JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

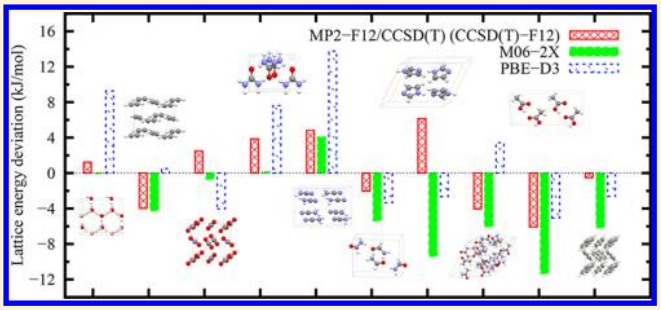
Accurate Prediction of Lattice Energies and Structures of Molecular Crystals with Molecular Quantum Chemistry Methods

Tao Fang, Wei Li, Fangwei Gu, and Shuhua Li*

School of Chemistry and Chemical Engineering, Key Laboratory of Mesoscopic Chemistry of Ministry of Education, Institute of Theoretical and Computational Chemistry, Nanjing University, Nanjing, 210093, People's Republic of China

Supporting Information

ABSTRACT: We extend the generalized energy-based fragmentation (GEBF) approach to molecular crystals under periodic boundary conditions (PBC), and we demonstrate the performance of the method for a variety of molecular crystals. With this approach, the lattice energy of a molecular crystal can be obtained from the energies of a series of embedded subsystems, which can be computed with existing advanced molecular quantum chemistry methods. The use of the field compensation method allows the method to take long-range electrostatic interaction of the infinite crystal environment into account and make the method almost translationally invariant. The computational cost of the present method scales linearly with the number of molecules in the unit cell. Illustrative applications demonstrate that the PBC-GEBF method with explicitly correlated quantum chemistry methods is capable of providing accurate descriptions on the lattice energies and structures for various types of molecular crystals. In addition, this approach can be employed to quantify the contributions of various intermolecular interactions to the theoretical lattice energy. Such qualitative understanding is very useful for rational design of molecular crystals.



1. INTRODUCTION

Accurate prediction of structures and lattice energies of molecular crystals from first-principles is of great significance in many areas of science. However, it still remains a challenging issue for *ab initio* calculations, mainly because the structures of molecular crystals result from the subtle balance of various types of noncovalent interactions such as hydrogen-bonding and dispersion interactions. Periodic density functional theory (DFT) is the most popular method for solid-state materials if appropriate exchange-correlation functionals are employed. In recent years, many popular functionals have been demonstrated to provide satisfactory descriptions for molecular crystals, especially when the dispersion interaction obtained with other methods is added to the DFT energy.[1−5] Nevertheless, the lattice energies obtained with such DFT-based methods for general molecular crystals may reach an accuracy of ∼5 kJ/mol if an appropriate functional is employed.[2] Periodic second-order Møller−Plesset perturbation theory (MP2) has been applied to some molecular crystals composed of small-sized molecules.[6−8] However, the applications of this approach are still limited, because of the high computational cost and its overestimation of intermolecular interactions in some cases.[9] The periodic version of coupled cluster with singles, doubles, and perturbative triples, CCSD(T), which is the "gold standard" method for molecules, has not been implemented for routine applications.

Different from periodic versions of traditional quantum chemistry methods, fragment-based methods provide an alternative approach for studying periodic systems. Many different approaches have been proposed,[10−20] which include, for example, the binary interaction method (BIM)[10,11] and the systematic fragmentation approach.[12,13] The principle behind these methods is that the lattice energy can be expanded as the many-body expansion, where "body" refers to a single monomer of the crystal (or a fragment of a monomer). Usually, the many-body expansion is truncated up to two- or three-body terms to make the calculation feasible. In some fragment-based methods,[10,11] subsystems containing monomers or dimers are calculated with existing molecular electronic structure methods. Such methods may provide reasonable descriptions for some crystals, but their accuracy should be less satisfactory for most crystals. The main advantage of fragment-based approaches is that advanced molecular electronic structure methods could be employed to provide highly accurate results for periodic systems. For example, a recent fragment-based approach[17] with up to four-body expansion shows that the crystalline lattice energy calculated with advanced wave function methods for the benzene crystal can reach the sub-kJ/mol accuracy. However, this and other fragment-based methods have not achieved such high accuracy for various types of molecular crystals, mainly because the effect of the crystal environment has not been treated with reasonable accuracy.

Received: September 13, 2014
Published: December 1, 2014

   dx.doi.org/10.1021/ct500833k | J. Chem. Theory Comput. 2015, 11, 91−98

In this work, we report an extension of the generalized energy-based fragmentation (GEBF) method,[21−26] which is designed to treat large molecules or molecular clusters, to molecular crystals with the periodic boundary condition (PBC), and test its performance on a variety of molecular crystals. In the PBC-GEBF approach, the total energy per unit cell of a molecular crystal can be evaluated with ground-state energies of a series of "electrostatically embedded" subsystems, each of which contains a few monomers. Our illustrative applications demonstrate that the PBC-GEBF method, combined with explicitly correlated methods[27−29] based on MP2 and CCSD(T), is capable of predicting accurate lattice energies and structures for various types of molecular crystals.

2. METHODOLOGY

2.1. The PBC-GEBF Method. In the GEBF approach proposed for molecules, the total energy of a large molecule is expressed as a linear combination of ground-state energies of various "embedded" subsystems, each of which consists of a few fragments (and capping hydrogen atoms if necessary) in the presence of background point charges generated by all other atoms in the system. When extending the GEBF approach to a periodic system, two new aspects should be considered. First, the translation symmetry introduced by periodicity should be utilized to judge whether the generated subsystems are identical or not. Second, each subsystem is embedded in the electrostatic field created by point charges on all other atoms in the crystal environment, which correspond to an infinite array of charges. To make the calculation on "embedded subsystems" practical, one can use the popular ONIOM-type technique[30] to simplify the calculation. The ground-state energy of a subsystem embedded in an infinite array of charges can be approximated as the energy of this subsystem embedded in a finite array of point charges of a super cell plus a classical charge−charge interaction term, as shown in the Supporting Information. On the basis of these two simplifications, we have derived a simple and elegant formula for the total energy per unit cell of a molecular crystal, as described below:

