RESEARCH ARTICLE

Comparison of approximate intermolecular potentials for ab initio fragment calculations on medium sized N-heterocycles

Bónis Barcza [ID] | **Ádám B. Szirmai** [ID] | **Katalin J. Szántó** [ID] | **Attila Tajti** [ID] | **Péter G. Szalay** [ID]

Institute of Chemistry, Laboratory of Theoretical Chemistry, ELTE Eötvös Loránd University, Budapest, Hungary

Correspondence
Péter G. Szalay and Attila Tajti, Institute of Chemistry, Laboratory of Theoretical Chemistry, ELTE Eötvös Loránd University, P. O. Box 32, H-1518, Budapest 112, Hungary.
Email: szalay@chem.elte.hu and attila.tajti@ttk.elte.hu

Abstract

The ground state intermolecular potential of bimolecular complexes of N-heterocycles is analyzed for the impact of individual terms in the interaction energy as provided by various, conceptually different theories. Novel combinations with several formulations of the electrostatic, Pauli repulsion, and dispersion contributions are tested at both short- and long-distance sides of the potential energy surface, for various alignments of the pyrrole dimer as well as the cytosine–uracil complex. The integration of a DFT/CCSD density embedding scheme, with dispersion terms from the effective fragment potential (EFP) method is found to provide good agreement with a reference CCSD(T) potential overall; simultaneously, a quantum mechanics/molecular mechanics approach using CHELPG atomic point charges for the electrostatic interaction, augmented by EFP dispersion and Pauli repulsion, comes also close to the reference result. Both schemes have the advantage of not relying on predefined force fields; rather, the interaction parameters can be determined for the system under study, thus being excellent candidates for ab initio modeling.

KEYWORDS
dispersion, effective fragment potential, embedding, intermolecular interactions, Pauli repulsion, QM/MM

1 | INTRODUCTION

Ab initio quantum chemistry is now an indispensable part of chemical research. Although a large arsenal of methods is available, the cost of calculations grows rapidly with system size; and, the more advanced the method, the faster the increase in cost. A possible way to overcome this problem is the fragmentation of the system into smaller components and applying an appropriate formulation to account for their interaction. The number of methods of this kind is tremendous, yet there are still many challenges. One of the biggest issues is how to cut chemical bonds during fragmentation.

A specific situation is presented by molecular complexes, whose fragments are molecules connected by non-covalent interactions. In this case the Hamiltonian of a two-component system is trivially partitioned as

$$\hat{H}(r_1,r_2) = \hat{H}_1(r_1) + \hat{H}_2(r_2) + \hat{V}(r_1,r_2), \quad (1)$$

with $\widehat{H}_i$ the Hamiltonian of the non-interacting fragments and $\widehat{V}(r_1, r_2)$ the interaction. This form evidently suggests perturbation theory as the approach of choice, with the zeroth order wave function taken as the product of the fragments' wave functions.

A large variety of methods with this ansatz are available today, of various sophistication. Many of them offer the division of the interaction energy into various, physically interpretable terms, including purely electrostatic forces, dispersion, Pauli repulsion and other smaller contribution s. The relative importance of the contributions depends on the systems under investigation, as well as their distance. The methods used for computing non-covalent interactions can be classified as follows:

1. Symmetry-Adapted Perturbation Theory (SAPT)[1,2] is in fact a hierarchical set of methods based on double perturbation theory, with respect to intermolecular interaction and electron correlation. SAPT allows the calculation of individual, physically distinguishable contributions, as well as the effect of electron correlation on them. These methods are often used as reference in benchmarking studies.
2. Classical force fields offer pre–calculated, transferable parameter sets for the most important contributions to the non-covalent interaction, such as electrostatic terms (partial charges, atomic multipoles), as well as Lennard-Jones-type terms for dispersion and Pauli repulsion.
3. Methods calculating the interaction energy directly from properties of the monomers. Among numerous methods of this type, the Effective Fragment Potential (EFP) methods[3] are the most prominent.
4. Various energy decomposition approaches (EDA) of different complexity and restrictions provide physically interpretable components of the interaction energy, ranging from the early suggestion of Morokuma,[4] the Block-Localized Wave function (BLW)[5] and Natural EDA (NEDA) schemes[6,7] to the Absolutely Localized MO (ALMO) method by Head-Gordon et al.[8,9]
5. Embedding type methods include the interaction explicitly in the Hamiltonian of the fragment(s). Most widespread are the quantum mechanics (QM)/Molecular Mechanics (MM) approaches,[10] but other, density-based embedding methods are also available.

The authors of this paper have long been interested in the description of the interaction of fragment–localized chromophores; preliminary calculations showed that the most often used, purely electrostatic approximations fail to provide the correct potential energy surface, because further contributions are significant.[10] This is particularly true when the short–distance part of the potential is of interest, for example for stacked interactions in DNA.

To find the best–suited approximation for these terms, the present paper investigates the ground states of non-covalently interacting bimolecular complexes and compares various methods for different contributions to the interaction energy. Keeping our focus on nitrogen–containing heterocycles, our test systems include pyrrole dimers (denoted hereafter as *Pyr-Pyr*) in different orientations, as well

as stacked cytosine-uracil base pairs (*Cyt-Ura*). These complexes have low-lying Frenkel-type pairs of excited states for which the proper handling of non-covalent interactions is needed to obtain accurate splitting.

The paper is organized as follows. Section 2 gives a short summary of the available methodologies, with emphasis on the differences in theoretical formulations and the possibility of incorporating them into ab initio calculations. Based on this analysis, in Section 3 we select a set of methods to be used in the tests. Section 4 describes the computational details, and Section 5 presents the results.

2 | METHODOLOGIES OF MODELING INTERMOLECULAR INTERACTIONS

Already at the dawn of quantum chemistry, it became clear that quantum effects give rise to important short–range forces.[11,12] By now, the treatment of intermolecular interactions has a broad theoretical background. Systematic overviews of the topic have been presented in comprehensive texts.[13–17]

The contributions to the interaction energy are usually referred to as[1,15]:

1. Electrostatic (eventually including polarization effects);
2. Dispersion;
3. Pauli repulsion (also called exchange repulsion);
4. Charge Transfer (CT);
5. and combinations of the above.

One way to distinguish these terms is offered by the SAPT theory[1,2] that, in different orders of perturbation theory, provides approximate definitions for the individual terms of interaction. Of course, one should always keep in mind that distinguishing these contributions is physically not strict since an exact (or accurate) calculation on the complete system would give just one single, interaction energy. Nevertheless, from the perspective of chemical and physical intuition, and also for developing reasonable approximations, this distinction is well justified.

In SAPT, the electrostatic energy enters in first order in the interaction operator, with sizable contributions from the correlation terms.[1,2] Thus, a careful modeling of this contribution is essential.

The electrostatic interaction is traditionally incorporated into the model via charges placed on the interacting molecules.[18] If these charges are at the atomic positions, the model corresponds to the chemical picture of atoms with partial charges. Also, this choice allows for transferable force fields since charges can be assigned to atom types, a concept used for other (intramolecular) parameters of the force fields. Traditional force fields like AMBER[19,20] or CHARMM[21] use such parametrization.

The ways of defining the partial charges have been reviewed recently by Janeček et al.[22] Here we just mention that the population analysis concept used in the past is nowadays replaced by fitting directly to ab initio results on interaction energies, dipole moments or

electrostatic potentials. For the present study CHELPG[23] is of special importance since it represents a readily available algorithm to calculate atomic charges for the particular monomer of interest. A recent advance in this field is the W-RESP potential of Janeček et al.[22]

A more flexible parametrization of the electrostatic interaction can be achieved using atomic multipoles, first suggested by Stone.[14,24] Pioneering work in applying this technique was done within the Effective Fragment Potential (EFP) framework.[3] A recent version of the method, termed EFP2[25,26] includes distributed multipoles up to octupoles at the atomic centers and also at bond midpoints, with polarizability included for localized molecular orbitals.[27] Modern force fields like AMOEBA (Atomic Multipole Optimized Energetics for Biomolecular Applications),[28] SIBFA (Sum of Interactions Between Fragments Ab initio computed),[29–31] DRF (Direct Reaction Field)[32] also use this parametrization including multipoles up to quadrupoles, often also with polarizable components.

For short intermolecular distances, however, the direct use of classical electrostatic interactions, whether in terms of charges or multipoles, does not work: the overlap of electronic densities cannot be described correctly by discrete charge distributions.[14] The importance of this effect in, for example, stacked nucleobases was shown by Sherill and co-workers[33,34] in a comparison to SAPT0: the slightly repulsive electrostatic interaction turns into attraction below 4 Å at the SAPT0 level, while AMBER and CHARMM remain repulsive. This effect plays important role in compensating, e.g., for Pauli repulsion (see later). To cover this artifact, damping terms have been introduced[14] in several model potentials like EFP,[35] AMOEBA,[36–38] SIBFA,[39] or as a stand-alone extension[40]; it has also been implemented in the Gaussian program package[41] for QM/MM type electrostatic modeling.

Two additional energy components decay fast with the distance of the fragments but proved to be important in SAPT theory.[1] One is the dispersion energy, a result of correlation of the respective electrons in the two fragments, which appears first in the second order perturbation terms, thus correlation is clearly essential in it. Second is the Pauli (exchange) repulsion, which arises when antisymmetrizing the wave function of the complete system.[42]

In molecular force fields, dispersion and Pauli repulsion are often considered together as one van der Waals potential, traced back to Lennard-Jones.[43] A comprehensive historical summary of the subject was published recently.[44] Lennard-Jones (LJ) (12–6) or Halgren's 14–7 potentials[45] are offered by various force fields. More sophisticated models are also available, using more rigorous theoretical formulation based on monomer properties. For example, S^2-dependent functions (with S being the orbital overlap) for Pauli repulsion is used in SIBFA[46] and EFP.[47–49]

For dispersion, EFP2 calculates the so-called C_6 coefficient of the London expansion as the instantaneous dipole–induced-dipole interaction, obtained from the frequency dependent polarizabilities; it also includes a C_8 term estimated as one-third of C_6.[50] The different versions of D corrections (D, D2, D3) on density functional theory (DFT), to include dispersion, were developed by Grimme et al.[51–53] for molecules, but these can also be used to obtain intermolecular dispersion energy. Similarly to the respective EFP2 term, it is based on a London

expansion up to C_8, with C_6 calculated from hydrides of the given atoms. It also includes a theoretically justifiable damping factor and three-body terms.

