

Noncovalent Interactions in Extended Systems Described by the Effective Fragment Potential Method: Theory and Application to Nucleobase Oligomers

Debashree Ghosh,[†] Dmytro Kosenkov,[‡] Vitalii Vanovschi,[†] Christopher F. Williams,[§] John M. Herbert,[§] Mark S. Gordon,^{||} Michael W. Schmidt,^{||} Lyudmila V. Slipchenko,^{*,‡} and Anna I. Krylov^{*,†}

Department of Chemistry, University of Southern California, Los Angeles, California 90089-0482, United States, Department of Chemistry, Purdue University, West Lafayette, Indiana 47907, United States, Department of Chemistry, The Ohio State University, Columbus, Ohio 43210, United States, and Department of Chemistry and Ames Laboratory, Iowa State University, Ames, Iowa 50011, United States

Received: August 10, 2010; Revised Manuscript Received: September 30, 2010

The implementation of the effective fragment potential (EFP) method within the Q-CHEM electronic structure package is presented. The EFP method is used to study noncovalent π - π and hydrogen-bonding interactions in DNA strands. Since EFP is a computationally inexpensive alternative to high-level ab initio calculations, it is possible to go beyond the dimers of nucleic acid bases and to investigate the asymptotic behavior of different components of the total interaction energy. The calculations demonstrated that the dispersion energy is a leading component in π -stacked oligomers of all sizes. Exchange-repulsion energy also plays an important role. The contribution of polarization is small in these systems, whereas the magnitude of electrostatics varies. Pairwise fragment interactions (i.e., the sum of dimer binding energies) were found to be a good approximation for the oligomer energy.

1. Introduction

Noncovalent interactions^{1–4} are at least an order of magnitude weaker than a typical chemical bond energy (100 kcal/mol), yet they govern such important processes as solvation, adsorption, condensation, and crystallization. They play a crucial role in many biological processes such as protein folding⁵ and structure,^{6–9} enzyme catalysis, drug binding, the function of DNA and RNA, protein–DNA interactions,^{10–12} self-assembled supramolecular architecture,^{13,14} and molecular recognition.^{15–17} For example, the structure of the DNA double helix is determined by hydrogen-bonding between the complementary nucleobases (adenine–thymine and guanine–cytosine) and π - π interactions between the stacked bases (see Figure 1). Noncovalent interactions also control the structure and function of intercalation compounds in DNA and RNA that are crucial for the understanding of replication and origin of life.^{10,18–20}

Noncovalent interactions most often involve closed-shell species in which electron sharing between the fragments is negligible (hence, noncovalent); however, they also play a role in weakly bound systems with some covalent character involving radicals (consider, for example, the (NO) dimer,²² solvated CN radical,²³ Al-ethylene complex²⁴).

Noncovalent interactions, such as hydrogen-bonding, π - π , and van der Waals, consist of electrostatic and dispersion forces. The strongest noncovalent interaction, hydrogen-bonding (5–18 kcal/mol), is dominated by electrostatics and also includes partial covalent character (electron sharing between the two fragments). On the other end of the spectrum there are van der Waals interactions, which originate in a correlated motion of electrons,

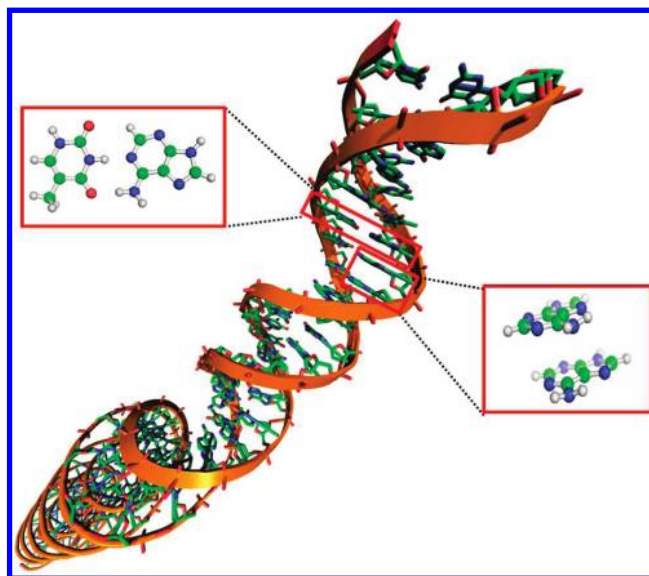


Figure 1. Hydrogen-bonding and π -stacking interactions are equally important for the structural integrity of the double helix. The hydrogen-bonding is responsible for keeping the strands together, and π -stacking determines the helical shape of the strands. Turning off either one of these interactions, as was done in a computational study of Hobza and co-workers,²¹ leads to the unfolding of the double helix.

ranging from several Kelvin (e.g., He₂) to several kcal/mol. The strength of π - π interactions, which is determined by an interplay between the electrostatics and dispersion, varies from 2–3 kcal/mol (as in the benzene dimer) to more than 10 kcal/mol (e.g., clusters of nucleobases).

Noncovalent interactions pose a challenge for computational methods.^{1,2} Their small magnitude calls for high accuracy. Dispersion is a purely electron correlation effect, and can only be captured by high-level correlated methods, such as the

* To whom correspondence should be addressed. E-mail: lslipchenko@purdue.edu (L. V. S.), krylov@usc.edu (A. I. K.).

[†] University of Southern California.

[‡] Purdue University.

[§] The Ohio State University.

^{||} Iowa State University.

coupled-cluster with single, double, and perturbative triple excitations [CCSD(T)] method. Moreover, large augmented atomic basis sets are needed for an accurate description of the interfragment region to mitigate basis set superposition errors (BSSE).

There are several computational strategies for modeling noncovalent interactions. For small systems, highly accurate results can be obtained using reliable ab initio methods.^{1,25–29} Although such calculations are of great value for benchmark purposes, they are not applicable to extended systems due to the steep computational scaling of wave function based methods. Different approaches are being developed to overcome this problem, ranging from linear scaling^{30–33} to energy additivity (ONIOM-like) and fragmentation techniques.^{2,34–38} Alternatively, weak intermolecular interactions can be treated as perturbations using wave functions (or electron densities) of the fragments, as done in symmetry-adapted perturbation theory (SAPT) and similar methods.^{26,39,40} More empirical approaches include dispersion-corrected density functional theory (DFT-D).^{41,42}

On another end of the spectrum, there are quantum mechanics/molecular mechanics (QM/MM) approaches using empirical force fields.^{43–46} The QM/MM methods are very efficient; however, the empirical parametrization of the MM part undermines their predictive power. Moreover, some effects are notoriously difficult to describe by traditional (nonpolarizable) force fields.

The effective fragment potential (EFP) method is developed as a nonempirical alternative to force-field based QM/MM.^{47–50} EFP, which is a QM-based potential, is a more sophisticated approach than simple parametrized MM methods. The first variant of EFP (EFP1) was designed to describe aqueous solvent effects. It included Coulombic and polarization terms in the same rigorous fashion as in the general EFP method⁵¹ (see Section 2); however, the remaining terms (mainly exchange-repulsion) were fitted by an exponential function to match the total ab initio energy of the water dimer potential energy surface. Depending on the level of theory used for fitting, there are Hartree–Fock-based⁴⁹ (EFP1/HF), MP2-based (EFP1/MP2), and DFT-based⁵² (EFP1/DFT) potentials. The fitted form of the exchange-repulsion term simplifies the coupling between the quantum and the EFP regions, and several QM/EFP1 models, with the quantum part described by HF, MP2, DFT, or MCSCF, have been implemented in GAMESS (general atomic and molecular electronic structure system).^{50,53} The functional form of EFP1 is similar to other polarizable embedding models; however, the parametrization is different. Recently, EFP1 has been integrated with configuration interaction singles (CIS),⁵⁴ time-dependent DFT (TD-DFT),⁵⁵ and multireference perturbation theory (MRPT).⁵⁶ EFP1 has been employed in studies of chemical reactions in solutions,^{57,58} clusters,^{52,59,60} an SN2 reaction,⁶¹ amino acid neutral/zwitterion equilibria,^{62,63} and electronic excitations of chromophores in bulk water.^{55,64}

To extend EFP beyond water, a more general approach with no empirically fitted parameters has been developed.^{48,65} This general EFP method is a computationally inexpensive way of modeling intermolecular interactions in noncovalently bound systems. The absence of fitted parameters and a natural partitioning of the interaction energy into Coulomb, polarization, dispersion, and exchange-repulsion terms make it an attractive choice for analysis and interpretation of intermolecular forces. Moreover, by construction, it is free from the BSSE that plagues ab initio calculations of weakly bound systems. The EFP Hamiltonian is pairwise; however, leading many-body effects are included through a self-consistent treatment of polarization.

The performance of the general EFP method⁵¹ (referred to as EFP2) has been benchmarked against high-level ab initio data for a variety of noncovalent systems, including those dominated by hydrogen-bonding, π – π , or mixed interactions. For example, in benzene dimers,⁶⁶ the EFP total interaction energies and the EFP energy components were shown to agree well with the high-level ab initio values computed by CCSD(T) and SAPT.^{25,67,68} For the three benzene dimer structures (sandwich, parallel-displaced, and T-shaped) the largest discrepancy between EFP and CCSD(T)/aug-cc-pVQZ is 0.4 kcal/mol in binding energies and 0.2 Å in the equilibrium structures interfragment distances [in ref 69 the EFP, SAPT and CCSD(T) potential energy curves for various benzene structures were computed using frozen fragment geometries]. The agreement of the EFP interaction energies with the CCSD(T) values was better than that of MP2. Accurate evaluation of the electrostatic energy was found to be crucial for this system. For example, in the sandwich and parallel-displaced structures, the attractive Coulomb interactions are solely due to the charge-penetration effect, and, therefore, Coulomb damping is essential. Calculations of substituted benzene dimers (OH, CH₃, F, and CN substituents) confirmed the high accuracy of EFP in predicting equilibrium structures and binding energies in π -stacked aromatic complexes and describing the trends in binding energies in dimers with electron-withdrawing or electron-donating substituents.⁷⁰

The structures and energetics of different conformations of the styrene dimer, which feature various π – π , H-bonding, and mixed π -H interacting patterns, are correctly captured by EFP, as demonstrated by comparison against MP2.⁷¹

The prediction of structures and binding energies in water–benzene complexes, in which an interplay between π – π and H– π interactions results in unique structural patterns, is a stringent test for any computational technique. In these complexes, π – π and H– π interactions between benzenes and between benzene and water molecules are similar in strength and compete with H-bonds of water. EFP provides reliable results for these complicated systems, as compared to the CCSD(T) and MP2 methods. Interestingly, benzene molecules in a water environment are polarized and participate in the hydrogen-bond network of water.⁷²

Beyond benchmarking EFP has already enabled a number of insightful computational studies of extended systems. Some of the examples are highlighted below.

EFP studies of small water–methanol clusters revealed an incomplete mixing of water and methanol at the molecular level; that is, heterogeneous structures were found to be consistently lower in energy than the homogeneously mixed structures. For the smallest clusters, these results were confirmed by MP2.⁷³

EFP was also used in a study of the solvation of alanine.^{74,75} Structures of nonionized and zwitterionic conformers of alanine in water obtained with Monte Carlo sampling were used for calculation of the free energy difference between the nonionized and zwitterionic forms at the MP2 and DFT levels of theory in combination with the EFP1 potential for water and the polarizable continuum model (PCM) description of the bulk.

EFP was employed in studies of bulk properties of liquids using molecular dynamics simulations.⁷⁶ Additionally, coarse-graining of the EFP potential has been explored on the basis of these simulations.⁷⁷

Coupling the EFP potential to the quantum region through the frozen localized molecular orbital approach⁷⁸ allowed calculations of pK_a of various amino acids in proteins.^{79,80}

Recently, EFP was interfaced with the equation-of-motion coupled-cluster singles and doubles (EOM-CCSD) method to describe electronically excited states of chromophores in solution.⁸¹ A number of studies on the electronically excited states of biological chromophores using EFP combined with multi-reference methods have also been reported.^{82–85}

In sum, numerous previous studies conducted with GAMESS^{50,53} have demonstrated that EFP is an accurate and robust approach to intermolecular interactions. The present work discusses implementation of the general EFP method in the Q-CHEM electronic structure package.⁸⁶ We apply EFP to investigate the asymptotic behavior of noncovalent interactions in the stacks of DNA bases. As discussed above, the relative contributions to noncovalent interactions have been investigated for a variety of small model systems (dimers and small clusters). The focus of this work is on the asymptotic behavior of different contributions in large systems. Our goal is to understand the physics behind noncovalent interactions, for example, how the relative importance of different contributions changes with the system size, what is the magnitude of nonadditive effects, etc. This knowledge is important for empirical methods development, as discussed in a recent study by Sherrill and co-workers.³ We present the results for stacks of thymine and adenine at configurations relevant to the native DNA structure. We analyze the convergence of different components of the interaction energy to their bulk values. In the stacking interactions, the dispersion and exchange-repulsion play a significant role. The contribution of electrostatics varies. Regarding the range of interactions, the exchange-repulsion and dispersion are essentially short-range interactions, whereas polarization and Coulomb terms have a much longer range. A detailed analysis of the interactions of heterocyclic aromatic compounds including DNA bases will appear in a separate publication.⁸⁷

The structure of this paper is as follows. The next section describes the theory behind EFP and gives general expressions for different components of the interaction energy. Calculation of the EFP parameters is described in Section 3. The EFP calculations of nucleic acid base (NAB) oligomers are presented in Section 4. Programmable expressions and further details of the EFP implementation are given in the Supporting Information. Concluding remarks are presented in Section 5.

