A Quantum Chemical Topology Picture of Intermolecular Electrostatic Interactions and Charge Penetration Energy

Fernando Jiménez-Grávalos* and Dimas Suárez*

Departamento de Química Física y Analítica, Universidad de Oviedo, E-33006, Oviedo, Spain.

E-mail: jimenezfernando@uniovi.es; dimas@uniovi.es

Abstract

Basing on the Interacting Quantum Atoms approach, we present herein a conceptual and theoretical framework of short-range electrostatic interactions, whose accurate description is still a challenging problem in molecular modeling. For all the non-covalent complexes in the S66 database, the fragment-based and atomic decomposition of the electrostatic binding energies is performed using both the charge density of the dimers and the unrelaxed densities of the monomers. This energy decomposition together with dispersion corrections gives rise to a pairwise approximation to the total binding energy. It also provides energetic descriptors at varying distance that directly address the atomic and molecular electrostatic interactions as described by point-charge or multipole-based potentials. Additionally, we propose a consistent definition of the charge penetration energy within quantum chemical topology, which is mainly characterized in terms of the intramolecular electrostatic energy. Finally, we discuss some practical implications of our results for the design and validation of electrostatic potentials.

1 Introduction

Molecular mechanics (MM) simulations take advantage of simple potential energy functions to tackle large molecular systems such as those involved in biochemical processes. These MM potentials, which are also named as force fields (FFs), are built upon certain physical models that lead to energy functions representing atomic or fragment contributions that provide a reliable global description with affordable computational requirements and depend on a set of parameters, which are commonly fitted to reference experimental and/or quantum mechanical (QM) data.[1]

The employment of MM methods is rooted in a hierarchy of physical assumptions.[2,3] First of all, the Born-Oppenheimer approximation is invoked to split electronic and nuclear motion, the MM functions being dependent only on nuclear coordinates. Secondly, the additivity assumption allows for the separation of the total energy into a sum of potentials according to their different physical sources. For example, bonded atoms may be described by two-body terms to account for stretching, three for bending or four for torsion. On the other hand, interactions comprising non-bonded atoms are usually represented by pairwise potentials such as the Lennard-Jones and the Coulomb ones. Finally, and probably one of the most appealing attributes of MM methods, is their transferability, the ability to correctly describe not only the set of model molecules/fragments, but also other different systems provided they are built upon similar chemical units.

Among the different physical terms in the MM potentials, electrostatics deserves special attention because it plays a key role, both at short and long range, on the stability and activity of large systems such as proteins, nucleic acids or lipids, among others.[4] Of course, the MM treatment of electrostatics must follow some approximations, since the calculation of the exact quantum mechanical electron density is intended to be avoided. The accuracy of the available methods changes when dealing with the short- or the long-range regime in which the intermolecular electrostatic interactions can be classified. At long range, the use of point charges or higher order multipoles has resulted in accurate electrostatics with significant

improvements to speed up and facilitate convergence of the calculations such as the Ewald summation and its variants to perform, for example, molecular simulations in solution under periodic boundary conditions.[5–10] At short-range, however, the approximations taken for long distances, which are mainly based on the multipole expansion, become less accurate or invalid[11] and a correct electrostatic description in this regime stills poses a challenge. Hence, there is a growing interest in improving the short-range electrostatics (e.g., for troublesome hydrogen-bonds), mainly focused on correcting the inherent error associated with the non-negligible interpenetration of densities (the so-called charge penetration (CP) error) that is not accounted for in traditional multipolar approximations. Thus, several investigations have been devoted in the last years to incorporate charge penetration terms into the MM electrostatic potentials.[12–15]

The separation of various physical terms as implemented in the MM potentials is somehow paralleled by the energy decomposition analysis (EDA) methods,[16–18] which either perform QM calculations in the fragments and subsequently evaluate interaction energy terms basing on perturbation theory or extract multiple energy terms from the full QM energy of molecular clusters. As a matter of fact, a major goal of any EDA approach is to ascertain the nature and type of the interactions among molecules as well as to rationalize their stabilizing or destabilizing roles, what may have implications for the design, parameterization and validation of the MM methods, in general, and the MM electrostatic potentials, in particular. However, many EDAs have been developed rooted in different approaches.[16] Hence, the symmetry-adapted-perturbation-theory (SAPT) method makes use of a perturbative approach to differentiate the distinct nature of the intermolecular interactions,[18,19] while the orbital-based EDAs exploit a stepped scheme to calculate the different energies according to some reference electronic states[17,20,21] and the interacting quantum atoms (IQA) method relies on a real space partition of the QM density matrices.[22,23]

According to recent studies, in spite of their crude approximations, it may be feasible to improve the classical MM potentials by utilizing the information provided by EDAs. More

specifically, it has been shown that the SAPT energy components (electrostatics, induction, exchange-repulsion and dispersion) can be modelled with relatively simple MM functions.[24,25] In particular, it has been demonstrated that the combination of empirical damping functions accounting for the CP error with point multipoles results in electrostatic energies at short range which are quite close to the SAPT ones. Actually, the SAPT electrostatic interactions provide the required reference data to parameterize and validate the CP-corrected potentials. However, different interpretations of short-range energetic effects involving the overlap of the electron densities of two or more fragments may be possible depending on the particular EDA method.[16] For this reason, we decided to reexamine the nature of electrostatic interactions under the prism of the real-space decomposition pursued by the IQA approach, which not only distinguishes between electrostatic or exchange-correlation components of the interaction energy, but also between intra- or inter-atomic (or fragment) contributions. Moreover, since IQA splits the total energy of a system and not only the interaction between selected fragments, it is capable of reconstructing (or dissecting) the energy ascribed to both covalent and non-covalent binding, unveiling, for instance, the critical role played by both electrostatics and covalency in molecular bonds.[26]