Beyond these *a posteriori* calculations, a more sophisticated way to treat intermolecular interactions is to explicitly include them into the Hamiltonian of the fragments. These models belong to the general family of embedding methods.[54] In these schemes the subsystem(s) is (are) described by higher levels of theory, with the environment approximated in a simpler model. The latter can be classical continuum,[55] MM[10,56] or lower level QM.[57]

The simplest embedding scheme, QM/MM, places the quantum system in a molecular mechanics environment. The QM Hamiltonian of the fragment(s) is augmented by (possibly one-electron) interaction terms usually taken from force fields. The point charges offered by the force fields can readily be used in this scheme to include the electrostatic contribution,[10,56] but the inclusion of multipoles is also possible, see for example the QM/EFP2 method.[26] An advantage of QM/MM is that one can use (in principle) any QM level to calculate the fragment's electronic structure under the influence of the environment[10] and even excited states can be treated in a simplified manner.

The inclusion of Pauli repulsion and dispersion is less straightforward in QM/MM. Often these are just added as *a posteriori* corrections to the energy. One should realize, however, that such potentials depend only on the distance (of the atoms), the parameters are predefined and do not reflect the actual electronic structure of the fragments. In particular, the potentials have been optimized for ground states, thus their use for excited states is not entirely justified.[58]

To overcome these problems, an embedding version of EFP2, termed QM/EFP2 or ab initio /EFP2 has been introduced recently.[25,26,59] The method treats dispersion[60] and Pauli exchange[58] in the form of special one-electron operators. Reference [26] compares individual terms obtained with EFP2 and QM/EFP2 with the corresponding SAPT values for the S66 test set.[61] It was found that QM/EFP2 does not necessarily deliver superior interaction potentials; this is most probably due to the lesser error compensation compared to the regular EFP model. First applications to excited states give promising results.[58,62] Implementations of QM/EFP are available in the GAMESS[63] and Q-Chem[64] program suites.

The embedding scheme also facilitates the treatment of all subsystems at QM level, with the advantage that the interaction with the electron density of the environment can directly be calculated, without discretization of the charge distribution or parametrization of other contributions. Usually, the density of the surroundings is calculated at a lower level of theory. The Kohn–Sham (KS) DFT offers an excellent framework for this QM based embedding as it scales favorably with the system size while providing a relatively accurate description of the environment's density. In a typical DFT embedding scheme the energy of the system is written as

$$E[\rho] = E[\rho_A] + E[\rho_B] + E_{int}[\rho_A, \rho_B], \tag{2}$$

where A is the active subsystem, B the approximated environment, and the third term their interaction; ρ denotes density. The main issue

TABLE 1 Overview of the methods and combinations used in this study

Method	Contribution				
	Electrostatics	Pauli repulsion	Dispersion	Polarization	CT
GAFF point charges (*GAFF QM/MM*)	•				
CHELPG point charges (*CHELPG QM/MM*)	•				
EFP2 multipole expansion (*EFP2 El*)	•				
Huzinaga embedding (*Embed*)	•	•			
EFP2 Pauli repulsion (*EFP2 Rep*)		•			
EFP2 dispersion (*EFP2 Disp*)			•		
D3 correction (*D3 Disp*)			•		
GAFF Lennard-Jones (*GAFF LJ*)		•	•		
GAFF Lennard-Jones C_6 contribution (*GAFF LJ Disp*)			•		
EFP2 polarization (*EFP2 Pol*)				•	
EFP2 charge transfer (*EFP2 CT*)					•
Total EFP2 (*EFP2 Tot*)	•	•	•	•	•
Combinations					
(*EFP2 El + EFP2 Rep*)	•	•			
(*EFP2 Rep + EFP2 Disp*)		•	•		
(*EFP2 El + EFP2 Rep + EFP2 Pol + EFP2 CT*)	•	•		•	•
(*EFP2 El + EFP2 Rep + EFP2 Disp*)	•	•	•		
(*CHELPG QM/MM + GAFF LJ*)	•	•	•		
(*CHELPG QM/MM + EFP2 Rep + EFP2 Disp*)	•	•	•		
(*Embed + D3 Disp*)	•	•	•		
(*Embed + EFP2 Disp*)	•	•	•		

with DFT embedding is the treatment of kinetic energy contributions appearing in the interaction term, which are not additive over the subsystems. This non-additivity can be eliminated by enforcing orthogonality between the orbitals of the subsystems. Several methods, including projector based approaches like the level shift method of Manby et al.[65] and the Huzinaga equation[66] or basis set orthogonalization[67] can be applied to this end. The embedding scheme based on the Huzinaga approach by Hégely and co-workers[68,69] is especially appealing since it provides a simple and formally exact framework. Also, for the active site(s) DFT can be replaced by wave function (WF) methods, allowing eventually the treatment of excited states. This technique employs a subtractive embedding scheme where a DFT calculation is performed on the entire system, which is then divided into an active and an environment part. The energy of an embedding calculation of this type can be given as

$$E_{\text{tot}}^{\text{WF−in−DFT}} = E_{\text{tot}}^{\text{DFT}} − E_{\text{A}}^{\text{DFT}} + E_{\underset{\sim}{\text{A}}}^{\text{WF}}, \qquad (3)$$

where $E_{\underset{\sim}{\text{A}}}^{\text{WF}}$ is calculated in the presence of the environment's density, including thereby the electrostatic interaction of the subsystems. Note that the exchange interaction between the two subsystems is retained from the original super system calculation and included in the final energy. However, since the underlying DFT calculation on the

super system does not describe dispersion correctly, this contribution needs to be added afterwards.

In this work we test the two main categories of methods for describing intermolecular interactions. First, *a posteriori* corrections sourced from representative force fields including EFP2, the latter considered as a system-specific FF. Second, embedding techniques, which, to our knowledge, were not yet studied systematically for intermolecular interactions beyond QM/MM. As comparison, SAPT interaction terms are also evaluated for the two stacked complexes. Note, however, that SAPT can only be used as a ground state benchmark since it is not yet set to be generally used for excited states, although one should acknowledge a recent promising effort by Hapka et al.[70]

3 | INTERACTION POTENTIALS

As outlined above, we want to test individual contributions to intermolecular non-covalent energy, as treated in various methods. The aim is to find a combination of contributions, which is accurate enough for ground states, can be combined with various QM methods, and eventually can also be extended to excited states. Our approach is pragmatic in the sense that instead of seeking general conclusions, we intend to identify the best potentials/procedures to

describe complexes of N-heterocycles at the vicinity of their equilibrium structures. Table 1 summarizes the methods we are going to explore.

4 | COMPUTATIONAL DETAILS

4.1 | Ab initio calculations

The reference interaction energies were calculated at the CCSD(T)/aug-cc-pVDZ level[71] applying counterpoise (CP) correction[72] for basis set superposition error (BSSE). The choice of this relatively small basis set seems justified by previous findings of Sinnokrot et al.[73]: for stacked interactions the aug-cc-pVDZ basis, if used with CP correction, performs well as compared to larger basis set results. The core electrons were frozen in the correlation treatment.

The QM/MM calculations were performed at the CCSD/aug-cc-pVDZ level using the CFOUR[74,75] program package, with the MM point charges incorporated into the one-electron Hamiltonian. The FF point charges, as well as the parameters of the Lennard-Jones potential were taken from the General Amber Force Field (GAFF).[20] The relevant FF parameters are listed in the Tables S1–S3 of the Supplementary material.

The determination of CHELPG point charges for the QM/MM calculations was done at the CCSD/aug-cc-pVDZ level, using GAMESS.[63] The electrostatic potential of each molecule was calculated on a set of grid points around the molecule, with a step size of 0.8 Å and the most distant point from any atom set to 3 Å. The point charges on the atoms were fitted to this potential, with the additional requirement to reproduce the dipole and quadrupole moments of the molecule. The CHELPG atomic charges are also found in Tables S1–S3 of the Supplementary material.

The EFP2 interaction potentials were calculated using the GAMESS built-in routine with default settings. The parameters include the fitted multipole moments up to octupoles, exponential damping functions for charge penetration effects, as well as dipole polarizabilities and unique sets of parameters for Pauli repulsion, dispersion and charge transfer. All of these were obtained from a restricted Hartree–Fock calculation on the individual molecules using the 6–311++G(3df,2p) basis set. The latter choice follows Slipchenko and Gordon[76] in their study on the benzene dimer with different orientations.

For the DFT/WF embedding calculations we used the Huzinaga embedding scheme of Hégely and co-workers.[68,69] In this approach a KS-DFT calculation is performed first on the entire system, which is then divided into an active part and the environment. The orbitals are localized on the subsystems and those on the active subsystem are reoptimized by solving the Huzinaga equation, the latter including the interaction with the environment represented by its occupied orbitals. These resulting orbitals were subsequently used in the wave function (WF) calculation on the active subsystem. The localization and partitioning of the orbital spaces plays an important role in this scheme: not only does it separate the occupied orbitals of the active subsystem and the environment, but the localization of virtuals also truncates the supersystem's orbital space to the orbital space of the active

subsystem, thus lowering the computational cost of the ab initio electron correlation calculations. This localization is done using the Subsystem Projected AO Decomposition (SPADE)[77] approach, which works in a black-box fashion, requiring only the atoms to be assigned to the respective subsystems.