2. The EFP Method

In EFP, one describes (relatively weak) intermolecular interactions using perturbation theory starting from the noninteracting (unperturbed) fragments. The intermolecular interactions can be represented as a series of short- and long-range terms. Long-range interactions, which are proportional to the distance according to $1/R^n$, include Coulomb, induction (polarization), and dispersion terms, whereas short-range interactions that depend on the interfragment density overlaps consist of exchange-repulsion, charge-transfer, and screening terms.⁸⁸

The total energy of a system containing both EFP and QM components consists of the interactions between the effective fragments ($E^{\text{ef-ef}}$) and the energy of the QM region in the field of the fragments. The former includes Coulomb, polarization, dispersion and exchange-repulsion contributions (the charge-transfer term, which is important for a description of ionic and highly polar species,⁸⁹ is omitted here):

$$E^{\text{ef-ef}} = E_{\text{Coul}} + E_{\text{pol}} + E_{\text{disp}} + E_{\text{ex-rep}} \quad (1)$$

The QM-EF interactions are computed as follows. The Coulomb and polarization parts of the EFP potential contribute to the quantum Hamiltonian via one-electron terms:

$$\hat{H}'_{pq} = \hat{H}_{pq} + \langle p | \hat{V}^{\text{Coul}} + \hat{V}^{\text{pol}} | q \rangle \quad (2)$$

whereas dispersion and exchange-repulsion QM-EF interactions are treated as additive corrections to the total energy, that is, similar to the fragment interactions. For fully quantum coupling, these contributions should also be included in the QM Hamiltonian, as, for example, has been done in ref 90.

In the present implementations, the fragments are rigid and their potentials can be generated from a set of ab initio calculations on each unique isolated fragment.⁹¹

The Q-CHEM implementation includes a library of standard fragments with precomputed potentials. The EF potential includes: (i) multipoles (produced by Stone's distributed multipolar analysis) for Coulomb and polarization terms; (ii) static polarizability tensors centered at localized molecular orbital (LMO) centroids (obtained from coupled-perturbed Hartree-Fock calculations), which are used for calculations of polarization; (iii) dynamic polarizability tensors centered on the LMOs that are generated by time-dependent HF calculations and used for calculations of dispersion; and (iv) the Fock matrix, basis set, and localized orbitals needed for the exchange-repulsion term. In sum, all the EFP parameters are obtained from ab initio calculations on an isolated fragment and contain no empirically fitted parameters.

Below we describe the theoretical background of the different energy terms in the EFP method, followed by the detailed formulation and programmable expressions.

In the long-range perturbation expansion, the zero-order Hamiltonian $\hat{H}^{(0)}$ is the Hamiltonian of the noninteracting fragments A and B, and the perturbation \hat{H}' is a particular (e.g., Coulomb) interaction between the fragments:

$$\begin{aligned} \hat{H}^{(0)}|mn\rangle &= (\hat{H}^A + \hat{H}^B)|mn\rangle = (E_m^A + E_n^B)|mn\rangle = E_{mn}^{(0)}|mn\rangle \\ \hat{H}' &= \int \frac{\rho^A(r)\rho^B(r')}{|r-r'|} d^3r d^3r' \end{aligned} \quad (3)$$

where $|m\rangle$ and $|n\rangle$ are the eigenfunctions of \hat{H}^A and \hat{H}^B , respectively. Following the nondegenerate Rayleigh-Schrödinger perturbation theory, the ground state energies ($m = n = 0$) of the closed-shell molecules are:

$$\begin{aligned} E &= E^{(0)} + E^{(1)} + E^{(2)} + \dots \\ E^{(0)} &= E_0^A + E_0^B \\ E^{(1)} &= \langle 00 | \hat{H}' | 00 \rangle \\ E^{(2)} &= - \sum_{mn} \frac{\langle 00 | \hat{H}' | mn \rangle \langle mn | \hat{H}' | 00 \rangle}{E_{mn}^{(0)} - E_0^{(0)}} \end{aligned} \quad (4)$$

where E^A and E^B are the energies of the individual isolated fragments, $E^{(n)}$ are the perturbative corrections of order n , and $E_{mn}^{(0)}$ is the energy of the state $|mn\rangle$ for the two noninteracting fragments.

The first-order perturbative correction $E^{(1)}$ is the expectation value of the Coulomb operator \hat{H}' for the ground state $|00\rangle$. EFP describes the Coulomb interaction by a distributed multipole expansion (Section 2.1). The second-order perturbative correction consists of polarization (induction) and dispersion terms:

$$\begin{aligned}
 E_{\text{ind}}^{\text{A}} &= - \sum_{m \neq 0} \frac{\langle 00 | \hat{H}^{\text{A}} | m0 \rangle \langle m0 | \hat{H}^{\text{A}} | 100 \rangle}{E_m^{\text{A}} - E_0^{\text{A}}} \\
 E_{\text{ind}}^{\text{B}} &= - \sum_{n \neq 0} \frac{\langle 00 | \hat{H}^{\text{B}} | 0n \rangle \langle 0n | \hat{H}^{\text{B}} | 100 \rangle}{E_n^{\text{B}} - E_0^{\text{B}}} \\
 E_{\text{disp}} &= - \sum_{m \neq 0, n \neq 0} \frac{\langle 00 | \hat{H}^{\text{A}} | mn \rangle \langle mn | \hat{H}^{\text{B}} | 100 \rangle}{E_m^{\text{A}} + E_n^{\text{B}} - E_0^{\text{A}} - E_0^{\text{B}}}
 \end{aligned} \quad (5)$$

The polarization component of EFP is accounted for by distributed polarizabilities (Section 2.2). The dispersion interaction is expressed as an inverse R dependence:

$$E_{\text{disp}} = - \sum_{n \geq 6} C_n R^{-n} \quad (6)$$

where coefficients C_n are derived from the frequency-dependent distributed polarizabilities (Section 2.3). The higher-order terms in the perturbative expansion, eq 4, are expected to have smaller values and are neglected in the current EFP formalism. One of the reasons for the success of EFP⁶⁹ is that some of the higher-order terms apparently cancel each other. For example, charge-transfer and exchange-induction are similar in magnitude, but have opposite signs and thus cancel out. The exchange-dispersion partially cancels higher-order dispersion terms (such as induced dipole-induced quadrupole) that are also omitted in the EFP formalism.^{49,69}

The perturbative treatment of intermolecular interactions described above works well at large separations between fragments, but breaks down when molecules approach each other such that their electronic densities overlap. EFP accounts for this effect by including short-range terms, that is, the exchange-repulsion interaction (Section 2.4) and charge-penetration corrections to the long-range terms.⁶⁹

The exchange-repulsion interaction is derived as an expansion in the intermolecular overlap, truncated at the quadratic term,^{92,93} which requires that each effective fragment carries a basis set that is used to calculate overlap and kinetic one-electron integrals for each interacting pair of fragments (see Section 2.4).

The long-range terms are modulated by damping (or screening) expressions. The classical point multipole model breaks down when fragments approach too closely, since then the actual electron density on the two fragments is not well approximated by point multipoles. Consequently, at short separations between fragments, the Coulomb interactions become too repulsive, whereas the induction and dispersion interactions become too attractive. Damping terms are discussed in Sections 2.1 and 2.3.

2.1. Coulomb Energy. The Coulomb component of the EFP energy accounts for Coulombic interactions. In molecular systems with hydrogen-bonds or polar molecules, this is the leading contribution to the total intermolecular interaction energy.⁹⁴

Buckingham⁹⁴ has shown that the Coulomb potential of a molecule at point x can be expressed using a multipole expansion around point k :

$$\begin{aligned}
 V_k^{\text{Coul}}(x) &= q_k T(r_{kx}) - \sum_a \mu_a^k T_a(r_{kx}) + \frac{1}{3} \sum_{a,b} \Theta_{ab}^k T_{ab}(r_{kx}) - \\
 &\quad \frac{1}{15} \sum_{a,b,c} \Omega_{abc}^k T_{abc}(r_{kx}) \quad (7)
 \end{aligned}$$

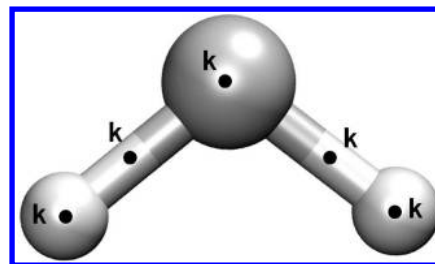


Figure 2. The distributed multipole expansion points for a water molecule. The Coulomb potential of an effective fragment is specified by charges, dipoles, quadrupoles, and octopoles placed at atomic centers and bond midpoints.

where q , μ , Θ , and Ω are the net charge, dipole, quadrupole, and octopole, respectively, and T , T_a , T_{ab} , and T_{abc} are the electrostatic tensors of the zero, first, second, and third ranks, respectively. This expression converges to the exact electrostatic potential when a complete (infinite) number of terms is included. However, the convergence is slow,⁹⁵ and a large number of terms is required to achieve reasonable accuracy, even when such expansions are assigned to the individual atoms using a Mulliken population analysis.⁹⁶ Thus, single-point or atomic-based representation of the molecular electrostatic potential is impractical.

An accurate representation of the electrostatic potential, which has been exploited in the EFP approach,^{47,49} is achieved by using a multipole expansion (obtained from Stone's distributed multipole analysis) around atomic centers and bond midpoints (i.e., the points with high electronic density) and truncating this expansion at octopoles.^{88,95,97} The expansion points for water are shown in Figure 2.

The fragment–fragment Coulomb interactions consist of charge–charge, charge–dipole, charge–quadrupole, charge–octopole, dipole–dipole, dipole–quadrupole, and quadrupole–quadrupole terms, as well as terms describing interactions of multipoles with the nuclei and nuclear repulsion energy. The corresponding equations are given in the Supporting Information.

The Coulomb interaction between an effective fragment and the QM part is described by a perturbation \hat{V} of the ab initio Hamiltonian:

$$\hat{H} = \hat{H}_0 + \hat{V} \quad (8)$$

The perturbation enters the one-electron part of the Hamiltonian as a sum of contributions from the expansion points k and nuclei I of the effective fragments A :

$$V = \sum_{p,q} \sum_A \left(\sum_{k \in A} d_{pq} \langle p | V_k^{\text{Coul}} | q \rangle + \sum_{I \in A} d_{pq} \left\langle p \left| \frac{Z_I}{R} \right| q \right\rangle \right) \quad (9)$$

where p and q are atomic orbitals in the ab initio region, Z_I is a nuclear charge, d_{pq} is an element of the atomic density matrix, and V_k^{Coul} is the Coulomb potential from the expansion point k .

The one-electron contribution to the Hamiltonian from the individual expansion point k consists of four terms originating from the electrostatic potential of the corresponding multipole:⁴⁷

$$\begin{aligned}
 \langle p | V_k^{\text{Coul}} | q \rangle &= \left\langle p \left| \frac{q^k}{R} \right| q \right\rangle + \left\langle p \left| \frac{\sum_a^{x,y,z} \mu_a^k}{R^3} \right| q \right\rangle \\
 &+ \left\langle p \left| \frac{\sum_{a,b}^{x,y,z} \Theta_{ab}^k (3ab - R^2 \delta_{ab})}{3R^5} \right| q \right\rangle \\
 &+ \left\langle p \left| \frac{\sum_{a,b,c}^{x,y,z} \Omega_{abc}^k (5abc - R^2 (a\delta_{bc} + b\delta_{ac} + c\delta_{ab}))}{5R^7} \right| q \right\rangle
 \end{aligned} \quad (10)$$

where R , a , b , and c are the distance and its Cartesian components between the electron and the expansion point k . The first integral in eq 10 has the same form as an electron–nucleus attraction integral; however, the remaining terms are not typically encountered in standard electronic structure calculations. Such integrals are readily evaluated in Gaussian basis sets using Obara–Saika recurrence relations,⁹⁸ which we have implemented within Q-CHEM’s general multipole one-electron integral code.

