Multipole Moments in the Effective Fragment Potential Method

Journal:	*The Journal of Physical Chemistry*
Manuscript ID	jp-2017-006828.R1
Manuscript Type:	Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Bertoni, Colleen; Iowa State University, Chemistry Slipchenko, Lyudmila; Purdue University, Chemistry Misquitta, Alston; University of Cambridge, Chemistry Gordon, Mark; Iowa State University, Chemistry

SCHOLARONE™
Manuscripts

Multipole moments in the Effective Fragment Potential method

Colleen Bertoni[1], Lyudmila V. Slipchenko[3], Alston J. Misquitta[2], Mark S. Gordon[1]

[1]Department of Chemistry, Iowa State University, and Ames Laboratory, Ames, IA

50011 USA

[2]School of Physics and Astronomy, Queen Mary, University of London, 327 Mile End

Road, London, E1 4NS England

[3]Department of Chemistry, Purdue University, West Lafayette, IN 47907 USA

Abstract

In the effective fragment potential (EFP) method the Coulomb potential is

represented using a set of multipole moments generated by the distributed

multipole analysis (DMA) method. Misquitta, Stone, and Fazeli recently developed a

basis space-iterated stockholder atom (BS-ISA) method to generate multipole

moments. This study assesses the accuracy of the EFP interaction energies using

sets of multipole moments generated from the BS-ISA method, and from several

versions of the DMA method (such as analytic and numeric grid-based), with

varying basis sets. Both methods lead to reasonable results, although using certain

implementations of the DMA method can result in large errors. With respect to the

CCSD(T)/CBS interaction energies, the mean unsigned error (MUE) of the EFP

method for the S22 data set using BS-ISA–generated multipole moments and DMA-

generated multipole moments (using a small basis set and the analytic DMA

procedure) is 0.78 and 0.72 kcal/mol, respectively. The MUE accuracy is on the

same order as MP2 and SCS-MP2. The MUEs are lower than in a previous study

benchmarking the EFP method without the EFP charge transfer term, demonstrating that the charge transfer term increases the accuracy of the EFP method. Regardless of the multipole moment method used, it is likely that much of the error is due to an insufficient short-range electrostatic term (i.e. charge penetration term), as shown by comparisons with symmetry-adapted perturbation theory.

1. Introduction

A main goal of quantum chemistry is to perform fast and accurate calculations on challenging systems, such as solvated proteins or reactions occurring in solution, and to provide insight into the interactions between molecules. Although there are methods that give highly accurate results for small molecules, it is difficult to extend these methods to larger species and still retain their accuracy. Thus, there has been considerable effort to develop more computationally efficient methods. In particular, interaction energy methods have had success in describing non-covalent interactions of large systems in a computationally efficient manner. Interaction energy methods have their roots in the splitting of a system into non-interacting fragments (usually molecules), and then using perturbation theory to calculate the interaction energy between the fragments. For long-range interactions, like Coulomb, polarization, and dispersion, the perturbation between the fragments is the Coulomb operator. At first-order in perturbation theory, the electrostatics and exchange-repulsion energies are obtained, while the polarization, charge-transfer, and dispersion energies are each part of the second-order energy. The Coulomb field is typically used in calculating

the Coulomb energy, and can also be used in other terms, like the polarization term. Since the Coulomb field can be used in multiple terms, it is essential to represent it accurately and in a computationally inexpensive manner.

To represent the Coulomb field, many interaction energy methods use a multipole moment expansion, which arises from a Taylor expansion of the classical Coulomb energy expression. However, using a multipole moment expansion in which each fragment has a single monopole, dipole, quadrupole, etc., has poor convergence properties. That is, if the fragments are too close together, the expansion no longer converges. Additionally, if the fragments are close, there is an attractive charge penetration energy that is not accounted for in the multipole moment expansion. To solve the convergence problem, a distributed multipole moment expansion can be used, where there is a monopole, dipole, quadrupole, etc., for a chosen number of sites distributed throughout each fragment. Then, the question is how to calculate the distributed multipole moments themselves. Calculating the distributed multipole moments typically depends on partitioning the molecular charge density among atom centers, bond midpoints, or other sites in the fragment. There has been much work on how to assign electronic charge densities to atoms. Several examples are: Mulliken charges[1], Stone's method for distributed multipole analysis (DMA) [2], the atoms-in-molecules method by Bader[3], the Hirschfeld stockholder method[4], the iterated Hirschfeld method[5], the atoms-in-molecules method by Popelier[6], the iterated Stockholder atom method by Lillestonen and Wheately[7], and the recently developed basis-space Iterated Stockholder Atoms [8] (BS-ISA) method by Misquitta, Stone, and Fazeli.

The Effective Fragment Potential (EFP) method is an interaction energy method that has been extensively developed. [9] [10] Several terms in the EFP method (Coulomb energy, polarization energy, charge transfer energy) use a set of multipole moments to represent the Coulomb field. Thus, an accurate set of multipole moments is important to ensure that the total interaction energy is accurate. Currently, the multipole moments are calculated with the Stone DMA. As discussed later, the DMA method can be unstable depending on the basis set, although a numerical version has been developed to overcome this problem. [11] The BS-ISA method has been shown to be accurate and to have promising convergence properties, so it is worthwhile to explore how the EFP method performs if the multipole moments generated by the BS-ISA method are used.

In this work, the EFP energy with BS-ISA-generated multipole moments (referred to here as EFP/ISA) and the EFP energy with DMA-generated multipole moments (referred to here as EFP/DMA) are compared. The structure of this paper is: Section 2 discusses the theory behind EFP, DMA, and BS-ISA; Section 3 discusses the computational details used in the comparisons; and Section 4 discusses the comparison and results.

2. Theoretical Background

This section summarizes the EFP method, with a particular emphasis on the terms that use multipole moments, and background on the DMA and the BS-ISA methods.

2.1 The Effective Fragment Potential method

The EFP method calculates the intermolecular interaction energy between molecules. In the EFP method, molecules are modeled with potentials that have functional forms derived from first principles, and parameters that are generated from an *ab initio* calculation.

There are five terms in the Effective Fragment Potential method: Coulomb, polarization, exchange repulsion, dispersion, and charge transfer. As shown in the equation below, polarization is a many-body term, while the other four terms are pairwise additive.

$$E_{AB}^{EFP} = E_{AB}^{Coul} + E_{AB}^{exchange\text{-}repulsion} + E_{AB}^{dispersion} + E_{AB}^{charge\text{-}transfer}$$
$$E_{total}^{EFP} = \sum_{A>B} E_{AB}^{EFP} + E_{total}^{polarization}$$
(1)

Each of the five terms depends on parameters generated from an *ab initio* calculation. The Coulomb, polarization, and charge transfer energy terms depend on a set of multipole moments to describe the electrostatic potential of the molecule. The polarization energy also depends on a set of distributed polarizability tensors generated from the Coupled Perturbed Hartree-Fock (CPHF) equation, which are distributed throughout the molecule using a set of localized molecular orbitals (LMOs). In addition to the multipole moments, the charge transfer energy depends on the basis set, the Fock matrix and a set of canonical virtual orbitals or valence virtual orbitals (VVOs).[12] [13] The exchange-repulsion energy depends on the set of LMOs, the basis set, and the Fock matrix. The dispersion energy depends on a set of distributed dynamic polarizability tensors generated from the dynamic analog of the CPHF equation and distributed throughout the molecule using a set of LMOs.

An EFP energy calculation requires two steps. The first is an *ab initio* calculation on an isolated molecule performed to generate the parameters for the molecule of interest. Then, the generated parameters are used in the EFP energy terms.

The next three sections consider the three EFP terms that depend explicitly on the set of multipole moments (Coulomb, polarization, and charge transfer).

2.1.1 Coulomb energy term

The Coulomb interaction energy term between two molecules A and B can be calculated by a distributed multipole moment expansion, as shown below. [35]

$$E_{AB}^{Coul} = \sum_{I}^{A}\sum_{J}^{B}\left[\begin{array}{l} q^J q^I T^{IJ} - \sum_{\alpha}^{x,y,z} q^J \mu_{\alpha}^I T_{\alpha}^{IJ} + \dfrac{1}{3}\sum_{\alpha,\beta}^{x,y,z} q^J \Theta_{\alpha\beta}^I T_{\alpha\beta}^{IJ} - \dfrac{1}{15}\sum_{\alpha,\beta,\gamma}^{x,y,z} q^J \Omega_{\alpha\beta\gamma}^I T_{\alpha\beta\gamma}^{IJ} \\[2ex] + \sum_{\alpha}^{x,y,z} \mu_{\alpha}^J q^I T_{\alpha}^{IJ} - \sum_{\alpha,\beta}^{x,y,z} \mu_{\alpha}^J \mu_{\beta}^I T_{\alpha\beta}^{IJ} + \dfrac{1}{3}\sum_{\alpha,\beta,\gamma}^{x,y,z} \mu_{\alpha}^J \Theta_{\beta\gamma}^I T_{\alpha\beta\gamma}^{IJ} \\[2ex] + \dfrac{1}{3}\sum_{\alpha,\beta}^{x,y,z} \Theta_{\alpha\beta}^J q^I T_{\alpha\beta}^{IJ} - \dfrac{1}{3}\sum_{\alpha,\beta,\gamma}^{x,y,z} \Theta_{\alpha\beta}^J \mu_{\gamma}^I T_{\alpha\beta\gamma}^{IJ} + \dfrac{1}{9}\sum_{\alpha,\beta,\gamma,\delta}^{x,y,z} \Theta_{\alpha\beta}^J \Theta_{\gamma\delta}^I T_{\alpha\beta\gamma\delta}^{IJ} \\[2ex] + \dfrac{1}{15}\sum_{\alpha,\beta,\gamma}^{x,y,z} \Omega_{\alpha\beta\gamma}^J q^I T_{\alpha\beta\gamma}^{IJ} \end{array}\right] \tag{2}$$

I (*J*) runs over all multipole moment expansion points in *A* (*B*), q^I is the monopole on site *I*, μ^I is the dipole on site *I*, Θ^I is the quadrupole on site *I*, Ω^I is the octopole on site *I* and $T_{\alpha\beta...\nu}^{IJ} = \nabla_{\alpha}\nabla_{\beta}...\nabla_{\nu}\dfrac{1}{R_{IJ}}$ is the multipole interaction tensor for sites *I* and *J*. R_{IJ} is the distance between sites *I* and *J*, where **R_{IJ}=R_J-R_I** in vector notation.