Herein, we study in detail the electrostatic interactions involved in non-covalent complexes with a twofold goal. One one hand, we aim to compare in a consistent and systematic manner the atomic and fragment contributions to the electrostatic energy as evaluated throughout a hierarchy of QM and MM approximations and at varying intermolecular distances. In this way, we seek to identify the best correspondence between the IQA and MM electrostatic terms. On the other hand, we critically examine the CP concept and propose a novel definition relying on a joint orbital and real-space decomposition scheme, which can give new insight into the CP energy. To help fulfill these goals, the rest of the manuscript is structured as follows. First, we present and describe the theoretical scaffold that holds our work, paying particular attention to the IQA —and its IQF variant— energy decomposition, followed by subsections concerning the zeroth-order QM approximation, the classical multi-

pole expansion, the electrostatic MM potentials, etc. The section on Theory and Methods ends with our assessment of the charge penetration error and the alternative definition proposed in this work. Subsequently, we describe some computational settings and the results of our test calculations, which were carried out on the S66 and S66x8 datasets[27,28] covering a wide range of non-covalent interactions frequently found in biomolecules. The various levels of description of the electrostatic interactions are then discussed basing on their statistical correlation with benchmark data, their dependence with the intermolecular separation, etc. The QM and IQA calculations yield further information, not only about the magnitude of the CP error, but more importantly about its different role in the IQA descriptors. Finally, we conclude that the aim of improving the electrostatic description in novel FFs is fulfilled at the expense of accounting for intramolecular effects not included in the pairwise classical MM potentials.

2 Theory and Methods

As stated in the Introduction, electrostatics in non-covalent complexes is the main concern behind this manuscript. With the aim of connecting the full QM description of the electrostatic energy with the classical MM potentials, we describe first the electrostatic contributions in the framework of the IQA partitioning of absolute and relative energies. Then the zeroth-order approximation and the multipole expansion commonly employed in the study of intermolecular complexes are succinctly described, allowing thus to state the basic approaches used by MM methods in order to evaluate short-range electrostatic energies. The limitations of the classical potentials are partly ascribed to the CP error, which is also discussed in the light of recent advancements. Finally, basing on the IQA concepts, we propose an alternative definition of the CP energy.

2.1 IQA energy decomposition

The interacting quantum atoms method is a robust and physically-sound approach to decompose the total energy into chemically meaningful components.[22,23] It is based upon a given partition of the first- and second-order reduced density matrices ($\rho_1(\mathbf{r}_1;\mathbf{r}'_1)$ and $\rho_2(\mathbf{r}_1,\mathbf{r}_2;\mathbf{r}'_1,\mathbf{r}'_2)$, respectively) such as the real space partition proposed by Bader and coworkers,[29] which is the most commonly used along with IQA (although other schemes may be adopted). Thus, the three dimensional space is decomposed into atomic regions Ω_I, which are the attraction basins of the gradient field of the electron density.

Given a global energy E of a system, IQA permits its decomposition into atomic components and pair interaction energies according to

$$E = \sum_I E^I_{net} + \sum_{I<J} E^{IJ}_{int}. \tag{1}$$

where E^I_{net} is called the atomic net energy and, under the Born-Oppenheimer approximation, represents the energy due to the kinetic energy of electrons plus all the interactions involved (i.e., electron-electron and electron-nucleus) inside the atomic basin of each atom I. Similarly, each E^{IJ}_{int} term comprises the interaction energy between the electrons (e) and nucleus (n) located in atom I with those ascribed to the other atoms, which can be separated into n-e, e-e and n-n contributions. In order to compute the the kinetic and n-e interaction energies the first-order reduced density, $\rho_1(\mathbf{r}_1;\mathbf{r}'_1)$, is required while only the diagonal $\rho_2(\mathbf{r}_1,\mathbf{r}_2)$ is needed for the e-e energy terms.

The pair density $\rho_2(\mathbf{r}_1,\mathbf{r}_2)$ can be decomposed into two contributions as

$$\rho_2(\mathbf{r}_1,\mathbf{r}_2) = \rho(\mathbf{r}_1)\rho(\mathbf{r}_2) + \rho_{xc}(\mathbf{r}_1,\mathbf{r}_2), \tag{2}$$

where $\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)$ is the non-correlated product of densities and $\rho_{xc}(\mathbf{r}_1,\mathbf{r}_2)$ is exchange-correlation (xc) correction. Equation 2 gives thus rise to two kinds of interatomic interaction energies

stemming from either the classical eletrostatic interaction E_{ele}^{IJ} or the quantum mechanical exchange-correlation E_{xc}^{IJ} as

$$E_{int}^{IJ} = E_{ele}^{IJ} + E_{xc}^{IJ}. \tag{3}$$

Such a decomposition of E_{int}^{IJ} is particularly relevant when assessing the nature of a given bond or interaction, since the electrostatic term is associated with ionicity and the exchange-correlation contribution with covalency.[23]

2.2 D3-IQF: the dispersion-corrected fragment version of IQA

The IQA decomposition admits the grouping of atomic terms into fragment contributions (e.g., functional groups, molecules). For practical purposes, the distinction between the atomic level analyses and the fragment-based ones was introduced as IQA for the former and IQF (interacting quantum fragments) for the latter.

In previous work,[30] it has been shown that IQF may be useful, for instance, to either reconstruct or dissect the formation energy of non-covalent complexes. Moreover, the IQA/IQF terms retrieved from the wave function have been augmented with the dispersion correction (D3) by Grimme in DFT and HF methods.[31] D3 is a semiclassical pairwise potential inspired by the London model than can, in turn, be corrected by a damping function to reproduce the correct asymptotic behavior of the dispersion. In this work, the pure D3 potential is combined with the Becke-Johnson damping function[32] (for the sake of simplicity, the potential is dubbed as D3).

Using D3-IQF, the formation energy of a molecular aggregate constituted by two fragments A and B ($\mathrm{A} + \mathrm{B} \longrightarrow \mathrm{A}\cdots\mathrm{B}$) is split as

$$\Delta E_{form} = \Delta E_{net}^{A} + \Delta E_{net}^{B} + E_{ele}^{AB} + E_{xc}^{AB} + E_{dis}^{AB} = E_{def}^{A} + E_{def}^{B} + E_{int}^{AB}, \tag{4}$$

where the deformation term E_{def} corresponds to the net energy variation ΔE_{net}, which collects both the intra- and interatomic IQA energies belonging to the corresponding frag-

ment. The interfragment interaction energy comprises the electrostatic (E_{ele}^{AB}), exchange-correlation (E_{xc}^{AB}) and the empirical dispersion (E_{dis}^{AB}), the latter being thus separated from the whole exchange-correlation interaction energy. Overall, the contribution of electrostatics and exchange-correlation to ΔE_{form} is actually split between the fragment deformation and the inter-fragment interactions.