These embedding calculations were performed with the MRCC program system,[78,79] using again the aug-cc-pVDZ basis set. For the KS-DFT part of the calculations the PBE[80] functional was employed, with the subsequent WF calculation on the active subsystem done at the CCSD level (frozen core).

The PBE-D3 dispersion correction[53] was evaluated using the DFTD3 program with Becke–Johnson (BJ) damping.[81] Only the intermolecular part of the dispersion was used; the relevant parameters are given in Tables S11 and S12 of the Supplementary material.

SAPT calculations have been performed using the SAPT2020 program[42,82] connected with ATMOL1024,[83] with the aug-cc-pVDZ basis set used on the fragments. Selecting the SAPT=T keyword, all contributions up to second order in the interaction potential[1] have been included in the total interaction energy, using correlation amplitudes from CCSD calculations performed on the fragments. The assignment of the various terms of SAPT to usual components (electrostatic, dispersion, Pauli exchange) has been done according to Equations (92)–(95) in Reference.[84]

For non–symmetric complexes, that is when the two monomer molecules are not equivalent, alternate choices for active/inactive fragments (treated at high level, vs. lower level) are possible. The two choices give minor differences in the QM/MM and embedding results and their arithmetic mean was accepted as the final result.

4.2 | Molecular systems

The test systems under study are illustrated in Figure 1. In the discussion below, the distance of monomers refers to that between the centers of masses.

For the pyrrole dimers six conformers were studied. Three of them are co-planar (side to side orientations), denoted *Pyr-Pyr (ILn)*. In *Pyr-Pyr (IL1)* the N–H bonds point in the same direction, with point group symmetry of C_{2v}; conformers *Pyr-Pyr (IL2)* and *Pyr-Pyr (IL3)* have D_{2h} symmetry, with the N-H groups pointing inward and outward, respectively. In addition, two T-shaped structures (*Pyr-Pyr (T1)*) and (*Pyr-Pyr (T2)*) (C_1 point group) and one stacked (*Pyr-Pyr (S)*) conformer was considered. In the latter, the fragments' dipoles point essentially in the opposite direction, but the symmetry was lowered by a 10° in-plane rotation away from the C_{2h} structure. The cytosine-uracil complex was investigated in a stacked (sandwich) setup (*Cyt-Ura (S)*), with oppositely oriented dipole moments.

The seven cases cover a broad spectrum of the nature of intermolecular interactions in the nitrogen heterocycles studied, with a large variety of equilibrium separations. We thus aspire to draw generic conclusions on the description of complexes of this type.

The equilibrium structures of the monomers, optimized at the MP2/6-31G* level, were taken from Ref. 85 and are documented in Tables S1–S3 of the Supplementary material along with the

Pyr-Pyr (IL1)

Pyr-Pyr (IL2)

Pyr-Pyr (IL3)

Pyr-Pyr (T1)

Pyr-Pyr (T2)

Pyr-Pyr (S)

Cyt-Ura (S)
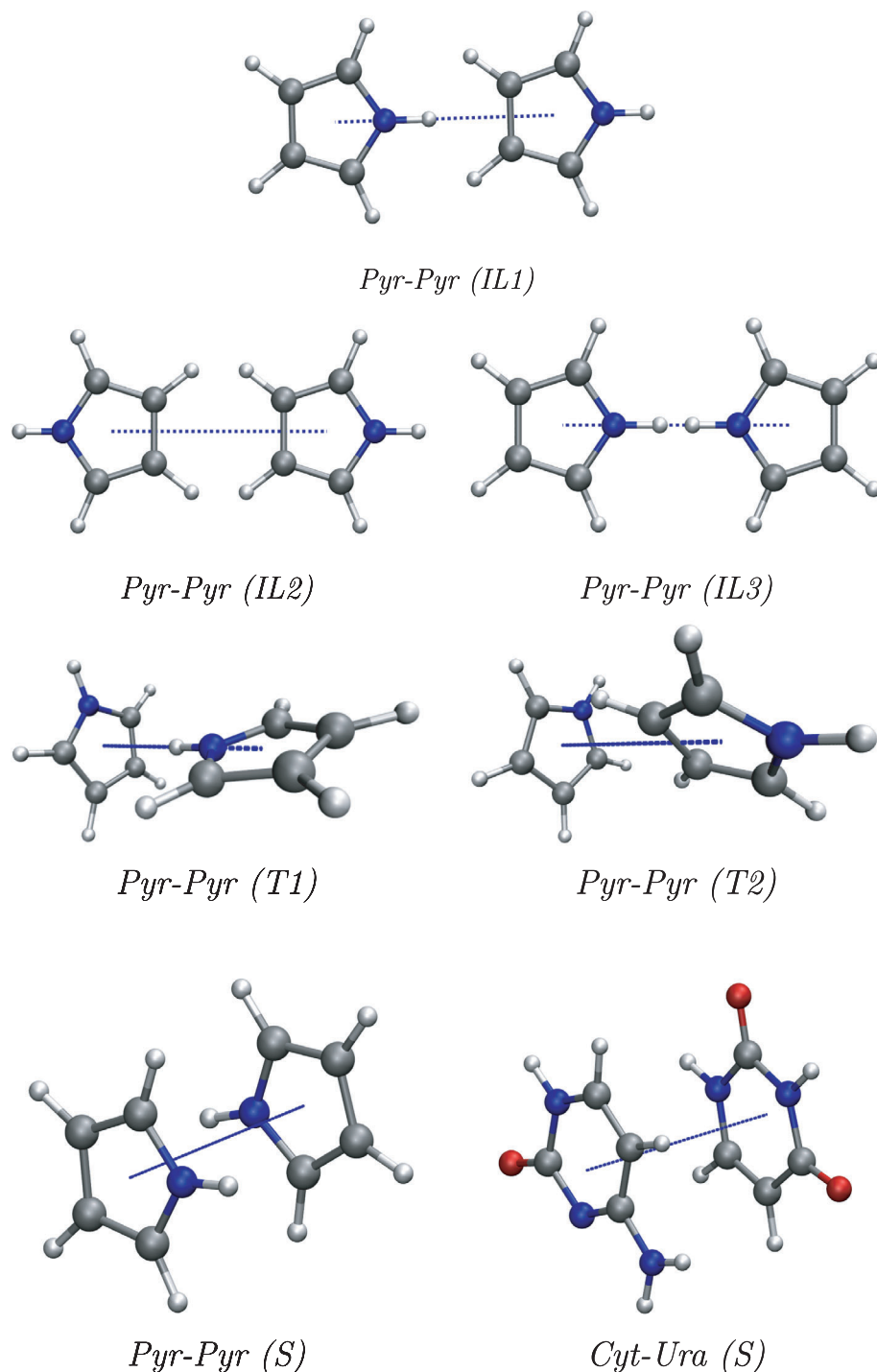
geometries of dimers at a representative distance in Tables S4–S10 of the Supplementary material.

5 | RESULTS AND DISCUSSION

5.1 | Dispersion and Pauli repulsion

In Figure 2 the various dispersion and combined dispersion plus Pauli repulsion potentials are compared. As of the dispersion potentials, the

EFP2 (*EFP2 Disp*, black), the D3 (*D3 Disp*, magenta) and the C_6 part of GAFF (*GAFF LJ Disp*, green) are considered, while for dispersion plus Pauli repulsion, the EFP2 ([(*EFP2 Rep* + *EFP2 Disp*)], red) and Lennard-Jones potential of GAFF (*GAFF LJ*, blue) are compared. For the pure dispersion, the EFP2 gives significantly steeper curves than D3, and GAFF is closer to the first one of the two. Consequently, for systems with a larger distance between the closest atoms of the different fragments, EFP2 and D3 will give similar results, while for systems with shorter separation the D3 dispersion will be smaller, resulting in less attractive or more repulsive total interaction curves (see later). In our
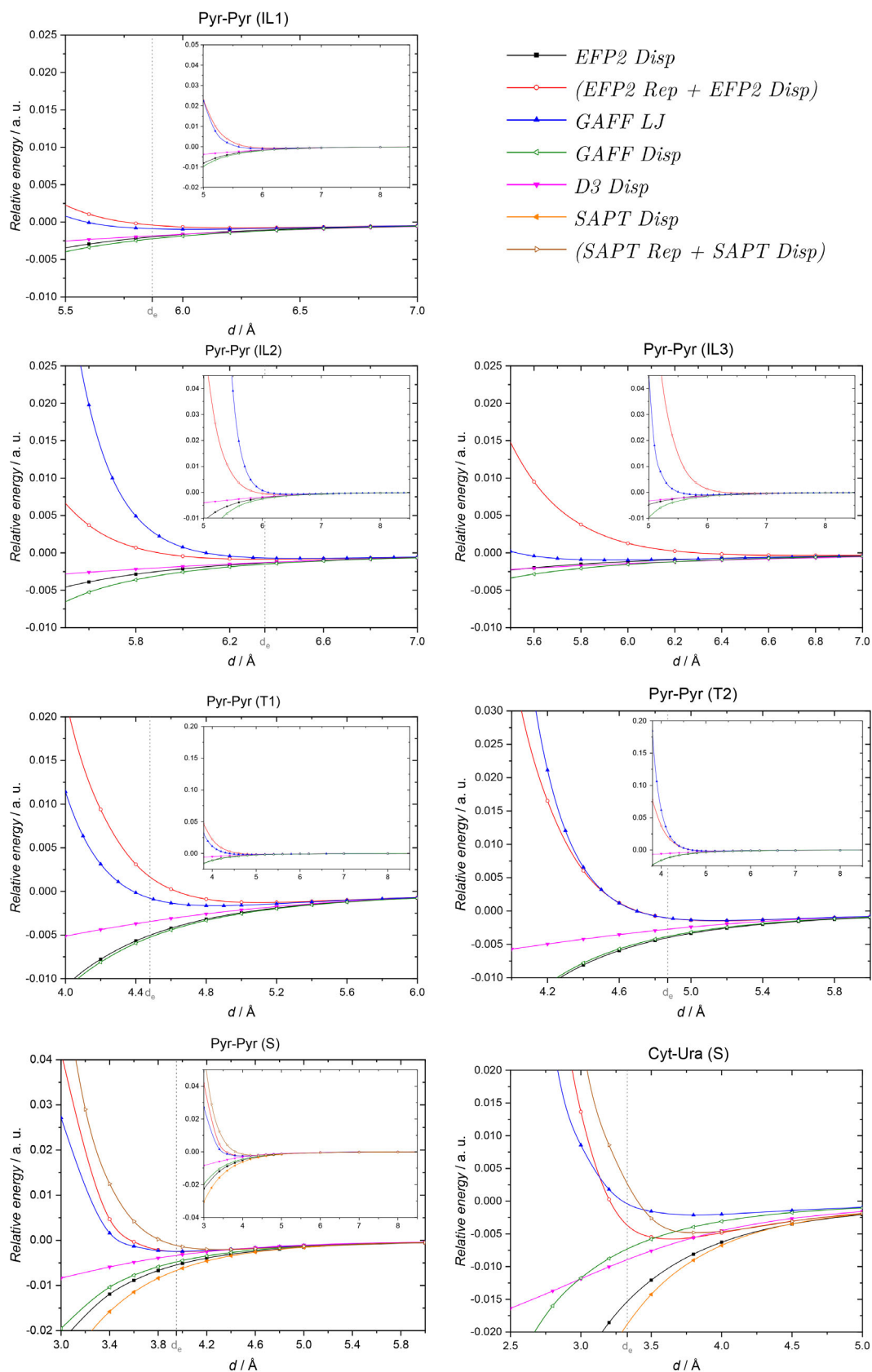
FIGURE 2 Distance dependence of various potentials describing dispersion and Pauli-repulsion contribution of the interaction energy. d_e marks the reference equilibrium distance. For *Cyt-Ura (S)*, due to technical difficulties, the SAPT dispersion curve only includes the $E_{disp}^{(20)}$ contribution, which results in a slight overestimation of the magnitude of this term (see Figure S1 of the Supplementary material.)