Charge penetration for the Coulomb energy term

Since the multipole moment expansion does not take into account the energy

stabilization that occurs when the charge densities of fragments overlap, a charge

penetration term or a damping term is added. There are two types of damping for

the Coulomb energy in the EFP method.[14] One is an exponential damping term,

which is not used here, and is not considered further. The second is based on the

overlap of LMOs on the two fragments, and is used to calculate an approximation to

the charge-penetration energy.[15] The charge-penetration energy for fragments A

and B is calculated as shown below.

$$E_{AB}^{chgpen} = \sum_{l}^{LMO \in A} \sum_{m}^{LMO \in B} \frac{-2S_{lm}^2}{R_{lm}} \sqrt{\left(\frac{1}{-2\ln|S_{lm}|} \right)} \qquad (3)$$

where

S_{lm} is the overlap integral between l and m

R_{lm} is the distance between the LMO centroid of l ($\langle l|x|l \rangle$ for the x-position) and the

LMO centroid of m ($\langle m|x|m \rangle$ for the x-position).

2.1.2 Polarization energy term

The polarization energy is a many-body energy term that is due to the

generation of induced dipoles on all of the fragments in the total electric field (static

and induced fields) of all the other fragments. Dipole polarizability tensors are

distributed onto the LMO centroids of the fragments. Then, in the presence of the

static and induced electric field on the other fragments, the dipole polarizability

tensors generate induced dipoles. The induced dipoles are iterated self-consistently,

and then used in the calculation of the polarization energy.

The induced dipole on fragment A on LMO centroid l in the β direction can

be written as:[35]

$$p_{l,\beta}^A = \sum_{\gamma}^{\{x,y,z\}} \alpha_{l,\beta\gamma} \left(E_{l,\gamma}^{0,A} + \sum_{B \neq A}^{\text{fragments}} \sum_{m}^{LMO \in B} \sum_{\kappa}^{\{x,y,z\}} T_{\gamma\kappa}^{lm} p_{m,\kappa}^B \right) \tag{4}$$

where

$T_{\gamma\kappa}^{lm}$ is the dipole-dipole interaction tensor for sites l and m

$\alpha_{l,\beta\gamma}$ is the dipole polarizability tensor on LMO l

$E_{l,\gamma}^{0,A}$ is the static electric field on fragment A on LMO centroid l in the γ direction

The static electric field, shown below, is expressed using the same distributed

multipole moments as in the Coulomb energy term.

$$E_{l,\gamma}^{0,A} = \sum_{B \neq A}^{\text{fragments}} \sum_{I}^{B} E_{lI,\gamma}^{0} = \sum_{B \neq A}^{\text{fragments}} \sum_{I}^{B} \left(q^I T_{\gamma}^{lI} + \sum_{\alpha}^{\{x,y,z\}} \mu_{\alpha}^I T_{\gamma\alpha}^{lI} + \frac{1}{3} \sum_{\alpha\beta}^{\{x,y,z\}} \Theta_{\alpha\beta}^I T_{\gamma\alpha\beta}^{lI} \right) \tag{5}$$

where I runs over the multipole moment expansion points in fragment B.

The polarization interaction energy term is shown below:

$$E^{pol} = \sum_{A}^{\text{fragments}} \left[-\frac{1}{2} \sum_{n}^{LMO \in A} \sum_{\alpha}^{\{x,y,z\}} E_{n,\alpha}^{0,A} p_{n,\alpha}^A \right] \tag{6}$$

Damping for the polarization energy term

The polarization energy is damped by a Tang-Toennies style Gaussian

formula[14][16]. The damping is accomplished by multiplying the multipole interaction

tensors by a damping function, and then rewriting the induced dipoles in terms of

the damped multipole interaction tensors. The damping function is

$F_{damp,ll}^{pol} = 1 - \exp\left(-R_{ll}^2\sqrt{fg}\right)\left(1 + R_{ll}^2\sqrt{fg}\right)$, where the terms f and g are constants often

set to 0.6. The damped polarization energy equations are similar to the non-damped

version but with damped multipole moment interaction tensors replacing regular

multipole moment interaction tensors. Defining $T_{\alpha\beta...\nu}^{ll,damped} \equiv F_{damp,ll}^{pol} T_{\alpha\beta...\nu}^{ll}$, the damped

static electric field can be written as:

$$E_{l,\gamma}^{0,A,damped} = \sum_{B\neq A}^{\text{fragments}} \sum_{I}^{B} E_{ll,\gamma}^{0,damped} = \sum_{B\neq A}^{\text{fragments}} \sum_{I}^{B} \left(q^I T_\gamma^{ll,damped} + \sum_\alpha^{\{x,y,z\}} \mu_\alpha^I T_{\gamma\alpha}^{ll,damped} + \frac{1}{3}\sum_{\alpha\beta}^{\{x,y,z\}} \Theta_{\alpha\beta}^I T_{\gamma\alpha\beta}^{ll,damped} \right)$$

$$(7)$$

Following the same substitution, the damped induced dipoles, $p_{l,\beta}^{A,damped}$, can then be

written as: $p_{l,\beta}^{A,damped} = \sum_\gamma^{\{x,y,z\}} \alpha_{l,\beta\gamma} \left(E_{l,\gamma}^{0,A,damped} + \sum_{B\neq A}^{\text{fragments}} \sum_m^{LMO\in B} \sum_\kappa^{\{x,y,z\}} T_{\gamma\kappa}^{lm,damped} p_{m,\kappa}^{B,damped} \right)$

2.1.3 Charge transfer energy term

The charge transfer energy can be thought of as a stabilizing energy due to

the interaction of occupied orbitals and virtual orbitals on two separate molecules.[13]

[17] The EFP charge transfer term was derived using a second-order perturbative

approach beginning with an antisymmetrized wavefunction at the Hartee-Fock level

of theory.

In the derivation, approximations are used to simplify the second-order

energy expression. One approximation is to represent the electrostatic potential as a

multipole moment expansion, using the same multipole moments as in the Coulomb

and polarization terms. The EFP charge transfer energy of molecule A induced by

molecule B is:

$$CT^{A(B)} = 2 \sum_{i}^{\substack{occ \\ CMO \in A}} \sum_{n}^{vir \in B} \frac{1}{1 - \sum_{m}^{\substack{all \\ MOs \in A}} S_{nm}^2} \frac{V_{in}^{EFPB} - \sum_{m}^{\substack{all \\ MOs \in A}} S_{nm} V_{im}^{EFPB}}{\left(F_{ii}^A - T_{nn}\right)} \times \left[V_{in}^{EFPB} - \sum_{m}^{\substack{all \\ MOs \in A}} S_{nm} V_{im}^{EFPB} + \sum_{j}^{\substack{occ \\ CMO \in B}} S_{ij} \left(T_{nj} - \sum_{m}^{\substack{all \\ MOs \in A}} S_{nm} T_{mj} \right) \right]$$

(8)

where T_{nj} is the kinetic energy integral between orbitals n and j, F_{ii}^A is the diagonal

Fock matrix element at orbital i in the canonical MO basis for fragment A, and V_{in}^{EFPB}

is the matrix element of the molecular electrostatic potential of fragment B between

orbitals i and n. This can written as:

$$\begin{aligned} V_{in}^{EFPB} &= \sum_{I}^{B} \langle i | \left(q^I \hat{T}^I - \sum_{\beta}^{\{x,y,z\}} \mu_{\beta}^I \hat{T}_{\beta}^I + \frac{1}{3} \sum_{\beta\gamma}^{\{x,y,z\}} \Theta_{\beta\gamma}^I \hat{T}_{\beta\gamma}^I \right) | n \rangle \\ &= \sum_{I}^{B} \int d\mathbf{r}_1 \chi_i^*(\mathbf{r}_1) \left(q^I T^{Ir_1} - \sum_{\beta}^{\{x,y,z\}} \mu_{\beta}^I T_{\beta}^{Ir_1} + \frac{1}{3} \sum_{\beta\gamma}^{\{x,y,z\}} \Theta_{\beta\gamma}^I T_{\beta\gamma}^{Ir_1} \right) \chi_n(\mathbf{r}_1) \end{aligned}$$

(9)

where I runs over the multipole moment expansion points in fragment B, $\mathbf{r}_1(= x_1, y_1,$

$z_1)$ is the position of the electron, $\chi_i(\mathbf{r}_1)$ is molecular orbital i. The right hand side of

Eq. (9) is evaluated in a similar manner to the standard nuclear attraction integral.

While there is not unanimous agreement regarding the importance of the

charge transfer interaction energy [10] [18], the EFP method predicts relatively large

charge transfer contributions for polar and ionic complexes, and systems with

hydrogen bonds[13]. [19]

2.2 Multipole moment methods

2.2.1 Distributed Multipole Analysis

In the DMA method, the molecular charge density is partitioned, and each piece of charge density is characterized by a multipole moment expansion. The partitioning can be done in basis function space or real space. Basis function space DMA is denoted here as DMA0 or analytical DMA.

For restricted Hartee-Fock (RHF), the molecular charge density can be written in terms of primitive Gaussians:

$$\rho(\mathbf{r}) = \sum_{\mu,\nu}^{AOs} P_{\mu\nu} \chi_\mu(\mathbf{r}-\mathbf{R}_I) \chi_\nu(\mathbf{r}-\mathbf{R}_J)$$

$$= \sum_{\mu,\nu}^{AOs} P_{\mu\nu} \sum_u^{PG \in \nu} \sum_t^{PG \in \mu} P'_{ut} \phi_u(\mathbf{r}-\mathbf{R}_I, \alpha_u) \phi_t(\mathbf{r}-\mathbf{R}_J, \alpha_t) \qquad (10)$$

$$= \sum_{\mu,\nu}^{AOs} \sum_u^{PG \in \nu} \sum_t^{PG \in \mu} \left[P_{\mu\nu} P'_{ut} \phi_{ut}(\mathbf{r}-\mathbf{R}_K, (\alpha_u+\alpha_t)) \right]$$

where $\chi_\mu(\mathbf{r}-\mathbf{R}_I)$ is a basis function composed of a sum of primitive Gaussians (PGs) centered on atom I, $\phi_u(\mathbf{r}-\mathbf{R}_I, \alpha_u)$ is a PG centered on atom I with contraction exponent α_u, P'_{ut} is the product of contraction coefficients for PGs u and t, $P_{\mu\nu}$ is the RHF density matrix element for AOs μ and ν, and $\mathbf{R}_K = \dfrac{\alpha_u \mathbf{R}_I + \alpha_t \mathbf{R}_J}{\alpha_u + \alpha_t}$.

As shown in the last equality in Eq. (10), the total electronic charge density is a sum over pieces of charge density (the term in the brackets) centered at the Gaussian overlap point R_k. A certain number of expansion sites are chosen in the molecule, such as all atom centers or all atom centers and bond midpoints. As discussed in Ref. 2, sets of multipole moment integrals can describe PG products

(that is, pieces of charge density). To only use multipole moment integrals centered at the chosen expansion points, the origins of multipole moment integrals for each piece of charge density are moved to the nearest chosen expansion site.

