9830 *J. Phys. Chem. A* **2007,** *111,* 9830−9837

Energy Screening for the Incremental Scheme: Application to Intermolecular Interactions

Joachim Friedrich,* Michael Hanrath, and Michael Dolg

Institut für Theoretische Chemie, Universität zu Koeln, Greinstrasse 4, D-50939 Koeln, Germany

Received: March 21, 2007; In Final Form: July 13, 2007

A systematic screening procedure for small contributions in the incremental expansion of the correlation energy is presented. The performance of the proposed scheme is checked for the calculation of intermolecular interactions in realistic test systems as large as a guanine−cytosine base pair. It is found that the computational cost for the incremental expansion can be reduced considerably without significant loss of accuracy. Typically, the errors of the systems investigated here amount to <3.4, 0.22, and 0.06% for second-, third-, and fourth-order expansions, respectively. For almost all cases, the error in the total correlation energy can be kept below 1 kcal/mol with respect to the canonical CCSD result if the incremental series is truncated in a proper way.

1. Introduction

The correct description of noncovalent interactions, such as hydrogen bonds, $\pi-\pi$ interactions, and aliphatic C−H/π interactions, are a challenge for theoretical models. These interactions are of high interest, because they play an important role in biological systems such as DNA or proteins.[1−3] Another interesting interaction is the aurophilic attraction, which has been studied by several groups.[4−6] Since the nature of this attraction is largely dominated by van der Waals forces,[7−9] standard density functional theory (DFT) fails to describe these interactions correctly. Such recently developed functionals as PWB6K,[10] BR,[11] XX,[11] BH&H,[12] M05-2X,[13] or vdW-DF[14] achieve an acceptable description of these systems but are not suitable for highly accurate calculations. An alternative approach to obtain more accurate interaction energies for intermolecular systems is the symmetry-adapted intermolecular perturbation theory,[15,16] which can also be combined with density functional theory.[17−19]

One of today's most accurate systematically improvable electronic structure methods is the coupled cluster ansatz (CC). The main drawback of the CC-based models is the strong dependence of the computational effort on the one-particle basis set, which limits the application of the coupled cluster singles and doubles method (CCSD) to small to medium-sized molecules. To overcome this problem, a lot of effort has been made in the development of local correlation methods during the past decade. Ground-state methods, such as LMP2,[20−28] LCCD, NLSCC, LCCSD, or LCCSD(T),[29−35] were developed in several groups. A conceptually different approach is the incremental scheme of Stoll,[36−38] which is based on the earlier ideas of Nesbet.[39] Within the framework of the incremental scheme, the occupied orbitals are localized and grouped into local domains. In the next step, correlation calculations are performed for all single domains, all pairs, etc. until the desired accuracy is reached. The correlation energy of the total system can be expanded as

$$E_{\mathrm{corr}} = \sum_i \Delta\epsilon_i + \frac{1}{2!}\sum_{ij}\Delta\epsilon_{ij} + \frac{1}{3!}\sum_{ijk}\Delta\epsilon_{ijk} + ... \qquad (1)$$

$$\Delta\epsilon_i = \epsilon_i \qquad \Delta\epsilon_{ij} = \epsilon_{ij} - \Delta\epsilon_i - \Delta\epsilon_j$$

Here, ϵ_i is the correlation energy of the subsystem i, and ϵ_{ij} is the correlation energy of the combined subsystems i and j. The prefactors of $(1/\mathcal{O}_i!)$ are introduced to avoid a double counting of contributions. If the summation indices are restricted in a way that every combination occurs only once, the prefactors can be dropped (vide infra). The general expression for an increment is given as

$$\Delta\epsilon_{i_1 i_2 ... i_t} = \epsilon_{i_1 i_2 ... i_t} - \sum_{I_{t-1}}\Delta\epsilon_{I_{t-1}} - \sum_{I_{t-2}}\Delta\epsilon_{I_{t-2}} - ... - \sum_{I_2}\Delta\epsilon_{I_2} - \sum_{I_1}\Delta\epsilon_{I_1} \qquad (2)$$

where the index I_{t-v} is defined as the elements of the power set of $\{i_1, i_2, ..., i_t\}$ ($\mathscr{P}(\{i_1, i_2, ..., i_t\})$) with the cardinality $t - v$ (v runs from 1 to ($t - 1$)). For molecules, the series in eq 1 terminates at an order equal to the number of domains and, provided no other approximations are made, always yields the exact correlation energy because the evaluation of the highest-order increment is equivalent to the full calculation. Thus, the application of eq 1 has no advantages. However, since local orbitals usually decay very rapidly, we can expect that the series can be truncated at low order[36,40−44]. A great advantage of the incremental scheme is that it can be adapted to the multireference case[45−49] as well as to periodic systems.[40,41,50−62]

Since many increments in eq 1 are negligibly small, it is very important to remove these terms before they are explicitly calculated. The goal of this work is to establish an efficient screening method to obtain a low-order scaling behavior with respect to the number of calculations. As test cases, we choose weak intermolecular interactions, which are especially challenging with regard to the required accuracy. We want to emphasize that the incremental scheme provides a general framework for the evaluation of correlation contributions for a large variety of systems extending significantly beyond the special cases considered in the present work (cf., e.g., refs 43 and 44).

2. Theory

2.1. Notation. In order to obtain a compact notation, we introduce the variable $\mathbb{X}$ as the summation index, which runs

10.1021/jp072256y CCC: $37.00 © 2007 American Chemical Society
Published on Web 09/13/2007

over all members of the power set of the set of domains $\mathscr{P}(\mathbb{D})$ ϵ_i up to a given cardinality. Using this notation, we can rewrite eq 1

$$E_{\mathrm{corr}} = \sum_{\substack{\mathrm{x} \\ \mathbb{X}\in\mathscr{P}(\mathbb{D})\wedge|\mathbb{X}|\leq\mathscr{O}}} \Delta\epsilon_{\mathbb{X}} \tag{3}$$

$$\mathbb{D} = \text{set of domains, } \mathbb{D} = \{D_1, D_2, ...\}$$

$$\mathscr{P}(\mathbb{D}) = \text{power set of the set of domains}$$

$$\mathscr{O} = \text{order of the expansion}$$

2.2. Obtaining Groups of Occupied Orbitals. Within our implementation, the occupied orbitals are localized with a Foster−Boys procedure[63] using the algorithm of Edmiston and Ruedenberg.[64] In the next step, we build the centers of charge for the localized orbitals using the diagonal elements of the dipole integrals in the LMO basis

$$\phi_a \rightarrow \vec{R}_a: = \begin{pmatrix} \langle\phi_a|x|\phi_a\rangle \\ \langle\phi_a|y|\phi_a\rangle \\ \langle\phi_a|z|\phi_a\rangle \end{pmatrix} = \begin{pmatrix} x_a \\ y_a \\ z_a \end{pmatrix} \tag{4}$$