Charge Penetration. The multipole representation of the Coulomb density of a fragment breaks down at close separations between the fragments. The multipole interactions become too repulsive due to significant overlap of the electronic densities and the charge-penetration effect. The magnitude of the charge-penetration effect is usually around 15% of the total Coulomb energy in polar systems; however, it can be as large as 200% of the Coulomb energy in systems with weak Coulomb interactions.⁶⁶ Several electrostatic screening functions have been considered.^{66,69,99} In the present implementation, we use a simple exponential damping of the charge–charge term, which was shown to efficiently account for most of the charge-penetration effect.⁹⁹ Screening of the higher-order multipoles, which can also be important, is not considered here. The charge–charge screened energy between the expansion points k and l is given by the following expression, where α_k and α_l are the damping parameters associated with the corresponding expansion points:

$$\begin{aligned}
 E_{kl}^{\text{ch-ch}} &= \\
 &\begin{cases} \left(1 - \left(1 + \frac{\alpha_k R_{kl}}{2} \right) e^{-\alpha_k R_{kl}} \right) \frac{q^k q^l}{R_{kl}} & \text{if } \alpha_k = \alpha_l \\ \left(1 - \frac{\alpha_l^2}{\alpha_l^2 - \alpha_k^2} e^{-\alpha_k R_{kl}} - \frac{\alpha_k^2}{\alpha_k^2 - \alpha_l^2} e^{-\alpha_l R_{kl}} \right) \frac{q^k q^l}{R_{kl}} & \text{if } \alpha_k \neq \alpha_l \end{cases}
 \end{aligned} \quad (11)$$

Generation of the damping parameters is discussed in Section 3.

Ab initio-EFP Coulomb interactions are calculated without damping corrections in the implementation presented here.

2.2. Polarization Energy. Polarization (induction) accounts for the intramolecular charge redistribution in response to an external electric field. It is the major component of many-body interactions responsible for cooperative molecular behavior. Because polarization cannot be parametrized by pairwise potentials, it is often neglected in molecular mechanics models. However, in hydrogen-bonded systems, up to 20% of the total intermolecular interaction energy is due to polarization.⁴⁷

The polarization interaction appears in the second-order of long-range perturbation theory, see eq 4. Expanding the

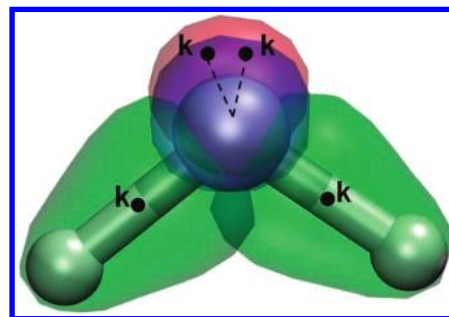


Figure 3. Distributed polarizabilities for a water molecule described by dipole polarizability tensors, see eq 12.

Coulomb potential V in a $(1/R)^n$ series and combining terms with the same electric field derivatives yields the following equation:

$$\begin{aligned}
 E^{\text{pol}} &= -\frac{1}{2} \alpha_{ab} F_a F_b - \frac{1}{6} \beta_{abc} F_a F_b F_c - \frac{1}{3} A_{a,bc} F_a F_{bc} \\
 &- \frac{1}{6} B_{ab,dc} F_a F_b F_{cd} - \frac{1}{6} C_{ab,dc} F_{ab} F_{cd} + \dots
 \end{aligned} \quad (12)$$

where F_a and F_{ab} are the first and second electric field derivatives; α is the static dipole polarizability tensor; β is the static dipole hyper-polarizability; and A , B , and C are quadrupole polarizability tensors.

The EFP method employs distributed polarizabilities placed at the centers of the valence LMOs. This allows the truncation of the expansion in eq 12 after the first term while retaining reasonable accuracy.⁴⁷ Unlike the isotropic total molecular polarizability tensor, the distributed polarizability tensors are anisotropic. Polarizability points for a water molecule are shown in Figure 3.

The polarization energy of a system consisting only of effective fragments is:

$$E^{\text{pol}} = -\frac{1}{2} \sum_k \sum_a^{x,y,z} \mu_a^k F_a^{\text{mult},k} \quad (13)$$

where μ_a^k are the Cartesian components of the induced dipoles at the distributed point k , and $F_a^{\text{mult},k}$ are the Cartesian components of the external field due to static multipoles and nuclei of other fragments. The induced dipoles at each polarizability point are computed as:

$$\mu_a^k = \sum_b^{x,y,z} \alpha_{ab}^k F_b^{\text{total},k} \quad (14)$$

where the total field comprises the static field plus the field due to other induced dipoles F_k^{ind} :

$$F_a^{\text{total},k \in A} = F_a^{\text{mult},k} + F_a^{\text{ind},k} = F_a^{\text{mult},k} + \sum_{l \in B, B \neq A} \sum_b^{x,y,z} T_{ab}^{kl} \mu_b^l \quad (15)$$

It follows from the above equation that the induced dipoles on fragment A depend on the values of the induced dipoles of all other fragments (B). $T_{ab}^{kl} = (3ab - R_{kl}^2 \delta_{ab}) / (R_{kl}^5)$ (R and a, b

are the distance and its Cartesian components) is the second-order Coulomb tensor. Therefore, the induced dipoles on fragments should be computed self-consistently. Equation 15 can be rewritten as:

$$F_a^{\text{mult},k} = \sum_l^{\text{fragments}} \sum_b^{x,y,z} D_{ab}^{kl} \mu_b^l \quad (16)$$

or

$$\mu_a^k = \sum_l^{\text{fragments}} \sum_b^{x,y,z} (D_{ab}^{kl})^{-1} F_b^{\text{mult},l} \quad (17)$$

where the tensor D_{ab}^{kl} is defined as:

$$D_{ab}^{kl} = \begin{cases} (\alpha_{ab}^k)^{-1} & \text{if } l = k, l \in A \\ 0 & \text{if } l \neq k, k, l \in A \\ -T_{ab}^{kl} & \text{if } k \in A, l \in B, A \neq B \end{cases} \quad (18)$$

Thus, to determine the induced dipoles, one formally needs the inverse of the matrix D_{ab}^{kl} . A more efficient approach used in the EFP method is based on solving eq 17 iteratively.

In a system consisting of effective fragments and an ab initio region, the total electric field $F_a^{\text{ai-total}}$ acting on an effective fragment consists of the static and induced fields due to other effective fragments, as well as fields due to the electron density $F_a^{\text{ai-elec}}$ and nuclei $F_a^{\text{ai-nuc}}$ of the ab initio part:

$$F_a^{\text{ai-total},k} = F_a^{\text{mult},k} + F_a^{\text{ind},k} + F_a^{\text{ai-elec},k} + F_a^{\text{ai-nuc},k} \quad (19)$$

The field due to the ab initio electron density is expressed through the electric field operator \hat{f}^{elec} as:

$$F_a^{\text{ai-elec},k} = \langle \Psi | \hat{f}_k^{\text{elec}} | \Psi \rangle \quad (20)$$

Thus, the induced dipoles on the effective fragments depend on the ab initio electron density, which, in turn, is affected by the field created by these induced dipoles through a one-electron contribution to the Hamiltonian:

$$\hat{V}^{\text{pol}} = -\frac{1}{2} \sum_k \sum_a^{x,y,z} \frac{(\mu_a^k + \bar{\mu}_a^k) a}{R^3} \quad (21)$$

where R and a are the distance and Cartesian components between an electron and the polarizability point k . $\bar{\mu}_a^k$ is the conjugate induced dipole:

$$\bar{\mu}_a^k = \sum_b^{x,y,z} (\alpha_{ab}^k)^T F_b^{\text{total},k} \quad (22)$$

where $(\alpha_{ab}^k)^T$ is the transposed polarizability tensor at the polarization point k .

The total polarization energy is computed self-consistently using a two-level iterative procedure illustrated in Figure 4. The objectives of

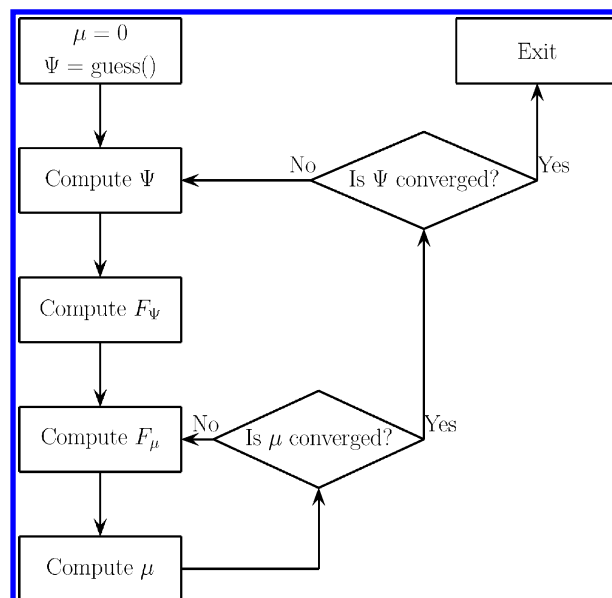


Figure 4. Computing the induced dipoles through the two-level self-consistent iterative procedure. Ψ is the ab initio wave function, μ is a set of induced dipoles, F_Ψ is the field due to the wave function, and F_μ is the field due to the induced dipoles.

the higher and lower levels are to converge the wave function and the induced dipoles for a given fixed wave function, respectively.

At every higher-level iteration, the wave function is updated based on the values of the induced dipoles from the previous step. Then the field from this new wave function acting on the polarizability points is recomputed. The convergence criterion is that the wave function parameters (e.g., molecular orbital coefficients) are within a predefined range from their values on the previous iteration.

At the lower level, the values of the induced dipoles are computed at every iteration based on the electric field from the ab initio part (which remains constant during the lower-level iterations), as well as the static and induced fields from the previous lower-level iteration. The convergence criterion is that the difference between the new induced dipoles and those from the previous iteration is within a predefined threshold.

Thus, the lower-level iterative procedure exits when the induced dipoles are self-consistent and are consistent with a frozen ab initio wave function. Convergence of the higher-level iterative procedure yields self-consistent induced dipoles and the ab initio wave function. If the system does not contain the ab initio part, only the lower-level procedure is executed.

Once the values of the induced dipoles are obtained, the total polarization energy of the system can be computed as:⁴⁷

$$E^{\text{pol}} = -\frac{1}{2} \sum_k \sum_a^{x,y,z} \mu_a^k (F_a^{\text{mult},k} + F_a^{\text{ai-nuc},k}) + \frac{1}{2} \sum_k \sum_a^{x,y,z} \bar{\mu}_a^k F_a^{\text{ai-elec},k} \quad (23)$$

The self-consistent treatment of the polarization accounts for many-body interaction effects. The present implementation does not include damping of the polarization energy.⁶⁹

2.3. Dispersion. Dispersion provides a leading contribution to stabilization in aromatic π -stacked compounds such as DNA strands. However, an accurate description of dispersion is a difficult task since it arises due to the correlated motion of electrons.

The exact expression for the dispersion energy appears in the second order of long-range perturbation theory. Expanding the electrostatic potential V in the series $(1/R)^n$ in eq 4, the dispersion energy can be expressed in the so-called London series:

$$E^{\text{disp}} = \frac{C_6}{R^6} + \frac{C_8}{R^8} + \frac{C_{10}}{R^{10}} + \dots \quad (24)$$

where the C_n coefficients correspond to induced dipole–induced dipole (C_6), induced dipole–induced quadrupole (C_8), induced quadrupole–induced quadrupole and induced dipole–induced octopole (C_{10}) interactions. Additionally, odd power terms (R^{-7} , R^{-9} , etc) appear in the expansion for noncentrosymmetric molecules or LMOs.¹⁰⁰

Similarly to the other EFP terms, the dispersion is treated in a distributed fashion, with expansion points located at the LMO centroids, that is, at the same centers as the polarization expansion points. Reasonable accuracy is achieved by truncating the London expansion, eq 24, after the first term and approximating the rest as 1/3 of this term.¹⁰¹

The induced dipole–induced dipole dispersion energy between the expansion points k and l can be computed by using the following equation:

$$E_{kl}^{\text{dip-dip}} = \sum_{abcd}^{x,y,z} T_{ab}^{kl} T_{cd}^{kl} \int_0^\infty \alpha_{ac}^k(iv) \alpha_{bd}^l(iv) dv \quad (25)$$

where α_{ac}^k and α_{bd}^l are dynamic polarizability tensors, iv is the imaginary frequency, T_{ab}^{kl} and T_{cd}^{kl} are the second-rank electrostatic tensors. The indices k and l refer to LMOs on the two interacting fragments, and the indices a, b, c , and d run over x, y , and z . The interaction energy described by eq 25 is anisotropic, distance-dependent, and computationally expensive.¹⁰¹ The approximation used in EFP includes only the trace of the polarizability tensors, which is isotropic and distance-independent.

