phys.* **2008**, *128*, 034103.
- (45) Subotnik, J. E.; Head-Gordon, M. Exploring the Accuracy of Relative Molecular Energies with Local Correlation Theory. *J. Phys.: Condens. Matter* **2008**, *20*, 294211.
- (46) Ziolkowski, M.; Jansik, B.; Jørgensen, P.; Olsen, J. Maximum Locality in Occupied and Virtual Orbital Spaces Using a Least-Change Strategy. *J. Chem. Phys.* **2009**, *131*, 124112.
- (47) Ziolkowski, M.; Jansik, B.; Kjærbørgard, T.; Jørgensen, P. Linear Scaling Coupled Cluster Method with Correlation Energy Based Error Control. *J. Chem. Phys.* **2010**, *133*, 014107.
- (48) Kristensen, K.; Ziolkowski, M.; Jansik, B.; Kjærbørgard, T.; Jørgensen, P. A Locality Analysis of the Divide-Expand-Consolidate Coupled Cluster Amplitude Equations. *J. Chem. Theor. Comput.* **2011**, *7*, 1677–1694.
- (49) Claeysens, F.; Harvey, J. N.; Manby, F. R.; Mata, R. A.; Mulholland, A. J.; Ranaghan, K. E.; Schütz, M.; Thiel, S.; Thiel, W.; Werner, H.-J. High-Accuracy Computation of Reaction Barriers in Enzymes. *Angew. Chem.* **2006**, *118*, 7010–7013.
- (50) Mata, R. A.; Werner, H.-J.; Thiel, S.; Thiel, W. Toward Accurate Barriers for Enzymatic Reactions: QM/MM Case Study on *p*-Hydroxybenzoate Hydroxylase. *J. Chem. Phys.* **2008**, *128*, 025104.
- (51) Dieterich, J. M.; Werner, H.-J.; Mata, R. A.; Metz, S.; Thiel, W. Reductive Half-reaction of Aldehyde Oxidoreductase Toward Acetaldehyde: Ab Initio and Free Energy Quantum Mechanical/Molecular Mechanical Calculations. *J. Chem. Phys.* **2010**, *132*, 035101.
- (52) Stoll, H. The Correlation Energy of Crystalline Silicon. *Chem. Phys. Lett.* **1992**, *191*, 548–552.
- (53) Stoll, H. On the Correlation Energy of Graphite. *J. Chem. Phys.* **1992**, *97*, 8449–8454.
- (54) Stoll, H. Correlation Energy of Diamond. *Phys. Rev. B* **1992**, *46*, 6700–6704.
- (55) Paulus, B.; Fulde, P.; Stoll, H. Electron Correlations for Ground-State Properties of Group-IV Semiconductors. *Phys. Rev. B* **1995**, *51*, 10572–10578.
- (56) Doll, K.; Dolg, M.; Fulde, P.; Stoll, H. Correlation Effects in Ionic Crystals: The Cohesive Energy of MgO. *Phys. Rev. B* **1995**, *52*, 4842–4848.
- (57) Paulus, B.; Fulde, P.; Stoll, H. Cohesive Energies of Cubic III–V Semiconductors. *Phys. Rev. B* **1996**, *54*, 2556–2560.
- (58) Nesbet, R. K. Atomic Bethe–Goldstone Equations. I. The Be Atom. *Phys. Rev.* **1967**, *155*, 51–55.
- (59) Nesbet, R. K. Atomic Bethe–Goldstone Equations. II. The Ne Atom. *Phys. Rev.* **1967**, *155*, 56–58.
- (60) Nesbet, R. K. Atomic Bethe–Goldstone Equations. III. Correlation Energies of Ground States of Be, B, C, N, O, F, and Ne. *Phys. Rev.* **1968**, *175*, 2–9.
- (61) Stoll, H.; Paulus, B.; Fulde, P. On the Accuracy of Correlation-Energy Expansions in Terms of Local Increments. *J. Chem. Phys.* **2005**, *123*, 144108.
- (62) Paulus, B. The Method of Increments—A Wavefunction-Based Ab Initio Correlation Method for Solids. *Phys. Rep.* **2006**, *428*, 1–52.
- (63) Kladko, K.; Fulde, P. On the Properties of Cumulant Expansions. *Int. J. Quantum Chem.* **1998**, *66*, 377–389.
- (64) Fulde, P. *Electron Correlations in Molecules and Solids*, 3rd ed.; Springer–Verlag: Berlin, 1995; pp 81–106.
- (65) Nitzan, A. *Chemical Dynamics in Condensed Phases*; Oxford University Press: New York, 2006; pp 241–242.
- (66) Chipot, C.; Pohorille, A., Eds. *Free Energy Calculations*; Springer: Berlin, 2007; pp 40–42.
- (67) Faddeev, L. D. Scattering Theory for a Three-Particle System. *Sov. Phys.—JETP* **1961**, *12*, 1014.
- (68) Faddeev, L. D.; Merkuriev, S. P. *Quantum Scattering Theory for Several Particle Systems*; Kluwer Academic Publishing: Dordrecht, The Netherlands, 1993; pp 39–60.
- (69) Chervanyov, A. Incremental Expansion of the Classical Partition Functions. *Phys. Lett. A* **2001**, *286*, 30–33.
- (70) Malek, J.; Flach, S.; Kladko, K. Incremental Expansions for the Ground-State Energy of the Two-Dimensional Hubbard Model. *Phys. Rev. B* **1999**, *59*, R5273–R5276.
- (71) Friedrich, J.; Hanrath, M.; Dolg, M. Fully Automated Implementation of the Incremental Scheme: Application to CCSD Energies for Hydrocarbons and Transition Metal Compounds. *J. Chem. Phys.* **2007**, *126*, 154110.