2.3 Electrostatic energy

Let us now consider the purely electrostatic energy for a given system with total charge density $\rho(\mathbf{r})$, which can be readily computed using the Coulomb law,

$$E_{ele} = \frac{1}{2}\int_{\infty}\int_{\infty}\frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}}d\mathbf{r}_1 d\mathbf{r}_2. \tag{5}$$

The total charge density, that includes both the electron density ρ_e derived from the electronic wave function and the nuclear charges Z_I, is

$$\rho_{tot}(\mathbf{r}) \equiv \rho(\mathbf{r}) = \sum_I Z_I \delta(\mathbf{r} - \mathbf{R}_I) - \rho_e(\mathbf{r}), \tag{6}$$

in which the positions of the nuclei $\mathbf{R}_I$ are fixed under the Born-Oppenheimer approximation.

The real space partition proposed by Bader results in the following atomic decomposition of the electrostatic energy

$$\begin{aligned} E_{ele} &= \frac{1}{2}\left[\int_{\Omega_I} d\mathbf{r}_1 \int_{\Omega_I} d\mathbf{r}_2 \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}} + \int_{\Omega_J} d\mathbf{r}_1 \int_{\Omega_J} d\mathbf{r}_2 \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}}\right. \\ &+ \left.\int_{\Omega_J} d\mathbf{r}_1 \int_{\Omega_J} d\mathbf{r}_2 \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}} + \int_{\Omega_J} d\mathbf{r}_1 \int_{\Omega_I} d\mathbf{r}_2 \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}}\right] \\ &= E_{ele}^I + E_{ele}^J + E_{ele}^{IJ}. \end{aligned} \tag{7}$$

Similarly, the fragment-based decomposition of E_{ele} is straightforward and analogous to the atomic one, that is, the electrostatic energy can be readily expressed in terms of intra-

and inter-fragment contributions. This allows a specific assessment of the total electrostatic component of the formation energy ΔE_{ele} in a two-fragment system AB as

$$\Delta E_{ele} = \Delta E_{ele}^{A} + \Delta E_{ele}^{B} + E_{ele}^{AB}, \tag{8}$$

where the intra-fragment variation in electrostatic energy (ΔE_{ele}^{A} and ΔE_{ele}^{B}) and the inter-fragment electrostatic interaction E_{ele}^{AB} may yield an insightful decomposition of electrostatic effects.

2.4 Zeroth-order electrostatic energy

In principle, equation 8 is evaluated over a continuous quantum electron density that can be obtained from the fully-relaxed wave function. However, within the context of MM methodologies, the electrostatic energy of non-covalent complexes formed by weakly-interacting molecular species is normally estimated at the zeroth-order level, that is, neglecting the mutual relaxation of the isolated molecular densities. Thus, for a complex AB composed of molecules A and B, $\rho^0(\mathbf{r}) = \rho_A^0(\mathbf{r}) + \rho_B^0(\mathbf{r})$. As a consequence, the zeroth-order electrostatic contribution to the formation energy ΔE_{ele}^0 equals the Coulomb interaction between the zeroth-order densities:

$$\Delta E_{ele}^0 = \frac{1}{2} \int_\infty \int_\infty \frac{\rho_A^0(\mathbf{r}_1)\rho_B^0(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2. \tag{9}$$

This energetic term corresponds to the so-called *first-order polarization energy* (or simply *electrostatic energy*) defined in the Symmetry-Adapted Perturbation Theory (SAPT),[19] which has been adopted as a benchmark electrostatic energy for the validation of short-range electrostatic potentials recently developed.

2.5 Multipole expansion

To avoid the usage of continuous charge distributions, the multipole expansion can be invoked to calculate (and interpret) electrostatic energies as well as to guide the development of various MM electrostatic potentials. Using spherical coordinates, the multipole expansion transforms Eq. 9 as

$$\Delta E^0_{ele,mp} = \sum_{l_1 m_1}^{\infty} \sum_{l_2 m_2}^{\infty} C_{l_1 m_1 l_2 m_2}(\hat{R}) \frac{Q^{0,A}_{l_1 m_1} Q^{0,B}_{l_1 m_1}}{R_{AB}^{l_1+l_2+1}}, \tag{10}$$

where $Q^{0,A}_{l_1 m_1}$ and $Q^{0,B}_{l_2 m_2}$ are the multipoles of the respective unrelaxed densities $\rho^0_A(\mathbf{r}_1)$ and $\rho^0_B(\mathbf{r}_2)$.[33] In Eq. 10, m_i runs from $-l_i$ to $+l_i$, $R_{AB} = |\mathbf{R}_B - \mathbf{R}_A|$ is the separation between the centers of both charge distributions and $\hat{R}$ is the angular coordinate that determines the orientation. $C_{l_1 m_1 l_2 m_2}(\hat{R})$ are known coefficients calculated for each $(l_1 m_1, l_2 m_2)$ quartet. The set of multipoles Q^0_{lm} is obtained from the continuous charge density through the real spherical harmonics S_{lm} following the integral

$$Q^0_{lm} = N_l \int_{\infty} r^l S_{lm}(\hat{r}) \rho^0(\mathbf{r}) d\mathbf{r}, \tag{11}$$

being N_l a constant dependent on l, and r and $\hat{r}$ denote the radial and the angular coordinates, respectively.

The usage of the multipole expansion to estimate the electrostatic energies and forces between non-bonded molecules is affected by a certain error with two different sources. On the one hand, the Taylor expansion that originates (10) must be truncated at some order. On the other one, when two or more molecules interact closely their electronic densities can overlap to a significant extent, regardless of whether or not their densities are allowed to relax. The energetic effect of the density overlap, which is entirely omitted by the multipole expansion, is generally associated with the charge penetration error.