The total dispersion energy for a system of effective fragments is:

$$E^{\text{disp}} = \frac{4}{3} \sum_{A,B} \sum_{k \in A} \sum_{l \in B} \frac{C_6^{kl}}{R_{kl}^6} \quad (26)$$

where A and B are effective fragments, k and l are LMO centroids, C_6^{kl} is the dispersion coefficient, and R_{kl} is the distance between two LMO centroids.

Intermolecular dispersion coefficients for each pair of expansion points k and l are computed using the 12 point Gauss–Legendre integration formula:

$$C_6^{kl} = \sum_{i=1}^{12} w_i \frac{2\nu_0}{(1-t_i)^2} \bar{\alpha}^k(iv_i) \bar{\alpha}^l(iv_i) \quad (27)$$

where $\bar{\alpha}^k(iv_i)$ and $\bar{\alpha}^l(iv_i)$ are one-third of the traces of the corresponding polarizability tensors.

For small distances between effective fragments, dispersion interactions must be corrected for charge penetration and the electron density overlap effect. This is ensured by the

Tang–Toennies damping formula¹⁰² with parameter $b = 1.5$, similarly to the original EFP implementation:

$$C_6^{kl} \rightarrow \left(1 - e^{-bR} \sum_{k=0}^{\infty} \frac{(bR)^k}{k!} \right) C_6^{kl} \quad (28)$$

We should note, however, that recent studies revealed that this formula overdamps the dispersion and a more appropriate damping function is based on the intermolecular overlap.⁶⁹

Ab initio–EFP dispersion interactions in the present implementation are treated as EFP–EFP interactions, that is, the QM region is represented as a fragment, and the dispersion QM–EFP interaction is evaluated using eq 26.

2.4. Exchange-Repulsion. Exchange-repulsion originates from the Pauli exclusion principle, which states that the wave function of two identical fermions must be antisymmetric. In traditional classical force fields, exchange-repulsion is introduced as a positive (repulsive) term, for example, $(1/R)^{12}$ in the Lennard–Jones potential. However, since the Pauli exclusion principle is intrinsically quantum mechanical, its classical description may not be sufficiently accurate. The EFP method uses a wave function-based formalism to account for the electron exchange.

The EFP exchange-repulsion energy term has been derived from the exact equation for the exchange-repulsion energy of two closed-shell molecules:^{65,92,93}

$$E^{\text{exch}} = \frac{\langle \Psi_A \Psi_B | \hat{A} \hat{H}_{AB} | \Psi_A \Psi_B \rangle}{\langle \Psi_A \Psi_B | \hat{A} | \Psi_A \Psi_B \rangle} - \langle \Psi_A \Psi_A | \hat{V}_{AB} | \Psi_B \Psi_B \rangle - E_A - E_B \quad (29)$$

where A and B are two molecules described by Ψ_A and Ψ_B wave functions, \hat{V}_{AB} is the intermolecular Coulomb operator, and \hat{A} is the antisymmetrization operator:

$$\hat{A} = P_0 - P_1 + P_2 - P_3 + \dots \quad (30)$$

where P_i is a permutation of i electron pairs consisting of one electron from each molecule.

Truncating sequence 30 after the second term and applying an infinite basis set and a spherical Gaussian overlap approximation led to the following expression for the exchange-repulsion energy in terms of LMOs:^{65,92}

$$E^{\text{exch}} = \frac{1}{2} \sum_A \sum_{B \neq A} \sum_{i \in A} \sum_{j \in B} E_{ij}^{\text{exch}} \quad (31)$$

$$E_{ij}^{\text{exch}} = -4 \sqrt{\frac{-2 \ln |S_{ij}|}{\pi}} \frac{S_{ij}^2}{R_{ij}} - 2S_{ij} \left(\sum_{k \in A} F_{ik}^A S_{kj} + \sum_{l \in B} F_{jl}^B S_{li} - 2T_{ij} \right) + 2S_{ij}^2 \left(\sum_{J \in B} -Z_J R_{iJ}^{-1} + 2 \sum_{l \in B} R_{il}^{-1} + \sum_{I \in A} -Z_I R_{Ij}^{-1} + 2 \sum_{k \in A} R_{kj}^{-1} - R_{ij}^{-1} \right) \quad (32)$$

where A and B are the effective fragments; i, j, k , and l are the LMOs; I and J are the nuclei; S and T are the intermolecular

overlap and kinetic energy integrals, respectively; and F is the Fock matrix.

The expression for E_{ij}^{exch} involves overlap and kinetic energy integrals between pairs of localized orbitals. In addition, since eq 32 is derived within an infinite basis set approximation, a reasonably large basis set is required for accuracy, that is, 6-31+G(d) is considered to be the smallest acceptable basis set, 6-311++G(3df,2p) is recommended. These factors make the exchange-repulsion the most computationally expensive part of EFP energy calculations in moderately sized systems (excluding charge-transfer⁸⁹).

Large systems require additional considerations. Equation 31 contains a sum over all fragment pairs and its computational cost formally scales as $O(N^2)$ with the number of effective fragments N . However, exchange-repulsion is a short-range interaction because the overlap and kinetic energy integrals decay exponentially with the interfragment distance. Therefore, by employing a distance-based screening, the number of essential overlap and kinetic energy integrals scales as $O(N)$. Consequently, for large systems the cost of exchange-repulsion will eventually become smaller than the evaluation of the long-range EFP components (such as Coulomb interactions).

The ab initio–EFP exchange-repulsion energy is currently calculated at the EFP–EFP level, by representing the quantum part as an EFP and using eq 32. In this way, the quantum Hamiltonian is not affected by the exchange potential of the fragments. The rigorous quantum coupling between the ab initio and EFP regions has been recently developed, but the computational cost is high.⁹⁰

3. Generation of the EF Potentials and the Fragment Library

The EF potential consists of the following parameters: coordinates of atoms, coordinates of the multipolar expansion points (typically, atoms and bond midpoints), the distributed multipole moments (up to octopoles), the electrostatic screening parameters, the coordinates of the LMO centroids, the static and dynamic polarizability tensors at the LMO centroids, the wave function and Fock matrix elements in the basis of the localized molecular orbitals, and the atomic labels of the EF atoms.

Coordinates of Effective Fragments. The position of a fragment is specified in terms of translation coordinates (x, y, z) and rotation by the Euler angles (α, β, γ) relative to the fragment frame, that is, coordinates of the standard fragment that are provided along with the other fragment parameters (see Figure 5). The rotation matrix corresponding to these angles is:

$$\mathbf{R} = \begin{pmatrix} \cos \alpha \cos \gamma - \sin \alpha \cos \beta \sin \gamma & -\cos \alpha \sin \gamma - \sin \alpha \cos \beta \cos \gamma & \sin \beta \sin \alpha \\ \sin \alpha \cos \gamma + \cos \alpha \cos \beta \sin \gamma & -\sin \alpha \sin \gamma + \cos \alpha \cos \beta \cos \gamma & -\sin \beta \cos \alpha \\ \sin \beta \sin \gamma & \sin \beta \cos \gamma & \cos \beta \end{pmatrix} \quad (33)$$

Coulomb Parameters. The distributed multipoles q, μ, Θ, Ω are obtained from an ab initio electronic density of the individual fragment using the distributed multipole analysis (DMA) procedure developed by Stone.^{95,97} First, a molecular wave function expressed in the basis of N gaussians is computed with an ab initio (typically, HF) method. Then, spherical multipole expansions are calculated at the centers of the Gaussian basis function products. These expansions are finite and include multipoles up to order $L^1 + L^2$ where L^1 and L^2 are angular momenta of the two gaussians. These multipoles are translated to the EFP expansion points (atomic centers and bond midpoints) as described in ref 95 and are truncated after the 4th term. This set of spherical multipoles is converted to traceless Buckingham multipoles using the equations provided in the Supporting Information.

Coulomb Damping Parameters. The screening parameters α [eq 11] are computed using a fitting procedure (described in details in refs 66 and 99) once for every type of fragment. This procedure optimizes the screening parameters so that the sum of the differences between the ab initio Coulomb potential (typically, the HF potential) and the EFP Coulomb potential with screening is minimized:

$$\Delta = \sum_p [V_{\text{ab initio}}^{\text{Coul}}(p) - V_{\text{EFP}}^{\text{Coul,damp}}(p)] \quad (34)$$

where the sum is taken over all points p of the grid. The screened EFP potential $V_{\text{EFP}}^{\text{Coul,damp}}$ has the following form:

$$V_{\text{EFP},k(x)}^{\text{Coul,damp}} = q_k T(r_{kx}) f_{\text{damp},k}(x) - \sum_a^{x,y,z} \mu_a^k T_a(r_{kx}) + \frac{1}{3} \sum_{a,b}^{x,y,z} \Theta_{ab}^k T_{ab}(r_{kx}) - \frac{1}{15} \sum_{a,b,c}^{x,y,z} \Omega_{abc}^k T_{abc}(r_{kx}) \quad (35)$$

where the damping function f_{damp} is:

$$f_{\text{damp},k}(x) = 1 - e^{-\alpha_k r_{kx}} \quad (36)$$

The grid is constructed such that it includes the regions of space that are important for intermolecular interactions, that is, between 0.7 and 3.0 van der Waals radii of each atom in the fragment. A projection of such a grid for a water molecule is shown in Figure 6.

Polarization Parameters. The EF polarization parameters are the centers of LMOs and the polarizability tensors. The LMOs are obtained with the Boys localization procedure based on the ab initio electronic density.^{103,104} Polarizabilities are obtained as the derivatives of the dipole moment (μ) with respect to electric field (F):

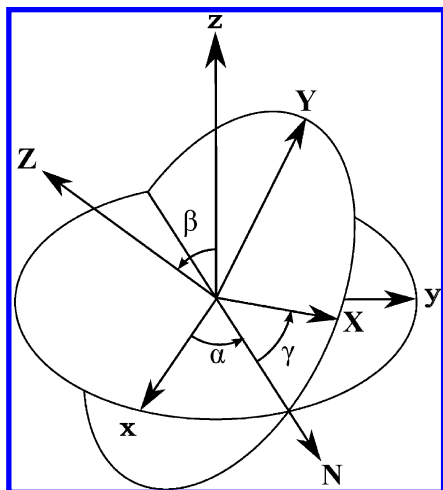


Figure 5. Representing the orientation with the Euler angles. xyz is the fixed system, XYZ is the rotated system, N is the intersection between the xy and XY planes called the *line of nodes*. Transformation between the xyz and XYZ systems is given by α rotation in the xy plane, followed by β rotation in the new $y'z'$ plane (around the N axis), followed by γ rotation in the new $x''y''$ plane.

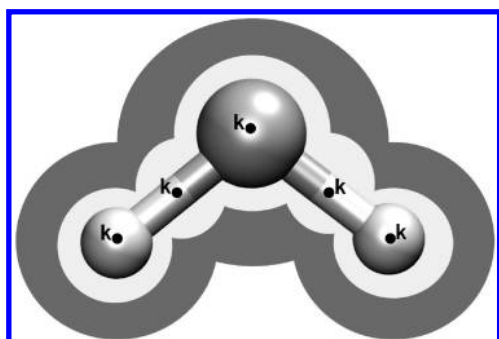


Figure 6. The shape of the grid (gray area) for fitting the screening parameters of a water molecule.

$$\alpha_{ab} = \frac{\partial \mu_a}{\partial F_b} \quad (37)$$

We use a finite difference procedure in which the dipole moment of an LMO is computed with and without an electric field and the polarizability tensor is computed as:

$$\alpha_{ab} = \frac{\mu_a(F_b) - \mu_a(0)}{F_b} \quad (38)$$

Dispersion Parameters. The dispersion parameters for an effective fragment are the coordinates of the LMO centroids and, for each LMO, the sets of traces of the dynamic polarizability tensors obtained at 12 frequencies corresponding to the Gauss–Legendre quadrature intervals. Dynamic polarizabilities can be computed using dynamic time-dependent Hartree–Fock (TDHF) or dynamic TD-DFT as described in ref 101.