$$E_{\text{unit-cell}} = \sum_{m}^{M} C_m \tilde{E}_m - \left(\sum_{m}^{M} C_m\right) \sum_{A\in K} \sum_{(B>A)\in K} \frac{Q_A Q_B}{R_{AB}} + E_{\text{Ewald-Sum}} \quad (1)$$

where $\tilde{E}_m$ stands for the ground-state energy of a given subsystem embedded in a finite array of point charges of a super cell (a cutoff distance 18 Å is used to define the super cell, denoted as K), C_m denotes the coefficient (a positive or negative integer) of this subsystem, Q_A represents the atomic charge at atom A (R_{AB} being the interatomic distance), and $E_{\text{Ewald-Sum}}$ stands for the classical charge−charge interaction in the central cell and the charge−charge interactions between the central cell and all image cells, which can be accurately computed with the Ewald summation method.[31]

The generation of subsystems and derivation of their coefficients (C_m) for molecular crystals are somewhat similar to those described previously for molecules.[23] The procedure contains the following steps:

(1) Each monomer in the unit cell of a molecular crystal is considered as a fragment.

(2) For each fragment located in the central cell (defined as a central fragment), a subsystem is constructed as a union of several fragments, which are close to the given central fragment with an interfragment distance less than a given threshold (ζ). Another parameter ($\gamma_{\max}$), the maximum number of fragments, is used to control the size of a subsystem. Only those fragments with the shortest interfragment distances are added to a central fragment to form a subsystem.

(3) For each spatially close fragment pair, in which at least one fragment is located in the central cell and the interfragment distance is less than ζ, a subsystem is also constructed as a union of several fragments. This subsystem contains this pair of fragments and a few neighboring fragments centering around the middle point of this pair. The same parameters as described above are used to control the size of the subsystem. For convenience, these subsystems constructed in steps 2 and 3, which are centered on a fragment or a pair of fragments, are called primitive subsystems, and their coefficients are assigned to be +1 by construction.

(4) The next step is to construct additional subsystems and determine their coefficients. It is well-known that the energy of a target system can be decomposed to n-fragment ($n \le \gamma_{\max}$) interaction terms, and the coefficient of each specific term should be +1. However, when summing up the n-fragment interaction terms of all primitive subsystems, one may notice that many multifragment terms are overcounted, because of the overlapping of some primitive subsystems. Hence, additional subsystems should be generated to remove these excessive terms, which are called derivative subsystems for convenience. The principle for constructing derivative subsystems and determining their coefficients is to guarantee that the summation of coefficients of any specific n-fragment ($n \le \gamma_{\max}$) interaction term from all subsystems (both primitive and derivative) is +1. In the first step, we construct n-fragment ($n = \gamma_{\max} - 1$) derivative subsystems. Assume that a specific ($\gamma_{\max} - 1$)-body term occurs k times in all primitive subsystems, then a derivative subsystem with these ($\gamma_{\max} - 1$) fragments is constructed, and its coefficient is determined to be $(1 - k)$. Once this derivative subsystem is generated, we update the list of subsystems, which contains all primitive and derivative subsystems (at the very beginning, this list contains only primitive subsystems). In the subsequent steps, derivative subsystems with n fragments ($n = \gamma_{\max} - 2, ..., 2, 1$) and their coefficients can be obtained with the procedure described above. More details of this procedure can be found in our previous work.[23] However, in constructing subsystems and deriving their coefficients for molecular crystals, the translation symmetry (due to PBC of molecular crystals) should be employed to count any specific n-fragment interaction term in this step. Because of the periodic conditions, when comparing two subsystems to find overlapping n-fragment terms, one should consider not only fragment indices, but also cell indices. Those n-fragment terms with same fragment indices and same relative cell indices are considered to be identical.

It is worth analyzing the computational scaling of the PBC-GEBF approach. For molecular crystals, the number of all subsystems (primitive and derivative) constructed in the PBC-GEBF approach increases linearly as the number of molecules in a unit cell increases. Since we truncate the size of a subsystem up to $\gamma_{\max}$ fragments, the size of a subsystem is independent of the number of molecules in a unit cell. Therefore, the computational cost of the PBC-GEBF method scales linearly with the number of molecules in a unit cell. This feature allows the PBC-GEBF approach to be applicable to those molecular crystals with a large number of molecules in a unit cell.

dx.doi.org/10.1021/ct500833k | J. Chem. Theory Comput. 2015, 11, 91−98

Like the GEBF method, the PBC-GEBF method is applicable at various theory levels, the abbreviation "PBC-GEBF-X" is used to stand for the PBC-GEBF calculation at the X theory level (X = DFT, MP2, and so on).

It is worth mentioning that the error of truncating an infinite array of point charges into a finite array of point charges (of the super cell) for each subsystem may be significant. To address this problem, we will employ the field compensation scheme for each subsystem.[32] The essence of this scheme is to add some compensating point charges at the boundaries of each unit cell (the values of these charges are determined by the dipole moment $\vec{\mu}$ and lattice vectors of the unit cell) to ensure that the overall dipole moment of the unit cell vanishes, as it should for a three-dimensional (3D) lattice. For an "embedded" subsystem containing several neighboring unit cells, one can observe that with this compensation field only compensating charges at the boundaries of this subsystem exist, since compensating charges inside the subsystem can cancel out each other, as illustrated in Figure 1 for a two-dimensional lattice. Thus, the influence of the rest of the infinite crystal (beyond a subsystem) can be approximately modeled as some compensating charges at the boundaries of this subsystem. Hence, an "embedded" subsystem here means that a subsystem is embedded into atomic charges on nuclei of all other atoms in a super cell and boundary charges discussed above. The details on how to generate subsystems and add atomic charges or boundary charges are provided in the Supporting Information. Although there are many other ways of taking the polarization effect and long-range interaction into account, the field compensation scheme is preferred, because of its simple and straightforward implementation. It should be pointed out that the field compensation scheme can only be applied to treat normal crystals without macroscopic electric polarization. This simple scheme is not directly applicable for crystals such as piezoelectric or ferroelectric crystals.