Equation 4 is used to map the set of occupied orbitals $\mathbb{O}$ to a set of vectors in real space. Now the distance matrix, D, of all vector pairs of the set is constructed. Furthermore, the connectivity matrix C for an edge-weighted graph of the occupied orbitals is defined as:

$$C_{ij} = \left\{ \begin{matrix} 10^8 & \text{if } D_{ij} \leq t \wedge \dfrac{w}{D_{ij}} \geq 10^8 \\ \dfrac{w}{D_{ij}} & \text{if } D_{ij} \leq t \wedge \dfrac{w}{D_{ij}} < 10^8 \\ 0 & \text{if } D_{ij} > t \end{matrix} \right\} \tag{5}$$

where t is a distance threshold and w is a constant stretching factor of 10^4. The factor of 10^8 enters as an approximation of infinity in the regime of 32 bit integers. The first two conditions control the value of the weight factor of the edge, and the third condition defines far distant orbital pairs as not connected. We use Metis graph partitioning[65] to divide the set of occupied orbitals into disjoint subsets $\mathbb{O}_i$. The domain_size_parameter dsp controls the number_of_parts parameter nop,

$$\text{nop} = \frac{\text{no. of occupied orbitals}}{\text{dsp}} \tag{6}$$

which is needed for the graph partitioning and roughly corresponds to the number of domains. Since Metis chooses an optimal set of domains, it might happen that a certain set, $\mathbb{O}_a$, is empty and has to be removed from the expansion. Therefore, the nop parameter is not necessarily equal to the number of parts. Since the Metis algorithm minimizes the sum of the cut edge weights, we control the locality of our domains by the definition of the connectivity matrix, C.

2.3. Excitation Spaces for One-Site Domains. Our goal is to obtain a virtual space, $\mathbb{V}_{\phi_a}$, for every occupied orbital, ϕ_a:

$$\phi_a \rightarrow \{\tilde{\phi}_{\mathrm{i}}^{\mathrm{AO}}\} \tag{7}$$

This is similar to the excitation domains in the well-established local correlation methods of Pulay et al.[23] or Werner et al.[24] Our virtual space is spanned by a set of projected atomic orbitals (PAO's) $\{\tilde{\phi}_{\mathrm{i}}^{\mathrm{AO}}\}$. The set of PAOs is constructed according to[23,29]

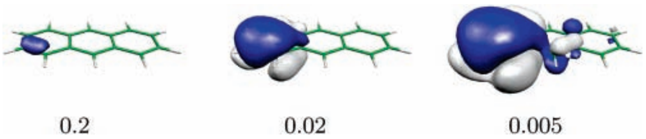

Figure 1. A localized orbital of anthracene, isosurface values in atomic units.

$$|\tilde{\phi}_i^{\mathrm{AO}}\rangle = |\phi_i^{\mathrm{AO}}\rangle - \sum_j^{\mathrm{occ}} \langle\phi_j^{\mathrm{MO}}|\phi_i^{\mathrm{AO}}\rangle|\phi_j^{\mathrm{MO}}\rangle \tag{8}$$

In local orbitals, we usually find that an AO function is important if its center is close to the center of charge, and the it is less important if its center is far away from the center of charge. Guided by the decay of localized occupied orbitals (cf., e.g., Figure 1), we can restrict the excitation space of an occupied orbital according to

$$\int(\phi_a^{\mathrm{MO}} - \bar{\phi}_a^{\mathrm{MO}})^2\, \mathrm{d}\tau \leq t_{\mathrm{dens}} \tag{9}$$

Since we are using local orbitals, we can apply the AO representation of $\bar{\phi}_a^{\mathrm{MO}}$ to find the most important AO functions in ϕ_a^{MO}.

$$\phi_a^{\mathrm{MO}} = \sum_i c_i \phi_i^{\mathrm{AO}}(\vec{r}, \vec{\lambda})$$

$$\bar{\phi}_a^{\mathrm{MO}} = \sum_i \delta_i c_i \phi_i^{\mathrm{AO}}(\vec{r}, \vec{\lambda})$$

$$\delta_i = \{0, 1\}, \text{ corresponding to the neglection or selection} \tag{10}$$

$\vec{\lambda}$ is the center of an AO function. The summation over i in ϕ^{MO} runs over all AO functions, whereas the summation over i in $\bar{\phi}^{\mathrm{MO}}$ runs only over a subset of the AO basis functions. We obtain the important AO functions in two steps: First, we order the AO basis functions according to the distance to the center of charge of ϕ_a^{MO} until eq 9 is fulfilled. All basis functions on a given center are included together. With this procedure, we obtain a $\bar{\phi}_a^{\mathrm{MO}}$, which contains a set of AO functions. From eq 8, we identify the mapping:

$$\phi_i^{\mathrm{AO}} \rightarrow \tilde{\phi}_i^{\mathrm{MO}} \tag{11}$$

The AO functions in $\bar{\phi}_a^{\mathrm{MO}}$ are mapped to their corresponding PAO's according to eq 11. At the end a local excitation space, $\mathbb{V}_{\phi_a}$ for the occupied orbital ϕ_a^{MO} is obtained.

The local excitation space, $\mathbb{V}_i$, for the one-site domain, i, is constructed by unification of the sets of PAOs that correspond to the occupied orbitals in the domain.

$$\mathbb{V}_i = \bigcup_{\phi_a\in\mathbb{O}_i} \mathbb{V}_{\phi_a} \tag{12}$$

2.4. Construction of the *n*-Site Domains. The *n*-site domains are constructed using simple set theory,

$$\mathbb{O}_{\mathrm{K}} = \bigcup_{\lambda\in\mathrm{e}\mathbb{K}} \mathbb{O}_\lambda \qquad \mathbb{V}_{\mathrm{K}} = \bigcup_{\lambda\in\mathbb{K}} \mathbb{V}_\lambda \tag{13}$$

where the set $\mathbb{K} = \{i_1, i_2, ..., i_n\}$ is a subset of the domains, $\mathbb{D}$ with the cardinality n.