Exchange-Repulsion Parameters. The parameters required for evaluation of the exchange-repulsion energy [eq 32] are the elements of the Fock matrix and the LMO information (orbital coefficients, basis set, and the LMO centroids) for each unique fragment. These parameters are extracted from a Hartree–Fock calculation followed by the orbital localization procedure (e.g., Boys localization).⁹³

Effective Fragment Library. To simplify EFP calculations, a library of standard fragments with precomputed effective

fragment potentials has been developed. Currently, the library includes 12 common organic solvents (see Supporting Information): acetone, carbon tetrachloride, dichloromethane, methane, methanol, ammonia, acetonitrile, water, dimethyl sulfoxide, benzene, phenol, and toluene, as well as fragment potentials for five nucleobases (adenine, thymine, cytosine, guanine, and uracil). The geometries of the solvent molecules were optimized using MP2/cc-pVTZ. Their EFP parameters were obtained using the HF method with the 6-31+G(d) (electrostatic multipoles and damping coefficients) and 6-311++G(3df,2p) (all other parameters) basis sets. For nucleobases, RI-MP2/cc-pVTZ optimized geometries were used. Their EFP parameters were obtained with the 6-31+G(d) (electrostatic multipoles and damping coefficients) and 6-311++G(3df,2p) (all other parameters) bases. The EFP parameters were obtained using the GAMESS program and were converted to the Q-CHEM format using a set of scripts (see Supporting Information for details).

4. Asymptotic Behavior of Noncovalent Interactions in Nucleobase Oligomers

NAB pairs and their stacked structures form the stable double-helical structure of DNA (see Figure 1). Several factors contribute to the stability of the structure of DNA,²¹ and their relative importance may depend on the size of the system. Our study addresses the following questions:

(1) How do relative contributions of various energy components (Coulomb, polarization, exchange-repulsion, and dispersion) depend on the size of the oligomers?

(2) Which components of the energy give rise to long-range interactions, and which components consist of only short-range interactions?

(3) How do far-field (non-nearest-neighbor) energy components depend on the cluster size? Can the sums of the dimer interaction energies be used to estimate the interaction energies for large oligomers with sufficient accuracy?

The term “fragment interaction energy” (or “interaction energy”) refers to the total EFP energy, which is the energy due to the interaction between the effective fragments.

In this work we follow the analysis of pairwise contributions in benzene oligomers by Sherrill and co-workers²⁷ and Tschumper and co-workers.¹⁰⁵ The goal is to investigate whether a simple sum of dimer energies is a good approximation for the interaction energies of the oligomer chains and to analyze the convergence of the different components of the interaction energy to their bulk values.

We begin by calibrating EFP against ab initio calculations using stacked and H-bonded adenine (AA) and thymine dimers (TT) (see Supporting Information). The agreement of the EFP stacked dimer and H-bonded dimer energies is within 1.5 and 3.5 kcal/mol, respectively, of the MP2 energies (note that the error bars of MP2 energies are of similar magnitude).^{106–108} We then proceed to investigate how the interaction energies depend on the relative orientation of the fragments (Section 4.1). The relative contributions of the various components of the EFP energies for adenine and thymine stacked structures are analyzed in Section 4.2. We then investigate the spatial extent of these interactions as well as the far-field component (i.e., the fraction that is not recovered by the energies of the neighboring dimers) of the various components (Sections 4.3 and 4.4) following the analysis in refs 27 and 105.

4.1. Effect of Fragment Rotation on Interaction Energy in the Adenine and Thymine Dimers. The experimental structure of the B-DNA helix can be described as the stacked AA/TT dimers with the fragments twisted by about 38° and

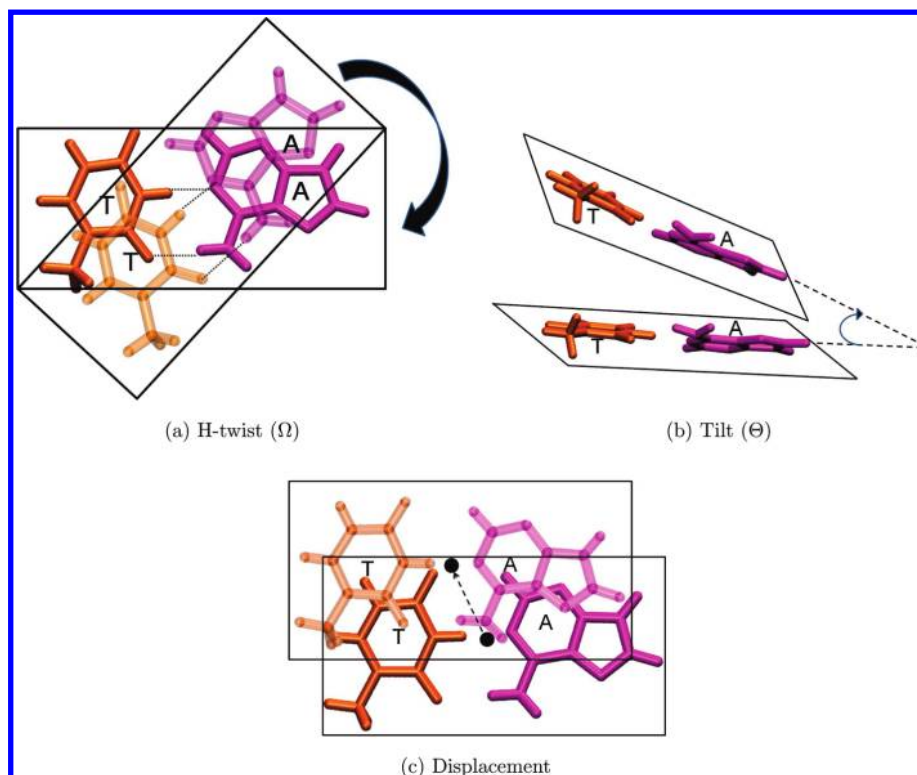


Figure 7. The h-twist (Ω), tilt (Θ), and displacement in the DNA helix. (a) The arrow indicates the rotation (h-twist) around the center of the base pair. This scissor-like motion results in a displacement in the XY plane along with a rotation of the adenine/thymine stacked dimers. (b) The arrow indicates the tilt angle between the planes of the adjacent base-pairs. (c) The arrow indicates the lateral displacement of the centers of the adjacent base pairs.

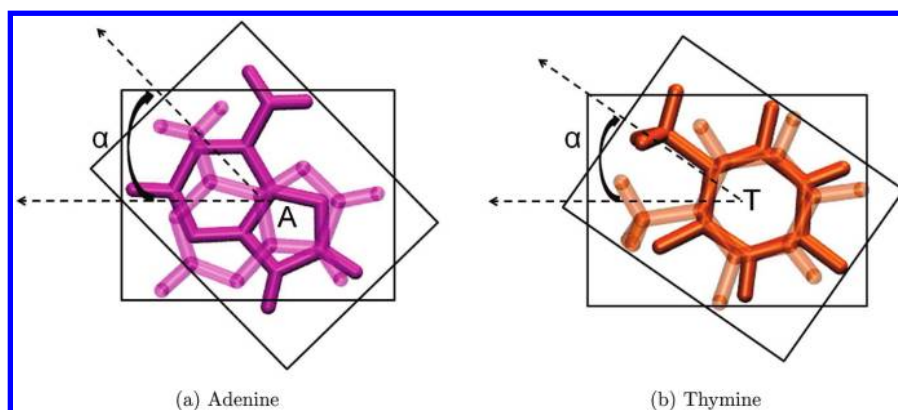


Figure 8. The rotation angle α in the stacked AA (left) and TT (right) dimers. The angle is defined as a rotation of the adenine or thymine monomer around its center of mass in the XY plane.

slightly displaced in the XY plane (see Figure 7). We begin by investigating the dependence of the interaction energy of the stacked dimers on the rotation around adenine or thymine centers of mass (see Figure 8). Since the distance between the AA/TT dimers in B-DNA is ≈ 3.33 Å, we investigate the effect of relative fragment rotation at the interfragment distance of ≈ 3.33 Å. The dependence of the interaction energy and its various components on this rotation angle α is shown in Figures 9 and 10, respectively.

As follows from Figure 9, the structure with two parallel stacked adenines with no rotation has low stabilization energy of ≈ 0.5 kcal/mol. The stabilization energy increases monotonically with the increase in α angle from 0 to 45° , and then decreases. Inspection of the individual components of the interaction energy (Figure 10, top) reveals that the maximum change in stabilization is due to the increase in Coulomb stabilization and the reduction of the exchange-repulsion

destabilization. The dispersion and polarization energies do not significantly contribute to the variations in stability due to the twist in the adenine dimers.

Figure 9 (bottom) shows the change in the fragment interaction energy with respect to the relative rotation between the thymine monomers in the TT stacked dimer. The trend is similar to that in the adenine dimers. The main difference is observed for the zero twist configuration, where the thymine dimers are not stable (the dimer has a destabilization energy of 6 kcal/mol) due to unfavorable charge-dipole interactions (the charge-dipole interaction in thymine is +1 kcal/mol, whereas it is -6 kcal/mol in adenine). The change in energy due to the rotation is also more pronounced in the case of thymine (≈ 14 kcal/mol) than in adenine (≈ 6 kcal/mol) mainly due to a more polar nature of thymine. Therefore, greater stabilization energy is gained due to the alignment of the polar NH and CO bonds

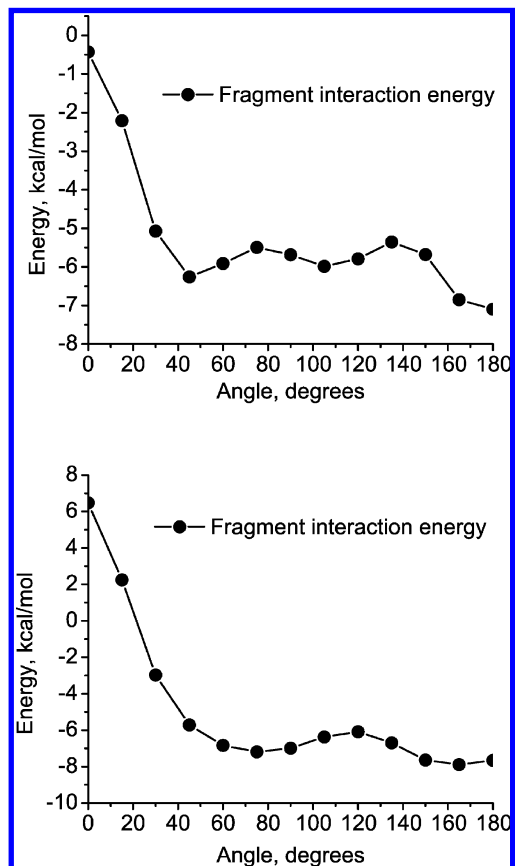


Figure 9. The dependence of the fragment interaction energy (in kcal/mol) on the rotation angle α between the fragments in the stacked AA (top) and TT (bottom) dimers. The fragments are parallel and the interfragment distance is fixed at 3.33 Å. There is no displacement in the XY plane (see Figure 8).

at a relative angle of 60°, as discussed in a recent study of TT and AA dimers.¹⁰⁹

Figure 10 (bottom), which summarizes the changes in the energy components due to the fragment rotation in the thymine dimer, shows that the components that exhibit substantial changes are Coulomb and exchange-repulsion. Near the rotation angle of $\alpha = 40^\circ$, which is close to that of the B-DNA structure, the total fragment interaction energy is ≈ -6 kcal/mol with the Coulomb component being ≈ -2.5 kcal/mol.

However, in addition to the relative rotation, the B-DNA structure also includes a displacement in the XY plane and the angle of rotation in the real B-DNA structure, known as the “h-twist”, is different from the simple rotation angle α discussed above. The effect of the h-twist (see Figure 7) causes a lateral displacement between the adenine/thymine molecules due to the axis of rotation being in the center of the adenine-thymine base pair. Thus, this scissor-like motion around the center of the base pair, changes the distance between the thymine dimers (in the XY plane) along with the relative angle between the dimers.

Table 1 shows the effect of the h-twist on the interaction energy components of the thymine dimers (see also Supporting Information). As one can see, the h-twist affects the exchange-repulsion and dispersion energies much more than a simple rotation around the center of the molecule. The change in the Coulomb energy along the h-twist angle plays a relatively small role compared to the corresponding change in exchange-repulsion.