It is well known that the DMA0 multipole moments are unstable with respect to expanding the basis set[11]. Although the multipole moments from different basis sets should produce similar electrostatic potentials sufficiently far from the expansion centers, the values for the multipole moments themselves can be basis set dependent. Consequently, the appropriate termination of the multipole expansion (e.g., at quadrupoles or octopoles) may depend on the basis set used. Thus, even though the electrostatic potential at sufficient distance from the expansion centers should be the same, the error due to the multipole truncation can vary. This is especially the case for basis sets with diffuse functions or with high angular momenta, since these functions tend to make a larger contribution to the higher moments and therefore the truncation point is important.[20, 2] Because of the instability with respect to basis set size, real space DMA was developed.

Real space DMA involves modifying the DMA0 algorithm such that if the exponent of a product of primitives (e.g., $\alpha_u + \alpha_t$) is smaller than a chosen cutoff, a grid-based numerical integration scheme is used to partition the contribution to the multipole moments. If the exponent is greater than the cutoff, DMA0 is used to partition the contribution to the multipole moments. Ref. 11 recommends a cutoff value of 4, and so the method is referred to here as DMA4.

It is also important to note that when the molecules are too close to each other, the multipole moment expansion of the Coulomb energy can diverge. How

close the molecules can get to each other before the expansion diverges depends on

how the charge density is partitioned, and on the expansion points chosen. The

greater the number of expansion points, the more accurately the multipole

expansion mimics the correct quantum density. So, the fewer the number of

expansions points used, the more likely it is that the expansion will diverge at a

given distance.

2.2.2 BS-ISA

In the implementation of the BS-ISA method used in this work, the molecular

charge density is partitioned among the atoms, and a set of multipole moments is

calculated for each atom. Instead of directly partitioning the density as in Eq. (10),

the BS-ISA approach is to first define atoms so that the atoms are as spherical as

possible while allowing for charge movement within the system.

The BS-ISA method has its origins in the Hirschfeld stockholder method for

calculating atomic densities. In the Hirschfeld stockholder method, the charge

density for each atom *a* is given as:

$$\rho^a(\mathbf{r}) = \rho(\mathbf{r}) \frac{w^a(\mathbf{r})}{\sum_b^{atoms} w^b(\mathbf{r})} \tag{11}$$

where $\rho(\mathbf{r})$ is the total molecular density and $w^a(\mathbf{r})$ is a function that describes the

shape of the atom *a*. The form of the shape function, $w^a(\mathbf{r})$, varies by method. An

insight by Lillestolen and Wheately was to use the spherical average of the atomic

density as the shape function, which avoids assuming a particular (usually free-

atom) shape function for each atom, and results in an equation that must be solved iteratively.[7] The BS-ISA method follows an analogous iteration scheme, but in basis space. That is, in the BS-ISA method, the terms in Eq. (11) are expanded in a basis, as shown below.

$$
\begin{aligned}
\rho^a(\mathbf{r}) &= \sum_k c_k^a \zeta_k^a(\mathbf{r}) \\
w^a(\mathbf{r}) &= \sum_{k \in s-\text{func}} c_k^a \zeta_{k,s}^a(\mathbf{r})
\end{aligned}
\tag{12}
$$

where c_k^a is a coefficient associated with atom a and is determined in the iterative procedure, $\zeta_k^a(\mathbf{r})$ is a basis function centered on atom a, k runs over all basis functions, $\zeta_{k,s}^a(\mathbf{r})$ is an s-type function on atom a, and $k \in s$ runs over all s-functions in the basis.

To determine the atomic density, the coefficients are calculated using an iterative procedure that minimizes a BS-ISA functional. The functional and minimization algorithm have been developed to make the procedure stable, accurate, and reliably convergent. Additionally, to ensure that the tail regions of the atomic densities are well described, the shape functions, $w^a(\mathbf{r})$, are modified so that they decay exponentially.

Once the atomic densities are obtained, multipole moment integrals can be computed. BS-ISA has many appealing properties, such as having a more systematic convergence with respect to multipole moment rank than DMA0 or DMA4. However, the DMA methods are more computationally efficient and algorithmically simpler. [8]

3. Computational details

As mentioned in the Background section, the EFP method has several parameters determined from an *ab initio* calculation. For the EFP/ISA and EFP/DMA calculations, all parameters except for the multipole moments are the same. That is, the static polarizability tensors, dynamic polarizability tensors, basis set, localized molecular orbitals, Fock matrix elements, and virtual molecular orbital coefficients are the same for the EFP/ISA and EFP/DMA calculations. All parameters except the multipole moments were generated using the 6-311++G(3df,2p) basis set. The EFP calculations and the *ab initio* calculations were done with the GAMESS[21][22] package. Several integral cutoffs were changed from the default values (ITOL was set to 24, ICUT to 12), and the SCF density convergence was tightened to 10^{-7}. Overlap-based damping was used to account for charge-penetration effects in the Coulomb energy. The localization method used was Boys[23][24] and the set of all canonical virtual orbitals was used for the charge transfer term.

The ISA multipole moments were generated with CamCASP 5.8.[25] The main basis set was aug-cc-pVTZ[26][27], the aux/ISA basis set was RI-MP2 aug-cc-pVTZ with ISA-set2[8] for s-functions (except for ethyne, which used RI-MP2 aug-cc-pVQZ with ISA-set2 for s-functions), and the ISA algorithm used was A+DF with ζ=0.1. Densities were from the PBE0 functional/AC. The asymptotic correction (AC) is the Casida-Salahub version of AC present in NWChem[28] with default (un-optimized) shift. NWChem was used for these calculations.

To consider the effects of basis set and the different DMA algorithms, the DMA multipole moments were generated in five different ways:

1. Following a previous paper that measured the accuracy of EFP against other force field methods[29], the DMA multipole moments were generated using HF/6-31+G(d) for non-aromatic molecules (ammonia, ethene, ethyne, formamide, formic acid, hcn, methane, water), and HF/6-31G(d) for aromatic molecules (2-aminopyridine, 2-pyridoxine, adenine, benzene, indole, phenol, pyrazine, thymine, uracil). The original analytic DMA procedure (DMA0) was used. This method is referred to as EFP/DMA0-small, since it uses a smaller basis set to generate the multipole moments. The set of atom centers and bond midpoints were used as expansion points for the multipole moments.

2. The DMA multipole moments are generated in the same manner as in EFP/DMA0-small, but only with the set of all atom centers as expansion points. This is referred to as EFP/DMA0-small-atoms.

3. DMA0 multiple moments were computed using HF/6-311++G(3df,2p), and the set of atom centers and bond midpoints were used as expansion points for the multipole moments. This method is referred to as EFP/DMA0.

4. DMA4 multiple moments were computed using the HF/6-311++G(3df,2p), and the set of atom centers and bond midpoints were used as expansion points for the multipole moments. This method is referred to as EFP/DMA4.

5. The DMA multipole moments were computed using HF/6-311++G(3df,2p), with DMA0 for non-aromatic molecules and DMA4 for aromatic molecules, and the set of atom centers and bond midpoints were used as expansion points for the multipole moments. This method is referred to as EFP/DMA-mixed.

The geometries at which ISA multipole moments and EFP potentials were generated are from the S22 dataset complexes. The geometry for ammonia, ethene, formic acid, phenol, pyrazine, water, and formamide is the geometry of the first monomer in the S22 dataset dimer for that molecule. The geometry of uracil is the geometry of the first monomer in the uracil H-bonded dimer. The geometry of benzene is the geometry of the first monomer in the benzene dimer T-shaped complex. The geometry of indole is the geometry of the indole in the benzene-indole T-shaped complex. The geometry of methane is the geometry of the methane in the benzene-methane dimer. The other molecules show up only once in the S22 dataset, and the S22 geometries are used for those molecules. Since the geometry of adenine and thymine in the Watson-Crick (WC) complex and the stacked complex differ significantly, the ISA multipole moments and EFP potentials were generated at both geometries, and used in the corresponding EFP calculations.

Six types of calculations are compared: EFP/ISA, EFP/DMA0-small, EFP/DMA0-small-atoms, EFP/DMA0, EFP/DMA4, and EFP/DMA-mixed.

4. Results

To test the accuracy of and to compare the methods, several comparisons are presented here. To compare predicted geometries, the S22 complexes were

geometry-optimized using all methods. The resulting geometries are compared to

the corresponding S22 geometries to assess the quality of geometry predictions.

Since the EFP fragments are internally frozen, the geometry optimization changes

only the angles and the distances between fragments. Next, the total interaction

energy at each optimized geometry is compared to the CCSD(T)/CBS binding energy

at the standard S22 geometry to assess the quality of the interaction energy

calculation for each method. In addition, the EFP energy components that depend on

the multipole moments (Coulomb energy, polarization energy, and charge transfer

energy) are compared to the corresponding SAPT2+(3)/aug-cc-pVTZ, CP (referred

to as "SAPT" in this work) energy components. The SAPT values are from the

Biofragment Database [30] [31] under "SAPT2+(3)/aug-cc-pVTZ, CP", and the values

correspond to those in the Addition/Correction to Ref. 29. The CCSD(T)/CBS values

correspond to the "S22A" reference set [32] in the Biofragment Database. [30] [31]

The equivalent SAPT terms used in the comparison are [See Ref. 33 and 34

for the notation]:

$$
\begin{aligned}
E_{SAPT}^{\text{Coulomb}} &= E_{\text{elst, resp}}^{(10)} + E_{\text{elst, resp}}^{(12)} + E_{\text{elst, resp}}^{(13)} \\
E_{SAPT}^{\text{exchange-repulsion}} &= E_{\text{exch}}^{(10)} + E_{\text{exch}}^{(11)} + E_{\text{exch}}^{(12)} \\
E_{SAPT}^{\text{induction}} &= E_{\text{ind, resp}}^{(20)} + E_{\text{exch-ind, resp}}^{(20)} + {}^{t}E_{\text{ind}}^{(22)} + {}^{t}E_{\text{exch-ind}}^{(22)} + \delta E_{\text{HF}}^{(2)} \\
E_{SAPT}^{\text{dispersion}} &= E_{\text{disp}}^{(20)} + E_{\text{disp}}^{(30)} + E_{\text{disp}}^{(21)} + E_{\text{disp}}^{(22)} + E_{\text{exch-disp}}^{(20)}
\end{aligned}
\tag{13}
$$

The sum of the EFP polarization and charge transfer energy is compared to the SAPT

induction energy. Note that the ISA multipole moments were computed using DFT

densities with the PBE0/AC functional; doing so could cause minor discrepancies

since the reference SAPT values were not computed using DFT.