Although the Taylor expansion would not be applicable whenever the densities over-

lap (i.e., when the distances from the origin of each multipole series intercept each other, $r_1 + r_2 \geq R_{AB}$), it is still largely useful in many cases as long as it converges. In fact, convergence of the multipole expansion can be enhanced by distributing multipoles throughout the molecule (e.g., located at atomic sites) instead of calculating them at a single molecular origin.[11,34,35] Therefore, the most typical strategies for the calculation of intermolecular electrostatic interactions using classical potentials rely on the use of distributed multipolar expansions at the expense of being affected by a certain error, which is normally assumed to be dominated by the CP error.

Within the Quantum Theory of Atoms in Molecules (QTAIM) by Bader and coworkers,[29] each atomic multipole is calculated using Eq. 11, but delimiting the integral to the boundaries of each basin. These atomic basins are, by definition, non-overlapping. However, this does not mean that $\Delta E^0_{ele,mp}$ calculated for those topological atoms is free from CP errors. In fact, the atomic basins are often sharp and the previous condition that $R_{AB} > r_1 + r_2$ becomes no longer fulfilled if the outermost distances r_1 from $\mathbf{R}_A$ and r_2 from $\mathbf{R}_B$, that define the convergence spheres, intercept each other. This explains the good agreement between the multipolar interatomic energies and those calculated by integration for relatively distant atoms, but not for bonded or very close pairs.[33,36,37]

2.6 Molecular mechanics electrostatic potentials

Given that a vast literature exists about the classical electrostatic problem,[1] we merely outline some basic features of the short-range potentials currently implemented in the most popular molecular simulation packages. We note first that the MM electrostatic potentials can be classified into two main families according to whether they rely exclusively on truncated multipolar expansions or they incorporate additional terms to mitigate their limitations.

2.6.1 Multipolar-based MM approaches

Regarding the order at which the multipole expansion is truncated —which is determined by l_{max}—, the MM electrostatic potentials are readily split into two groups. On one side, MM methods such as AMBER,[38] CHARMM,[39] GROMOS,[40] OPLS,[41] adopt simple electrostatic formulas with point charges (i.e., monopoles, with $l_{max} = 0$). It is important to remind that, in this approach, the values of the atomic charges are not derived directly from the charge density, but obtained upon fitting procedures against the molecular electrostatic potential (ESP) computed with QM methods. On the other, more sophisticated methods, such as NEMO,[42] AMOEBA[43] or the QTAIM-based FFLUX,[44,45] include higher order multipoles (frequently up to the quadrupoles, $l_{max} = 2$) in order to capture the anisotropy of the distribution of electrons in space. These potentials are generaly built from the QM density matrix of the molecule of interest by means of the distributed multipole analysis (DMA) or similar methods that derive first all the multipolar contributions from the products of Gaussian basis set functions and, subsequently, apply geometric criteria to reallocate them throughout the molecule at the atomic sites and, sometimes, other points of interest such as bond centers. In addition, some methods (e.g., AMOEBA or NEMO) refine the DMA multipoles[35] to better reproduce the QM ESP values. In this way, the resulting charges/multipoles may include in an effective way both high-order multipolar effects and CP effects. Actually, the performance of the MM potentials is examined statistically as a whole (i.e., using the full MM potential including all bonded and non-bonded terms) by various energetic and structural validation tests. A quite different approach is followed by the FFLUX force field. It makes use of QTAIM multipoles —as those corresponding to the Bader's atomic basins previously commented— in contrast to the more widespread DMA methodologies, and estimates them by means of a machine learning thechnique depending on each atom's environment.

2.6.2 CP-corrected potentials

In comparison with the atomic/multipolar methods that are massively employed in current simulation packages, the electrostatic MM potentials that go beyond the multipolar approximation are much less consolidated. In this category, we find different methodologies as SIBFA (Sum of Interactions between fragments ab initio computed),[46] EFP (Effective Fragment Potential)[47] and AMOEBA+[24] that complement the parent electrostatic multipolar formulas with other functions that are primarily designed to capture very-short range electrostatics and to remove the CP error. Thus, they usually apply a damping function to the multipolar V_{mp} potential to estimate a CP-corrected electrostatic potential $\tilde{V}_{ele}$ as

$$\tilde{V}_{ele} = V_{mp} \cdot f_{damp}(r). \tag{12}$$

The functional form of the damping function is generally based on the correction to the hydrogen-like atomic potential that the expansion of r^{-1} does not include[11,12,15,48]

$$f_{damp}(r) = 1 - ce^{-\alpha r}, \tag{13}$$

where c may be a constant or even a polynomial[13] dependent on the atomic separation r, and α is another parameter. To determine these parameters, a different paradigm in the parameterization strategy is required with respect to the multipolar-based methods. Thus, the newest potentials V_{ele} seek to reproduce ΔE^0_{ele} as evaluated by SAPT or similar methodologies on a set of test molecules.[24]

2.7 Charge Penetration Energy

The literature defines the electrostatic CP energy (E_{pen}) between two molecules A and B as the difference between their zeroth-order electrostatic energy ΔE^0_{ele} computed with the continuous charge densities and its analogue $\Delta E^0_{ele,mp}$ computed by means of multipole

expansion,

$$E_{pen} = \Delta E^0_{ele} - \Delta E^0_{ele,mp}. \tag{14}$$

Conceptually, this straightforward definition of E_{pen} is satisfactory. It also shows that E_{pen} is not only an inter-fragment quantity, but rather an energy that presents also intramolecular contributions according to the IQA partitioning of the whole ΔE^0_{ele}. In this respect, the energetic definition suggests that the CP error is not limited to *the change in the electrostatic interaction between two atoms due to their electron cloud overlap and the associated loss of nuclear screening.*[48]

The rigorous evaluation of E_{pen} for different systems at varying intermolecular separations would allow a deeper analysis of electrostatics and, eventually, the development of more accurate potentials. However, as noticed by Dominiak et al.,[49] *different methods of obtaining multipole moments lead to different radii of (pseudo)convergence, different levels of multipole expansions at which (pseudo)convergence is achieved and different values of penetration energy.* Therefore, the value of the CP energy as evaluated with Eq. 14 may depend on the particular method used to derive the multipoles. Moreover, the usage of truncated expansions introduces some truncation error in equation 14 so that both truncation and penetration effects become somewhat mixed in the resulting E_{pen} values, unless a well converged multipolar energy is available, what depends on the intermolecular separation.[49]

An alternative to evaluate E_{pen} has been roposed by Kayris and Jensen.[50] Having noticed the relationship between the CP error and the magnitude of the orbital overlap, they attempt to recover such effect from scratch, with a derivation of E_{pen} independently from the multipolar model used to estimate electrostatics at first stage. However, the authors find that the inherent dependence on the set of molecular orbitals used may lead to different CP values.