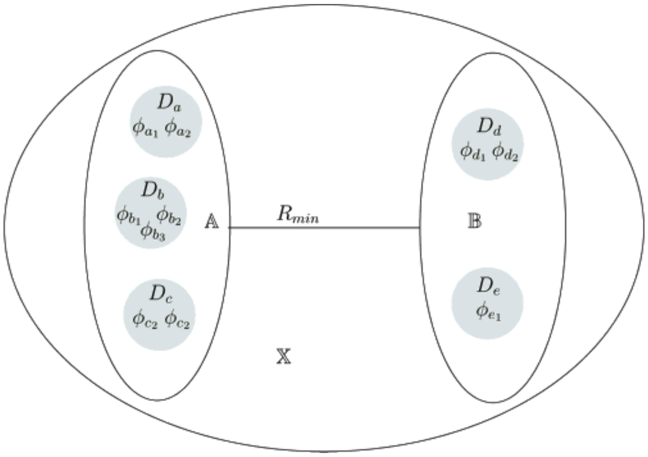

Figure 2. Exemplary domains of the increment $\Delta\epsilon_{\mathbb{X}}$ with the two sets of domains, $\mathbb{A}$ and $\mathbb{B}$, with a minimal distance $R_{\min}$.

2.5. Obtaining the Correlation Energies. We account for the linear dependencies and the nonorthogonality in the PAO space according to a linear transformation that includes a symmetric orthogonalization:

$$D^{-1/2}U^{\dagger}C^{T}S^{AO}CUD^{-1/2} = 1 \tag{14}$$

C is the MO coefficient matrix; S^{AO}, the overlap matrix in AO basis; $D = U^{\dagger}C^{T}S^{AO}CU$, a diagonal matrix; and $\tilde{U}$, the matrix that diagonalizes $\tilde{S} = C^{T}S^{AO}C$. The matrix U is obtained by restricting $\tilde{U}$ to those eigenvectors that correspond to an eigenvalue greater than 10^{-10} in $\tilde{U}^{\dagger}\tilde{S}\tilde{U}$. The MO matrix for the correlation calculation is constructed from the localized occupied orbitals and the new orthonormalized PAOs. All occupied orbitals that are not in the set $\mathbb{O}_{\mathrm{K}}$ are frozen, and all virtual orbitals that are not in $\mathbb{V}_{\mathrm{K}}$ are deleted in the following CCSD calculation. The total CCSD correlation energy is obtained according to eq 1.

2.6. Further Approximations. If the one-site domains of a given two-site domain are far apart, we find that due to the local nature of electron correlation, the corresponding incremental energy correction is small, since the correlation energy of the two-site domain is given as

$$\epsilon_{ij} = \epsilon_i + \epsilon_j + \epsilon(R_{ij}) \tag{15}$$

In the limit of an infinite distance, R_{ij}, between the two domains i and j, the correction $\epsilon(R_{ij})$ vanishes exactly, and the increment $\Delta\epsilon_{ij}$ in eq 16 becomes exactly zero:

$$\begin{aligned}\Delta\epsilon_{ij} &= \epsilon_{ij} - \Delta\epsilon_i - \Delta\epsilon_j\\ &= \epsilon_i + \epsilon_j + \epsilon(R_{ij}) - \Delta\epsilon_i - \Delta\epsilon_j\\ &= \epsilon(R_{ij}) = 0 \text{ for } R_{ij} = \infty\end{aligned} \tag{16}$$

This can be used to truncate the higher order increments by a distance threshold $R_{\min}$ in a similar way. We neglect all increments $\Delta\epsilon_{\mathbb{X}}$ where we can form at least two subsets of $\mathbb{X}$ ($\mathbb{A}$ and $\mathbb{B}$), which have a larger distance than $R_{\min}$. An example is given in Figure 2.

1. Truncation Procedure. We can reduce the question of the importance of a given n-site increment for the total energy to a question of graph connectivity. Let $\mathbb{X}$ be an n-site domain, $\mathbb{V}_{\mathbb{X}}$ be a set of vectors with

$$\mathbb{V}_{\mathbb{X}} = \left\{\vec{R}_a|\phi_a \in \bigcup_{\lambda\in\mathbb{X}} D_\lambda\right\} \tag{17}$$

Further, we define the graph $\mathcal{G}(\mathbb{V}_{\mathbb{X}}, \mathbb{E})$, where the threshold $R_{\min}$ defines the adjacency of $\mathcal{G}(\mathbb{V}_{\mathbb{X}}, \mathbb{E})$.

$$\mathbb{E} = \{\{\vec{x}, \vec{y}\}|\ \vec{x}, \vec{y} \in \mathbb{V}_{\mathbb{X}} \wedge 0 < |\vec{x} - \vec{y}| < R_{\min}\} \tag{18}$$

If it is possible to establish a path from any vertex to any other vertex of a graph, the graph is said to be connected; otherwise, the graph is disconnected.[66] We neglect all n-site domains ($n > 1$) for which $\mathcal{G}(\mathbb{V}_{\mathbb{X}}, \mathbb{E})$ is disconnected. We note that eqs 16 and 18 require compact, one-site domains, for example, the orbitals in the one-site domains must be close in space.

Furthermore, we implemented a dynamic distance threshold for the different orders, $\mathcal{O}_i$, of the expansion to have more degrees of freedom for the truncation. Since high-order increments are usually smaller than low-order increments, we use a distance truncation according to $\left(\frac{f}{\mathcal{O}_i}\right)$. In this case, f is an adjustable parameter, and we test the performance of this dynamic screening for several values of f.

2. Energy Screening. Another way to reduce the total number of calculations is to calculate the energy increments with a lower level method and neglect all terms that are smaller than a given threshold at this level,

$$E_{\mathrm{corr}} = \sum_{\substack{\mathbb{X}\\ \mathbb{X}\in\mathcal{P}(\mathbb{D})\wedge|\mathbb{X}|\leq\mathcal{O}\\ |\Delta\epsilon_{\mathbb{X}}|>E_{\mathrm{thres}}}} \Delta\epsilon_{\mathbb{X}} \tag{19}$$

where $\mathbb{X}$ runs over all members of the power set of the set of the domains $\mathcal{P}(\mathbb{D})$, up to a certain cardinality, as in eq 3.

2.7. Formal Scaling. The formal scaling of the incremental expansion in eq 1 is determined by the number of individual calculations $\mathcal{N}_{\mathrm{calc}}$ and the time for the calculations in the subspaces. The number of calculations is given as

$$\mathcal{N}_{\mathrm{calc}} = \sum_{i=1}^{\mathcal{O}}\binom{n}{i} \tag{20}$$

If we can neglect all energy increments in which the domains are separated by a given distance, the number of calculations per order scales with the size of the system. This means that the total number of calculations scales linearly for a given order and distance threshold, $R_{\min}$. We check the performance of such a distance threshold with respect to the accuracy and computational savings in this work. Another aspect is the truncation of the virtual space for arbitrary domains. This is necessary to obtain virtual subspaces that are independent of the total size of the system.