To help gain insight into the differences in the dimers in the S22 dataset, the

dimers are split into hydrogen bonding, dispersion dominated, and mixed types,

following a previous EFP study. [29]

Predicted Geometries of the S22 Complexes

In the S22 dataset, the T-shaped benzene dimer is constrained to C_{2V}

symmetry, so this prescription is followed for the EFP methods. Table 1 shows the

differences relative to the S22 values for specific atom-atom distances. The mean

unsigned error (MUE) is also given in the table. In Table 1, X⸱⸱RD denotes the

distance between the atom X and the center of the plane made by the benzene ring.

(The plane is calculated using the first three atoms of the benzene in the dimer.) R1

and R2 are the vertical and horizontal distances between the planes of the rings,

respectively. (See Figure 1.) The notation in Table 1 is based on, and very similar to

that in Ref. 29. Section 1 of the Supporting Information contains the coordinates of

the dimers and gives the atom numbering used in the distance comparisons in Table

1.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 1: Differences (Å) in distance between the *ab initio* S22 geometry and the EFP geometries after optimization

	distance[a]	ΔEFP/ISA	ΔEFP/DMA0-small	ΔEFP/DMA0-small-atoms	ΔEFP/DMA0	ΔEFP/DMA4	ΔEFP/DMA-mixed
Hydrogen-Bonded Complexes							
ammonia dimer	N1⋯N5	-0.02	-0.10	-0.17	0.00	0.16	0.00
water dimer	O1⋯O4	0.00	-0.07	-0.12	0.05	0.04	0.05
formic acid dimer	O2⋯O8	0.11	0.05	-0.04	0.09	0.20	0.09
formamide dimer	O2⋯N9	0.10	-0.03	-0.03	0.04	0.20	0.04
uracil H-bonded dimer	N1⋯O23	0.08	0.04	0.01	0.02	0.18	0.18
2-pyridoxine 2-aminopyridine	N1⋯N15	0.04	-0.18	-[b]	-0.06	0.22	0.22
adenine–thymine WC	N1⋯N20	0.00	-0.18	-[b]	-0.06	0.23	0.23
MUE for Hydrogen-Bonded Complexes		0.05	0.09	0.07[e]	0.05	0.18	0.12
Dispersion-Dominated Complexes							
methane dimer	C1⋯C6	-0.12	-0.08	-0.12	-0.12	0.04	-0.12
ethene dimer	C1⋯C7	-0.07	-0.09	-0.16	-0.10	0.10	-0.10
benzene–methane	C1⋯RD[c]	0.21	0.23	0.21	-0.11	0.21	0.28
benzene stack	R1/R2[d]	0.43/-0.42	0.44/-0.30	0.41/-0.15	0.44/0.11	0.48/-0.59	0.48/-0.59
pyrazine dimer	R1/R2[d]	0.28/-0.11	0.30/-0.23	0.27/-0.12	0.35/-0.32	0.33/-1.04	0.33/-1.04
uracil stack	R1/R2[d]	0.18/-0.02	0.14/-0.02	0.13/-0.01	0.06/0.92	0.19/0.03	0.19/0.03
indole–benzene stack	R1/R2[d]	0.38/0.00	0.35/0.28	0.33/0.26	-[b]	0.44/-0.36	0.44/-0.36
adenine–thymine stack	R1/R2[d]	0.24/-0.20	0.22/-0.07	0.20/0.00	0.02/0.18	0.22/-0.22	0.22/-0.22
MUE for Dispersion-Dominated Complexes		0.20	0.21	0.18	0.25[e]	0.33	0.34

Mixed Complexes							
ethene–ethyne	C8···C2	0.16	0.07	0.08	0.06	0.19	0.06
benzene–water	O1···RD[c]	0.10	0.07	-0.04	-0.02	0.16	0.19
benzene–ammonia	N···RD[c]	0.15	0.17	0.11	-0.02	0.16	0.01
benzene–HCN	C14···RD[c]	0.24	0.15	0.09	0.17	0.21	0.25
benzene dimer T-shaped	C1··· RD[c]	0.30	0.30	0.30	0.25	0.30	0.30
indole–benzene T-shaped	N21···RD[c]	0.23	0.21	0.15	-0.05	0.26	0.26
phenol dimer	O7···O20	0.07	0.03	-0.01	0.07	0.00	0.00
MUE for Mixed Complexes		0.18	0.14	0.11	0.06	0.14	0.11
Overall MUE		0.16	0.16	0.14 [e]	0.15 [e]	0.25	0.23

[a]Atoms are numbered as in Section 1 of the Supporting Information. [b]The geometry optimization did not complete, since the induced dipole procedure failed to converge. [c] The distance between the atom X and the center of the plane made by the benzene ring, where the plane is calculated using the first three atoms of the benzene.[d] R1 and R2 are the vertical and horizontal distances between the planes of the rings, respectively. [e] The MUE is computed without the cases for which the induced dipole procedure does not converge.

ΔX is the difference between method X and the *ab initio* result. The values of the distances are in Section 1 of the Supporting Information.

Among the hydrogen-bonding complexes, the individual errors for all methods and complexes are less than 0.24 Å, which is in good agreement with the S22 geometries. The EFP/ISA, EFP/DMA4, and EFP/DMA-mixed methods have mostly positive differences, meaning that they overestimate the intermolecular

separation. The methods that used a smaller basis set to calculate the multipole

moments, the EFP/DMA0-small and EFP/DMA0-small-atoms methods, mostly have

negative differences, meaning that they underestimate the intermolecular distance.

A previous study used EFP/DMA0-small without charge transfer to optimize the S22

set, and found that the intermolecular separations were overestimated, on average

by 0.10 Å. Thus the effect of including the EFP charge transfer is to decrease the

distances, which is expected, since charge transfer is typically an attractive

interaction. Use of the smaller basis set might make the interaction too attractive,

since the distance is underestimated. The induced dipole procedure did not

converge when using the EFP/DMA0-small-atoms method for two aromatic

complexes, possibly because there are not enough expansion points.

In the dispersion-dominated complexes, the distances in the methane and

ethene dimers are underestimated by all methods except for EFP/DMA4, which

overestimates the distance. All of these errors are less than 0.3 A, which is in good

agreement with the S22 geometries. In the aromatic ring complexes, all methods

overestimate the distance between the ring planes, which implies that at least the

sign of this distance is not dependent on the multipole moments used. For the

EFP/ISA method and the methods that used a smaller basis set for the DMA

multipole moments, the parallel shift of the ring planes (R2) is underestimated for

all complexes except for indole-benzene. The induced dipole procedure did not

converge when using the EFP/DMA0 method for the indole-benzene stack complex.

Although overall the error is low, the methods using DMA multipole moments

generated from large basis sets performed the worst for the aromatic complexes.

The EFP/DMA4 and EFP/DMA-mixed methods (which are the same in this case), predict the R2 value for pyrazine dimer to be about 1 Å different from the S22 geometry, and the EFP/DMA0 method predicts the R2 value for the uracil dimer to be 0.9 Å different from the S22 geometry.

In the mixed complexes, the EFP/ISA, EFP/DMA0-small, EFP/DMA4, and EFP/DMA4-mixed methods slightly overestimate the distances for all complexes. The EFP/DMA0 and EFP/DMA0-small-atoms methods underestimate the distances for certain complexes and overestimate the distances for others. Overall, all methods are in good agreement with the S22 dataset, with the maximum error not exceeding 0.3 Å.

For all methods studied here, the overall mean unsigned error is under 0.26 Å. In two of the methods, EFP/DMA0-small-atoms and EFP/DMA0, the self-consistent induced-dipole procedure does not converge during one step of the geometry optimization for at least one complex. The possible causes for the divergence will be discussed in a later section. As can be seen in Section 1 of the SI, there were several geometries that differed from the S22 geometry by a small rotation, but the difference in the CCSD(T) energy between the different rotations are also very small. From the methods that are stable for all complexes, EFP/ISA and EFP/DMA0-small show the best agreement with S22 reference geometries, with deviations of 0.16 Å. This is only slightly better than the accuracy of the EFP/DMA0-small scheme without inclusion of charge-transfer (0.17 Å), as reported in Ref. 29.

Total interaction energies of the S22 Complexes

The total interaction energies of the methods at the optimized geometries are compared to CCSD(T)/CBS values [32,29] to test the accuracy of the energy calculations. The total EFP energy values are provided in Section 2 of the Supporting Information.

Figures 2, 3, and 4 show the differences in interaction energies between each method and the CCSD(T)/CBS values for each category of interaction energy. Note that if a geometry optimization failed due to non-convergence of the induced dipole procedure, the corresponding interaction energies are not shown in the figures. To summarize the results, the MUE for each method is shown in Table 2.

Table 2: MUE for all methods with respect to CCSD(T)/CBS (kcal/mol)

	EFP/ISA	EFP/DMA0-small	EFP/DMA0-small-atoms	EFP/DMA0	EFP/DMA4	EFP/DMA-mixed
MUE(HB)	1.297	1.672	2.854*	0.715	3.741	2.771
MUE(DISP)	0.429	0.331	0.279	0.957*	0.555	0.666
MUE(MIXED)	0.668	0.199	0.413	0.723	0.214	0.301
MUE(overall)	0.781	0.716	0.970*	0.800*	1.460	1.220

* The cases for which the induced dipole procedure does not converge are omitted

For the hydrogen-bonding complexes, the EFP/DMA0 method has the smallest MUE, while the EFP/DMA0-small-atoms has the lowest MUE for the dispersion-dominated species. The smallest MUE for the mixed systems is obtained with both the EFP/DMA0-small and the EFP/DMA4 methods. For the dispersion-dominated species and the mixed species, all of the MUEs are below 1 kcal/mol, so all methods work very well for these two types of dimers. The MUEs in interaction energies for the hydrogen bonded species range from 0.7 kcal/mol (DMA0) to 3.7 kcal/mol (DMA4). For the hydrogen-bonding complexes, the EFP/ISA and EFP/DMA4 methods consistently overestimate the energy, while the EFP/DMA0-small and EFP/DMA0-small-atoms methods consistently underestimate the energy. The EFP/DMA0 and EFP/DMA-mixed methods both underestimate and overestimate the energy for various complexes. The EFP/DMA4 and EFP/DMA-mixed methods have the largest individual error among all complexes,

overestimating the energy by 6.8 kcal/mol on the adenine-thymine Watson-Crick complex, a hydrogen-bonding complex, in particular.