case this means that the choice between D3 and EFP2 is less relevant for the co-planar structures, while for the T-shaped and sandwich configurations the difference can go up to 0.002 a.u. or even larger (0.01 a.u.), as in the case of *Cyt-Ura (S)*.

The dispersion plus Pauli repulsion potential of EFP2 (*EFP2 Rep +- EFP2 Disp*) corresponds, as also seen in Table 1, to the effects covered by the GAFF Lennard-Jones potential. The behavior of the *GAFF LJ* potential (blue curves on Figure 2) with respect to the corresponding EFP2 curves (red curve) is not systematic. While EFP2 is generally more repulsive, for *Pyr-Pyr (IL2)* the opposite is true. Although the general shape of the curves appears to be similar, substantial energy differences can occur even around the equilibrium distances of some structures. For the *Pyr-Pyr (IL1)* and *Pyr-Pyr (IL 2)*, as well as for *Pyr-Pyr (S)* the difference is very small at the corresponding equilibrium distance, becomes, however, substantial on the repulsive side of the curve. In contrast, for *Pyr-Pyr (IL3)* and *Pyr-Pyr (T 1)* where the proton attached to the N atom is close to the other molecule, the discrepancy is remarkably large. *Cyt-Ura (S)* is special in the sense that the two curves cross each other: at the equilibrium separation the (*EFP2 Rep + EFP2 Disp*) shows a relatively strong binding (0.005 a.u.) compared to the other examples in this study, while the *GAFF LJ* curve is less bound with an interaction energy closer to that of other structures (−0.002 a.u.). This is because for this system the dispersion contribution in *GAFF LJ* is small, similar to D3, at least near equilibrium (see the corresponding curves in Figure 2). This finding warrants the favorization of EFP2 potentials over Lennard-Jones type ones for the dispersion contribution in similar systems. One has to keep in mind, however, that other force fields can possibly show a better performance than *GAFF LJ* parameters.

Figure 2 also includes SAPT results for the stacked systems, which support many of the conclusions above. Not only appears EFP2 dispersion as preferred over D3 but also when combined with the Pauli repulsion, the EFP2 follows the SAPT curve better than *GAFF LJ* does. Nevertheless, the repulsive nature kicks in faster with SAPT than with EFP2.

5.2 | Electrostatic interaction

In Figure 3 the distance dependence of the pure electrostatic contributions obtained from the QM/MM approach with GAFF (*GAFF QM/MM*, black), as well as CHELPG charges (*CHELPG QM/MM*, red) and the corresponding EFP2 part (*EFP2 El*, blue) are seen, along with the curve from the embedding calculation (which includes also Pauli repulsion, see above) (*Embed*, orange) and the EFP2 curve corresponding to it ([*EFP2 El + EFP2 Rep*], magenta). The figure also includes the EFP2 curves augmented with the polarization and CT contributions ([*EFP2 El + EFP2 Rep + EFP2 Pol + EFP2 CT*], green).

These latter terms do not turn out to be important in any case. In fact, with the exception of *Pyr-Pyr (IL1)* where their inclusion reduces the repulsive nature of the surface to some extent, the difference they make is practically negligible.

When considering the pure electrostatic contributions (*EFP2 El* and QM/MM curves) one should discuss the attractive and repulsive

cases separately. Attractive are *Pyr-Pyr (IL1)*, *Pyr-Pyr (T1)* and *Pyr-Pyr (T2)* as well as *Cyt-Ura (S)* curves. In these cases, even if QM/MM electrostatic attraction is stronger at longer separations (*Pyr-Pyr (IL1)*, *Pyr-Pyr (T2)*, *Cyt-Ura (S)*) EFP2 always predicts stronger interaction at shorter distances. This latter effect is most probably due to the damping/screening effect built into EFP2 but not in our QM/MM approach. In the two cases possessing repulsive electrostatic interactions (*Pyr-Pyr (IL2)* and *Pyr-Pyr (S)*), the *EFP2 El* curve shows a pronounced turn-back at shorter distances, probably also caused by the damping/screening formula employed in this model. At larger distances, before this turn-back occurs, the *EFP2 El* interaction is considerably stronger than that predicted by the QM/MM models. The largest relative discrepancy between the electrostatic curves is observed for *Pyr-Pyr (S)*, but note that the electrostatic interaction here is very small compared to other cases. Still, this discrepancy will influence even qualitatively the total interaction curve (see later). For a better understanding of this effect, Figure 3 also shows the SAPT electrostatic curves (in brown) for the *Pyr-Pyr (S)* and *Cyt-Ura (S)* systems. These support the above finding: screening/damping is important, and its absence makes the QM/MM curves too repulsive or less attractive at short distances. On the other hand, for *Pyr-Pyr (S)* the EFP2 curve is overly repulsive at all distances, resulting in a quite bad total energy curve for this system (see below). Similar discrepancy is also observable for *Cyt-Ura (S)*, but it is less severe.

There is no substantial difference between the curves obtained in the QM/MM approach, using either CHELPG and GAFF point charges. Therefore, we chose CHELPG in our further studies as the parameters can be obtained from an ab initio calculation for the molecules in question.

Since the embedding model (*Embed*, orange curve of Figure 3) automatically includes the Pauli repulsion contribution (see above), the electrostatic part of the embedding cannot be compared directly to the pure electrostatic curves. Instead, comparisons have to be made to the (*EFP2 El + EFP2 Rep*) curves. In this regard, a very good agreement is observed for *Cyt-Ura (S)*, *Pyr-Pyr (T1)*, *Pyr-Pyr (T2)*, and *Pyr-Pyr (IL3)*, while for *Pyr-Pyr (IL1)* and *Pyr-Pyr (IL2)* the curves cross, indicating a different distance dependence. The worst agreement is observed for *Pyr-Pyr (S)* where embedding predicts a significantly weaker repulsion, which, in light of the discussion on the pure electrostatic component above, is likely to be closer to reality. Near the equilibrium, substantial (relative) difference is observed for the latter three arrangements.

For the two sandwich configurations, the comparison of the *Embed* (orange) curve to other electrostatic ones reveals the ultimate importance of Pauli repulsion, changing the electrostatic interaction qualitatively, even resulting in a minimum in case of the *Cyt-Ura (S)* complex.