2.7.1 A novel IQA-like definition of Charge Penetration Energy

Within the IQF formalism, it is possible to combine both the Bader partitioning scheme ($\mathbb{R}^3 = \Omega_A + \Omega_B$) with a total zeroth-order density decomposition ($\rho^0_{AB} = \rho^0_A + \rho^0_B$) in order to obtain the following energy terms:

i) the interaction due to the same monomer density (ρ^0_A or ρ^0_B) inside a given molecular basin (Ω_A or Ω_B) leading to $E^A_{ele}(\rho^0_A, \rho^0_A)$, $E^B_{ele}(\rho^0_A, \rho^0_A)$, $E^A_{ele}(\rho^0_B, \rho^0_B)$ and $E^B_{ele}(\rho^0_B, \rho^0_B)$.

ii) the interaction between ρ^0_A and ρ^0_B in either Ω_A or Ω_B basins ($E^A_{ele}(\rho^0_A, \rho^0_B)$ and $E^B_{ele}(\rho^0_A, \rho^0_B)$).

iii) the electrostatic energy between the two basins comprising the same density ($E^{AB}_{ele}(\rho^0_A, \rho^0_A)$ and $E^{AB}_{ele}(\rho^0_B, \rho^0_B)$).

iv) the interaction between ρ^0_A and ρ^0_B densities, each one being located in a different molecular basin ($E^{AB}_{ele}(\rho^0_A, \rho^0_B)$ and $E^{BA}_{ele}(\rho^0_A, \rho^0_B)$).

Hence, the total IQF electrostatic energy of a complex AB can be written as

$$\begin{aligned} E^0_{ele} &= E^A_{ele}(\rho^0_A, \rho^0_A) + E^A_{ele}(\rho^0_B, \rho^0_B) + E^A_{ele}(\rho^0_A, \rho^0_B) \\ &+ E^B_{ele}(\rho^0_A, \rho^0_A) + E^B_{ele}(\rho^0_B, \rho^0_B) + E^B_{ele}(\rho^0_A, \rho^0_B) \\ &+ E^{AB}_{ele}(\rho^0_A, \rho^0_A) + E^{AB}_{ele}(\rho^0_B, \rho^0_B) + E^{AB}_{ele}(\rho^0_A, \rho^0_B) + E^{BA}_{ele}(\rho^0_A, \rho^0_B). \end{aligned} \tag{15}$$

In the notation used above the two interacting densities are encompassed by parentheses, while the basins they are integrated in are identified by the corresponding superscripts in the given order. Hence, for instance, the term $E^{BA}_{ele}(\rho^0_A, \rho^0_B)$ stands for $\int_{\Omega_B} d\mathbf{r}_1 \int_{\Omega_A} d\mathbf{r}_2 \rho^0_A(\mathbf{r}_1)\rho^0_B(\mathbf{r}_2) r_{12}^{-1}$. Note that only one superscript signifies a double integration over the same basin, according to which $E^B_{ele}(\rho^0_A, \rho^0_A)$ corresponds to $\int_{\Omega_B} d\mathbf{r}_1 \int_{\Omega_B} d\mathbf{r}_2 \rho^0_A(\mathbf{r}_1)\rho^0_A(\mathbf{r}_2) r_{12}^{-1}$.

When the above double decomposition is applied to the electrostatic energies of the separate fragments, such as A, in the final complex, the electrostatic energy of the original

species becomes

$$E_{ele}(\rho_A^0, \rho_A^0) = E_{ele}^A(\rho_A^0, \rho_A^0) + E_{ele}^B(\rho_A^0, \rho_A^0) + E_{ele}^{AB}(\rho_A^0, \rho_A^0) + E_{ele}^{BA}(\rho_A^0, \rho_A^0). \tag{16}$$

Note that this partitioning is achieved by using the real space decomposition into two molecular basins derived from the AB wave function and that geometry relaxation effects are not considered. By subtracting from Eq. 16 the fragment energies, the corresponding electrostatic contribution to the formation energy of the complex is obtained,

$$\Delta E_{ele}^0 = E_{ele}^A(\rho_A^0, \rho_B^0) + E_{ele}^B(\rho_A^0, \rho_B^0) + E_{ele}^{AB}(\rho_A^0, \rho_B^0) + E_{ele}^{BA}(\rho_A^0, \rho_B^0). \tag{17}$$

Among the surviving terms in Eq. 17, $E_{ele}^{AB}(\rho_A^0, \rho_B^0)$, which accounts for the purely interaction energy between the ρ_A^0 density located in Ω_A and the ρ_B^0 density located in Ω_B, reveals itself as the ordinary interaction term between the two monomers A and B. It matches ΔE_{ele}^0 at long distances, while the other three terms would present a similar behavior of increasing in magnitude when shortening the intermolecular distances R_{AB} and cancelling out in the opposite situation. Thus, those three terms can be directly related with the interpenetration of molecular densities and grouped in an IQF-like electrostatic charge penetration energy

$$E_{ele,pen}^{IQF} = E_{ele}^A(\rho_A^0, \rho_B^0) + E_{ele}^B(\rho_A^0, \rho_B^0) + E_{ele}^{BA}(\rho_A^0, \rho_B^0), \tag{18}$$

that collects the electrostatic energy coming from the penetration of each molecular density into the opposite molecular basin. This term fulfils $\lim_{R_{AB}\to\infty} E_{ele,pen}^{IQF} = 0$ (and so its three components), while $\lim_{R_{AB}\to\infty} E_{ele}^{AB}(\rho_A^0, \rho_B^0) = \Delta E_{ele}^0$. Figure 1 represents the previous four terms between the partitioned ρ_A^0 and ρ_B^0 adding up to ΔE_{ele}^0, and compares them to the zeroth-order pairwise term $E_{ele}^{0,AB}$ between the total density in each basin.

We will see that this alternative definition of the charge penetration energy, in the context of the quantum chemical topology, constitutes an insightful description of the electrostatic

effects played by the mutual interpenetration of molecular charge densities.