Overall, the hydrogen bonding complexes are the major source of error for most methods. For the EFP/DMA0-small-atoms and EFP/DMA0 methods a small number of the induced dipoles do not converge during the optimizations. For the two failed optimizations with the EFP/DMA0-small-atoms method, the fragments approached very close to each other. Thus, the divergence is likely due to the EFP Coulomb and/or polarization interaction overestimating the attraction without being sufficiently balanced by the repulsive EFP exchange-repulsion term. The overestimation of the attraction could be due to errors in the multipole moment expansion at short range, or in the charge penetration term. For the failed optimization of the indole-benzene stack with the EFP/DMA0 method, the angle between the indole and benzene changes from approximately 180° (parallel) to ~45°. It is possible that the electrostatic potential is not described correctly for indole or benzene using the EFP/DMA0 method, and that more terms are needed in the multipole moment expansion to get a more accurate representation overall of the electrostatic potential.

EFP energy components at S22 geometry

To gain insight into the interaction energy errors, the EFP energy decomposition at the initial S22 geometry for each method is compared to the SAPT energy decomposition at the S22 geometry. Tables 3, 4, and 5 present the MUEs for the Coulomb term, the sum of the polarization and charge transfer terms, and the

total interaction energy term, respectively. Figures 5, 6, and 7 show the energy

differences between each method and the SAPT energies. Section 2 of the

Supporting Information contains the SAPT and EFP interaction energy components

for the S22 complexes.

Table 3: MUE for the EFP Coulomb term (kcal/mol)

	EFP/ISA	EFP/DMA0-small	EFP/DMA0-small-atoms	EFP/DMA0	EFP/DMA4	EFP/DMA-mixed
MUE(HB)	2.485	1.631	1.806	0.863	5.453	3.596
MUE(DISP)	2.560	2.475	2.431	3.105	1.514	1.487
MUE(MIXED)	0.960	0.553	0.614	0.806	0.816	0.897
MUE(overall)	2.027	1.595	1.654	1.677	2.545	1.970

Overall, the three EFP/DMA0 methods have the smallest Coulomb energy

MUEs, all within 0.1 kcal/mol of each other and below 2 kcal/mol. The EFP/ISA and

EFP/DMA-mixed methods have Coulomb energy MUEs that are only slightly larger,

and the MUE for the EFP/DMA4 method is about 0.5 kcal/mol larger than the others.

The latter is still reasonable.

For the hydrogen-bonded dimers, the general trend for all methods is to

overestimate the Coulomb term. The EFP/DMA0 method has the lowest Coulomb

energy MUE for the hydrogen bonding complexes, with a value less than 0.9

kcal/mol. The error in the Coulomb energy could be from the multipole moment

expansion or the charge penetration term. The largest errors are likely due to the

charge penetration term not accounting for all of the charge penetration, especially

for particularly strong interactions. It should be noted that the charge-penetration

correction is identical in all considered EFP methods, because it is computed using

inter-fragment overlap integrals and is not based on the multipole expansion.

For the dispersion-dominated dimers, all methods have individual errors of

less than 0.9 kcal/mol for the complexes without aromatic systems, agreeing well

with SAPT. However, all methods have large positive errors for the aromatic

systems. For most of the methods and complexes, the positive error can be

explained by an insufficient charge penetration term. Although the multipole

moment expansion part of the EFP Coulomb term is often positive, the SAPT

Coulomb energy is negative, so the charge penetration term is necessary to change

the sign of the EFP Coulomb energy. The largest individual error is that for the

indole-benzene stacked structure, with the EFP/DMA0 method. As mentioned above,

during the geometry optimization of the benzene-indole stacked structure using the

EFP/DMA0 method, the induced dipole procedure did not converge. A reason for the

non-convergence could be the large error in the Coulomb term seen here.

For the mixed complexes, the EFP Coulomb energy is similar to the SAPT

Coulomb energy, with the MUE for all methods less than 1 kcal/mol. The Coulomb

energies are relatively small for the mixed complexes.

There are several interesting comparisons to make. As may be seen by

comparing the EFP/DMA0-small and EFP/DMA0-small-atoms methods, having

expansion points on only atoms results in similar Coulomb energies to having

expansion points on atoms and bond midpoints. The numeric EFP/DMA4 method

generally has similar or smaller errors than the EFP/DMA0 method, except for the

hydrogen-bonded complexes, for which the reverse is true. The EFP/ISA method

consistently slightly overestimates the Coulomb energy.

Table 4: MUE for the EFP Polarization and Charge transfer term (kcal/mol)

	EFP/ISA	EFP/DMA0-small	EFP/DMA0-small-atoms	EFP/DMA0	EFP/DMA4	EFP/DMA-mixed
MUE(HB)	2.098	0.705	0.871	1.641	3.366	2.416
MUE(DISP)	0.395	0.366	0.177	1.097	0.973	0.916
MUE(MIXED)	0.313	0.196	0.228	0.418	0.639	0.494
MUE(overall)	0.911	0.420	0.414	1.054	1.628	1.259

Now, consider the polarization + charge transfer (P+CT) term (Table 4).

Overall, the MUEs for the EFP/ISA, EFP/DMA0-small-atoms and EFP/DMA0-small

methods are all less than 1 kcal/mol, and the MUE for the EFP/DMA0 method is only

slightly larger than 1 kcal/mol.

For the hydrogen-bonded complexes, all of the methods except EFP/DMA0-

small and EFP/DMA0-small-atoms generally overestimate the P+CT interaction

energy. The overestimation could be due to an underestimation of the charge

penetration energy. Since the polarization term uses the static electric field

generated by the multipole moments, and since the multipole moment expansion is

not accurate at short distances, the error might be due to the multipole moment

expansion not properly describing short-range interactions. While the EFP Coulomb

term includes a charge penetration term to offset this problem, the EFP polarization

term includes a multiplicative damping term, and the EFP charge transfer term does

not include any damping. It is possible that the polarization damping term does not

account for all of the effects of charge penetration and that the lack of CT damping

results in an underestimation of the energy. Almost all methods have large errors

for the formic acid dimer, with the EFP/DMA4 method having the largest error. The

large errors produced by the EFP/DMA4 method might be due to the multipole

expansion produced by DMA4 being truncated too soon. The potential could be

improved by including higher-rank multipoles.

For the dispersion-dominated dimers, all methods are in good agreement

with the SAPT induction term for the complexes without ring systems, with the

error being less than 0.5 kcal/mol for all methods. The errors are larger for the ring-

systems.

All methods agree very well with the SAPT induction energy for the mixed

complexes. The individual errors are generally small, less than 1.2 kcal/mol for all

methods and complexes.

Overall, all methods have relatively small errors when compared to the SAPT

induction term, with the overall MUEs less than 1.7 kcal/mol for all methods. As in

the previous section, there are several interesting comparisons to make. The

EFP/DMA0-small method gives consistently better results than EFP/DMA0, which

might be due to a basis set effect, as mentioned above, or due to the multipole

moment expansion for EFP/DMA0 being truncated too soon. The EFP/ISA method

has consistent small overestimations, unlike any of the other methods.

Table 5: MUE for the total EFP energies (kcal/mol)

	EFP/ISA	EFP/DMA0-small	EFP/DMA0-small-atoms	EFP/DMA0	EFP/DMA4	EFP/DMA-mixed
MUE(HB)	1.955	1.315	2.105	0.612	6.215	3.773
MUE(DISP)	2.109	2.022	1.852	2.069	1.504	1.619
MUE(MIXED)	1.272	0.535	0.399	0.825	0.998	1.308
MUE(overall)	1.794	1.324	1.470	1.210	2.842	2.205

Finally, consider the total EFP interaction energies. As described in Ref. 29, and seen here, there is some cancellation of error. The exchange-repulsion term is generally underestimated, partially cancelling out the overestimation of the Coulomb and polarization term, while the dispersion interaction energy is generally similar to the SAPT dispersion energy. The EFP/DMA4 method has the largest errors in the total interaction energies, mostly due to overestimating the interaction energy in the Coulomb and induction terms in the hydrogen-bonded dimers. The EFP/DMA0 method has the lowest overall MUE, partially due to error cancelation. Most methods have the largest errors in the hydrogen-bonded and dispersion-dominated complexes, and for most of the complexes and methods, the total interaction energy is too repulsive.

5. Conclusion

An important strength of the EFP method is that, because there are no empirically fitted parameters, the method can systematically be improved. As demonstrated in this work, it is straightforward to use a different set of multipole moments in the calculation, and still get accurate and reasonable results. As long as a set of multipole moments is provided that describes the electrostatic potential of a molecule reasonably well, the EFP method will provide reasonable results.

Of the methods considered here, the EFP/ISA and EFP/DMA0-small methods have the lowest overall error compared to the CCSD(T)/CBS results. The MUE for the S22 complexes is 0.78 and 0.72 kcal/mol for EFP/ISA and EFP/DMA0-small, respectively. The MUEs are similar to the MUEs for MP2 and SCS-MP2 when compared to CCSD(T)/CBS (0.88 and 0.8 kcal/mol, respectively). A similar study of the EFP/DMA0-small accuracy in which the charge transfer term was not included had a MUE of 0.9 kcal/mol, so including charge transfer increased the accuracy. [29]

Using a larger basis set to calculate the multipole moments with the DMA0 or DMA4 method results in a higher MUE than the DMA0-small method, but overall provides reasonable results, with MUEs of 0.800, 1.460, and 1.220 kcal/mol compared to the CCSD(T)/CBS results for the EFP/DMA0, EFP/DMA4, and EFP/DMA-mixed methods, respectively. In the case of the EFP/DMA0 method, the induced dipole procedure did not converge during the course of the indole-benzene stack geometry optimization. The lack of convergence is thought to be because the DMA0 multipole moment expansion is truncated too soon for indole or benzene with the basis set used, which might be remedied by including higher multipole moment ranks in the multipole moment expansion.

Computing the multipole moments using the smaller basis set and expansion points only on atom centers (EFP/DMA0-small-atoms) results in a similar MUE to using bond midpoints and atom centers as expansion points, but in two cases results in the induced dipole procedure not converging during the geometry optimization. Fewer expansion points can lead to divergence of the multipole moment expansion at short ranges. Although the EFP/ISA method also only uses atom centers as expansion points, it does not have induced dipole procedure divergence. Thus, the convergence of the multipole moments in the ISA procedure seems to be more robust than in DMA, as noted in Ref. 8.

Overall, the EFP/ISA method is a promising method. As noted in Ref. 8, the ISA multipole moments tend to systematically converge the multipole moment expansion at a lower term than DMA methods, which is likely why the EFP/ISA method has low errors, and consistently slightly overestimates the SAPT components. The main downside to using ISA multipole moments is that the procedure to generate them is much more computationally expensive than the procedure used to generate the DMA multipole moments. As mentioned above, some of the errors might occur because the ISA multipole moments were computed using DFT densities with the PBE0/AC functional, while the reference SAPT values used here were from SAPT2+(3)/aug-cc-pVTZ, CP.