3. Computational Details

3.1. Geometries. If nothing else is stated, we optimized the geometries with the RI-BP86/SVP method in the TURBOMOLE 5.6[67] quantum chemistry package. Stationary points were characterized by analyzing the Hessian matrix.

3.2. Incremental Calculations. The current implementation contains interfaces to the MOLPRO[68] and DALTON[69] quantum chemistry packages to obtain the molecular orbital coefficient matrix, the overlap matrix in AO basis, and the dipole integrals in AO basis from the previous SCF calculation. After extraction of this data, a Foster−Boys localization[63] with unitary 2 × 2 rotations in the occupied space is performed. The localization is carried out according to the original procedure of Edmiston and Ruedenberg[64] with a threshold of 10^{-12} in $D_{\max}(u_i, u_j) - D(\varphi_i, \varphi_j)$ (for details we refer to eq 26 in ref 64 and eq 15 in ref 70).

TABLE 1: Comparison of the Incremental Energies for the $(H_2O)_8$ Cluster in Figure 3 with the Full CCSD Calculations and Comparison of the Basis Set Effect[a]

basis	order i	ith order correction [au]	$E_{corr}(i)$ [au]	error [kcal/mol]	% E_{corr}
6-31G**	1	−1.648 410	−1.648 410	35.61	96.67
	2	−0.056 524	−1.704 934	0.14	99.99
	3	−0.000 207	−1.705 142	0.01	100.00
	4	−0.000 007	−1.705 149	0.00	100.00
⋮					
exact CCSD			−1.705 151		
cc-pVTZ	1	−2.244 489	−2.244 489		
	2	−0.077 810	−2.322 298		
	3	0.000 111	−2.322 188		

[a] Eight domains, core = 0.

TABLE 2: Convergence Behavior of the Incremental Scheme with Respect to the Density Parameter of Equation 9[a]

$(H_2O)_8$	order i	ith order correction [au]	$E_{corr}(i)$ [au]	error [kcal/mol]	% E_{corr}
density threshold 0.1					
	1	−1.573 940	−1.573 940	32.17	96.85
	2	−0.053 823	−1.627 763	−1.60	100.16
	3	−0.000 571	−1.628 335	−1.96	100.19
⋮					
exact CCSD			−1.625 212		
density threshold 0.01					
	1	−1.575 988	−1.575 988	30.89	96.97
	2	−0.053 820	−1.629 808	−2.88	100.28
	3	0.004 666	−1.625 142	0.04	100.00
⋮					
exact CCSD			−1.625 212		
density threshold 0.001					
	1	−1.578 597	−1.578 597	29.25	97.13
	2	−0.046 443	−1.625 040	0.11	99.99
	3	−0.000 053	−1.625 093	0.07	99.99
⋮					
exact CCSD			−1.625 212		

[a] dsp = 5, 8 domains, core = 0, calculation in 6-31G* basis set of Pople[72] at the Ri-BP86/SVP geometry.

TABLE 3: Performance of the Approximation of the Incremental Scheme with Respect to a Dynamic Distance Threshold $R_{min}(\mathcal{O}_i) = f/\mathcal{O}_i$ for the CCSD/6-31G energy of $(H_2O)_{11}$**[a]

N_{calc}/ otal N_{calc}	f	order i	ith order correction [a.u.]	$E_{corr}(i)$ [a.u.]	error [kcal/mol]	% E_{corr}
11/11	12	1	−2.235 023	−2.235 023	33.35	97.68
16/55		2	−0.049 996	−2.285 019	1.98	99.86
27/165		3	−0.003 239	−2.288 258	−0.05	100.00
0/330		4	0.000 000	−2.288 258	−0.05	100.00
54/561						
11/11	16	1	−2.235 023	−2.235 023	33.35	97.68
41/55		2	−0.052 651	−2.287 674	0.32	99.98
31/165		3	−0.000 246	−2.287 920	0.16	99.99
53/330		4	−0.000 443	−2.288 363	−0.12	100.01
136/561						
11/11	∞	1	−2.235 023	−2.235 023	33.35	97.68
55/55		2	−0.052 874	−2.287 897	0.18	99.99
165/165		3	−0.000 260	−2.288 157	0.01	100.00
330/330		4	−0.000 025	−2.288 182	0.00	100.00
561/561						

[a] Eleven domains, core = 11.

TABLE 4: Performance of the Approximation Scheme Equation 19 for the CCSD/6-31G Energy of the $(H_2O)_{11}$ Cluster Due to an Energy Selection**[a]

E_{thres}	n cal	E_{corr}	error	% E_{corr}
10^{-8}	513	−2.288 182	0.00	100.00
10^{-7}	269	−2.288 179	0.00	100.00
10^{-6}	130	−2.288 167	0.01	100.00
10^{-5}	70	−2.287 968	0.13	99.99
10^{-4}	37	−2.286 425	1.10	99.92

[a] For simplicity, we used the exact incremental energies to estimate the individual contributions of the increments.

TABLE 5: Performance of the Dynamic Distance Threshold $R_{min}(\mathcal{O}_i) = f/\mathcal{O}_i$ with Respect to the Computational Saving for $(H_2O)_{11}$

	∞		16		12	
f order	N calc [%]	error [kcal/mol]	N calc [%]	error [kcal/mol]	N calc [%]	error [kcal/mol]
2	100	0.18	75	0.32	29	1.98
3	100	0.01	19	0.16	16	−0.05
4	100	0.00	16	−0.12	0	−0.05
total	100		24		10	
total cpu [%]	100		19		5	

Orbitals that are treated as frozen in the correlation calculation are excluded from the unitary transformations. In the next step, we calculate the CCSD correlation energies of the domains with MOLPRO or DALTON. The DALTON calculations were performed to check the convergence of the incremental series when the virtual space is truncated according to eq 9. The MOLPRO calculations were performed with the complete virtual space.

Since the number of calculations increases quite fast according to eq 20, we had to tighten the threshold for the SCF energies (1.D-10 a.u), the threshold for the CCSD energies (1.00D-08 a.u), and coefficients (1.00D-5) to prevent the numerical errors affecting the correlation energies.