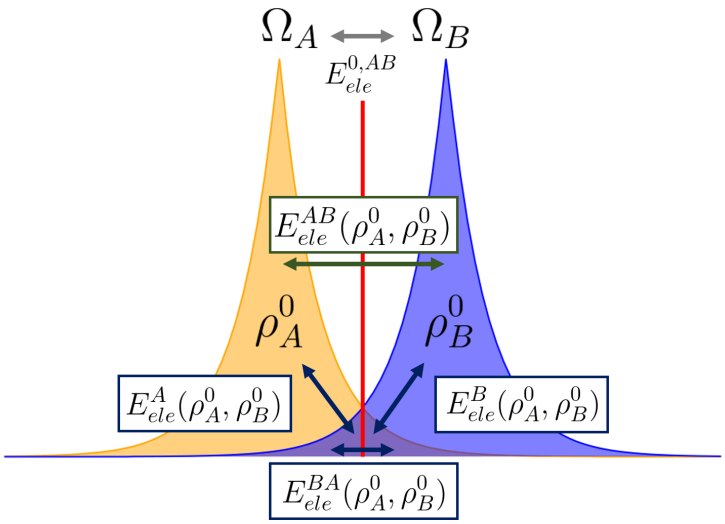


Figure 1: Graphical scheme of the four contributions giving rise to ΔE_{ele}^{0}, where three of them (in dark blue) comprise the IQF electrostatic penetration term and the remaining one (dark green) accounts for the interaction of ρ_A^0 and ρ_B^0 lying in the molecular basins Ω_A and Ω_B, respectively. The zeroth-order IQF pairwise term $E_{ele}^{0,AB}$ has been also included to remark its difference with the previous $E_{ele}^{AB}(\rho_A^0, \rho_B^0)$, as it accounts for an interaction between total densities inside each basin (the original ρ_A^0 or ρ_B^0 and the tail from the other that has penetrated into another domain).

3 Computational details

3.1 Molecular geometries and reference interaction energies

All the QM and classical electrostatic calculations were performed on the molecular geometries retrieved from the S66 database,[27] which contains a set of 66 complexes featuring the most common non-covalent interactions in biomolecules. In addition, a selection of 12 representative complexes from the S66x8 database,[28] which is an extension of 66 to eight different fractions of the equilibrium intermolecular distance, were also considered. The benchmark CCSD(T)/CBS interaction energies collected in S66 were employed as the reference values for comparative purposes.

3.2 HF-D3 calculations

Single-point Hartree-Fock (HF) calculations with the triple-zeta cc-pVTZ basis set[51] were carried out on the S66 and the S66x8 geometries using the `GAMESS-US` package.[52] The Grimme's dispersion potential (D3) as implemented in the `DFT-D3` code[53] was employed to incorporate the dispersion energy. Additionally, in order to correctly reproduce the asymptotic behavior of the dispersion energy at small distances, the Becke-Johnson damping function was chosen.[54]

We selected HF as the QM method of choice because it lacks entirely of dispersion energy and thereby yields a straight physical partitioning of energy in combination with the D3 potential. We also note in passing that HF-D3 has been shown to describe correctly and efficiently the structure and energetics of biomolecules[55] and that a variant of the DFT-SAPT EDA method has been also developed in which the costly ab initio dispersion calculations are replaced by a reparameterized D3 potential.[56]

3.3 IQA calculations

The IQA energetic analyses were performed with the `PROMOLDEN` code.[57] The settings comprised β-spheres with radii of a 60% of the distance between each nucleus and its closest critical point. Within them, high-quality Lebedev angular grids with 974 points were used, along with Euler-McLaurin radial quadratures with 382 radial points. A bipolar expansion of r_{12}^{-1} was selected, being its L_{max} of 6. On the other hand, the outer part of the basins (i.e., outside the β-spheres) employed same angular and radial quadratures, albeit increasing their respective angular and radial points up to 5810 and 512, with a maximum radius of 15 au. In this case, r_{12}^{-1} was expanded by means of a Laplace expansion with $L_{max} = 10$.

3.4 Point-charge and multipolar calculations

Atomic charges were computed for the separate monomers in the S66 structures by means of the restrained electrostatic potential (RESP) method following the General Amber Force Field (GAFF)[58] prescriptions. Thus, the electrostatic potential in the gas-phase was computed at the HF/6-31G* level using the `Gaussian03`[59] suite while the `antechamber` program[60] was used to derive the RESP atomic charges.

Two different sets of atomic multipoles were employed in the classical multipolar calculations. On one hand, atomic multipoles were derived up to the quadrupoles ($l_{max} = 2$) following the AMOEBA Force Field parameterization protocol.[43,61] The molecular geometries of the S66 monomers were optimized at the MP2/6-311G(d,p) level of theory using the `Gaussian09` package.[62] The GDMA program[63] was employed to perform a distributed multipole analysis of the MP2/6-311G(d,p) density from which an initial set of atom type multipoles was extracted. The initial multipoles were refined in a two-step process by fitting to reference electrostatic potentials calculated at the MP2/6-311G(d,p) and MP2/aug-cc-pVTZ levels of theory with `Gaussian09`. This task was done with the software utilities `POLEDIT` and `POTENTIAL`, which are included in the `TINKER` toolbox for molecular modeling.[64] On the other hand, using the `PROMOLDEN` program, the charge multipoles within each QTAIM atomic basin were computed according to the spherical harmonic formalism up to the specified $l_{max} = 10$. In this way, we also derived the HF/cc-pVTZ QTAIM atomic multipoles of the S66 monomers. All the electrostatic interaction energies for the AMOEBA and the QTAIM multipoles were obtained using the `MPOLINT` code.[65]

Additionally, a set of S66x8 complexes was tested under the AMOEBA+ CP-corrected potentials.[24] For this, `TINKER` was again invoked to calculate the respective CP energies as the difference between the CP-corrected multipoles and the multipolar energies previously derived. In this case, the parameters of the dumping functions were directly taken from reference [24].

3.5 Graphs and statistical analyses

Octave[66] and GNUplot[67] have been used to perform the statistical analyses and the correlation plots, while Python's Matplotlib[68] has been chosen for the rest of the representations.