Analyzing the energy components at the S22 geometry shows that many of the methods predict that the total interaction energies are too repulsive. Thus, it is clear that the short-range penetration effects (charge penetration term, the electric field damping) might be underestimated in the EFP method. Additionally, for certain

molecules, the multipole moment expansion generated with DMA0 or DMA4 for the

larger basis sets does not seem to be converged for the level of truncation used,

which could be remedied by including additional multipole moments in the

multipole moment expansion.

Supporting Information Available: Tables of coordinates and distance information for optimized geometries, and interaction energies for optimized geometries and S22 geometries.

This material is available free of charge via the Internet at http://pubs.acs.org.

References

1.   Mulliken, R. S., Electronic Population Analysis on LCAO–MO Molecular Wave Functions. I. *The Journal of Chemical Physics* **1955,** *23* (10), 1833-1840.
2.   Stone, A. J., *The Theory of Intermolecular Forces*. Oxford University Press: New York, 1997.
3.   Bader, R. F. W., Atoms in Molecules. Clarendon Press: Oxford, 1990.
4.   Hirshfeld, F. L., Bonded-atom fragments for describing molecular charge densities. *Theoretica chimica acta* **1977,** *44* (2), 129-138.
5.   Bultinck, P.; Van Alsenoy, C.; Ayers, P. W.; Carbó-Dorca, R., Critical analysis and extension of the Hirshfeld atoms in molecules. *The Journal of Chemical Physics* **2007,** *126* (14), 144111.
6.   Popelier, P. L., Atoms in Molecules: An Introduction. Prentice Hall: Harlow, 2000.
7.   Lillestolen, T. C.; Wheatley, R. J., Redefining the atom: atomic charge densities produced by an iterative stockholder approach. *Chemical Communications* **2008,** (45), 5909-5911.

8.   Misquitta, A. J.; Stone, A. J.; Fazeli, F., Distributed Multipoles from a Robust Basis-Space Implementation of the Iterated Stockholder Atoms Procedure. *Journal of Chemical Theory and Computation* **2014,** *10* (12), 5405-5418.

9.   Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D., An effective fragment method for modeling solvent effects in quantum mechanical calculations. *The Journal of Chemical Physics* **1996,** *105* (5), 1968-1986.

10.   Gordon, M. S.; Smith, Q. A.; Xu, P.; Slipchenko, L. V., Accurate First Principles Model Potentials for Intermolecular Interactions. *Annual Review of Physical Chemistry* **2013,** *64* (1), 553-578.

11.   Stone, A. J., Distributed Multipole Analysis: Stability for Large Basis Sets. *Journal of Chemical Theory and Computation* **2005,** *1* (6), 1128-1132.

12.   Xu, P.; Gordon, M. S., Charge transfer interaction using quasiatomic minimal-basis orbitals in the effective fragment potential method. *The Journal of Chemical Physics* **2013,** *139* (19), 194104.

13.   Li, H.; Gordon, M. S.; Jensen, J. H., Charge transfer interaction in the effective fragment potential method. *The Journal of Chemical Physics* **2006,** *124* (21), 214108.

14.   Slipchenko, L. V.; Gordon, M. S., Damping functions in the effective fragment potential method. *Molecular Physics* **2009,** *107* (8-12), 999-1016.

15.   Kairys, V.; Jensen, J. H., Evaluation of the charge penetration energy between non-orthogonal molecular orbitals using the Spherical Gaussian Overlap approximation. *Chemical Physics Letters* **1999,** *315* (1–2), 140-144.

16.   Tang, K. T.; Toennies, J. P., An improved simple model for the van der Waals potential based on universal damping functions for the dispersion coefficients. *The Journal of Chemical Physics* **1984,** *80* (8), 3726-3741.

17.   Jensen, J. H., Intermolecular exchange-induction and charge transfer: Derivation of approximate formulas using nonorthogonal localized molecular orbitals. *The Journal of Chemical Physics* **2001,** *114* (20), 8775-8783.

18.   Stone, A. J.; Misquitta, A. J., Charge-transfer in Symmetry-Adapted Perturbation Theory. *Chemical Physics Letters* **2009,** *473* (1–3), 201-205.

19.   Gordon, M. S., *Fragmentation: Toward Accurate Calculations on Complex Molecular Systems*. John Wiley and Sons: In Press.

20.   Slipchenko, L. V.; Gordon, M. S., Electrostatic energy in the effective fragment potential method: Theory and application to benzene dimer. *Journal of Computational Chemistry* **2007,** *28* (1), 276-291.

21.   Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A., General atomic and molecular electronic structure system. *Journal of Computational Chemistry* **1993,** *14* (11), 1347-1363.

22.   Gordon , M. S.; Schmidt , M. W., Advances in Electronic Structure Theory: GAMESS a Decade Later . In *Theory and Applications of Computational Chemistry: The First Forty Years*, Elsevier B.V.: Amsterdam, The Netherlands, 2005; pp 1167 – 1189.

23.   Boys, S. F., Construction of Some Molecular Orbitals to Be Approximately Invariant for Changes from One Molecule to Another. *Reviews of Modern Physics* **1960,** *32* (2), 296-299.

24.   Edmiston, C.; Ruedenberg, K., Localized Atomic and Molecular Orbitals. *Reviews of Modern Physics* **1963,** *35* (3), 457-464.

25.   Misquitta, A. J.; Stone, A. J. *CamCASP: a program for studying intermolecular interactions and for the calculation of molecular properties in distributed form*, http://www-stone.ch.cam.ac.uk/programs.html - CamCASP.

26.   Dunning, T. H., Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *Journal of Chemical Physics* **1989,** *90* (2), 1007.

27.   Kendall, R. A.; Dunning, T. H.; Harrison, R. J., Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *The Journal of Chemical Physics* **1992,** *96* (9), 6796-6806.

28.   Valiev, M.; Bylaska, E. J.; Govind, N.; Kowalski, K.; Straatsma, T. P.; Van Dam, H. J. J.; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T. L.; de Jong, W. A., NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Computer Physics Communications* **2010,** *181* (9), 1477-1489.

29.   Flick, J. C.; Kosenkov, D.; Hohenstein, E. G.; Sherrill, C. D.; Slipchenko, L. V., Accurate Prediction of Noncovalent Interaction Energies with the Effective Fragment Potential Method: Comparison of Energy Components to Symmetry-Adapted Perturbation Theory for the S22 Test Set. *Journal of Chemical Theory and Computation* **2012,** *8* (8), 2835-2843.

30.   Sherrill, D. C.; Merz, K., M, Jr., Quantum Mechanical Methods for Quantifying and Analyzing Non-Covalent Interactions and for Force-Field Development. In *Many-Body Effects and Electrostatics in Biomolecules,* Cui, Q.; Meuwly, M.; Ren, P., Eds. CRC Press: 2016.

31.   The Biofragment Database. http://vergil.chemistry.gatech.edu/bfdb_private/bfdb/cgi-bin/bfdb.py.

32.   Takatani, T.; Hohenstein, E. G.; Malagoli, M.; Marshall, M. S.; Sherrill, C. D., Basis set consistent revision of the S22 test set of noncovalent interaction energies. *The Journal of Chemical Physics* **2010,** *132* (14), 144104.

33.   Jeziorski, B.; Moszynski, R.; Szalewicz, K., Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van der Waals Complexes. *Chemical Reviews* **1994,** *94* (7), 1887-1930.

34.   Hohenstein, E. G.; Sherrill, C. D., Wavefunction methods for noncovalent interactions. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012,** *2* (2), 304-326.

35.   Bertoni, C.; Gordon, M. S., Analytic Gradients for the Effective Fragment Molecular Orbital Method. *Journal of Chemical Theory and Computation* **2016,** *12* (10), 4743-4767.
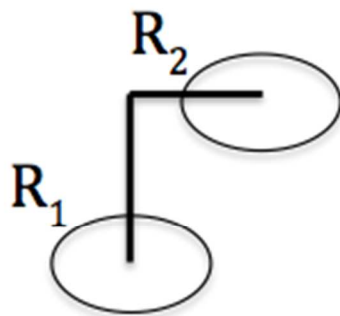
List of Figures

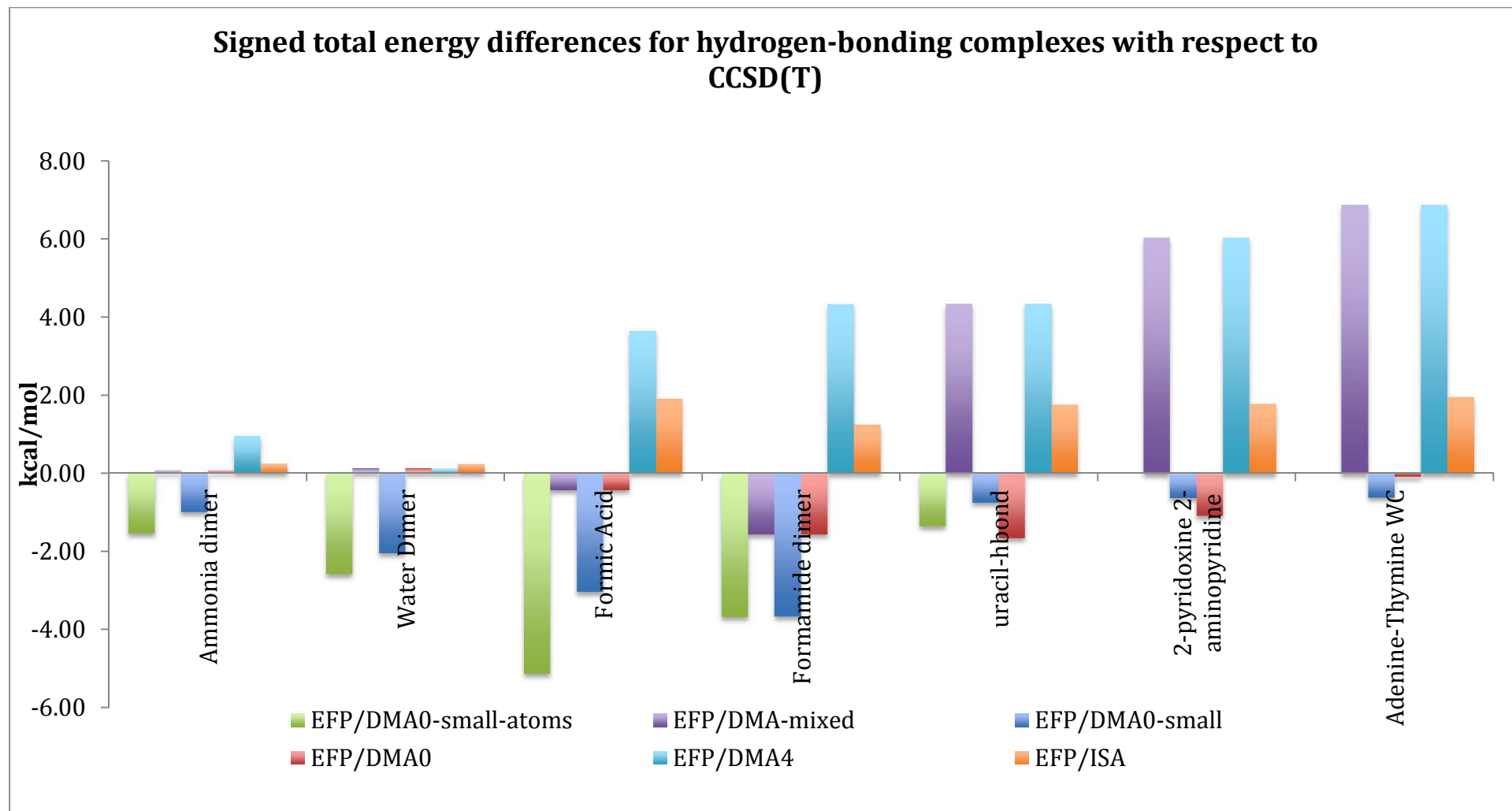**Figure 1**: A definition of the R_1 and R_2 values