5.3 | Total interaction energy

In Figure 4 the total interaction energy of the investigated complexes are shown. The black curve is the reference energy calculated for the
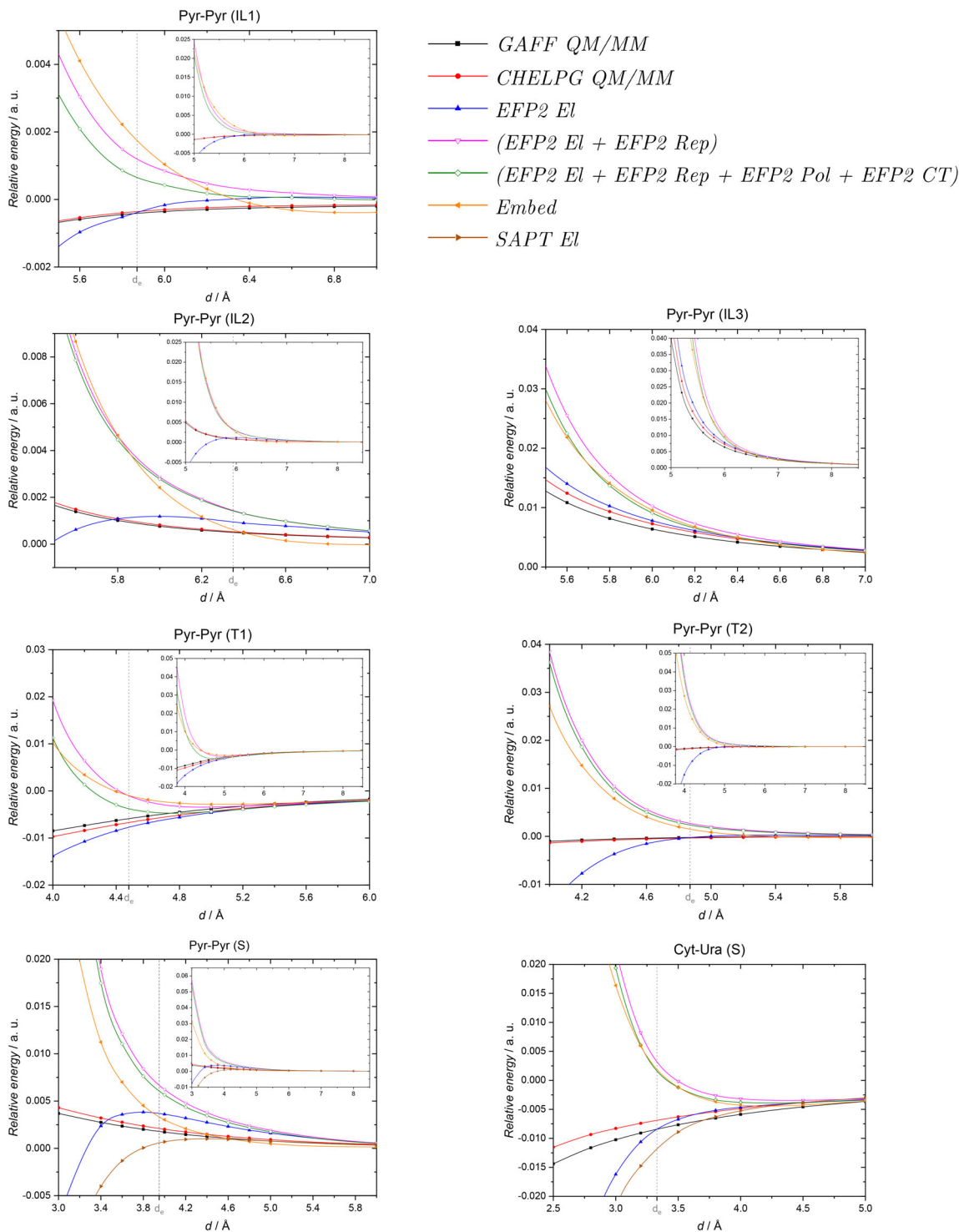
FIGURE 3 Distance dependence of various potentials describing electrostatic and Pauli-repulsion contribution of the interaction energy. d_e marks the reference equilibrium distance

complex using the CCSD(T) method and aug-cc-pVDZ basis with CP correction. The total EFP interaction energy *EFP2 Tot* is shown in red, the embedding curves augmented with two different dispersion corrections in green (*Embed + D3 Disp*) and orange (*Embed +- EFP2 Disp*), while the CHELPG QM/MM electrostatics, augmented with either the GAFF Lennard-Jones potential (*CHELPG QM/MM + GAFF LJ*) or the EFP2 dispersion and Pauli repulsion terms (*CHELPG QM/MM + EFP2 Rep + EFP2 Disp*), are displayed in blue and magenta, respectively. The substantial differences in the shape of the presented curves warrant the individual discussion of the results for each test system. Additional information for this analysis is given in Table 2 where the different contributions to the

total interaction energy are listed at the respective equilibrium distance of the complex under study.

Pyr-Pyr (IL1) structure: Despite the fact that the dipole moments of the two monomers are aligned in this system, there is a very weak interaction between the monomers. Contrary to e.g. the sandwich systems discussed below, here all three contributions to the interaction energy are small (see Table 2). Both the *EFP2 Tot* and the (*CHELPG QM/MM + GAFF LJ*) curves agree very well with the reference one, while the (*CHELPG QM/MM + EFP2 Rep + EFP2 Disp*), as well as the dispersion corrected embedding curves, show a weaker agreement. The latter ones, in particular, have a minimum at a considerably longer distance, although the interaction energy at the minimum is similar to that of the reference. Contrary to the cases discussed below, here a replacement of the D3 dispersion correction with the EFP2 counterpart results in a negligible improvement. On the other hand, by pairing the *CHELPG QM/MM* term with EFP2 dispersion and Pauli repulsion instead of the *GAFF LJ*, the agreement with the reference curve becomes worse and the result is closer to the embedding cases.

Pyr-Pyr (IL2) structure: The total interaction energy is even smaller in this structure than in *Pyr-Pyr (IL1)*. This is partly because the electrostatic part is repulsive, thus compensating dispersion. Here the best agreement with the reference is provided by (*Embed + EFP2 Disp*), although the minimum is at a somewhat larger distance. The (*Embed + D3 Disp*) combination provides a better location of the minimum point, but the difference is not relevant, and the short-distance repulsion is clearly overestimated by this model. The *EFP2 Tot* potential remains repulsive throughout the entire curve, thus failing to agree even qualitatively with the reference. The (*CHELPG QM/MM + GAFF LJ*) model is less bound than the reference, showing just the half of the interaction energy, however, the position of the minimum reasonably agrees with the (*Embed + EFP2 Disp*) combination. Note that this is the system where a large discrepancy between the *GAFF LJ* and the (*EFP2 Rep + EFP2 Disp*) curves was observed above, the former being more repulsive.

Pyr-Pyr (IL3) structure: The total interaction curve for this system is repulsive in the investigated region, which is being reproduced by all methods. The best agreement with the reference is provided by the (*CHELPG QM/MM + GAFF LJ*) combination. However, contrary to all other cases where the short-distance repulsion is clearly overestimated, this curve crosses the reference around 6 Å separation and shows a considerably weaker repulsion below this point. This is in line with the finding that the *GAFF LJ* curve considerably underestimates the corresponding EFP2 curve (see Figure 2). The (*Embed + EFP2 Disp*) and (*Embed + D3 Disp*) combinations, as well as *EFP2 Tot* all agree very well, showing a similar error of parallelity to the reference: note that in this case the dispersion parts of the interactions were also found to agree above.

Pyr-Pyr (T1) structure: The total interaction energy in this system is the second largest among the investigated structures, with a strong attractive electrostatic component (see Table 2). (Note that since the two dipoles are perpendicular to each other, this part should be zero in the dipole approximation.) The absolute values of the Pauli

repulsion and dispersion are also sizeable. The *EFP2 Tot* curve shows the best agreement with the reference, slightly overestimating the binding energy. The (*CHELPG QM/MM + GAFF LJ*) curve also comes close to the reference one and underestimates the binding by just 15%, outperforming the (*CHELPG QM/MM + EFP2 Rep + EFP2 Disp*) combination. This deviation is clearly caused by the discrepancy of the *GAFF LJ* potential and the (*EFP2 Rep + EFP2 Disp*) terms discussed in Section 5.1. With the combination of embedding and D3 (*Embed + D3 Disp*) the interaction energy is too weak by about 50%, a large part of which can be corrected by replacing the D3 term with the EFP2 dispersion (*Embed + EFP2 Disp*).

Pyr-Pyr (T2) structure: The interaction in this structure is dominated by the dispersion, while the Pauli repulsion is mostly compensated by the electrostatic attraction. A good agreement is seen between most approximate methods, but they give only about half of the interaction energy. Embedding can again be improved by replacing D3 with EFP2 dispersion, resulting in an almost perfect agreement with the reference. Since *GAFF LJ* and (*EFP2 Rep + EFP2 Disp*) were found to agree very well in this case (see Section 5.1), the *CHELPG QM/MM* result cannot be improved considerably by changing one to the other.

Pyrrole-pyrrole sandwich(Pyr-Pyr (S)): Describing the interaction in this system is a special challenge for theoretical methods.[86] Near the equilibrium, dispersion is the strongest contribution to the interaction energy (-5 mE$_h$), but it is largely canceled by the Pauli repulsion and the (also repulsive) electrostatics, resulting in a very small total interaction energy. It is thus an example where good result can only be achieved with a method where a delicate balance between the different contributions is present. Indeed, as Figure 4 shows, there is a substantial discrepancy between the two embedding curves, with the one including D3 (*Embed + D3 Disp* 3) being less bonding. It is (*Embed + EFP2 Disp*) which has the smaller absolute error (also the overall shape of the curve being similar to the reference), while the former is similar to both QM/MM curves. The agreement of the QM/MM curves is a consequence of the similarity of the *GAFF LJ* and (*EFP2 Rep + EFP2 Disp*) potentials, discussed earlier in Section 5.1. Still, these methods yield only about half of the reference total interaction energy. Contrary to other cases, the *EFP2 Tot* curve is even qualitatively incorrect, showing a constantly repulsive intermolecular interaction. This unexpected behavior can be attributed to the overly strong electrostatic repulsion provided by the EFP2 method. Both the weaker performance of the QM/MM approaches and the failure of the EFP2 method was understood when comparing the electrostatic contributions to SAPT in 5.2. Speaking of the latter, Figure 4 also shows the corresponding curve in brown: the excellent agreement of (*Embed + EFP2 Disp*) with the SAPT curve is very reassuring. This system is thus an illustration of the extraordinary difficulty of modeling weak interaction in stacked π-π complexes by approximate methods. In this regard, the good performance of the combination of embedding and EFP2 dispersion is remarkable.