4 Results and Discussion

4.1 S66 dataset

According to the main interaction governing binding, the S66 complexes can be classified into three main groups, namely H-bond (with 23 complexes), dispersion-dominated (23 dimers) and mixed complexes (the remaining 20 systems). For further details about their specific classification, see Table S1 in the Supporting Information (SI). Similarly, the interactions between pairs of atoms can be grouped in polar, non-polar and mixed basing on the GAFF atom types. According to this, in Figures 1-4 each type of complex/interaction is color coded: pink for H-bond/polar, yellow for mixed and blue for dispersion/non-polar.

Regarding the level of theory chosen for our QM calculations, it has been shown in previous work[30] that the HF/cc-pVTZ formation energies in conjunction with the D3 dispersion energy correlate satisfactorily with the benchmark CCSD(T)/CBS energies from the S66 database (see also Figure S1). Not only is the global correlation quite suitable ($R^2 = 0.991$) but also the three categories are in good agreement with the reference values, the root mean square (RMS) errors lying in a small range from 0.82 to 1.80 kcal mol^{-1}. Besides being computationally efficient, another major advantage of using HF-D3 is the fact that it permits, when applying IQA, a trustworthy distinction of the relative classical and non-classical contributions. Since HF is a correlation-free method, HF-D3 also provides a pure split of the dispersion energy from other sources. For all these reasons we selected HF-D3/cc-pVTZ for our study.

4.2 Electrostatic contributions to binding energies

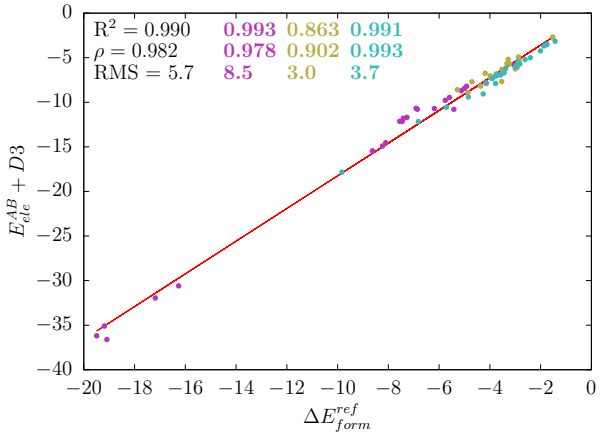
As non-covalent binding processes are in multiple cases strongly influenced by electrostatics, we analyzed first the intra- or inter-fragment character of the electrostatic contributions to the formation energy ΔE_{ele} basing on the IQF assessments (Eq. 8). In this analysis, we opted to augment the electrostatic energy with the pairwise D3 dispersion so as to compare the benchmark binding energies ΔE^{ref}_{form} with the dispersion-augmented electrostatic formation energies ($\Delta E_{ele} + D3$) and the intermolecular electrostatic interactions ($E^{AB}_{ele} + D3$).

Table 1: Comparison between the benchmark formation energies and the IQF electrostatic contributions to formation, enhanced with D3 dispersion. It comprises the coefficient of determination R^2, the Spearman correlation coefficient (ρ) and the root mean square error (RMS/kcal mol^{-1})

Energy term	Complex type	R^2	ρ	RMS
$\Delta E_{ele} + D3$	Global	0.888	0.719	17.2
	H-bond	0.970	0.930	23.9
	Mixed	0.593	0.820	12.8
	Dispersion	0.851	0.910	11.7
$E^{AB}_{ele} + D3$	Global	0.990	0.982	5.7
	H-bond	0.993	0.978	8.4
	Mixed	0.863	0.902	3.0
	Dispersion	0.991	0.993	3.7

.

Figure 2: Correlation between ΔE^{ref}_{form} and the pairwise term $E^{AB}_{ele} + D3$ (energies in kcal mol^{-1}).



As can be appreciated in Table 1, the global electrostatic formation energy together with

the D3 dispersion shows a moderate correlation ($R^2 = 0.880$) with the reference formation energies. Among the three classes of systems, the H-bond complexes exhibit the best fitting ($R^2 = 0.970$), although being accompanied by the highest root mean square (RMS) error ($RMS = 23.9$ kcal mol^{-1}). The dispersion complexes show a poorer agreement, with the mixed group having a low R^2 value of 0.593. In contrast, the IQF pairwise energies in conjunction with the D3 dispersion $E_{ele}^{AB} + D3$ are quite well correlated (see Figure 2), not only globally ($R^2 = 0.990$) or in the case of the H-bond group, but also within the dispersion complexes ($R^2 = 0.991$), and with better performance than $\Delta E_{ele} + D3$ in the case of the mixed class ($R^2 = 0.863$). Also, the RMS errors are remarkably lower (8.4 kcal mol^{-1} for H-bond complexes, 3.0 and 3.6 kcal mol^{-1} for mixed and dispersion, respectively).

The reasonable correlation of the $E_{ele}^{AB} + D3$ terms with the reference energies can be considered as an IQF validation of the additive & pairwise approximations adopted by classical potentials. As a matter of fact, the IQF partitioning shows that the adoption of such approximation for the electrostatic contribution does not require any perturbation-theory treatment. Combining the *global* electrostatics with the pairwise D3 energy yields an unbalanced description given that ΔE_{ele} comprises also the intramolecular electrostatic deformation energies, which, in turn, tend to cancel to a large extent with the intramolecular exchange-correlation effects (absent in ΔE_{ele}) as shown in the detailed IQF analysis of the S66 binding energies.[30] Therefore, E_{ele}^{AB} arises as the most relevant IQF descriptor of the electrostatic effects determining non-covalent binding.

4.3 Zeroth-order approximation in IQF electrostatics

As stated in the Theory section, for treating weakly-interacting non-covalent complexes, the zeroth-order approximation is generally invoked. The $\Delta E_{ele} + D3$ and E_{ele}^{AB} descriptors, which were also evaluated using the unrelaxed charge densities of the monomers (Table 2), allow us to further assess this approach.