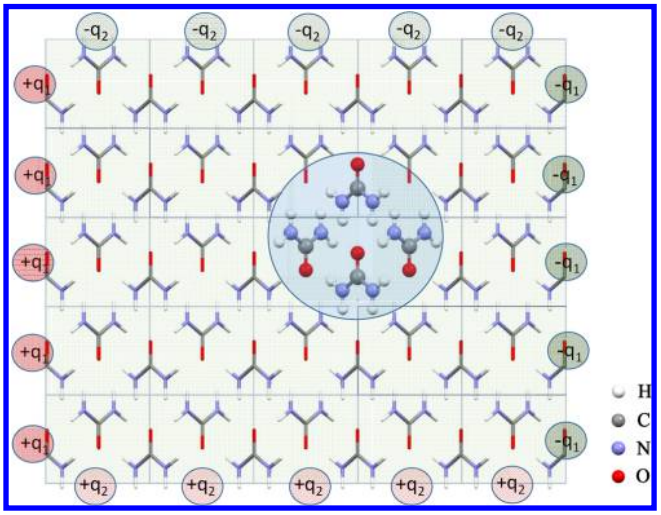


Figure 1. An illustrative picture for an "electrostatic embedding" subsystem, which contains four explicit urea molecules (denoted as ball-and-stick in the circle) and background point charges in neighboring cells and compensation charges on the boundaries. Point charges are placed in nuclear centers of all the other molecules (excluding four explicit urea molecules) denoted as sticks. The dipole moment of a unit cell, evaluated with atomic charges and atomic coordinates, is decomposed along the directions of three translation vectors. Then, pairs of compensating charges with opposite sign are then placed on the edges of boundary cells to neutralize the corresponding components of the dipole moment of the super cell.

The PBC-GEBF method has been applied to a variety of molecular crystals. The calculated cell energies have been demonstrated to be in good agreement with the values from conventional PBC calculations, with the deviations generally less than 2−3 mH (where mH denotes millihartrees). These stringent comparisons demonstrate that the PBC-GEBF method is able to provide accurate results for molecular crystals.

2.2. The Gradients of the PBC-GEBF Cell Energy. The analytical gradients of the PBC-GEBF cell energy, with respect to the coordinates of all atoms in a unit cell, and the lattice vectors can be obtained once the gradients of each "embedded" subsystem are available with the existing quantum chemistry programs such as Gaussian09[33] and Molpro[34,35] packages. The corresponding gradients can be formulated as described below:

$$\frac{\partial E_{\text{unit-cell}}}{\partial \mathbf{r}_{\text{A},0}} = \sum_{i\in\text{Box}}\left[\sum_m C_m\left(\frac{\partial \tilde{E}'_m}{\partial \mathbf{r}_{\text{A},i}} - \mathbf{F}_{m,a,i}Q_a - \sum_{j\in\text{Box},b}\mathbf{f}_{ai,bj}\right) + \left(\sum_m C_m\right)\sum_{j\in\text{Box},b}\mathbf{f}_{ai,bj}\right] + \frac{\partial E_{\text{Ewald-Sum}}}{\partial \mathbf{r}_{A,0}} \quad (2)$$

$$\frac{\partial E_{\text{unit-cell}}}{\partial \mathbf{L}} = \sum_{i\in\text{Box},A}\mathbf{i}_L\left[\sum_m C_m\left(\frac{\partial \tilde{E}'_m}{\partial \mathbf{r}_{A,i}} - \mathbf{F}_{m,a,i}Q_a - \sum_{j\in\text{Box},b}\mathbf{f}_{ai,bj}\right) + \left(\sum_m C_m\right)\sum_{j\in\text{Box},b}\mathbf{f}_{ai,bj}\right] + \frac{\partial E_{\text{Ewald-Sum}}}{\partial \mathbf{L}} \quad (3)$$

Here, A denotes a real atom in a given subsystem, a and b denote the point charge centers located at atoms A and B, and i, j denotes the cell index. $\boldsymbol{F}_{m,a,i}$ represents the electric field on the center a in the ith cell generated by subsystem m, and $\mathbf{f}_{ai,bj}$ is the Coulombic force between charge on a in the ith cell and charge on b in the jth cell. With these gradients, one can routinely optimize the crystal structures with the PBC-GEBF method at a given theory level.

2.3. Implementation of the PBC-GEBF Method. In our implementation of the PBC-GEBF method, we employ existing programs such as Gaussian09 and Molpro to perform DFT and wave-function-based *ab initio* calculations on embedded subsystems. In addition, we employ a quasi-Newton BFGS[36] optimization technique for crystal structure optimizations. To obtain the most-accurate lattice energies, we have performed subsystem calculations at the CCSD(T)-F12a[37,38] level with the aug-cc-pVTZ basis set (abbreviated as CCSD(T)-F12/avtz) for some crystals. For other molecular crystals, in which CCSD-(T)-F12/avtz calculations are not affordable, we employ the hybrid MP2-F12/CCSD(T) method instead. In this hybrid approach, we first perform MP2-F12[39] calculations with the aug-cc-pVTZ (or aug-cc-pVDZ) basis set to evaluate the lattice energies. Then, a higher-order electron correlation (or post-MP2) correction term is added to the MP2-F12 lattice energies to get more-accurate estimates. The post-MP2 term is estimated as the lattice energy difference between the CCSD(T) and MP2 results at the aug-cc-pVDZ (or cc-

dx.doi.org/10.1021/ct500833k | J. Chem. Theory Comput. 2015, 11, 91−98

pVDZ) basis set (defined as $\Delta^{CCSD(T)} = E^{CCSD(T)} - E^{MP2}$).[19,40,41] The basis set superposition error (BSSE) is estimated with the counterpoise method.[42] Our calculations show that the BSSE corrections from the MP2-F12 method with the aug-cc-pVDZ basis set are quite small, being no more than a few kJ/mol. It is worth pointing out that the post-MP2 term can only be calculated with moderate basis sets (aug-cc-pVDZ or cc-pVDZ) for systems under study since CCSD(T) calculations are very expensive for larger systems with larger basis sets. However, previous works[40,41] have shown that this post-MP2 term is relatively insensitive to basis set improvement when calculating the binding energies of weakly bound molecules.