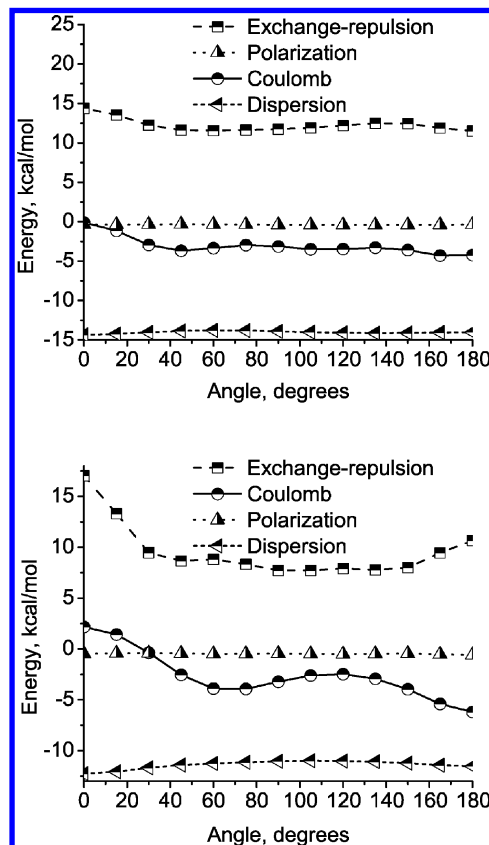


Figure 10. The dependence of the different energy components (in kcal/mol) on the rotation angle α between the fragments in the stacked AA (top) and TT (bottom) dimers. The fragments are parallel and the interfragment distance is fixed at 3.33 Å. There is no displacement in the XY plane (see Figure 8).

TABLE 1: Effect of the h-Twist (Ω , see Figure 7) on the Fragment Interaction Energies (in kcal/mol) of the Stacked Thymine Dimer

frag. angle	interaction energy	coulomb energy	exch-repulsion energy	polarization energy	dispersion energy
0	5.356	1.338	17.214	-0.467	-12.728
15	-1.205	2.259	6.694	-0.367	-9.790
30	-3.583	0.774	2.639	-0.300	-6.696
45	-3.475	-0.009	1.007	-0.261	-4.212

4.2. Relative Contribution of the Components of the Interaction Energy. Numerous theoretical studies have investigated the nature of noncovalent interactions.^{2,110–117} Owing to the high cost of ab initio calculations, the bulk of these calculations have been performed for dimers, with a handful of trimer and tetramer studies.^{1–3} The various components of interaction energy in small model clusters have also been studied.^{3,25,66} The goal of the present calculations is to understand how the components of the total fragment interaction energy vary with the size of the oligomers.

To understand the effect of long-range interactions of stacked nucleobases in DNA, we consider model oligomers of up to 20 stacked bases. The stacks were constructed by using the twist, tilt, and displacements (as shown in Figure 7) from the crystal structure of B-DNA, that is, the parameters for AA and TT dimers were extracted from the crystal structure and then 10 such identical dimers were replicated.

Figure 11 shows the percentage contribution of the different components of the interaction energy for stacked adenine/thymine oligomers as a function of the oligomer size. The

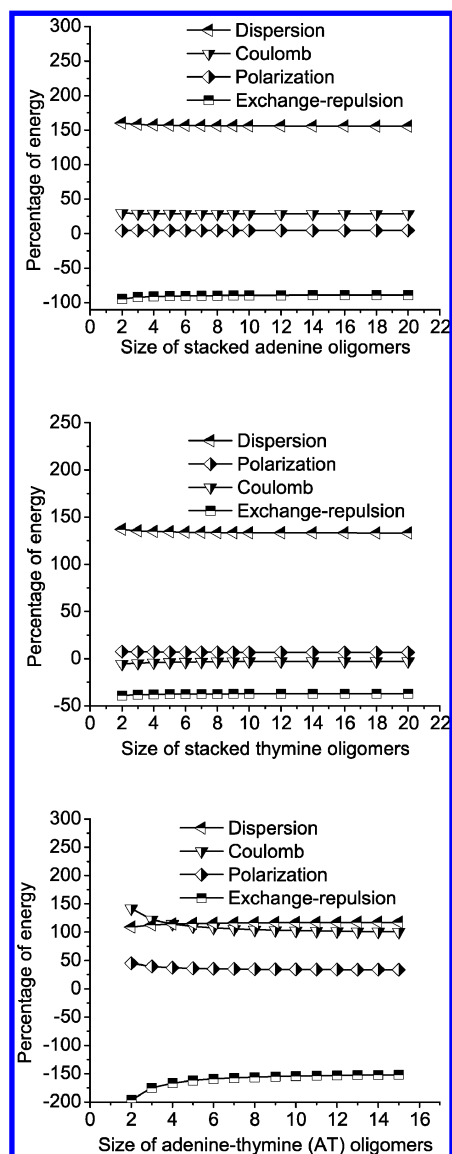


Figure 11. Percentage contribution of each component to the fragment interaction energy of adenine (top), thymine (middle), and adenine–thymine (bottom) stacked oligomers of various lengths.

relative contributions are rather insensitive to the length, which is somewhat surprising. Dispersion is the leading component in oligomers of all sizes in both adenine and thymine stacks. The attractive dispersion and Coulomb energies are balanced by the destabilization due to exchange-repulsion. The polarization contribution is relatively small in all of the oligomers. The only difference between the adenine and thymine stacks is the magnitude of the Coulomb component. Unexpectedly, the Coulomb component in the thymine stacks is quite small (almost negligible) due to the combined effect of the displacement in the *XY* plane and the twist of 38° (as shown in Table 1).

In the case of the adenine–thymine oligomers, where a dimer is defined as $(AT)_2$, the Coulomb and dispersion components are comparable to each other and provide the major part of the stabilization energy. This is due to the predominance of Coulomb in the H-bonded interactions in adenine–thymine pairs. The importance of polarization, though smaller than the Coulomb and dispersion terms, is significantly higher than in the case of individual adenine or thymine stacks due to H-bonded nature of the dimers.

These results suggest that the importance of the different components of the fragment interaction energy does not vary

TABLE 2: Different Components of the Fragment Interaction Energy (in kcal/mol) for Stacked Adenine Oligomers

length of oligomers	frag. interaction energy	coulomb energy	exch-repulsion energy	polarization energy	dispersion energy
2	-6.74	-2.02	6.38	-0.29	-10.80
3	-13.85	-4.04	12.74	-0.63	-21.92
4	-21.01	-6.06	19.10	-0.97	-33.08
5	-28.18	-8.10	25.46	-1.30	-44.24
6	-35.37	-10.14	31.84	-1.64	-55.42
7	-42.56	-12.19	38.20	-1.98	-66.59
8	-49.75	-14.24	44.57	-2.31	-77.76
9	-56.95	-16.30	50.93	-2.65	-88.93
10	-64.14	-18.35	57.30	-2.99	-100.11
12	-78.53	-22.44	70.02	-3.67	-122.44
14	-92.93	-26.54	82.75	-4.35	-144.79
16	-107.3	-30.64	95.49	-5.04	-167.13
18	-121.7	-34.75	108.22	-5.72	-189.47
20	-136.1	-38.85	120.95	-6.41	-211.82

TABLE 3: Different Components of the Fragment Interaction Energy (in kcal/mol) for Stacked Thymine Oligomers

length of oligomers	frag. interaction energy	coulomb energy	exch-repulsion energy	polarization energy	dispersion energy
2	-3.75	0.21	1.47	-0.28	-5.15
3	-7.68	0.35	2.93	-0.56	-10.41
4	-11.63	0.48	4.40	-0.83	-15.68
5	-15.60	0.59	5.87	-1.09	-20.96
6	-19.57	0.69	7.34	-1.36	-26.24
7	-23.57	0.76	8.81	-1.61	-31.52
8	-27.55	0.84	10.28	-1.87	-36.80
9	-31.53	0.93	11.75	-2.12	-42.09
10	-35.51	1.02	13.22	-2.38	-47.37
12	-43.44	1.23	16.16	-2.90	-57.93
14	-51.37	1.45	19.10	-3.42	-68.50
16	-59.32	1.64	22.04	-3.94	-79.06
18	-67.26	1.85	24.98	-4.46	-89.62
20	-75.19	2.05	27.91	-4.98	-100.17

significantly with the system size. The dispersion and exchange-repulsion always play the dominant role in these π – π interactions, while the polarization contribution is small. However, the Coulomb part of the interaction energy can vary significantly depending on the system and relative fragment orientation.

4.3. Long-Range Interaction. To understand the effect of long- and short-range interactions and the relative importance of different components, we examined the same 20 units long oligomer stacks of adenine and thymine. Tables 2 and 3 show the total fragment interaction energies and their components for different oligomers. The absolute values of all energy components increase monotonically with the stack length due to the increasing number of interactions. Intuitively, the most important interactions are between the nearest neighbors. For the stack of N fragments, there are $N - 1$ such two-body interactions. To quantify the change in energy per interacting pair with the size of the oligomer, we considered a more relevant quantity, that is, the fragment interaction energy scaled by $1/(N - 1)$. Figure 12 shows the scaled components of the fragment interaction energy. As one can see, each of the components scaled per number of the nearest-neighbor two-body interactions is fairly size-independent. This confirms that the leading contribution indeed comes from the nearest-neighbor two-body interactions.

Let us now consider the derivative of the plots from Figure 12 scaled to the inverse of the component itself, that is, $(1/E)(dE/dN)$, where E is the energy component, dE is the change in the energy component, and dN is the change in the length of the stacked oligomers (Figure 13). These plots provide a measure of the rate of change of the energy components with the size of

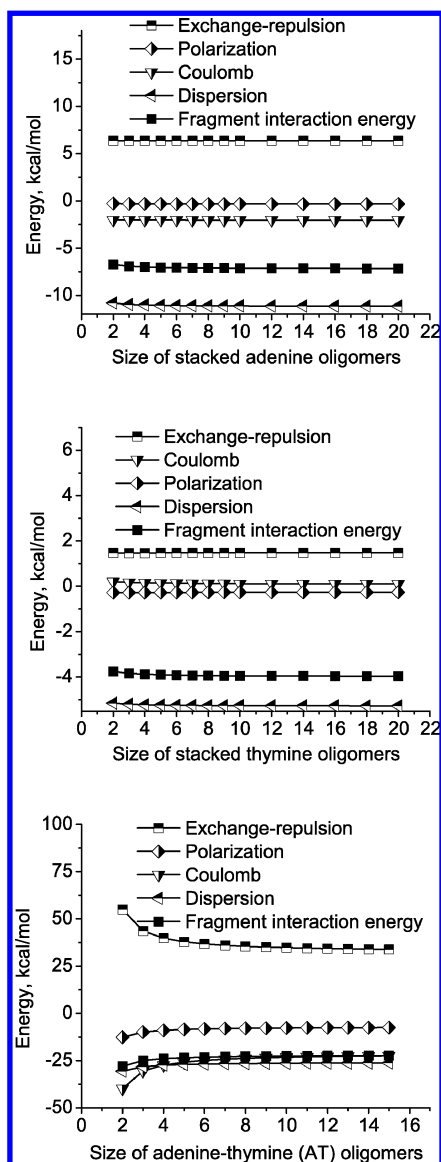


Figure 12. Different components of the interaction energy (in kcal/mol) scaled by the number of the nearest two-body interaction as a function of the oligomer size for adenine (top), thymine (middle), and adenine–thymine (bottom) stacks.

the oligomers. The asymptotic behavior of the derivative plot (insets in Figure 13) shows the length scale over which each interaction occurs and contributes to the respective component of the energy.

For both adenine and thymine, the derivative for the exchange-repulsion energy is nearly zero for all sizes of oligomers, which is consistent with the short-range pairwise character of this term. At the asymptotic limit, the maximum value in the derivative plots for adenine occurs for polarization followed by Coulomb and dispersion. This is somewhat counterintuitive since the Coulomb energy includes the charge–charge interactions that scale as $1/R$ and are expected to exhibit the slowest decay. However, the polarization decays more slowly. A closer examination of the different components of the Coulomb energy provides an explanation. The leading contribution to the Coulomb terms is not the charge–charge term but the dipole–dipole and charge–dipole terms. This results in a faster than $1/R$ decay of the Coulomb interaction energy. Overall, the $1/R$ decay holds for charged species only, whereas for neutrals the leading Coulomb term is almost always dipole–dipole and decreases as $1/(R^3)$.