4. Applications

4.1. Water Clusters. For the correlation energy of systems with intermolecular interactions, such as water clusters, we find that the incremental scheme performs better than for the hydrocarbon compounds or transition metal compounds discussed in ref 43. In the 6-31G** basis, one obtains for $(H_2O)_8$ 99.99% of the correlation energy already at second order. The convergence behavior is similar for the calculations in the larger cc-pVTZ basis in which the full calculation is already infeasible on less than 1.35 GB machines (Table 1).

In Table 2, we compare for $(H_2O)_8$ the convergence behavior of the incremental scheme with respect to the density parameter t_{dens} of eq 9. Naturally, the parameter affects the convergence behavior of the series if the excitation space is restricted too rigorously. With proper values of this parameter, we can still obtain fast convergence in the series, as we can see for the last two examples in Table 2.

Tables 3 and 4 show for $(H_2O)_{11}$ in 6-31G** basis the performance of an energy screening procedure using a dynamic distance threshold $R_{min}(\mathcal{O})$ and an energy threshold, respectively. The number of calculations can be reduced significantly if we use the above distance truncation, as we can see from Table 3. If we compare the reduction of calculations according to an energy threshold from Table 4 with the performance of the truncation, on the basis of graph theory, we see that they behave quite similarly. Note that we used the exact energies to simulate

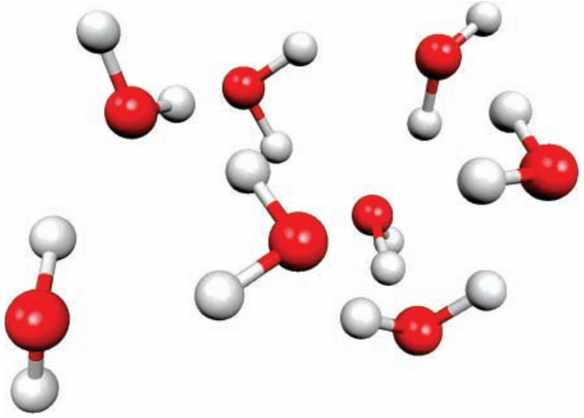

Figure 3. RI-BP86/SVP optimized structure of a set of eight water molecules.

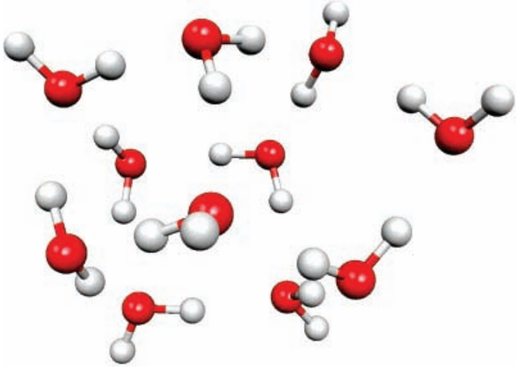

Figure 4. $(H_2O)_{11}$ cluster taken from Bulusu et al.[71]

the energy screening. If approximate values are used, it would be necessary to use a lower threshold to avoid discarding too many relevant contributions due to the approximation error. Table 5 presents the saving of the CPU time for $(H_2O)_{11}$. For this example, we find that we can reduce the calculation time for the incremental calculation to 5−25% of the full incremental calculation. Furthermore, the ratio between the number of calculations of the distance approximated incremental calculation and the full incremental calculation gives an upper bound to the computational time, as compared to the time for the full incremental calculation.

$$\frac{\mathcal{N}_{\mathrm{calc}}(R_{\mathrm{min}}, \mathcal{O})}{\mathcal{N}_{\mathrm{calc}}(\mathcal{O})} > \frac{\text{cpu-time}(R_{\mathrm{min}}, \mathcal{O})}{\text{cpu-time}(\mathcal{O})} \tag{21}$$

This is, in general, true because $R_{\mathrm{min}}(\mathcal{O})$ removes more high-order contributions, which are more time-consuming to evaluate than the low-order contributions.

4.2. π−π/CH Interactions. π−π interactions are very important, since they also can affect the structure of DNA and proteins. The benzene dimer as a model of π−π interactions has been studied by several groups.[3,14,73−78] Another important intermolecular interaction is the CH−π interaction.[79−81] For our proposes, we have chosen the indole−methane complex as studied by Ringer et al.[81]

For the intermolecular interactions between two benzene molecules (Figure 5), we obtain at third-order level almost the exact CCSD energy (Table 6). For the CH−π interaction in Figure 6, we obtain a reasonable energy at third-order level and almost the exact CCSD energy at fourth-order level. We point out that we have fast convergence for the compact 6-31G** basis as well as for the diffuse 6-31++G** basis set.

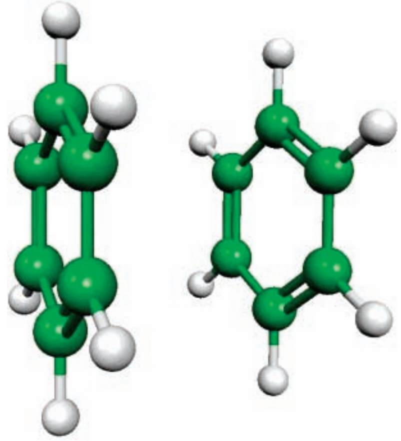

Figure 5. Sandwich structure of the benzene dimer at a benzene−benzene distance of 3.75 Å using the monomer C−H and C−C distances of Gauss and Stanton.[82]

TABLE 6: Comparison of the Incremental Energies with the Full CCSD Calculations for the Molecules in Figures 5 and 6

system	order i	ith order correction [au]	$E_{\mathrm{corr}}(i)$ [au]	error [kcal/mol]	% E_{corr}
$(C_6H_6)_2$[a]					
	1	−1.100 083	−1.100 083	345.59	66.64
	2	−0.573 180	−1.673 263	−14.09	101.36
	3	0.022 393	−1.650 871	−0.04	100.00
	4	0.000 185	−1.650 686	0.08	99.99
⋮					
exact CCSD			−1.650 808		
benzene[b]					
	1	−0.549 574	−0.549 574	171.32	66.81
	2	−0.283 411	−0.832 985	−6.52	101.26
	3	0.010 222	−0.822 762	−0.10	100.02
	4	0.000 235	−0.822 528	0.04	99.99
⋮					
exact CCSD			−0.822 596		
indole-methane complex[c]					
	1	−1.012 639	−1.012 639	274.42	69.84
	2	−0.459 677	−1.472 315	−14.03	101.54
	3	0.022 988	−1.449 327	0.39	99.96
	4	−0.000 546	−1.449 874	0.05	99.99
⋮					
exact CCSD			−1.449 955		
indole[d]					
	1	−0.826 255	−0.826 255	271.37	65.64
	2	−0.453 952	−1.280 206	−13.49	101.71
	3	0.021 990	−1.258 216	0.31	99.96
	4	−0.000 411	−1.258 627	0.05	99.99
⋮					
exact CCSD			−1.258 713		

[a] In 6-31G** basis set of Pople[72,83] (10 domains, core = 12). [b] Five domains, core = 6. [c] 6-31++G** basis set of Pople[72,83] (8 domains, core = 10). [d] Seven domains, core = 9.