We will take an ice crystal (phase II) as an example, to demonstrate the effect of the field compensation method and determine the suitable values of two parameters (ζ and γ_{max}) in the PBC-GEBF method. On the basis of our previous works on molecules, ζ = 4.0 Å will be used as the default parameter. For periodic systems, the dipole moment of the unit cell is dependent on the choice of unit cell. One can expect that, for those unit cell motifs with larger dipole moments, the effect of compensation field will be more significant. For the phase II ice crystal, whose crystal structure is taken from the literature,[43] we have performed calculations with the B3LYP functional[44,45] at the 6-31G(d,p) basis set with the PBC-GEBF method, with two choices of unit cell (choices I and II). The unit cell of choice II is constructed by shifting the unit cell of choice I by 0.3 times the box length along the *c*-axis direction. The results obtained with PBC-GEBF-B3LYP/6-31G(d,p) calculations with different γ_{max} values, in the presence (or absence) of the compensation field, are collected in Table 1. It can be seen from Table 1 that,

Table 1. Unit Cell Energy in the Phase-II Ice Crystal Calculated at the PBC-GEBF−B3LYP/6-31G(d,p) Level with and without the Compensation Field, Respectively[a]

	With the Compensation Field		Without the Compensation Field	
γ_{max}	energy (a.u.)	deviation (mH)	energy (a.u.)	deviation (mH)
		Unit Cell Choice I		
2	−917.437577	−31.77	−917.472594	−66.79
3	−917.424826	−19.02	−917.459163	−53.36
4	−917.407655	−1.85	−917.419273	−13.47
5	−917.405701	0.10	−917.440479	−34.67
		Unit Cell Choice II		
2	−917.438053	−32.25	−917.438969	−33.16
3	−917.425005	−19.20	−917.436959	−31.15
4	−917.409763	−3.96	−917.416774	−10.97
5	−917.406064	−0.26	−917.386662	19.14
		PBC-B3LYP/6-31G(d,p)		
		−917.405806		

[a]The results obtained with the conventional PBC-B3LYP/6-31G(d,p) method are listed for comparison. PBC-GEBF results with different parameters (γ_{max}) are given for comparison.

with both choices of unit cell, the PBC-GEBF-B3LYP cell energy obtained with the compensation field gradually converges to the conventional periodic B3LYP value, when the maximum number of fragments increases to 5. The cell energy differences obtained with two choices of unit cell are all <2.2 mH for different γ_{max} values. Since each unit cell contains 12 water molecules, the lattice energies (per molecule) obtained with different unit cell motifs are within 0.5 kJ/mol, when the compensation field is employed. In contrast, without the compensation field, calculations with both choices of unit cell do not give convergent results as the value of γ_{max} increases. The calculated cell energy is strongly dependent on the choice of unit cell. Take the calculation with γ_{max} = 5 as an example. The magnitude of the calculated dipole moment is 34.6 D for unit cell choice I, but 30.7 D for unit cell choice II. As a result, the values of background point charges are significantly different. The differences in background point charges are responsible for the different cell energies obtained with two choices of unit cell. One can expect that the oscillating behaviors of the calculated cell energies can be noticeably reduced when the size of the supercell is sufficiently large. However, with the compensation field, the size of the supercell chosen with a cutoff distance of 18 Å is enough to provide almost consistent results.

With the compensation field, the results from PBC-GEBF calculations are almost independent of the choice of unit cell (or nearly translationally invariant). For this and several other systems, our calculations demonstrated that the parameters (ζ = 4.0 Å and γ_{max} = 4) may be set as the default parameters for PBC-GEBF calculations, to balance the accuracy and the computational cost.

3. RESULTS AND DISCUSSIONS

3.1. Optimized Structures and Lattice Energies of Molecular Crystals. The PBC-GEBF method will be employed to calculate the lattice energies and structures for 10 molecular crystals containing small molecules. The systems we choose include a variety of types of intermolecular interactions. A benchmark set of experimental lattice energies have been derived from the experimentally measured sublimation enthalpies, zero-point vibrational energies, and thermal corrections.[2,3,19] These experimental data can be used to evaluate the performance of the PBC-GEBF method. For all systems under investigation, we have fully optimized their crystal structures (including atomic coordinates and lattice parameters) with the PBC-GEBF-M06-2X method at the 6-311++G(d,p) level (with their X-ray crystal structures as initial structures). The M06-2X functional, which can give satisfactory descriptions for various noncovalent interactions in a wide range of molecular systems,[46] is employed for crystal structure optimizations. A comparison of calculated and experimental crystal structural parameters is listed in Table 2. One can see from the table that the calculated lattice parameters are in excellent accordance with the experimental values. The average deviation for unit cell lengths is ∼1.3%, and the average deviation for unit-cell angles is <0.5%. For some crystals such as C_2H_2 and ice, the crystal structure and cell volume from the M06-2X prediction are very close to the experimental values. However, one can also observe that the M06-2X calculations tend to underestimate the cell volume for some crystals. For example, for acetic acid and CO_2 crystals, the deviations are as large as 9.9%, and 16%, respectively. For those crystals in which the deviation of the calculated cell volume, with respect to the experimental value, is >7.0%, we also perform PBC-GEBF optimizations at the MP2/6-311++G(2d,2p) level. A comparison of the optimized lattice parameters and the experimental data shows that the MP2-optimized crystal structures are significantly improved for these systems. For instance, the errors on cell volume predictions are reduced from 16% to 7.5% for CO_2, and from 9.9% to 3.1% for acetic acid. In addition, several interatomic distances of acetic acid and CO_2 crystals obtained with GEBF-MP2 are displayed in Figure 2.