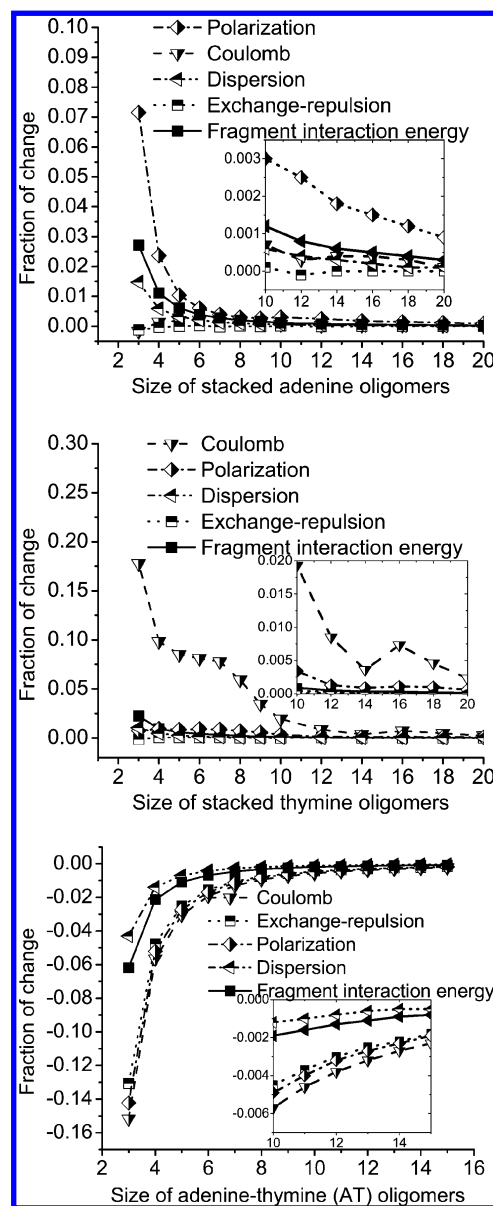


Figure 13. Derivatives of the interaction energy components per number of the nearest two-body interactions scaled to the component as a function of the oligomer length. The inset zooms into the data for largest size oligomers.

The derivative plots for thymine oligomers are somewhat different—the Coulomb component exhibits very large changes even at longer oligomer sizes, showing the long-range nature of the Coulomb component consistent with intuitive expectations based on $1/R$ decay of the predominant charge–charge term in the Coulomb component. However, due to the small absolute values of the Coulomb component, the long-range nature of Coulomb interactions in the thymine oligomers is of minor importance. The shoulder in the Coulomb component of the derivative plot of the thymine oligomers is possibly an artifact of the very small Coulomb energy and the sensitive cancellations between the various components of the Coulomb energy.

The derivative plots for the adenine–thymine oligomers show a trend that is very different from the individual adenine and thymine oligomers. The rate of change is much higher than in the pure stacks and so is the order in which the individual components differ. This is mainly due to the fact that in each of the $(AT)_n$ oligomers there are varying numbers of different interactions (H-bonded AT, cross H-bonded AT, and stacking

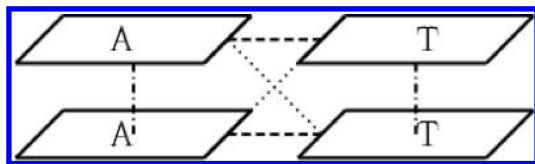


Figure 14. Different types of interactions in the adenine–thymine dimer: H-bond between adjacent adenine and thymine (dash), H-bond between the adenine and thymine molecules that are one step away from each other (dot), and stacking interactions between adenines and thymines (dash-dot). In the case of $(AT)_n$, if we consider only nearest-neighbor interactions, there are $(n - 1)$ adenine and thymine stacked interactions (dash-dot), n AT H-bonding interactions (dash) and $2n$ cross H-bonding interactions (dot).

interactions for AA and TT, as shown in Figure 14), while we are considering the change with respect to the stacking interactions only.

To summarize, the exchange-repulsion energy is a short-range interaction that can be accurately approximated by nearest-neighbor energies. The dispersion interaction ($\propto 1/(R^6)$) has a longer-range than exchange-repulsion [$\propto \exp(-aR)$]; however, it falls off within a 10-mer. The Coulomb and polarization are the longest-range interactions; however, the actual length scale over which they provide significant contributions is system-specific due to the intricate interplay between different components (i.e., charge–charge, charge–dipole, charge–quadrupole, etc).

4.4. Far-field Interactions and Their Dependence on the Size of the Oligomers. The next problem that we address is the relative importance of the nonpairwise, or, more precisely, far-field (non-nearest-neighbor) part of the interaction energy and its dependence on the size of the oligomers. This analysis assesses the feasibility and accuracy of representing multiple noncovalent π – π interactions by a sum of pairwise interactions.

For the simplest example, a trimer, we can define the total two-body energy $E_2(\text{total})$ as:

$$E_2(\text{total}) = E_2(1, 2) + E_2(1, 3) + E_2(2, 3) \quad (39)$$

where $E_2(N, M)$ is the interaction energy for the NM dimer. The many-body interaction (i.e., nonpairwise) is defined as the energy that is not accounted for by the sum of these two-body interactions. The many-body interaction for the trimer is:

$$E_3 = E_{\text{total}} - E_2(\text{total}) \quad (40)$$

and in a general case of N oligomers it becomes:

$$E_{\text{mb}} = E_{\text{total}} - E_2(\text{total}) \quad (41)$$

Total pairwise energy, $E_2(\text{total})$, can be approximated by considering only the interactions between the nearest-neighbors (or next nearest-neighbors), as was done in refs 27 and 105. Such analysis allows one to quantify nonadditive component of the far-field part of the interaction energy.

The many-body energy in eq 41 consists only of intrinsic nonpairwise contributions. In EFP, the only nonpairwise term is polarization. One can characterize the importance of the many-body (or nonadditive) contribution by considering its percentage fraction to the interaction energy:

$$\Delta = \frac{E_{\text{mb}}}{E_{\text{total}}} \times 100\% \quad (42)$$

Using the EFP interaction energies for the adenine and thymine oligomers, the far-field (non-nearest-neighbor) part of the interaction energy is less than 5% of the total interaction energy. Taking into account the next nearest neighbors decreases this value to less than 1%. Additional details of the analysis are provided in Supporting Information.

5. Conclusion

We report the implementation of the EFP method in the Q-CHEM electronic structure program.⁸⁶ EFP,⁴⁸ an ab initio-based model potential, is an inexpensive alternative to a full ab initio description of noncovalent interactions. EFP can be described as a polarizable force field (or polarizable embedding model) that has no empirically fitted parameters. All parameters are obtained from ab initio calculations on isolated fragments.

We employed the EFP method to investigate noncovalent interaction energies in large oligomers of adenine and thymine representing a simple model of DNA. For the dimers, the EFP energies of π – π and H-bonding interactions are in good agreement with high-level ab initio calculations. The low computational cost of EFP allowed us to go beyond the dimers^{108,114,118,119} and to investigate asymptotic behavior of different components of interaction energies.

Our calculations demonstrate that the dispersion energy is the leading component in the π – π interaction for all oligomer sizes. The exchange-repulsion energy also plays an important role. The contribution of polarization is small in these systems, whereas the magnitude of the Coulomb term varies. The importance of various components obtained from our calculations can be compared to the results obtained by SAPT calculations^{2,116} where the dispersion and exchange-repulsion were found to be the major components in the parallel-slipped structures of benzene and substituted benzene dimers. The total interaction energies and the individual components are also in agreement with the results obtained by Leszczynski and co-workers.¹¹⁷

The rate of change of different energy components with the size of oligomers shows that exchange-repulsion has the shortest-range followed by the dispersion energy. The polarization and Coulomb are the longest-range interactions.

To assess whether a simple sum of the dimer energies can accurately represent the total interaction energy in extended systems, we analyzed the far-field part of the EFP energies. The polarization energy has a high nonadditive fraction. Other energy components have a significantly lower percentage of far-field contributions.

The accuracy and low computational cost of the EFP method are very encouraging. However, to take full advantage of EFP, analytic gradients are necessary, and this work is underway.

Acknowledgment. This work was supported by an NIH-SBIR grant with Q-CHEM, Inc. (A. I. K. and L. V. S.) and by grants from the Basic Energy Science Division of the Department of Energy to the Ames Laboratory Chemical Physics Program (M. S. G.). L.V.S. and J.M.H. acknowledge additional support from the NSF-CAREER grants (CHE-0955419 and CHE-0748448, respectively). We thank Dr. Yihan Shao and Dr. Jing Kong for their generous help with the implementation. We are also grateful to Prof. Alexander V. Nemukhin and Dr. Bella

Grigorenko for valuable discussions. In addition, we acknowledge contributions of Ilya Kaliman.

Supporting Information Available: Supporting Information is available free of charge via the Internet at <http://pubs.acs.org>. The Supporting Information contains details of the equations used in the fragment-fragment Coulomb interactions, Buckingham multipoles as well as notes about the effective fragment library of common solvents. The benchmark results for adenine and thymine dimers are presented. The effects of h-twist on the interaction energy in thymine dimer and far-field interactions in stacked adenine and thymine dimers are discussed. The geometries of the adenine and thymine monomers and oligomers are also given.