4.3. DNA Base Pair. Accurate calculation of the guanine−cytosine base pair is very important to benchmark lower level methods, such as force fields or DFT, since highly accurate wavefunction-based methods are not applicable to large DNA molecules. From a theoretical point of view, these systems are interesting because of the large number of π electrons. Since a localization of conjugated π orbitals yields usually extended localized orbitals, we checked the performance of our local approach for this case. The results of the incremental calculations are given in Table 7. We find that a truncation according to $(16/\mathcal{O}_i)$ yields quite accurate results for this system. At third-

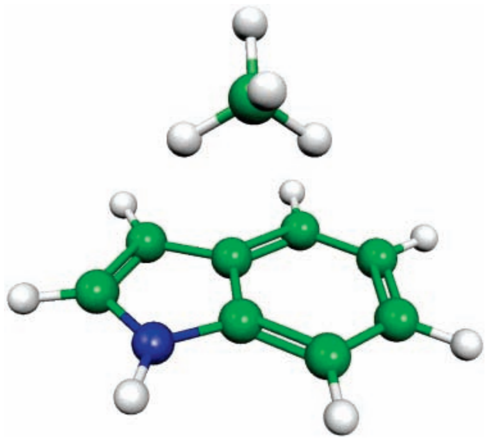

Figure 6. Indole methane complex optimized by Ringer et al.[81]

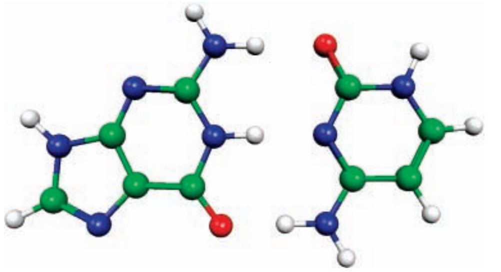

Figure 7. RI-BP86/SVP optimized guanine−cytosine base pair.

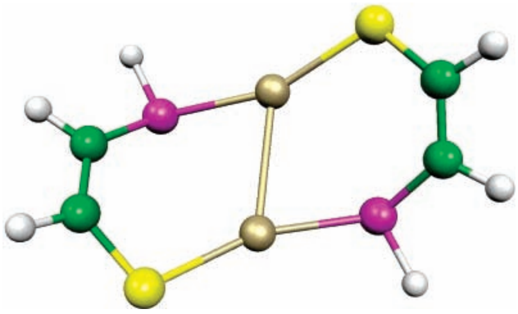

Figure 8. MP2 optimized structure of $Au_2(PH-C_2H_2-S)_2$ using the 6-31G** basis set of Pople[72,83] for P,C,H,S and ECP60MDF/(8s6p5d)/[7s3p4d][84] for Au. In order to polarize the d shell of the gold atoms, we added two f-type polarization functions ($\alpha_f = 0.20, 1.19$).[6]

TABLE 7: Comparison of the Incremental Energies with the Full CCSD Calculations for the Guanine−Cytosine in Figure 7 Using the 6-31G Basis Set of Pople[72,83a]**

guanine−cytosine base pair				
order i	ith order correction [au]	$E_{corr}(i)$ [au]	error [kcal/mol]	% E_{corr}
1	−1.916 622	−1.916 622	592.65	66.99
2	−0.979 382	−2.896 004	−21.92	101.22
3	0.031 816	−2.864 189	−1.96	100.11
4	0.002 382	−2.861 807	−0.46	100.03
exact CCSD		−2.861 067		

[a] dsp = 3, 16 domains, core = 19.

order level, we obtain 100.11% of the correlation energy, whereas we obtain 100.03% at fourth-order level. This corresponds to an absolute error of only −0.46 kcal/mol with respect to the exact CCSD energy. Note that the full CCSD calculation needs 2.5 GB of RAM and 18.5 GB of disk space. The incremental calculations, however, took at most 1.4 GB of RAM and 7.4 GB of disk space. This example demonstrates again the power of the automatic distance truncation because the ratio ($\mathcal{N}_{calc}(R_{min}, \mathcal{O})/\mathcal{N}_{calc}(\mathcal{O})$) is 0.17, that is, 2095 CCSD calculations for $f = 16$ at fourth order are avoided.

TABLE 8: Comparison of the Incremental Energies with the Full CCSD Calculations for $Au_2(PH-C_2H_2-S)_2$ in Figure 8 Using the 6-31G Basis Set of Pople[72,83] for P,C,H,S and ECP60MDF/(8s6p5d)/[7s3p4d][84] for Au[a]**

$Au_2(PH-C_2H_2-S)_2$				
order i	ith order correction [au]	$E_{corr}(i)$ [au]	error [kcal/mol]	% E_{corr}
		dsp 3		
1	−0.904 022	−0.904 022	317.17	64.14
2	−0.553 462	−1.457 485	−30.13	103.41
3	0.0511 70	−1.406 315	1.98	99.78
4	−0.003 270	−1.409 585	−0.07	100.01
		dsp 5		
1	−1.111 903	−1.111 903	186.72	78.89
2	−0.318 708	−1.430 611	−13.27	101.50
3	0.0225 91	−1.408 020	0.91	99.90
4	−0.001 396	−1.409 416	0.03	100.00
exact CCSD		−1.409 467		

[a] dsp = 3, 10 domains, core = 32; dsp = 5, 6 domains, core = 32.

TABLE 9: Binding Energies Per Water for $(H_2O)_{11}$, Interaction Energy for the Benzene Dimer, and the Interaction Energy of the Indole−Methane Dimer[a]

order	ΔE [kcal/mol]	error [kcal/mol]
	$(H_2O)_{11}$	
1	−9.1	3.0
2	−12.0	0.2
3	−12.1	0.0
4	−12.1	0.0
⋮		
exact CCSD	−12.1	
	benzene dimer	
1	2.57	2.94
2	−1.42	−1.05
3	−0.20	0.17
4	−0.38	−0.01
⋮		
exact CCSD	−0.37	
	indole−methane dimer	
1	0.6	3.0
2	−3.0	−0.5
3	−2.4	0.1
4	−2.5	0.0
⋮		
exact CCSD	−2.5	

[a] The errors are given with respect to the exact CCSD calculation.