Cytosine-uracil complex(Cyt-Ura (S)): Among the studied systems, this complex shows the strongest interaction between the two
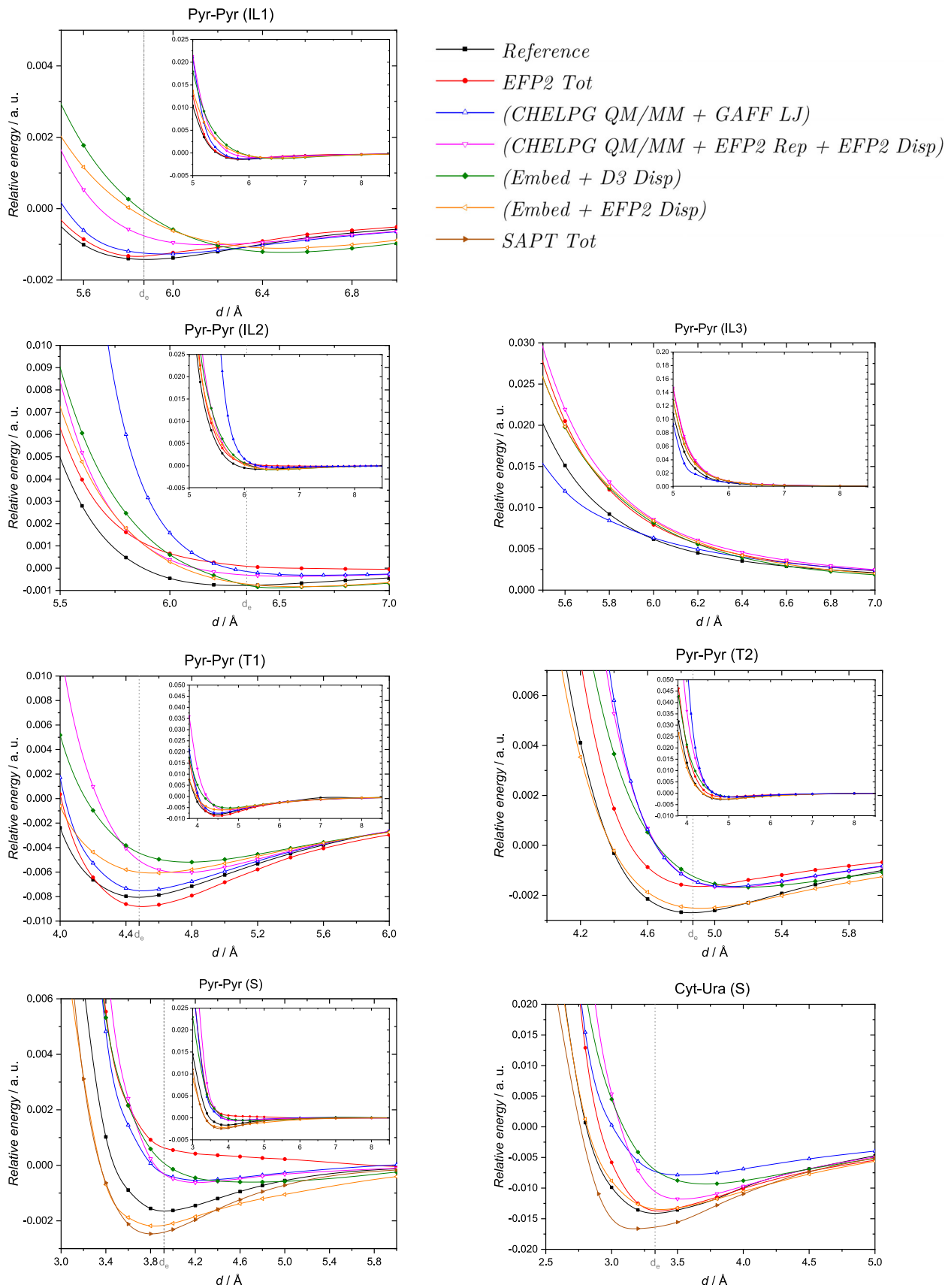
FIGURE 4   Total interaction energy of the investigated complexes calculated at various levels of theory. d_e marks the reference equilibrium distance. For *Cyt-Ura (S)*, due to technical difficulties, the symmetry-adapted perturbation theory curve only includes the $E_{disp}^{(20)}$ dispersion contribution which results in a slight overestimation of the attraction between the two molecules (see Figure S1 of the Supplementary material.)

	Pyr-Pyr						Cyt-Ura
	(IL1)	(IL2)	(IL3)	(T1)	(T2)	(S)	(S)
d_e/a.u.	5.9	6.3	(6.0)[a]	4.5	4.9	3.9	3.3
Embed	1.7	0.6	9.6	−1.0	1.5	3.3	1.8
CHELPG QM/MM	−0.4	0.5	7.3	−6.8	−0.4	2.1	−6.9
EFP2 El	−0.4	0.9	7.8	−7.7	−0.3	3.7	−8.4
EFP2 Rep	1.6	0.5	2.5	6.6	2.9	3.0	11.6
EFP2 Disp	−2.0	−1.4	−1.2	−5.0	−4.0	−5.4	−15.3
D3 Disp	−1.8	−1.4	−1.4	−3.4	−2.7	−3.4	−9.0
GAFF LJ	−0.9	−0.7	−0.9	−0.8	−1.1	−2.4	−0.4
(EFP2 Rep + EFP2 Disp)	−0.4	−0.9	1.3	1.6	−1.1	−2.4	−3.7
EFP2 Tot	−1.3	0.1	7.9	−8.8	−1.6	0.6	−13.7
(Embed + EFP2 Disp)	−0.2	−0.7	8.4	−6.0	−2.5	−2.2	−13.4
(Embed + D3 Disp 3) 3	−0.1	−0.8	8.2	−4.5	−1.2	0.1	−7.0
Reference	−1.4	−0.8	6.2	−8.1	−2.7	−1.7	−14.1

TABLE 2 Energy contributions (in mE_h) of different terms and combinations at the reference equilibrium point d_e

[a]Selected point as there is no minimum in the investigated range.

molecules ($−14$ mE_h). Similarly to *Pyr-Pyr (S)*, the interaction is dominated by the dispersion ($−12$ mE_h) which is about twice as large as the size of the electrostatic and Pauli repulsion. Contrary to *Pyr-Pyr (S)* however, the electrostatic part is attractive, compensating the Pauli repulsion almost entirely. In comparison to the reference curve, the *EFP2 Tot* as well as (*Embed* + *EFP2 Disp*) potentials turn out to be the best choices. At longer distances also the (*CHELPG QM/MM* + *EFP2 Rep* + *EFP2 Disp*) curve comes close to the reference, but the repulsion shown at the short-distance side of the minimum is clearly too strong due to the missing penetration contribution (see Section 5.2). It is thus seen that for this system the ab initio methods using EFP2 components clearly outperform those with D3 or LJ corrections.

In summary, despite the very different bonding situations in the model systems, most of the methods investigated here give reasonable results.

The *EFP2 Tot* interaction curves agree very well with the reference curve in most cases, however this model fails even qualitatively for the *Pyr-Pyr (S)* and *Pyr-Pyr (IL2)* systems, which are loosely bound by other methods but not with *EFP2 Tot*.

CHELPG QM/MM, with either the *GAFF LJ* or the (*EFP2 Rep* +- *EFP2 Disp*) term gives similar results for the stacked systems and *Pyr-Pyr (T2)*, while in other cases the curves with *GAFF LJ* correction are substantially closer to the reference curves, except for *Pyr-Pyr (IL2)* where *GAFF LJ* fails badly. Typically, the bonding provided by this combination is a bit weaker than in the reference, representing a quality comparable to *EFP2 Tot* and somewhat less accurate than embedding methods.

The dispersion corrected embedding curves (*Embed* + *EFP2 Disp*) and (*Embed* + *D3 Disp*) are in most cases closer to the reference than other methods investigated here. From these two, the former with the EFP2 dispersion clearly gives the best agreement, outperforming the latter, which uses D3 dispersion.

6 | CONCLUSIONS

Various approximate methods for the description of non-covalent intermolecular interactions were investigated on certain conformers of pyrrole-pyrrole and cytosine-uracil complexes as test systems. Specifically, the significance of various terms in the models and combinations thereof were analyzed. Despite the approximate, in some sense arbitrary definition of some contributions, we confirmed their general importance. In fact, it was found that all contributions of a particular model need to be included even for *qualitatively* correct potential energy curves.

With future studies in mind, we paid special attention to models that can be combined with any electronic structure methods. In this respect, QM/MM type formulations with point charges are ideal since they can easily be implemented in the one-electron Hamiltonian, without affecting the execution of any level of calculation. However, despite its widespread, successful applications to, e.g., solvent effects, QM/MM is not suited *per se* for weakly bonded dimers. In our test systems, other terms, such as dispersion or Pauli repulsion give non-negligible, often dominating contributions. Thus, the electrostatic part needs to be augmented with the latter. Most force fields contain a Lennard-Jones type potential for this purpose; our calculations indicate, however, that more sophisticated methods, e.g. the Effective Fragment Potential technique, may be needed.

At the same time, one has to keep in mind that different models address the division of interaction into individual terms by different strategies. Therefore, it is not surprising that combinations of different approaches tend to give inconsistent or controversial results. Still, we found certain setups of promising performance, approaching the reference CCSD(T) and, where available, also the SAPT results. In particular, a DFT/CCSD embedding model covering electrostatic and Pauli repulsion effects, if augmented with dispersion via the respective EFP2 term, proved to be a reasonable

working strategy. It has shown remarkable accuracy specifically for the stacked interactions. A QM/MM style electrostatics using CHE-LPG charges, with the point charges determined at the partner molecule's nuclei is also a viable alternative to the widely adopted force field charges, also liberating the description from the dependence on precalculated parameters and atom types. Combination with either the EFP2 dispersion plus Pauli repulsion or with a Lennard-Jones potential of the GAFF force field gave similar results, the latter being slightly better in some examples; but the latter finding is probably specific to the choice of this particular FF and may not be achievable for systems with different atom types. Also, the LJ contribution is disturbingly random: depending on the specific arrangement, it may severely overestimate, in other cases underestimate the interaction. Thus, its application would be difficult to support and we feel that the idea of a general, system-adopted description provided by EFP2 is comforting enough to be favored over the LJ approach of any force field.

The D3 dispersion correction, a strategy also founded on transferable atomic parameters, was clearly outperformed by EFP2 in our results (This finding is, of course, not related to the original applicability of D-type corrections in DFT).