dx.doi.org/10.1021/ct500833k | J. Chem. Theory Comput. 2015, 11, 91−98

Table 2. Comparison of Optimized Lattice Parameters from PBC-GEBF−M06-2X/6-311++G(d,p) Calculations with Those Obtained from Experimental Crystal Structures

		a (Å)	b (Å)	c (Å)	α (deg)	β (deg)	γ (deg)	cell volume
C_2H_2[a]	GEBF-M06-2X	6.198	6.014	5.538	90.0	90.0	89.9	206.5 (0.0%)
	X-ray	6.193	6.005	5.551	90.0	90.0	90.0	206.4
urea	GEBF-M06-2X	5.435	5.436	4.639	89.8	90.0	90.0	137.0 (−5.5%)
	X-ray	5.565	5.565	4.684	90.0	90.0	90.0	145.1
cyanamide	GEBF-M06-2X	6.625	6.472	9.124	89.3	91.2	89.2	391.0 (−6.0%)
	X-ray	6.856	6.628	9.147	90.0	90.0	90.0	415.7
ice	GEBF-M06-2X	9.001	7.791	7.346	90.0	90.0	90.0	515.1 (0.0%)
	X-ray	9.000	7.794	7.348	90.0	90.0	90.0	515.4
acetamide	GEBF-M06-2X	11.454	11.442	12.648	89.8	90.2	120.2	1433.4 (−2.7%)
	X-ray	11.492	11.492	12.892	90.0	90.0	120.0	1474.5
formamide	GEBF-M06-2X	3.357	8.950	6.914	90.0	101.8	90.5	203.4 (−9.2%)
	GEBF-MP2[b]	3.572	8.978	6.968	90.2	100.2	90.5	219.9 (−1.8%)
	X-ray	3.604	9.041	6.994	90.0	100.5	90.0	224.1
imidazole	GEBF-M06-2X	7.346	5.111	9.723	90.1	120.8	89.3	313.5 (−9.7%)
	GEBF-MP2[b]	7.404	5.129	9.715	90.4	120.3	88.6	318.4 (−8.3%)
	X-ray	7.569	5.366	9.785	90.0	119.1	90.0	347.3
benzene	GEBF-M06-2X	7.167	9.216	6.503	90.3	90.2	90.1	429.5 (−7.1%)
	GEBF-MP2[b]	7.303	9.338	6.653	90.1	90.1	90.1	453.7 (−1.9%)
	X-ray	7.360	9.375	6.703	90.0	90.0	90.0	462.5
CO_2	GEBF-M06-2X	5.268	5.315	5.334	90.2	89.7	90.1	149.3 (−16%)
	GEBF-MP2[b]	5.496	5.471	5.473	89.8	89.9	90.0	164.6 (−7.5%)
	X-ray	5.63	5.63	5.63	90.0	90.0	90.0	177.9
acetic acid	GEBF-M06-2X	13.075	3.619	5.660	90.4	90.1	90.3	267.8 (−9.9%)
	GEBF-MP2[b]	13.082	3.858	5.712	88.1	88.8	90.7	288.0 (−3.1%)
	X-ray	13.151	3.923	5.762	90.0	90.0	90.0	297.3

[a]The aug-cc-pVTZ basis set is employed. [b]The 6-311++G(2d,2p) basis set is employed in the PBC-GEBF-MP2 optimization.

For intermolecular atomic distances, the OH···O distances in acetic acid differ from the experimental value by merely 0.01 Å, while the two types of the nearest O···O distances in CO_2 are ∼0.1 Å shorter than those in the X-ray structure.[47,48] One can see that, for the CO_2 crystal of the dispersion type, the MP2 theory with even larger basis sets may provide further improvement. Considering the fact that the experimental structures measured at finite temperatures undergo thermal expansion to some extent, relative to the zero temperature structures obtained here, one can conclude that the PBC-GEBF-MP2 method (with large basis sets) provides satisfactory descriptions on the crystal structures of molecular crystals.

The calculated lattice energies for all systems under study, together with those experimental data, are listed in Table 3 and are schematically shown in Figure 3. It can be observed that the deviations of the lattice energies obtained at the CCSD(T)-F12a or CCSD(T)/MP2-F12 level, with respect to the reference data (for those crystals with two experimental lattice energies, the average value is taken as the experimental lattice energy), are all <6 kJ/mol, with the mean absolute error being only 3.5 kJ/mol. Some of the deviations between calculated and experimental lattice energies may result from the basis set incompleteness error (because of the use of the aug-cc-pVDZ basis set in MP2-F12 calculations, and the use of the cc-pVDZ basis set in computing the post-MP2 correction terms). For ice Ih (a typical hydrogen-bonded system), the predicted lattice energy differs by ∼1 kJ/mol from the experimental data. For more-complex crystals (such as formamide and cyanamide), where multiple interactions coexist, the predicted lattice energies are all in good agreement with the experimental data. By analyzing the data listed in Table 3, one can note that the post-MP2 correction term is negligible for CO_2, formamide, and acetamide. However, for imidazole and benzene crystals, the post-MP2 correction term decreases the lattice energy by more than 10 kJ/mol, indicating that the MP2-F12 method overestimate the $\pi-\pi$ stacking interaction in these two crystals. Thus, for some molecular crystals, accurate estimate of dynamic correlation (which may be obtained with explicitly correlated CCSD(T)-F12a or hybrid CCSD(T)/MP2-F12 calculations) is essential for predicting accurate lattice energies. On the other hand, we have also compared the performance of M06-2X and PBE-D3[49] functionals on the prediction of lattice energies for all the crystals under investigation. The BSSE-corrected lattice energies obtained from PBC-GEBF-M06-2X and PBC-GEBF-PBE-D3 calculations at the GEBF-M06-2X-optimized crystal structures are also displayed in Table 3. It can be seen that the