4.4. The Aurophilic Attraction. To check the performance of the presented approach for the aurophilic intermolecular interaction, we chose the test molecule $Au_2(PH-C_2H_2-S)_2$ which was studied by Mendizabal and Pyykkö[6] ($d^{10}-d^{10}$ interaction). Compared to the other systems in this study we find a relatively slow convergence, especially for small domains. We still have a fairly large error, up to 2 kcal/mol, for the domain size in Table 8 at third-order level. At fourth-order level, we are very close to the exact CCSD energy for both cases. We note that an analysis of the single energy contributions analogous to Table 4 shows that we could reduce the number of calculations significantly without loss of accuracy.

4.5. Interaction Energies. Since the physical quantity of main interest for intermolecular systems is the interaction energy, we calculated the binding energy per water molecule for $(H_2O)_{11}$ and the binding energy for the benzene dimer and for the indole−ethane dimer (Table 9). We find in all cases that the convergence of the incremental scheme for this quantity is fast. This can be expected, since the incremental series for the absolute energies is already very accurate at third or fourth order. For the CCSD/6-31G** binding energy of the benzene dimer, we find that a incremental expansion up to fourth order is required to get the quantitative binding energy. Note that the absolute error is small in this case, too, but the CCSD/6-31G** binding energy is only 0.37 kcal/mol, and therefore, a higher accuracy for the incremental series is required to obtain a quantitative binding energy. Since 10 domains where used for the benzene dimer, we find that a fourth order incremental expansion is still a low-order truncation as compared to the full incremental expansion up to 10th order.

5. Conclusions

We showed that the incremental scheme provides a way to obtain accurate correlation energies for systems for which standard approaches fail due to too high hardware requirements. It was also demonstrated that intermolecular interactions can be recovered accurately at relatively low order. We established a general and automatic procedure to calculate only important increments. This procedure significantly reduces the calculation time in the incremental calculations without leading to significant loss of accuracy and should also be useful for systematic investigations of even larger molecular systems as well as for crystalline compounds.

Acknowledgment. Financial support of the German Science Foundation (DFG) through the priority program 1145 is gratefully acknowledged.

References and Notes

(1) Hunter, C. A.; Singh, J.; Thornton, J. M. *J. Mol. Biol.* **1991**, *218*, 837.

(2) McGaughey, G. B.; Gages, M.; Rappé, A. K. *J. Biol. Chem*. **1998**, *273*, 15458.

(3) Hill, J. G.; Platts, J. A.; Werner, H.-J. *Phys. Chem. Chem. Phys*. **2006**, *8*, 4072.

(4) Pyykkö, P. *Chem. Rev*. **1997**, *97*, 597.

(5) Schmidbaur, H. *Gold Bull*. **2000**, *33*, 1.

(6) Mendizabal, F.; Pyykkö, P. *Phys. Chem. Chem. Phys.* **2004**, *6*, 900.

(7) Pyykkö, P.; Runeberg, N.; Mendizabal, F. *Chem.—Eur. J*. **1997**, *3*, 1451.

(8) Pyykkö, P.; Mendizabal, F. *Chem.—Eur. J.* **1997**, 1458.

(9) Runeberg, N.; Schütz, M.; Werner, H.-J. *J. Chem. Phys*. **1999**, *110*, 7210.

(10) Zhao, Y.; Truhlar, D. G. *Phys. Chem. Chem. Phys*. **2005**, *7*, 2701.

(11) Becke, A. D.; Johnson, E. R. *J. Chem. Phys*. **2005**, *123*, 154101.

(12) Waller, M. P.; Robertazzi, A.; Platts, J. A.; Hibbs, D. E.; Williams, P. A. *J. Comput. Chem*. **2006**, *27*, 491.

(13) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. J. *Chem. Theory Comput*. **2006**, *2*, 364.

(14) Puzder, A.; Dion, M.; Langreth, D. C. *J. Chem. Phys.* **2006**, *124*, 164105.

(15) Jeziorski, B.; Moszynski, R.; Szalewicz. *Chem. Rev.* **1994**, *94*, 1887.

(16) Patkowski, K.; Jeziorski, B.; Szalewicz *J. Chem. Phys*. **2006**, *125*, 154107.

(17) Hesselmann, A.; Jansen, G. *Phys. Chem. Chem. Phys*. **2003**, *5*, 5010.

(18) Hesselmann, A.; Jansen, G.; Schütz, M. *J. Chem. Phys.* **2005**, *122*, 014103.

(19) Hesselmann, A.; Jansen, G.; Schütz, M. *J. Am. Chem. Soc.* **2006**, *128*, 11730.

(20) Pulay, P.; Saebø, S. *Theor. Chim. Acta* **1986**, *69*, 357.

(21) Saebø, S.; Pulay, P. *J. Chem. Phys.* **1987**, *86*, 914.

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

(23) Boughton, J. W.; Pulay, P. J. *Comput. Chem.* **1993**, *14*, 736.

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

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

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

(27) Maslen, P. E.; Lee, M. S.; Head-Gordon, M. *Chem. Phys. Lett.* **2000**, *319*, 205.

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

(29) Hampel, C.; Werner, H.-J. *J. Chem. Phys.* **1996**, *104*, 6286.

(30) Schütz, M. J. *Chem. Phys*. **2000**, *113*, 9986.

(31) Schütz, M.; Werner, H.-J. *J. Chem. Phys.* **2001**, *114*, 661.

(32) Flocke, N.; Bartlett, R. J. *J. Chem. Phys.* **2004**, *121*, 10935.

(33) Subotnik, J. E. M.; Head-Gordon. *J. Chem. Phys.* **2005**, *123*, 64108.

(34) Subotnik, J. E.; Sodt, A.; Head-Gordon, M. *J. Chem. Phys.* **2006**, *125*, 74116.

(35) Auer, A.; Nooijen, M. J. *Chem. Phys*. **2006**, *125*, 24104.

(36) Stoll, H. *Chem. Phys. Lett.* **1992**, *91*, 548.

(37) Stoll, H. *Phys. Rev. B: Condens. Matter Mater. Phys*. **1992**, *46*, 6700.

(38) Stoll, H. J. *Chem. Phys.* **1992**, *97*, 8449.

(39) Nesbet, R. K. *Adv. Chem. Phys.* **1969**, *14*, 1.

(40) Shukla, A.; Dolg, M.; Fulde, P.; Stoll, H. *Phys. Rev. B: Condens. Matter Mater. Phys*. **1999**, *60*, 5211.