Figure 2: Comparison of the total energy differences for the hydrogen bonding complexes at the optimized geometry
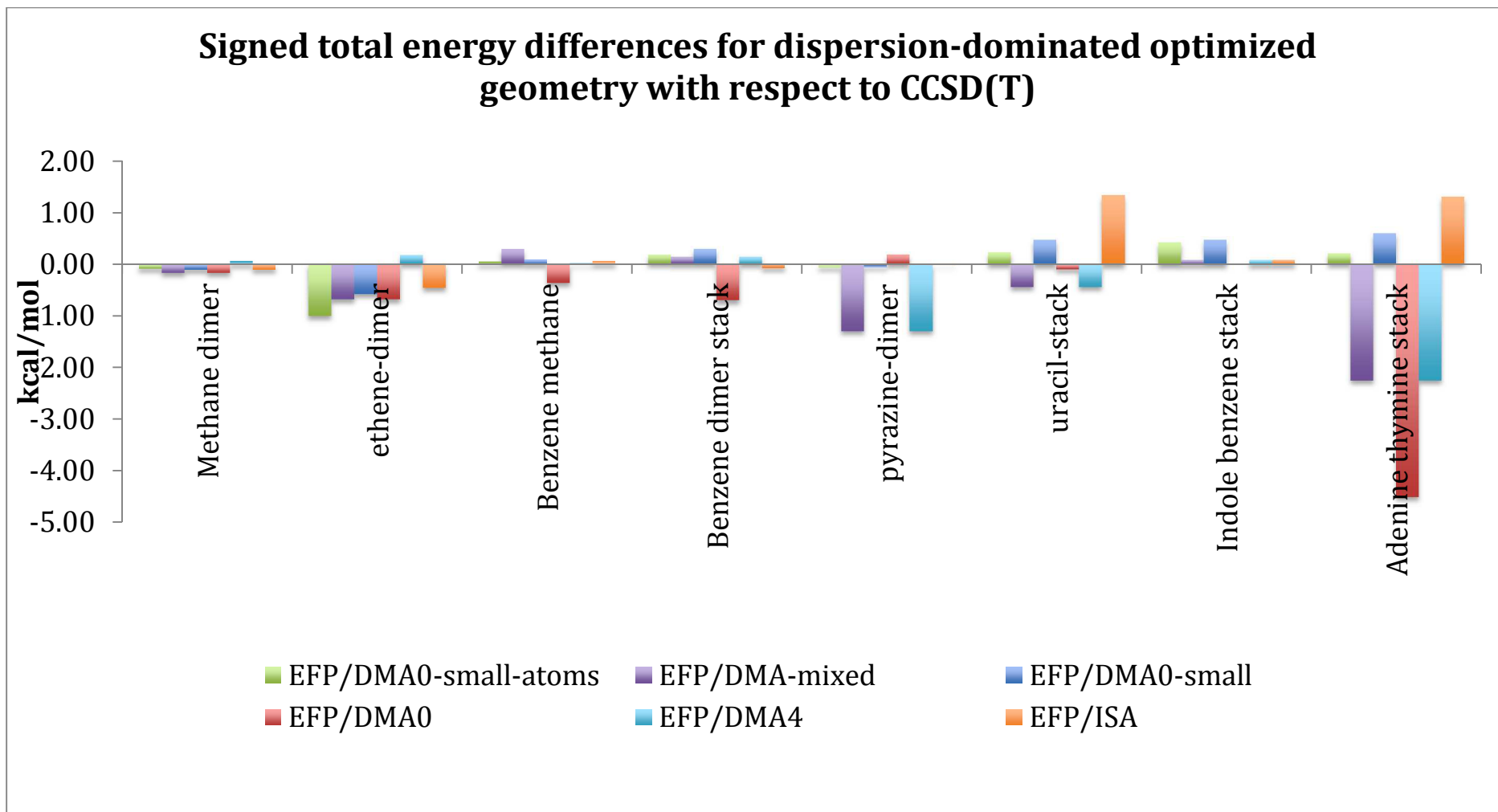
Figure 3: Comparison of the total energy differences for the dispersion-dominated complexes at the optimized geometry
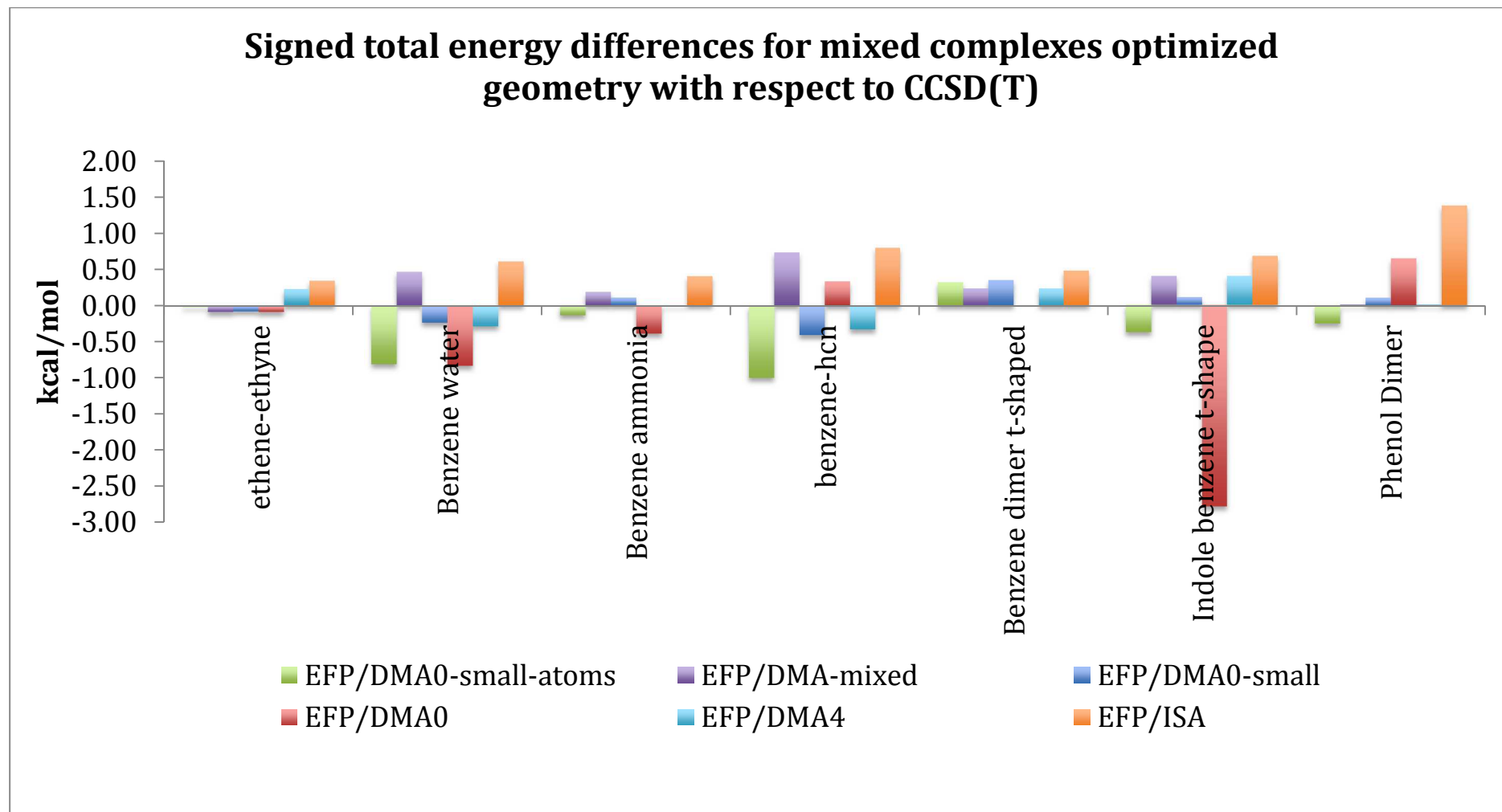
Figure 4: Comparison of the total energy differences for the mixed complexes at the optimized geometry