dx.doi.org/10.1021/ct500833k | J. Chem. Theory Comput. 2015, 11, 91−98

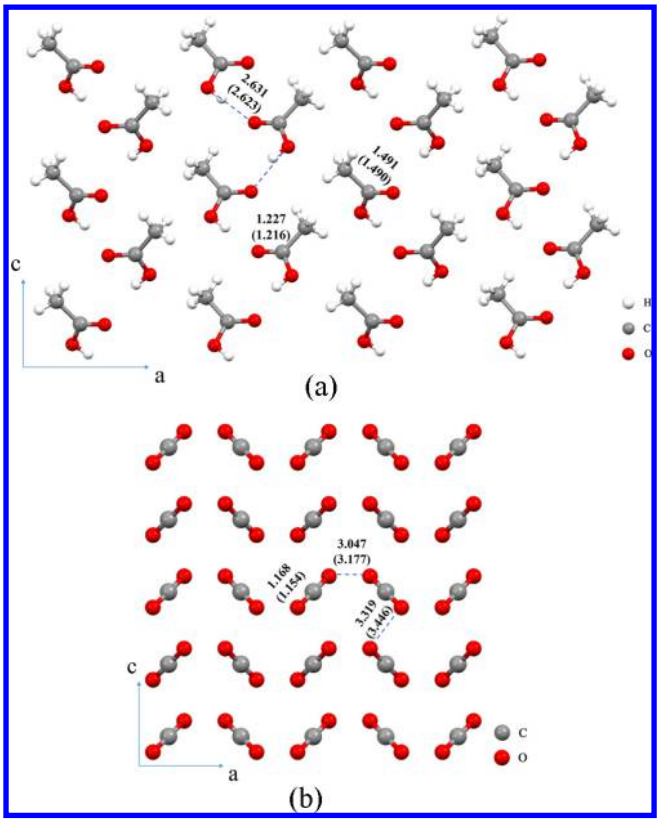


Figure 2. Optimized crystal structures of the crystals of (a) acetic acid and (b) CO_2. The PBC-GEBF-MP2 optimizations are done with the 6-311++G(2d,2p) basis set. The optimized intramolecular and intermolecular atomic distances are displayed, and the values shown in parentheses are taken from the X-ray crystal structures.

M06-2X and PBE-D3 methods with the cc-pVTZ basis set can also give fairly satisfactory descriptions on the lattice energies, with the mean absolute errors being 4.7 kJ/mol (for M06-2X) and 5.2 kJ/mol (for PBE-D3). The overall accuracy of both functionals is inferior to CCSD(T)-F12a or CCSD(T)/MP2-F12. For the M06-2X functional, it underestimates the lattice energies to some extent for imidazole, acetic acid, and benzene crystals. For instance, the lattice energy of acetic acid calculated

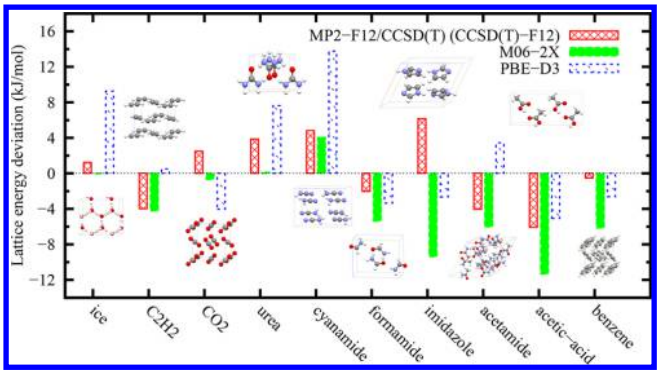


Figure 3. Deviation of lattice energies calculated with GEBF-MP2-F12/CCSD(T) (or GEBF-CCSD(T)-F12) and GEBF-M06-2X, with respect to the experimental lattice energies (for those crystals with two experimental estimates, the average value is taken as the experimental lattice energy).

with M06-2X is lower than the experimental value (71.8 kJ/mol), by 11.3 kJ/mol. For the PBE-D3 functional, one can see that it overestimates the lattice energy by more than 8 kJ/mol for some hydrogen-bonded molecular crystals such as ice, urea, and cyanamide. In summary, the overall performance of the M06-2X and PBE-D3 methods indicates that the cost-effective M06-2X method or dispersion-corrected DFT methods is suitable for providing lattice energies and structures of general molecular crystals with medium accuracy.

It is worthwhile to compare the cell energies obtained from PBC-GEBF calculations and those obtained from conventional PBC calculations. In conventional periodic HF or DFT calculations, the use of large Gaussian basis sets (such as cc-pVTZ) usually suffers from the convergence problem. Nevertheless, this problem is avoided in the PBC-GEBF calculations (since only subsystems are calculated). We have compared the cell energies obtained with conventional PBC M06-2X and PBC-GEBF-M06-2X calculations for two crystals (C_2H_2 and formamide). The cc-pVDZ basis set is employed in the calculations. The differences between the PBC M06-2X cell energies and those from PBC-GEBF-M06-2X calculations are merely 2.0 mH for formamide and 0.6 mH for C_2H_2. Considering the fact that there are four molecules in the unit

Table 3. Lattice Energies for Several Molecular Crystals Calculated with the PBC-GEBF-X Method (X = CCSD(T)-F12a or Hybrid MP2-F12/CCSD(T))[a]

crystal	method	lattice energy[b] (kJ/mol)	$\Delta^{CCSD(T)c}$	best estimate	M06-2X	PBE-D3	experiment value
ice	CCSD(T)-F12/avtz	60.24	0.0	60.24	58.96	68.27	59[d]
C_2H_2	CCSD(T)-F12/avtz	24.52	0.0	24.52	24.28	29.00	28.5[e]
CO_2	MP2-F12/avtz	29.45	0.85	30.30	27.12	23.77	27.8[f]
urea	MP2-F12/avtz	106.22	−2.95	103.27	99.52	107.03	99.4[f]
cyanamide	MP2-F12/avtz	91.22	−7.20	84.02	83.25	92.97	79.2[f]
formamide	MP2-F12/avtz	77.00	1.32	78.32	74.99	77.00	82,[d] 78.7[f]
imidazole	MP2-F12/avdz	107.97	−13.47	94.50	78.98	85.70	91 ± 4,[d] 85.7[f]
acetamide	MP2-F12/avdz	83.17	−1.22	81.96	79.98	89.44	86 ± 2[d]
acetic acid	MP2-F12/avtz	66.72	−0.99	65.73	60.48	66.73	71.8[f]
benzene	MP2-F12/avdz	65.01	−14.33	50.68	45.06	48.58	52 ± 3,[d] 50.4,[f] 55.3[g]

[a]The best theoretical estimates are from CCSD(T)-F12a or MP2-F12/CCSD(T) calculations. The BSSE correction is included in computing lattice energies. The results obtained with PBC-GEBF−M06-2X and PBC-GEBF−PBE-D3 approaches at the cc-pVTZ basis set are also shown for comparison. [b]The calculations are performed with $\gamma_{max} = 3$ for benzene, imidazole, and acetamide, and with $\gamma_{max} = 4$ for all other systems. [c]The post-MP2 term is done with $\gamma_{max} = 2$ for benzene, imidazole, and acetamide, and with $\gamma_{max} = 3$ for all other systems. The cc-pVDZ basis set is used for all systems except CO_2, formamide, and cyanamide (aug-cc-pVDZ is used). [d]Data taken from ref 19. [e]Data taken from ref 3. [f]Data taken from ref 2. [g]Data taken from ref 17.

dx.doi.org/10.1021/ct500833k | J. Chem. Theory Comput. 2015, 11, 91−98

cell of both crystals, one can conclude that the lattice energies obtained from the PBC-GEBF-M06-2X approach only deviate from those from conventional PBC-M06-2X calculations by no more than 1.5 kJ/mol.