References and Notes

- Sherrill, C. D. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Cundari, T. A., Eds.; Jon Wiley & Sons: Hoboken, New Jersey, 2009; p 1.
- Tschumper, G. S. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Cundari, T. A., Eds.; Jon Wiley & Sons: Hoboken, New Jersey, 2009; p 39.
- Tauer, T. P.; Sherrill, C. D. *J. Phys. Chem. A* **2005**, *109*, 10475.
- Hunter, C. A.; Sanders, J. K. M. *J. Am. Chem. Soc.* **1990**, *112*, 5525.
- Privalov, P. L.; Makhatadze, G. I. *J. Mol. Biol.* **1992**, *224*, 715.
- Burley, S. K.; Petsko, G. A. *Science* **1985**, *229*, 23.
- Hunter, C. A.; Singh, J.; Thornton, J. M. *J. Mol. Biol.* **1991**, *218*, 837.
- Morgan, R. S.; Tatsh, C. E.; Gushard, R. H.; Moadon, J. M.; Warme, P. K. *Int. J. Pept. Prot. Res.* **1978**, *11*, 209.
- Zauhar, R. J.; Colbert, C. L.; Morgan, R. S.; Welsh, W. J. *Biopolymer* **2000**, *53*, 233.
- Brana, M. F.; Cacho, M.; Gradillas, A.; Pascual-Teresa, B.; Ramos, A. *Curr. Pharm. Des.* **2001**, *7*, 1745.
- Wintjens, R.; Lievin, J.; Rooman, M.; Buisine, E. *J. Mol. Biol.* **2000**, *302*, 395.
- Anbarasu, A.; Anand, S.; Babu, M. M.; Sethumadhavan, R. *Macromolecules* **2007**, *41*, 251.
- Claessens, C. G.; Stoddart, J. F. *J. Phys. Org. Chem.* **1997**, *10*, 254.
- Gazit, E. *Chem. Soc. Rev.* **2007**, *36*, 1263.
- Meyer, E. A.; Castellano, R. K.; Diederich, F. *Angew. Chem., Int. Ed.* **2003**, *42*, 1210.
- Askew, B.; Ballester, P.; Buhr, C.; Jeong, K. S.; Jones, S.; Parris, K.; Williams, K.; Rebek, J., Jr. *J. Am. Chem. Soc.* **1989**, *111*, 1082.
- Hunter, C. A. *Chem. Soc. Rev.* **1994**, *23*, 101.
- Hud, N.; Anet, F. A. L. *J. Theor. Biol.* **2000**, *205*, 543.
- Řeha, D.; Kabeláč, M.; Ryjáček, F.; Šponer, J.; Šponer, J. E.; Elstner, M.; Suhai, S.; Hobza, P. *J. Am. Chem. Soc.* **2002**, *124*, 3366.
- Ferguson, L. R.; Denny, W. A. *Mutat. Res.* **2007**, *623*, 14.
- Rezac, J.; Hobza, P.; Harris, S. A. *Biophys. J.* **2010**, *98*, 101.
- Gessner, O.; Lee, A. M. D.; Shaffer, J. P.; Reisler, H.; Levchenko, S.; Krylov, A.; Underwood, J. G.; Shi, H.; East, A. L. L.; Wardlaw, D. M.; Chrysostom, E. T.-H.; Hayden, C. C.; Stollow, A. *Science* **2006**, *311*, 219.
- Pieniazek, P. A.; Bradforth, S. E.; Krylov, A. I. *J. Phys. Chem. A* **2006**, *110*, 4854.
- Cristian, A. M. C.; Krylov, A. I. *J. Chem. Phys.* **2003**, *118*, 10912.
- Podeszwa, R.; Szalewicz, K. *J. Chem. Phys.* **2007**, *126*, 194101.
- Podeszwa, R.; Bukowski, R.; Szalewicz, K. *J. Phys. Chem. A* **2006**, *110*, 10345.
- Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2006**, *110*, 10656.
- Tsuzuki, S.; Uchimarui, T.; Matsumura, K.; Mikami, M.; Tanabe, K. *Chem. Phys. Lett.* **2000**, *319*, 547.
- Ye, X.; Li, Z.; Wang, W.; Fan, K.; Xu, W.; Hua, Z. *Chem. Phys. Lett.* **2004**, *397*, 56.
- Bowler, D. R.; Miyazaki, T.; Gillan, M. J. *J. Phys.: Condens. Matter* **2002**, *14*, 2781.
- Schütz, M.; Hetzer, G.; Werner, H. *J. Chem. Phys.* **1999**, *111*, 5691.
- Scuseria, G. E.; Ayala, P. Y. *J. Chem. Phys.* **1999**, *111*, 8330.
- White, C. A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M. *Chem. Phys. Lett.* **1996**, *253*, 268.
- Deev, V.; Collins, M. A. *J. Chem. Phys.* **2005**, *122*, 154102.
- Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, *313*, 701.
- Fedorov, D. G.; Kitaura, K. *J. Phys. Chem. A* **2007**, *111*, 6904.
- Fedorov, D. G.; Kitaura, K. *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*; CRC Press: Boca Raton, 2009.
- Gordon, M. S.; Mullin, J. M.; Pruitt, S. R.; Roskop, L. B.; Slipchenko, L. V.; Boatz, J. A. *J. Phys. Chem. B* **2009**, *113*, 9646.
- Sanz-Garcia, A.; Wheatley, R. J. *ChemPhysChem* **2003**, *5*, 801.
- Korona, T. *J. Chem. Phys.* **2008**, *128*, 224104.
- Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463.
- Chai, J.-D.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615.
- Allunger, N. L.; Yuh, Y. H.; Lii, J. H. *J. Am. Chem. Soc.* **1989**, *111*, 8551.
- Mayo, S. L.; Olafson, B. D.; Goddard, W. A. *J. Phys. Chem.* **1990**, *94*, 8897.
- Casewit, C. J.; Colwell, K. S.; Rappe, A. K. *J. Am. Chem. Soc.* **1992**, *114*, 10035.
- MacKerell, A. D.; Bashford, D.; Bellott, Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.
- Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. *J. Chem. Phys.* **1996**, *105*, 1968.
- Gordon, M. S.; Slipchenko, L.; Li, H.; Jensen, J. H. In *Annual Reports in Computational Chemistry*; Spellmeyer, D. C., Wheeler, R., Eds.; Annual Reports in Computational Chemistry; Elsevier: Oxford, U.K., 2007; Vol. 3, p 177.
- Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. *J. Phys. Chem. A* **2001**, *105*, 293.
- Schmidt, M. W.; Baldrige, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Mastunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347.
- The term EFP2 is sometimes used to refer to the general EFP method discussed here.
- Adamovic, I.; Freitag, M. A.; Gordon, M. S. *J. Chem. Phys.* **2003**, *118*, 6725.
- Gordon, M. S.; Schmidt, M. W. In *Theory and Applications of Computational Chemistry*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E., Eds.; Elsevier: Amsterdam, Netherlands, 2005; Ch. 41; p 1167.
- Arora, P.; Slipchenko, L. V.; Webb, S. P.; Defusco, A.; Gordon, M. S. *J. Phys. Chem. A* **2010**, *114*, 6742.
- Yoo, S.; Zahariev, F.; Sok, S.; Gordon, M. S. *J. Chem. Phys.* **2008**, *129*, 8.
- DeFusco III, A. A.; Gordon, M. S., unpublished.
- Chen, W.; Gordon, M. S. *J. Chem. Phys.* **1996**, *105*, 11081.
- Webb, S. P.; Gordon, M. S. *J. Phys. Chem. A* **1999**, *103*, 1265.
- Merrill, G. N.; Gordon, M. S. *J. Phys. Chem. A* **1998**, *102*, 2650.
- Day, P. N.; Pachter, R.; Gordon, M. S.; Merrill, G. N. *J. Chem. Phys.* **2000**, *112*, 2063.
- Adamovic, I.; Gordon, M. S. *J. Phys. Chem. A* **2005**, *109*, 1629.
- Bandyopadhyay, P.; Gordon, M. S. *J. Chem. Phys.* **2000**, *113*, 1104.
- Bandyopadhyay, P.; Gordon, M. S.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **2002**, *116*, 5023.
- Kina, D.; Nakayama, A.; Noro, T.; Taketsugu, T.; Gordon, M. S. *J. Phys. Chem. A* **2008**, *112*, 9675.
- Jensen, J. H. *J. Chem. Phys.* **1996**, *104*, 7795.
- Slipchenko, L. V.; Gordon, M. S. *J. Comput. Chem.* **2007**, *28*, 276.
- Sinnokrot, M. O.; Sherrill, C. D. *J. Am. Chem. Soc.* **2004**, *126*, 7690.
- Podeszwa, R.; Bukowski, R.; Szalewicz, K. *J. Phys. Chem. A* **2006**, *110*, 10345.
- Slipchenko, L. V.; Gordon, M. S. *Mol. Phys.* **2009**, *107*, 999.
- Smith, T.; Slipchenko, L. V.; Gordon, M. S. *J. Phys. Chem. A* **2008**, *112*, 5286.
- Adamovic, I.; Li, H.; Lamm, M. H.; Gordon, M. S. *J. Phys. Chem. A* **2006**, *110*, 519.
- Slipchenko, L. V.; Gordon, M. S. *J. Phys. Chem. A* **2009**, *113*, 2092.
- Adamovic, I.; Gordon, M. S. *J. Phys. Chem. A* **2006**, *110*, 10267.
- Mullin, J. M.; Gordon, M. S. *J. Phys. Chem. B* **2009**, *113*, 8657.
- Mullin, J. M.; Gordon, M. S. *J. Phys. Chem. B* **2009**, *113*, 14413.
- Mullin, J. M. Slicing, Dicing, Washing, Evolving. All in the Name of Capturing an Electron (or Two). Ph.D. Thesis, Iowa State University, 2009.
- Pranami, G.; Slipchenko, L. V.; Lamm, N. H.; Gordon, M. S. In *Multi-scale Quantum Models for Biocatalysis*; Lee, T.-S., York, D. M., Eds.;

Challenges and Advances in Computational Chemistry and Physics; Springer: Netherlands, 2009.

- (78) Kairys, V.; Jensen, J. H. *J. Phys. Chem. A* **2000**, *104*, 6656.
- (79) Li, H.; Hains, A. W.; Everts, J. E.; Robertson, A. D.; Jensen, J. H. *J. Phys. Chem. B* **2002**, *106*, 3486.
- (80) Jensen, J. H.; Li, H.; Robertson, A. D.; Molina, A. *J. Phys. Chem. A* **2005**, *109*, 6634.
- (81) Slipchenko, L. V. *J. Phys. Chem. A* **2010**, *114*, 8828.
- (82) Polyakov, I.; Epifanovsky, E.; Grigorenko, B. L.; Krylov, A. I.; Nemukhin, A. V. *J. Chem. Theory Comput.* **2009**, *5*, 1907.
- (83) Nemukhin, A. V.; Grigorenko, B. L.; Rogov, A. V.; Topol, I. A.; Burt, S. K. *Theor. Chem. Acc.* **2004**, *111*, 36.
- (84) Grigorenko, B. L.; Nemukhin, A. V.; Topol, I. A.; Burt, S. K. *J. Phys. Chem. A* **2002**, *106*, 10663.
- (85) Nemukhin, A. V.; Grigorenko, B. L.; Bochenkova, A. V.; Topol, I. A.; Burt, S. K. *J. Molec. Struct. (Theochem)* **2002**, *581*, 167.
- (86) Shao, Y.; Molnar, L. F.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, S.; Gilbert, A. T. B.; Slipchenko, L. V.; Levchenko, S. V.; O'Neil, D. P.; Distasio, R. A., Jr; Lochan, R. C.; Wang, T.; Beran, G. J. O.; Besley, N. A.; Herbert, J. M.; Lin, C. Y.; Van Voorhis, T.; Chien, S. H.; Sodt, A.; Steele, R. P.; Rassolov, V. A.; Maslen, P.; Korambath, P. P.; Adamson, R. D.; Austin, B.; Baker, J.; Bird, E. F. C.; Daschel, H.; Doerksen, R. J.; Drew, A.; Dunietz, B. D.; Dutoi, A. D.; Furlani, T. R.; Gwaltney, S. R.; Heyden, A.; Hirata, S.; Hsu, C.-P.; Kedziora, G. S.; Khalliulin, R. Z.; Klunziger, P.; Lee, A. M.; Liang, W. Z.; Lotan, I.; Nair, N.; Peters, B.; Proynov, E. I.; Pieniazek, P. A.; Rhee, Y. M.; Ritchie, J.; Rosta, E.; Sherrill, C. D.; Simmonett, A. C.; Subotnik, J. E.; Woodcock, H. L., III; Zhang, W.; Bell, A. T.; Chakraborty, A. K.; Chipman, D. M.; Keil, F. J.; Warshel, A.; Herberich, W. J.; Schaefer, H. F., III; Kong, J.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172.
- (87) Smith, T.; Slipchenko, L. V.; Gordon, M. S., unpublished.
- (88) Stone, A. J. In *The Theory of Intermolecular Forces*; Oxford University Press: New York, 1997; p 50.
- (89) Li, H.; Gordon, M. S.; Jensen, J. H. *J. Chem. Phys.* **2006**, *124*.
- (90) Kemp, D. D.; Rintelman, J. M.; Gordon, M. S.; Jensen, J. H. *Theor. Chim. Acta* **2010**, *125*, 481.
- (91) E.g., the "makefp" run in the GAMESS suite of programs performs this task.
- (92) Jensen, J. H.; Gordon, M. S. *Mol. Phys.* **1996**, *89*, 1313.
- (93) Jensen, J. H.; Gordon, M. S. *J. Chem. Phys.* **1998**, *108*, 4772.
- (94) Buckingham, A. D. *Q. Rev. Chem. Soc.* **1959**, *13*, 183.
- (95) Stone, A. J. *Chem. Phys. Lett.* **1981**, *83*, 233.
- (96) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.
- (97) Stone, A. J.; Alderton, M. *Mol. Phys.* **1985**, *56*, 1047.
- (98) Obara, S.; Saika, A. *J. Chem. Phys.* **1986**, *84*, 3963.
- (99) Freitag, M. A.; Gordon, M. S.; Jensen, J. H.; Stevens, W. J. *J. Chem. Phys.* **2000**, *112*, 7300.
- (100) Magnasco, V.; Costa, C.; Figari, G. *J. Molec. Struct. (Theochem)* **1990**, *206*, 235.
- (101) Adamovic, I.; Gordon, M. S. *Mol. Phys.* **2005**, *103*, 379.
- (102) Tang, K. T.; Toennies, J. P. *J. Chem. Phys.* **1984**, *80*, 3726.
- (103) Boys, S. F. *Rev. Mod. Phys.* **1960**, *32*, 296.
- (104) Boys, S. F. In *Quantum Theory of Atoms, Molecules, and the Solid State*; Academic Press: New York, 1966; p 253.
- (105) Hopkins, B. W.; Tschumper, G. S. *Chem. Phys. Lett.* **2005**, *407*, 362.
- (106) Sponer, J.; Leszczynski, J.; Hobza, P. *Biopolymers* **2002**, *61*, 3.
- (107) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 5590.
- (108) Jurečka, P.; Hobza, P. *J. Am. Chem. Soc.* **2003**, *125*, 15608.
- (109) Bravaya, K. B.; Kostko, O.; Ahmed, M.; Krylov, A. I. *Phys. Chem. Chem. Phys.* **2010**, *12*, 2292.
- (110) Černý, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2007**, *9*, 5291.
- (111) Rappé, A. K.; Bernstein, E. R. *J. Phys. Chem. A* **2000**, *104*, 6117.
- (112) Müller-Dethlefs, K.; Hobza, P. *Chem. Rev.* **2000**, *100*, 143.
- (113) Engkvist, O.; Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Chem. Phys.* **1999**, *110*, 5758.
- (114) Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2004**, *108*, 10200.
- (115) Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2006**, *110*, 10656.
- (116) Bates, D. M.; Anderson, J. A.; Oloyede, P.; Tschumper, G. S. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2775.
- (117) Hill, G.; Forde, G.; Hill, N.; Lester, W. A., Jr; Sokalski, W. A.; Leszczynski, J. *Chem. Phys. Lett.* **2003**, *381*, 729.
- (118) Hobza, P.; Zahradnik, R.; Müller-Dethlefs, K. *Collect. Czech. Chem. Commun.* **2006**, *71*, 443.
- (119) Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Phys. Chem.* **1996**, *100*, 18790.

JP107557P