(41) Mukhopadhyay, A. B.; Dolg, M. J. *Chem. Phys.* **2004**, *120*, 8734.

(42) Stoll, H.; Paulus, B.; Fulde, P. J. *Chem. Phys.* **2005**, *123*, 144108.

(43) Friedrich, J.; Hanrath, M.; Dolg, M. J. *Chem. Phys.* **2007**, *126*, 154110.

(44) Friedrich, J.; Hanrath, M.; Dolg, M. *Chem. Phys.* **2007**, *338*, 33.

(45) Stoll, H. *Ann. Phys.* **1996**, *5*, 355.

(46) Mödl, M.; Dolg, M.; Fulde, P.; Stoll, H. *J. Chem. Phys.* **1997**, *106*, 1836.

(47) Paulus, B. *Chem. Phys. Lett.* **2003**, *371*, 7.

(48) Bezugly, V.; Birkenheuer, U. *Chem. Phys. Lett.* **2004**, *399*, 57.

(49) Bezugly, V. *Wavefunction-based method for excited-state electron correlations in periodic systems*; Doktorarbeit: Dresden, 2004.

(50) Abdurahman, A.; Shukla, A.; Dolg, M. *J. Chem. Phys.* **2000**, *112*, 4801.

(51) Abdurahman, A.; Shukla, A.; Dolg, M. *Chem. Phys. Lett*. **2000**, *257*, 301.

(52) Abdurahman, A.; Shukla, A.; Dolg, M. *Phys. Rev. B: Condens. Matter Mater. Phys*. **2002**, *65*, 116106.

(53) Buth, C.; Paulus, B. *Phys. Rev. B: Condens. Matter Mater. Phys*. **2006**, *74*, 45122.

(54) Doll, K.; Dolg, M.; Fulde, P.; Stoll, H. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1997**, *55*, 10282.

(55) Kalvoda, S.; Dolg, M.; Flad, H.-J.; Fulde, P.; Stoll, H. *Phys. Rev. B: Condens. Matter Mater. Phys*. **1998**, *57*, 2127.

(56) Albrecht, M.; Paulus, B.; Stoll, H. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1997**, *56*, 7339.

(57) Paulus, B. *Chem. Phys. Lett.* **2003**, *371*, 7.

(58) Paulus, B.; Rosciszewski, K.; Stoll, H. P. *Fulde. Phys. Chem. Chem. Phys.* **2003**, *5*, 5523.

(59) Willnauer, C.; Birkenheuer, U. *J. Chem. Phys.* **2004**, *120*, 11910.

(60) Voloshina, E.; Paulus, B. *Theor. Chem. Acc.* **2005**, *114*, 259.

(61) Voloshina, E.; Paulus, B. *J. Chem. Phys.* **2006**, *124*, 234711.

(62) Gaston, N.; Paulus, B.; Rosciszewski, K.; Schwerdtfeger, P.; Stoll, H. *Phys. Rev. B: Condens. Matter Mater. Phys*. **2006**, *74*, 94102.

(63) Foster, J. M.; Boys, S. F. *Rev. Mod. Phys.* **1960**, *32*, 300.

(64) Edmiston, C.; Ruedenberg, K. *Rev. Mod. Phys.* **1963**, *35*, 457.

(65) Karypis, G.; Kumar, V. *SIAM J. Sci. Comput*. **1998**, *20* (1), 359.

(66) Glossary of graph theory at http://en.wikipedia.org/wiki/Glossary_of_graph_theory.

(67) Ahlrichs, R.; Bär, M.; Baron, H.-P.; Bauernschmitt, R.; Böcker, S.; Ehrig, M.; Eichkorn, K.; Elliott, S.; Furche, F.; Haase, F.; Häser, M.; Horn, H.; Huber, C.; Huniar, U.; Kölmel, C.; Kollwitz, M.; Ochsenfeld, C.; Öhm, H.; Schäfer, A.; Schneider, U.; Treutler, O.; von Arnim, M.; Weigend, F.; Weis, P.; Weiss, H. *Turbomole 5*; Institut für Physikal. Chemie, Universität Karlsruhe, Kaiserstr. 12, D-76128, Karlsruhe, 2002.

(68) Amos, R. D.; Bernhardsson, A.; Berning, A.; Celani, P.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Hampel, C.; Hetzer, G.; Knowles, P. J.; Korona, T.; Lindh, R.; Lloyd, A. W.; McNicholas, S. J.; Manby, F. R.; Meyer, W.; Mura, M. E.; Nicklass, A.; Palmieri, P.; Pitzer, R.; Rauhut, G.; Schütz, M.; Schumann, U.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T.; Werner, H.-J. MOLPRO, a package of ab initio programs designed by H.-J. Werner and P. J. Knowles, version 2002; Tech. Report, University of Birmingham, 2002.

(69) DALTON, a molecular electronic structure program, release 2.0; 2005; see http://www.kjemi.uio.no/software/dalton/dalton.html.

(70) Pipek, J.; Mezey, P. G. *J. Chem. Phys.* **1989**, *90*, 4916.

(71) Bulusu, S.; Yoo, S.; Apra, E.; Xantheas, S.; Zeng, X. C. *J. Phys. Chem. A* **2006**, *110*, 11781.

(72) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257.

(73) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. *J. Am. Chem. Soc.* **2002**, *124*, 104.

(74) Sinnokrot, M. O.; Valeev, E. F.; Sherrill, C. D. *J. Am. Chem. Soc.* **2002**, *124*, 10887.

(75) Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2004**, *108*, 10200.

(76) Zhikol, O.; Shishkin, O. V.; Lyssenko, K. A.; Leszczynski, J. *J. Chem. Phys.* **2005**, *122*, 144104.

(77) Park, Y. C.; Lee, J. S. *J. Phys. Chem. A* **2006**, *110*, 5091.

(78) Podeszwa, R.; Bukowski, R.; Szalewicz, K. *J. Phys. Chem. A* **2006**, *110*, 10354.

(79) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. *J. Am. Chem. Soc*. **2000**, *122*, 3746.

(80) Shibasaki, K.; Fujii, A.; Mikami, N.; Tsuzuki, S. *J. Phys. Chem. A* **2006**, *110*, 4397.

(81) Ringer, A. L.; Figgs, M. S.; Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2006**, *110*, 10822.

(82) Gauss, J.; Stanton, J. F. *J. Phys. Chem. A* **2000**, *104*, 2865.

(83) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213.

(84) Schwerdtfeger, P.; Dolg, M.; Schwarz, W.; Bowmaker, G.; Boyd, P. J. *Chem. Phys.* **1989**, *91*, 1762.