- (72) Friedrich, J. Incremental Scheme for Intermolecular Interactions: Benchmarking the Accuracy and the Efficiency. *J. Chem. Theory Comput.* **2012**, *8*, 1597–1607.
- (73) Friedrich, J.; Dolg, M. Implementation and Performance of a Domain-Specific Basis Set Incremental Approach for Correlation Energies: Applications to Hydrocarbons and a Glycine Oligomer. *J. Chem. Phys.* **2008**, *129*, 244105.
- (74) Friedrich, J.; Dolg, M. Fully Automated Incremental Evaluation of MP2 and CCSD(T) Energies: Application to Water Clusters. *J. Chem. Theory Comput.* **2009**, *5*, 287–294.
- (75) Friedrich, J.; Tew, D. P.; Klopper, W.; Dolg, M. Automated Incremental Scheme for Explicitly Correlated Methods. *J. Chem. Phys.* **2010**, *132*, 164114.
- (76) Friedrich, J.; Walczak, K. Incremental CCSD(T)(F12)IMP2-F12—A Method to Obtain Highly Accurate CCSD(T) Energies for Large Molecules. *J. Chem. Theory Comput.* **2013**, *9*, 408–417.
- (77) Friedrich, J.; Hanrath, M.; Dolg, M. Energy Screening for the Incremental Scheme: Application to Intermolecular Interactions. *J. Phys. Chem. A* **2007**, *111*, 9830–9837.
- (78) Friedrich, J.; Hanrath, M.; Dolg, M. Using Symmetry in the Framework of the Incremental Scheme: Molecular Applications. *Chem. Phys.* **2008**, *346*, 266–274.
- (79) Friedrich, J.; Walczak, K.; Dolg, M. Evaluation of Core and Core-Valence Correlation Contributions Using the Incremental Scheme. *Chem. Phys.* **2009**, *356*, 47–53.
- (80) Friedrich, J.; Coriani, S.; Helgaker, T.; Dolg, M. Implementation of the Incremental Scheme for One-electron First-order Properties in Coupled-Cluster Theory. *J. Chem. Phys.* **2009**, *131*, 154102.
- (81) Yang, J.; Dolg, M. Evaluation of Electronic Correlation Contributions for Optical Tensors of Large Systems Using the Incremental Scheme. *J. Chem. Phys.* **2007**, *127*, 084108.
- (82) Friedrich, J.; Perl, E.; Roatsch, M.; Spickermann, C.; Kirchner, B. Coupled Cluster in Condensed Phase. Part I: Static Quantum Chemical Calculations of Hydrogen Fluoride Clusters. *J. Chem. Theory Comput.* **2011**, *7*, 843–851.
- (83) Spickermann, C.; Perl, E.; von Domaros, M.; Roatsch, M.; Friedrich, J.; Kirchner, B. Coupled Cluster in Condensed Phase. Part II: Liquid Hydrogen Fluoride from Quantum Cluster Equilibrium Theory. *J. Chem. Theory Comput.* **2011**, *7*, 868–875.
- (84) Fedorov, D. G.; Kitaura, K. The Importance of Three-Body Terms in the Fragment Molecular Orbital Method. *J. Chem. Phys.* **2004**, *120*, 6832–6840.
- (85) Matsuda, H. Physical Nature of Higher-Order Mutual Information: Intrinsic Correlations and Frustration. *Phys. Rev. E* **2000**, *62*, 3096–3102.
- (86) Elrod, M. J.; Saykally, R. J. Many-Body Effects in Intermolecular Forces. *Chem. Rev.* **1994**, *94*, 1975–1997.
- (87) Foster, J. M.; Boys, S. F. Canonical Configurational Interaction Procedure. *Rev. Mod. Phys.* **1960**, *32*, 300–302.
- (88) Forgy, E. Cluster Analysis of Multivariate Data: Efficiency Versus Interpretability of Classifications. *Biometrics* **1965**, *21*, 768–780.
- (89) MacQueen, J. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Vol. I, Statistics; LeCam, L. M., Neyman, J., Eds.; University of California Press: Berkeley, CA, 1965.
- (90) Downs, G. M.; Barnard, J. M. Clustering Methods and Their Uses in Computational Chemistry. *Rev. Comput. Chem.* **2002**, *18*, 1–40.
- (91) Edmiston, C.; Ruedenberg, K. Localized Atomic and Molecular Orbitals. *Rev. Mod. Phys.* **1963**, *35*, 457–464.
- (92) Karypis, G.; Kumar, V. A Fast and High Quality Multilevel Scheme for Partitioning Irregular Graphs. *SIAM J. Sci. Comput.* **1998**, *20*, 359–392.
- (93) Jurgens-Lutovsky, R.; Almlöf, J. Dual Basis Sets in Calculations of Electron Correlation. *Chem. Phys. Lett.* **1991**, *178*, 451–454.