In summary, among the many alternative formulations arising from the synthesis of different theories we have found combinations, which not only offer good performance, but also retain the generality and theoretical desirability of ab initio modeling. For future studies, especially our planned work on *excited* states, the sophisticated description offered by DFT/CCSD embedding and the easily implementable QM/MM approach with CHELPG charges, both augmented with missing terms from EFP, are the recommended tools.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplementary material of this article

ORCID

Bónis Barcza https://orcid.org/0000-0002-7021-7665
Ádám B. Szirmai https://orcid.org/0000-0003-2272-5574
Katalin J. Szántó https://orcid.org/0000-0003-4767-0987
Attila Tajti https://orcid.org/0000-0002-7974-6141
Péter G. Szalay https://orcid.org/0000-0003-1885-3557

REFERENCES

[1] B. Jeziorski, R. Moszynski, K. Szalewicz, *Chem. Rev.* **1994**, *94*, 1887.
[2] K. Szalewicz, *WIREs Comput. Mol. Sci.* **2012**, *2*, 254.
[3] P. N. Day, J. H. Jensen, M. S. Gordon, S. P. Webb, W. J. Stevens, M. Krauss, D. Garmer, H. Basch, D. Cohen, *J. Chem. Phys.* **1996**, *105*, 1968.
[4] K. Morokuma, *J. Chem. Phys.* **1971**, *55*, 1236.
[5] Y. Mo, P. Bao, J. Gao, *Phys. Chem. Chem. Phys.* **2011**, *13*, 6760.
[6] E. D. Glendening, A. Streitwieser, *J. Chem. Phys.* **1994**, *100*, 2900.
[7] E. D. Glendening, *J. Am. Chem. Soc.* **1996**, *118*, 2473.
[8] R. Z. Khaliullin, E. A. Cobar, R. C. Lochan, A. T. Bell, M. Head-Gordon, *J. Phys. Chem. A* **2007**, *111*, 8753.
[9] P. R. Horn, E. J. Sundstrom, T. A. Baker, M. Head-Gordon, *J. Chem. Phys.* **2013**, *138*, 134119.
[10] H. M. Senn, W. Thiel, *Angew. Chem. Int. Ed.* **2009**, *48*, 1198.
[11] R. Eisenschitz, F. London, *Z. Phys.* **1930**, *60*, 491.
[12] W. Heitler, F. London, *Z. Phys.* **1927**, *44*, 455.
[13] H. Margenau, N. R. Kestner, *Theory of Intermolecular Forces*, Pergamon Press, Oxford **1969**.
[14] A. J. Stone, *The Theory of Intermolecular Forces*, OUP Oxford, Oxford **2013**.
[15] M. S. Gordon, J. H. Jensen, *Wavefunctions and Chemical Bonding*, Interpretation (Wiley and Sons Inc., Online **2002**, p. 3198-3214.
[16] P. Arrighini, *Intermolecular Forces and Their Evaluation by Perturbation Theory, Lecture Notes in Chemistry*, Springer, Berlin Heidelberg, New York **2012**.
[17] I. Kaplan, *Intermolecular Interactions: Physical Picture, Computational Methods and Model Potentials, Wiley Series in Theoretical Chemistry*, Wiley, Chichester, West Sussex, England **2006**.
[18] A. Leach, *Molecular Modelling: Principles and Applications*, Prentice Hall, Harlow, Essex, England **2001**.
[19] J. W. Ponder, D. A. Case, *Protein Simulations* of *Advances in Protein Chemistry*, Vol. 66, Academic Press, San Diego, CA **2003**, p. 27-85.
[20] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157.
[21] B. R. Brooks, C. L. Brooks III, A. D. Mackerell Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, M. Karplus, *J. Comput. Chem.* **2009**, *30*, 1545.
[22] M. Janeček, P. Kührová, V. Mlýnský, M. Otyepka, J. Šponer, P. Banáš, *J. Chem. Theory Comput.* **2021**, *17*, 3495.
[23] C. M. Breneman, K. B. Wiberg, *J. Comput. Chem.* **1990**, *11*, 361.
[24] A. Stone, *Chem. Phys. Lett.* **1981**, *83*, 233.
[25] M. S. Gordon, Q. A. Smith, P. Xu, L. V. Slipchenko, *Annu. Rev. Phys. Chem.* **2013**, *64*, 553.
[26] T. Sattasathuchana, P. Xu, M. S. Gordon, *J. Phys. Chem. A* **2019**, *123*, 8460.
[27] R. M. Minikis, V. Kairys, J. H. Jensen, *J. Phys. Chem. A* **2001**, *105*, 3829.
[28] J. W. Ponder, C. Wu, P. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio, et al., *J. Phys. Chem. B* **2010**, *114*, 2549.
[29] G. Cisneros, T. Darden, N. Gresh, J. Pilmé, P. Reinhardt, O. Parisel, J.-P. Piquemal, *Design of next generation force fields from AB initio computations: Beyond Point Charges Electrostatics*, Springer, Netherlands, Dordrecht **2009**, p 137-172.
[30] J.-P. Piquemal, G. A. Cisneros, P. Reinhardt, N. Gresh, T. A. Darden, *J. Chem. Phys.* **2006**, *124*, 104101.
[31] N. Gresh, G. A. Cisneros, T. A. Darden, J.-P. Piquemal, *J. Chem. Theory Comput.* **2007**, *3*, 1960.
[32] M. Swart, P. van Duijnen, *Mol. Simul.* **2006**, *32*, 471.

[33] C. D. Sherrill, B. G. Sumpter, M. O. Sinnokrot, M. S. Marshall, E. G. Hohenstein, R. C. Walker, I. R. Gould, *J. Comput. Chem.* **2009**, *30*, 2187.

[34] T. M. Parker, E. G. Hohenstein, R. M. Parrish, N. V. Hud, C. D. Sherrill, *J. Am. Chem. Soc.* **2013**, *135*, 1306.

[35] M. A. Freitag, M. S. Gordon, J. H. Jensen, W. J. Stevens, *J. Chem. Phys.* **2000**, *112*, 7300.

[36] C. Liu, J.-P. Piquemal, P. Ren, *J. Chem. Theory Comput.* **2019**, *15*, 4122.

[37] Q. Wang, J. A. Rackers, C. He, R. Qi, C. Narth, L. Lagardere, N. Gresh, J. W. Ponder, J.-P. Piquemal, P. Ren, *J. Chem. Theory Comput.* **2015**, *11*, 2609.

[38] J. A. Rackers, Q. Wang, C. Liu, J.-P. Piquemal, P. Ren, J. W. Ponder, *Phys. Chem. Chem. Phys.* **2017**, *19*, 276.

[39] J.-P. Piquemal, N. Gresh, C. Giessner-Prettre, *J. Phys. Chem. A* **2003**, *107*, 10353.

[40] A. S. Werneck, T. M. R. Filho, L. E. Dardenne, *J. Phys. Chem. A* **2008**, *112*, 268.

[41] B. Wang, D. G. Truhlar, *J. Chem. Theory Comput.* **2010**, *6*, 3330.

[42] J. Garcia, R. Podeszwa, K. Szalewicz, *J. Chem. Phys.* **2020**, *152*, 184109.

[43] J. E. Lennard-Jones, *Proc. Phys. Soc.* **1931**, *43*, 461.

[44] J. A. Rackers, J. W. Ponder, *J. Chem. Phys.* **2019**, *150*, 084104.

[45] T. A. Halgren, *J. Am. Chem. Soc.* **1992**, *114*, 7827.

[46] J.-P. Piquemal, H. Chevreau, N. Gresh, *J. Chem. Theory Comput.* **2007**, *3*, 824.

[47] J. H. Jensen, M. S. Gordon, *Mol. Phys.* **1996**, *89*, 1313.

[48] J. H. Jensen, M. S. Gordon, *J. Chem. Phys.* **1998**, *108*, 4772.

[49] D. D. Kemp, J. M. Rintelman, M. S. Gordon, J. H. Jensen, *Theor. Chem. Accounts* **2010**, *125*, 481.

[50] Q. A. Smith, K. Ruedenberg, M. S. Gordon, L. V. Slipchenko, *J. Chem. Phys.* **2012**, *136*, 244107.

[51] S. Grimme, *J. Comput. Chem.* **2004**, *25*, 1463.

[52] S. Grimme, *J. Comput. Chem.* **2006**, *27*, 1787.

[53] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.

[54] A. S. P. Gomes, C. R. Jacob, *Annu. Rep. Prog. Chem., Sect. C: Phys. Chem.* **2012**, *108*, 222.

[55] J. Tomasi, B. Mennucci, R. Cammi, *Chem. Rev.* **2005**, *105*, 2999.

[56] D. Bakowies, W. Thiel, *J. Phys. Chem.* **1996**, *100*, 10580.

[57] Q. Sun, G. K.-L. Chan, *Acc. Chem. Res.* **2016**, *49*, 2705.

[58] C. I. Viquez Rojas, J. Fine, L. V. Slipchenko, *J. Chem. Phys.* **2018**, *149*, 094103.

[59] J. H. Jensen, *J. Chem. Phys.* **2001**, *114*, 8775.

[60] L. V. Slipchenko, M. S. Gordon, K. Ruedenberg, *J. Phys. Chem. A* **2017**, *121*, 9495.

[61] J. Řezáč, K. E. Riley, P. Hobza, *J. Chem. Theory Comput.* **2011**, *7*, 2427.

[62] C. I. Viquez Rojas, L. V. Slipchenko, *J. Chem. Theory Comput.* **2020**, *16*, 6408.

[63] G. M. J. Barca, C. Bertoni, L. Carrington, D. Datta, N. De Silva, J. E. Deustua, D. G. Fedorov, J. R. Gour, A. O. Gunina, E. Guidez, et al., *J. Chem. Phys.* **2020**, *152*, 1089.