3.2. Component Analysis of the Theoretical Lattice Energy. Within the PBC-GEBF approach, it is straightforward to perform the energy component analysis to qualitatively understand the contributions of some factors to the theoretical lattice energy. For example, the contribution of the four-body interaction to the lattice energy can be defined as the difference between the lattice energy obtained from GEBF calculations with up to three-body interaction ($\gamma_{max} = 3$) and that from the standard GEBF calculations with $\gamma_{max} = 4$. The results obtained at the M06-2X/cc-pVTZ level are displayed in Figure 4a. One can see that, for all systems under study, the average absolute four-body interactions are ~1.34 kJ/mol, with the maximum value of 6.0 kJ/mol. This result indicates that the four-body interaction does have significant effect on the lattice energy for some crystals. It should be mentioned that, in PBC-GEBF calculations with $\gamma_{max} = 3$, the n-body interaction ($n \leq 3$) is calculated with electrostatic embedding, which is different from the direct n-body interaction. In fact, in the PBC-GEBF calculations with $\gamma_{max} = 3$, the lattice energy also contains the contribution from the interaction of three explicit molecules with other molecules, represented as point charges. On the other hand, the contribution of long-range electrostatic and polarization interactions may be estimated as the difference between the lattice energies obtained from GEBF calculations with and without electrostatic embedding. The results obtained at the M06-2X/cc-pVTZ level are displayed in Figure 4b. It can be seen that, for CO_2 and C_2H_2 (dispersion type), the long-range electrostatic and polarization interactions are insignificant, being <1.0 kJ/mol. In contrast, for systems such as imidazole and acetamide (mainly bound by hydrogen bonds), the long-range electrostatic and polarization interactions contribute >10 kJ/mol to the theoretical lattice energy. This result indicates that point charges in the infinite crystal environment of hydrogen-bonded molecular crystals introduce strong polarization. Thus, the inclusion of long-range electrostatic and polarization interactions via the background and boundary point charges is essential for general molecular crystals.

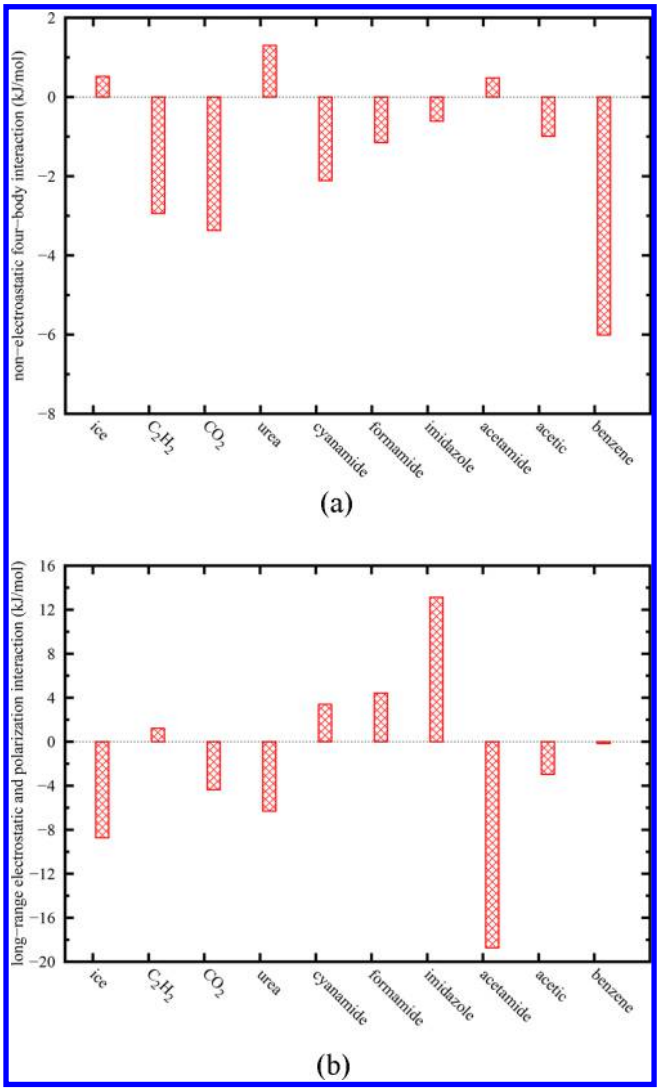


Figure 4. Contributions of (a) the four-body interaction to the lattice energy and (b) the long-range electrostatic and polarization interactions for a series of molecular crystals at the PBC-GEBF-M06-2X/cc-pVTZ level.

4. CONCLUSIONS

We have proposed and implemented the generalized energy-based fragmentation approach under periodic boundary conditions (PBC-GEBF) for molecular crystals. This approach is conceptually very simple, allowing various types of molecular crystals to be theoretically studied with widely used molecular quantum chemistry methods. The use of the field compensation method allows the method to take long-range electrostatic interaction into account, so that various types of molecular crystals can be properly described. With the PBC-GEBF method, one can obtain accurate crystal structures and lattice energies for general molecular crystals using advanced molecular quantum chemistry methods. In addition, this approach can be used to quantify the contributions of various factors to the theoretical lattice energy. Illustrative applications demonstrate that the PBC-GEBF method with explicitly correlated quantum chemistry methods is capable of providing accurate descriptions on lattice energies and structures for a wide range of molecular crystals. The PBC-GEBF method is expected to become an efficient theoretical method for investigating various types of molecular crystals. The combination of various global optimization algorithms with the present PBC-GEBF method may provide a promising theoretical tool for predicting crystal structures of molecular crystals.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The ONIOM-type technique for estimating the energies of embedded subsystems and the implementation details of the PBC-GEBF method are described. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: shuhua@nju.edu.cn.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Grant Nos. 21333004, 21361140376, and 21473087) and the National Basic Research Program (Grant No. 2011CB808501).