- (94) Havriliak, S.; King, H. F. Rydberg Radicals. I. Frozen-Core Model for Rydberg Levels of the Ammonium Radical. *J. Am. Chem. Soc.* **1983**, *105*, 4–12.
- (95) Havriliak, S.; Furlani, T. R.; King, H. F. Rydberg Levels of the Sodium Atom and the Ammonium Radical Calculated by Perturbation Theory. *Can. J. Phys.* **1984**, *62*, 1336–1346.
- (96) Wolinski, K.; Pulay, P. Second-Order Møller–Plesset Calculations with Dual Basis Sets. *J. Chem. Phys.* **2003**, *118*, 9497–9503.
- (97) Deng, J.; Gill, P. M. W. Communication: A New Approach to Dual-Basis Second-Order Møller–Plesset Calculations. *J. Chem. Phys.* **2011**, *134*, 081103.
- (98) Steele, R. P.; Head-Gordon, M.; Tully, J. C. Ab Initio Molecular Dynamics with Dual Basis Set Methods. *J. Phys. Chem. A* **2010**, *114*, 11853–11860.
- (99) Löwdin, P.-O. On the Non-Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals. *J. Chem. Phys.* **1950**, *18*, 365–375.
- (100) Carlson, B. C.; Keller, J. M. Orthogonalization Procedures and the Localization of Wannier Functions. *Phys. Rev.* **1957**, *105*, 102–103.
- (101) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623–11627.
- (102) Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.
- (103) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; 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.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, Revision C.02*. Gaussian, Inc.: Wallingford, CT, 2004.
- (104) Řezáč, J.; Riley, K. E.; Hobza, P. Benchmark Calculations of Noncovalent Interactions of Halogenated Molecules. *J. Chem. Theory Comput.* **2012**, *8*, 4285–4292.
- (105) Yoo, S.; Aprá, E.; Zeng, X. C.; Xantheas, S. S. High-Level Ab Initio Electronic Structure Calculations of Water Clusters (H₂O)₁₆ and (H₂O)₁₇: A New Global Minimum for (H₂O)₁₆. *J. Phys. Chem. Lett.* **2010**, *1*, 3122–3127.
- (106) Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M.; Celani, P.; Korona, T.; Lindh, R.; Mitrushenkov, A.; Rauhut, G.; Shamasundar, K. R.; Adler, T. B.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Goll, E.; Hampel, C.; Hesselmann, A.; Hetzer, G.; Hrenar, T.; Jansen, G.; Köppl, C.; Liu, Y.; Lloyd, A. W.; Mata, R. A.; May, A. J.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; O'Neill, D. P.; Palmieri, P.; Pflüger, K.; Pitzer, R.; Reiher, M.; Shiozaki, T.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T.; Wang, M.; Wolf, A. MOLPRO, version 2010.1, a package of ab initio programs. see <http://www.molpro.net>.
- (107) Holliday, J. D.; Rodgers, S. L.; Willett, P.; Chen, M.-Y.; Mahfouf, M.; Lawson, K.; Mullier, G. Clustering Files of Chemical Structures Using the Fuzzy *k*-Means Clustering Method. *J. Chem. Inf. Model.* **2004**, *44*, 894–902.
- (108) Politzer, P.; Murray, J. S.; Clark, T. Halogen Bonding: an Electrostatically-driven Highly Directional Noncovalent Interaction. *Phys. Chem. Chem. Phys.* **2010**, *12*, 7748–7757.

(109) Schmittel, M.; Strittmatter, M.; Kiau, S. Switching from the Myers Reaction to a New Thermal Cyclization Mode in Enyne Allenes. *Tetrahedron Lett.* **1995**, 36, 4975–4978.

(110) Schmittel, M.; Strittmatter, M.; Vollmann, K.; Kiau, S. Intramolecular Formal Diels–Alder Reaction in Enyne Allenes. A New Synthetic Route to Benzofluorenes and Indeno[1,2- g]quinolines. *Tetrahedron Lett.* **1996**, 37, 999–1002.

(111) Schmittel, M.; Keller, M.; Kiau, S.; Strittmatter, M. A Surprising Switch from the Myers– Saito Cyclization to a Novel Biradical Cyclization in Enyne Allenes: Formal Diels–Alder and Ene Reactions with High Synthetic Potential. *Chem.–Eur. J.* **1997**, 3, 807–816.

(112) Bekele, T.; Christian, C. F.; Lipton, M. A.; Singleton, D. A. Concerted” Transition State, Stepwise Mechanism. Dynamics Effects in C₂–C₆ Enyne Allene Cyclizations. *J. Am. Chem. Soc.* **2005**, 127, 9216–9223.