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

Figure 5: Comparison of the Coulomb energy differences for the S22 geometries

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
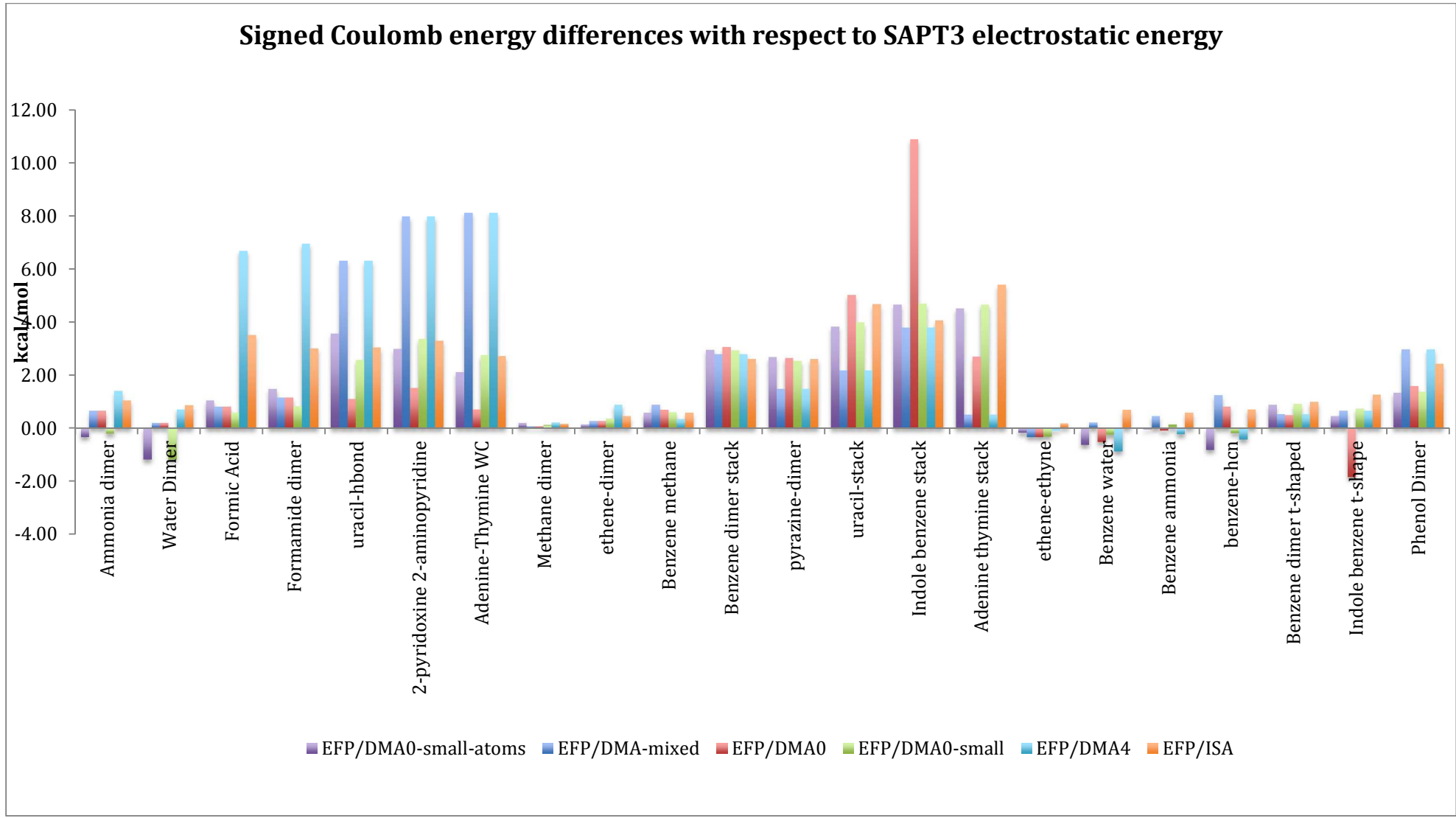30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

Figure 6: Comparison of the polarization and charge transfer energy differences for the S22 geometries

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

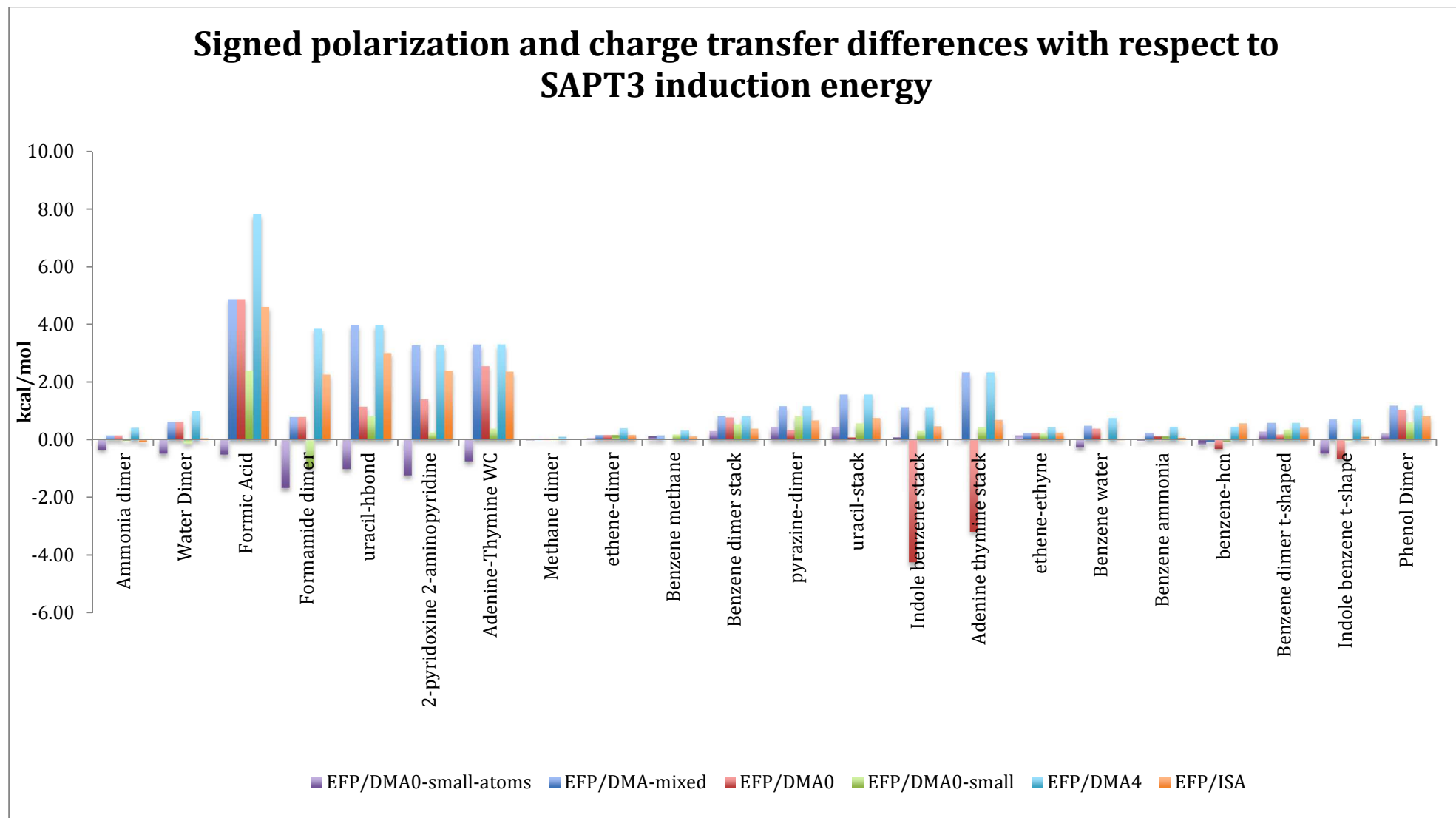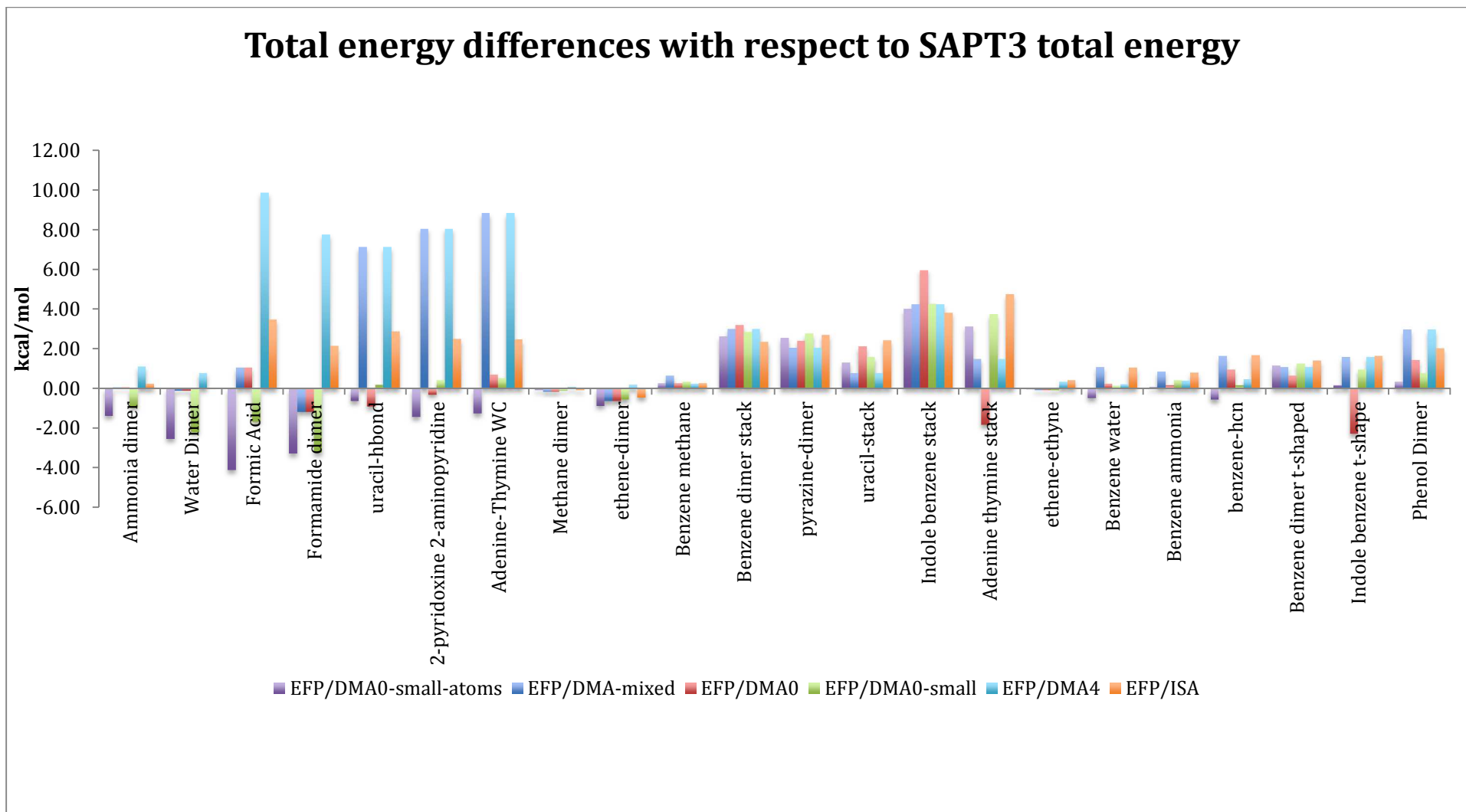**Figure 7:** Comparison of the total energy differences for the S22 geometries

TOC Graphic

$$E_{\text{total}}^{EFP} = E^{\text{Coulomb}}\left(q\ ,\mu_{\alpha},\Theta_{\alpha\beta},...\right) + E^{\text{exchange-repulsion}} + E^{\text{dispersion}} + E^{\text{charge-transfer}}\left(q\ ,\mu_{\alpha},\Theta_{\alpha\beta},...\right) + E^{\text{polarization}}\left(q\ ,\mu_{\alpha},\Theta_{\alpha\beta},...\right)$$

Distributed multipole moments
BS-ISA and DMA