■ REFERENCES

(1) Moellmann, J.; Grimme, S. *J. Phys. Chem. C* **2014**, *118*, 7615−7621.
(2) Otero-de-la-Roza, A.; Johnson, E. R. *J. Chem. Phys.* **2012**, *137*, 054103.
(3) Civalleri, B.; Zicovich-Wilson, C. M.; Valenzano, L.; Ugliengo, P. *CrystEngComm* **2008**, *10*, 405−410.
(4) Yousaf, K. E.; Brothers, E. N. *J. Chem. Theory Comput.* **2010**, *6*, 864−872.
(5) DiStasio, R. A.; Car, R.; Scheffler, M.; Tkatchenko, A. *Phys. Rev. Lett.* **2012**, *108*, 236402.
(6) Constans, P.; Ayala, P. Y.; Scuseria, G. E. *J. Chem. Phys.* **2000**, *113*, 10451−10458.
(7) Pisani, C.; Maschio, L.; Casassa, S.; Halo, M.; Schütz, M.; Usvyat, D. *J. Comput. Chem.* **2008**, *29*, 2113−2124.
(8) Usvyat, D.; Manby, F. R.; Casassa, S.; Pisani, C.; Schütz, M.; Maschio, L. *Phys. Rev. B* **2007**, *76*, 075101.
(9) Riley, K. E.; Pitoňak, M.; Jurečka, P.; Hobza, P. *Chem. Rev.* **2010**, *110*, 5023−5063.
(10) Hirata, S. *J. Chem. Phys.* **2008**, *129*, 204104.
(11) Sode, O.; Hirata, S. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7765−7779.
(12) Deev, V.; Collins, M. A. *J. Chem. Phys.* **2005**, *122*, 154102.
(13) Netzloff, H. M.; Collins, M. A. *J. Chem. Phys.* **2007**, *127*, 134113.
(14) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, *313*, 701−706.
(15) Fujita, T.; Nakano, T.; Tanaka, S. *Chem. Phys. Lett.* **2011**, *506*, 112−116.
(16) Rice, B. M.; Szalewicz, K.; Podeszwa, R. *Phys. Rev. Lett.* **2008**, *101*, 115503.
(17) Yang, J.; Hu, W.; Usvyat, D.; Matthews, D.; Schuetz, M.; Chan, G. K. *Science* **2014**, *345*, 640−643.
(18) Bygrave, P. J.; Allan, N. L.; Manby, F. R. *J. Chem. Phys.* **2012**, *137*, 164102.
(19) Beran, G. J. O.; Nanda, K. *J. Phys. Chem. Lett.* **2010**, *1*, 3480−3487.
(20) Wen, S.; Nanda, K.; Huang, Y.; Beran, G. J. O. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7578−7590.
(21) Li, S.; Li, W.; Fang, T. *J. Am. Chem. Soc.* **2005**, *127*, 7215−7226.
(22) Li, W.; Li, S.; Jiang, Y. *J. Phys. Chem. A* **2007**, *111*, 2193−2199.
(23) Hua, S.; Hua, W.; Li, S. *J. Phys. Chem. A* **2010**, *114*, 8126−8134.
(24) Li, W. *J. Chem. Phys.* **2013**, *138*, 014106.
(25) Wang, K.; Li, W.; Li, S. *J. Chem. Theory Comput.* **2014**, *10*, 1546−1553.
(26) Li, S.; Li, W.; Ma, J. *Acc. Chem. Res.* **2014**, *47* (9), 2712−2720.
(27) Kutzelnigg, W.; Klopper, W. *J. Chem. Phys.* **1991**, *94*, 1985−2001.
(28) Ten-no, S. *Chem. Phys. Lett.* **2004**, *398*, 56−61.
(29) Kutzelnigg, W. *Theor. Chim. Acta* **1985**, *68*, 445−469.
(30) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357−19363.
(31) Ewald, P. P. *Ann. Phys.—Berlin* **1921**, *64*, 253−287.
(32) Piela, L. *Ideas of Quantum Chemistry*; Elsevier: Amsterdam, 2007.
(33) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, Revision D.01*; Gaussian, Inc.: Wallingford, CT, 2009.
(34) Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M. *WIREs Comput. Mol. Sci.* **2012**, *2*, 242−253.
(35) 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.; Peng, D.; Pflüger, K.; Pitzer, R.; Reiher, M.; Shiozaki, T.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T.; Wang, M. MOLPRO, Version 2012.1; Institut für Theoretische Chemie, Universität Stuttgart: Stuttgart, Germany, 2012. (See http://www.molpro.net.)
(36) Shanno, D. F. *Math. Comput.* **1970**, *24*, 647−656.
(37) Adler, T. B.; Knizia, G.; Werner, H.-J. *J. Chem. Phys.* **2007**, *127*, 221106.
(38) Knizia, G.; Adler, T. B.; Werner, H.-J. *J. Chem. Phys.* **2009**, *130*, 054104.
(39) Werner, H.-J.; Adler, T. B.; Manby, F. R. *J. Chem. Phys.* **2007**, *126*, 164102.
(40) Sinnokrot, M. O.; Valeev, E. F.; Sherrill, C. D. *J. Am. Chem. Soc.* **2002**, *124*, 19887−10893.
(41) Šponer, J.; Jurečka, P.; Hobza, P. *J. Am. Chem. Soc.* **2004**, *126*, 10142−10151.
(42) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553−566.
(43) Kambara, O.; Takahashi, K.; Hayashi, M.; Kuo, J. *Phys. Chem. Chem. Phys.* **2012**, *14*, 11484−11490.
(44) Yang, W.; Parr, R. G.; Lee, C. *Phys. Rev. B* **1988**, *37*, 785−789.
(45) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(46) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215−241.
(47) Simon, A.; Peters, K. *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem.* **1980**, *36*, 2750−2751.
(48) Boese, R.; Blaser, D.; Latz, R.; Baumen, A. *Acta Crystallogr., Sect. C: Cryst. Struct. Commun.* **1999**, *55*, 9900001.
(49) Grimme, S.; Ehrlich, S.; Goerigk, L. *J. Comput. Chem.* **2011**, *32*, 1456−1465.