[64] E. Epifanovsky, A. T. B. Gilbert, X. Feng, J. Lee, Y. Mao, N. Mardirossian, P. Pokhilko, A. F. White, M. P. Coons, A. L. Dempwolff, Z. Gan, D. Hait, P. R. Horn, L. D. Jacobson, I. Kaliman, J. Kussmann, A. W. Lange, K. U. Lao, D. S. Levine, J. Liu, S. C. McKenzie, A. F. Morrison, K. D. Nanda, F. Plasser, D. R. Rehn, M. L. Vidal, Z. Q. You, Y. Zhu, B. Alam, B. J. Albrecht, A. Aldossary, E. Alguire, J. H. Andersen, V. Athavale, D. Barton, K. Begam, A. Behn, N. Bellonzi, Y. A. Bernard, E. J. Berquist, H. G. A. Burton, A. Carreras, K. Carter-Fenk, R. Chakraborty, A. D. Chien, K. D. Closser, V. Cofer-Shabica, S. Dasgupta, M. de Wergifosse, J. Deng, M. Diedenhofen, H. Do, S. Ehlert, P. T. Fang, S. Fatehi, Q. Feng, T. Friedhoff, J. Gayvert, Q. Ge, G. Gidofalvi, M. Goldey, J. Gomes, C. E. González-Espinoza, S. Gulania, A. O. Gunina, M. W. D. Hanson-Heine, P. H. P. Harbach, A. Hauser, M. F. Herbst, M. Hernández Vera, M. Hodecker, Z. C. Holden, S. Houck, X. Huang, K. Hui, B. C. Huynh, M. Ivanov, Á. Jász, H. Ji, H. Jiang, B. Kaduk, S. Kähler, K. Khistyaev, J. Kim, G. Kis, P. Klunzinger, Z. Koczor-Benda, J. H. Koh, D. Kosenkov, L. Koulias, T. Kowalczyk, C. M. Krauter, K. Kue, A. Kunitsa, T. Kus, I. Ladjánszki, A. Landau, K. V. Lawler, D. Lefrancois, S. Lehtola, R. R. Li, Y. P. Li, J. Liang, M. Liebenthal, H. H. Lin, Y. S. Lin, F. Liu, K. Y. Liu, M. Loipersberger, A. Luenser, A. Manjanath, P. Manohar, E. Mansoor, S. F. Manzer, S. P. Mao, A. V. Marenich, T. Markovich, S. Mason, S. A. Maurer, P. F. McLaughlin, M. F. S. J. Menger, J. M. Mewes, S. A. Mewes, P. Morgante, J. W. Mullinax, K. J. Oosterbaan, G. Paran, A. C. Paul, S. K. Paul, F. Pavošević, Z. Pei, S. Prager, E. I. Proynov, Á. Rák, E. Ramos-Cordoba, B. Rana, A. E. Rask, A. Rettig, R. M. Richard, F. Rob, E. Rossomme, T. Scheele, M. Scheurer, M. Schneider, N. Sergueev, S. M. Sharada, W. Skomorowski, D. W. Small, C. J. Stein, Y. C. Su, E. J. Sundstrom, Z. Tao, J. Thirman, G. J. Tornai, T. Tsuchimochi, N. M. Tubman, S. P. Veccham, O. Vydrov, J. Wenzel, J. Witte, A. Yamada, K. Yao, S. Yeganeh, S. R. Yost, A. Zech, I. Y. Zhang, X. Zhang, Y. Zhang, D. Zuev, A. Aspuru-Guzik, A. T. Bell, N. A. Besley, K. B. Bravaya, B. R. Brooks, D. Casanova, J. D. Chai, S. Coriani, C. J. Cramer, G. Cserey, A. E. DePrince III, R. A. DiStasio Jr., A. Dreuw, B. D. Dunietz, T. R. Furlani, W. A. Goddard III, S. Hammes-Schiffer, T. Head-Gordon, W. J. Hehre, C. P. Hsu, T. C. Jagau, Y. Jung, A. Klamt, J. Kong, D. S. Lambrecht, W. Z. Liang, N. J. Mayhall, C. W. McCurdy, J. B. Neaton, C. Ochsenfeld, J. A. Parkhill, R. Peverati, V. A. Rassolov, Y. Shao, L. V. Slipchenko, T. Stauch, R. P. Steele, J. E. Subotnik, A. J. W. Thom, A. Tkatchenko, D. G. Truhlar, T. van Voorhis, T. A. Wesolowski, K. B. Whaley, H. L. Woodcock III, P. M. Zimmerman, S. Faraji, P. M. W. Gill, M. Head-Gordon, J. M. Herbert, A. I. Krylov, *J. Chem. Phys.* **2021**, *155*, 084801.

[65] F. R. Manby, M. Stella, J. D. Goodpaster, T. F. Miller, *J. Chem. Theory Comput.* **2012**, *8*, 2564.

[66] S. Huzinaga, A. A. Cantu, *J. Chem. Phys.* **1971**, *55*, 5543.

[67] J. Pu, J. Gao, D. G. Truhlar, *J. Phys. Chem. A* **2004**, *108*, 632.

[68] B. Hégely, F. Bogár, G. G. Ferenczy, M. Kállay, *Theor. Chem. Accounts* **2015**, *134*, 132.

[69] B. Hégely, P. R. Nagy, G. G. Ferenczy, M. Kállay, *J. Chem. Phys.* **2016**, *145*, 64107.

[70] M. Hapka, M. Przybytek, K. Pernal, *J. Chem. Theory Comput.* **2021**, *17*, 5538.

[71] R. A. Kendall, T. H. Dunning, R. J. Harrison, *J. Chem. Phys.* **1992**, *96*, 6796.

[72] S. Boys, F. Bernardi, *Mol. Phys.* **1970**, *19*, 553.

[73] M. O. Sinnokrot, C. D. Sherrill, *J. Phys. Chem. A* **2006**, *110*, 10656.

[74] D. A. Matthews, L. Cheng, M. E. Harding, F. Lipparini, S. Stopkowicz, T.-C. Jagau, P. G. Szalay, J. Gauss, J. F. Stanton, *J. Chem. Phys.* **2020**, *152*, 214108.

[75] J. F. Stanton, J. Gauss, L. Cheng, M. E. Harding, D. A. Matthews, and P. G. Szalay, CFOUR, Coupled-Cluster techniques for Computational Chemistry, a quantum-chemical program package, With contributions from A.A. Auer, R.J. Bartlett, U. Benedikt, C. Berger, D.E. Bernholdt, Y.J. Bomble, O. Christiansen, F. Engel, R. Faber, M. Heckert, O. Heun, M. Hilgenberg, C. Huber, T.-C. Jagau, D. Jonsson, J. Jusélius, T. Kirsch, K. Klein, W.J. Lauderdale, F. Lipparini, T. Metzroth, L.A. Mück, D.P. O'Neill, D.R. Price, E. Prochnow, C. Puzzarini, K. Ruud, F. Schiffmann, W. Schwalbach, C. Simmons, S. Stopkowicz, A. Tajti, J. Vázquez, F. Wang, J.D. Watts and the integral packages MOLECULE (J. Almlöf and P.R. Taylor), PROPS (P.R. Taylor), ABACUS (T. Helgaker, H.J. Aa. Jensen, P. Jorgensen, and J. Olsen), and ECP routines by A. V. Mitinand C. van Wullen. For the current version, see, http://www.cfour.de (accessed: April 2022).

[76] L. V. Slipchenko, M. S. Gordon, *J. Comput. Chem.* **2007**, *28*, 276.

[77] D. Claudino, N. J. Mayhall, *J. Chem. Theory Comput.* **2019**, *15*, 1053.

[78] M. Kállay, P. R. Nagy, D. Mester, Z. Rolik, G. Samu, J. Csontos, J. Csóka, P. B. Szabó, L. Gyevi-Nagy, B. Hégely, I. Ladjánszki, L. Szegedy, B. Ladóczki, K. Petrov, M. Farkas, P. D. Mezei, Á. Ganyecz, *J. Chem. Phys.* **2020**, *152*, 74107.

[79] M. Kállay, P. R. Nagy, Z. Rolik, D. Mester, G. Samu, J. Csontos, J. Csóka, B. P. Szabó, L. Gyevi-Nagy, I. Ladjánszki, L. Szegedy, B. Ladóczki, K. Petrov, M. Farkas, P. D. Mezei, Á. Ganyecz, *MRCC, a quantum chemical program suite*, See also Z. Rolik, L. Szegedy, I. Ladjánszki, B. Ladóczki, and M. Kállay, *J. Chem. Phys.* **2013**, *139*, 94105. www.mrcc.hu (accessed: April 2022).

[80] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865.

[81] S. Grimme, S. Ehrlich, L. Goerigk, *J. Comput. Chem.* **2011**, *32*, 1456.

[82] R. Bukowski, W. Cencek, J. Garcia, P. Jankowski, M. Jeziorska, B. Jeziorski, S. A. Kucharski, V. F. Lotrich, M. P. Metz, A. J. Misquitta, et al., *SAPT2020: An Ab Initio Program for Many-Body Symmetry-Adapted Perturbation Theory Calculations of Intermolecular Interaction Energies*, University of Delaware and University of Warsaw, **2020**. http://www.physics.udel.edu/szalewic/SAPT/SAPT.html (accessed: April 2022).

[83] V. Saunders, M. Guest, *ATMOL Program Package*, SERC Daresbury Laboratory, Daresbury, UK.

[84] K. Szalewicz, K. Patkowski, B. Jeziorski, *Intermolecular Forces and Clusters II*, Springer, Berlin, Heidelberg **2005**, pp. 43-117.

[85] M. Schreiber, M. R. J. Silva, S. P. A. Sauer, W. Thiel, *J. Chem. Phys.* **2008**, *128*, 134110.

[86] T. Sierański, *Struct. Chem.* **2016**, *27*, 1107.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.