Comparing Tables 1 and 2, we see that the trends in the correlation coefficients and in

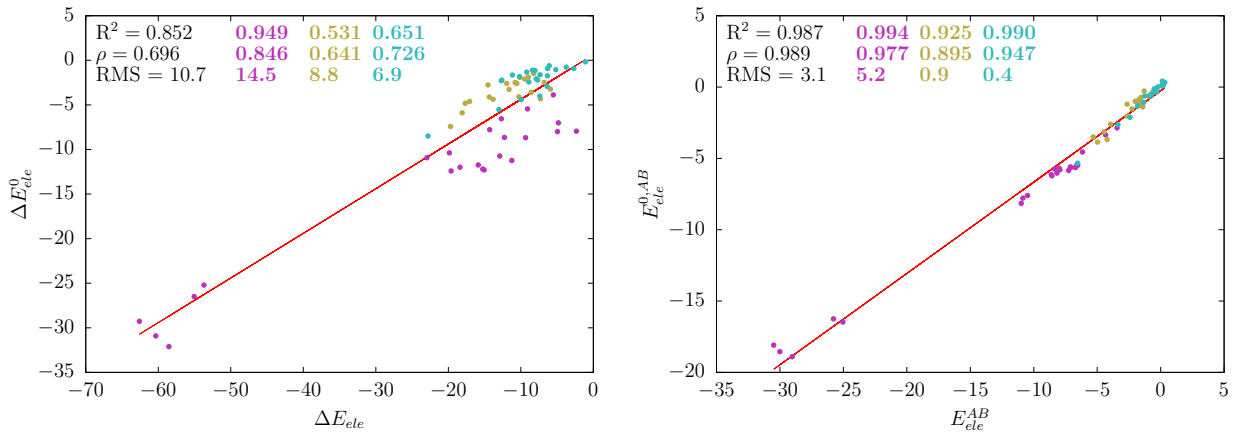
the RMS errors between the IQF terms and the reference energies are similar. The zeroth-order approximation improves the overall performance of the $\Delta E_{ele} + D3$ energies, increasing their R^2 coefficients and reducing the RMS errors to a large extent. However, the degree of accordance of the mixed complexes (R^2=0.121 in Table 2) is significantly deteriorated by using the zeroth-order density, what may be indicative of the particular importance of charge-relaxation and polarization effects in such kind of complexes. More importantly, the unrelaxed monomer densities result in purely-pairwise $E_{ele}^{0,AB}$ descriptors that correlate in a similar fashion with the benchmark binding energies than their E_{ele}^{AB} counterparts, reducing also the RMS errors. Therefore, the inter-fragment $E_{ele}^{0,AB}$ energy raises again as the most trustful electrostatic descriptor.

Table 2: Comparison between ΔE_{form}^{ref} and the D3 dispersion-corrected zeroth-order IQF electrostatic energies.

Energy term	Complex type	R^2	ρ	RMS
$\Delta E_{ele}^{0} + D3$	Global	0.956	0.898	7.090
	H-bond	0.994	0.983	9.442
	Mixed	0.121	0.684	5.109
	Dispersion	0.862	0.888	5.692
$E_{ele}^{0,AB} + D3$	Global	0.971	0.956	3.101
	H-bond	0.989	0.971	3.469
	Mixed	0.755	0.845	2.241
	Dispersion	0.988	0.992	3.346

A direct comparison between the corresponding E_{ele}^{AB} energies derived from a full QM calculation and a zeroth-order one (Figure 3) reveals a good agreement between them, with a global R^2 value of 0.987 that ranges from 0.925 to 0.994 when addressing the complex type. The errors are of a few kcal mol^{-1} due to the higher RMS of 5.2 kcal mol^{-1} for H-bond complexes, which have relatively large energies and present wider oscillations, whereas the other two groups show $RMS < 1$ kcal mol^{-1}. On the other hand, the same comparison for the full electrostatic contribution ΔE_{ele} shows a weaker correlation (still successful in the case of the H-bond complexes), with much higher errors and worse correlation coefficients (being the global R^2 of 0.852 and RMS of 10.649 kcal mol^{-1}). Therefore, the full QM relaxation

Figure 3: Electrostatic descriptors according to the zeroth-order approximation as compared with the QM relaxed energies (all in kcal mol^{-1}).



(deformation) of the charge density at the dimer geometry mainly affects the intramolecular electrostatics that collects the majority of the polarization and charge penetration effects.

For the sake of completeness, we also tested an intermediate description obtained from the antisymmetrized product of the monomer wave functions. The results can be found in the Supporting Information (Table S2). The IQF pairwise energies are approximately equivalent to those provided by the zeroth-order description. Consequently, the simpler zeroth-order approximation suffices to roughly describe the electrostatic effects in the weakly interacting dimers.

4.4 Performance of classical electrostatic potentials

The previous results point out that the interfragment IQF's E^{AB}_{ele} in combination with a dispersion potential explains quite well the binding energy of the non-covalent S66 complexes. Moreover, this electrostatic term can be reasonably approximated by the corresponding interaction energy between unperturbed fragments (i.e., $E^{0,AB}_{ele}$). Hence, we propose the latter quantity as the most suitable IQF descriptor at which approximated potentials can be addressed in order to assess their ability to capture the pairwise electrostatics of continuous charge distributions.

As noticed in Theory and Methods, the approximate MM potentials rely on the multipolar expansion of the charge density of the separate monomers, but the way of establishing the set of multipoles in a molecule is not unique. Hence, we considered on the one hand two widely-used electrostatic potentials: the AMOEBA potential based on the DMA method[35] up to the quadrupoles ($l_{max} = 2$), and the Coulombic potential evaluated with the RESP atomic charges as those used in the GAFF force field. The two potentials share the constraint of reproducing the molecular electrostatic potential. Additionally, we evaluated the performance of the regular QTAIM multipoles, which result from the multipole expansion carried out in each atomic basin at each nuclear position and are directly linked with the IQF energetic quantities. As commented previously, the requisite for multipolar convergence (i.e., the maximum radius from the origin of one multipolar distribution must not enter another one) is not always fulfilled.[33,36] Thus, only QTAIM monopoles ($l = 0$), dipoles ($l = 1$) and quadrupoles ($l = 2$) were used (an expansion up to $l = 10$ was also tested, but with dramatically differing results for strongly interacting complexes; see SI). This selection of $l_{max} = 2$ is also congruent with the AMOEBA electrostatic potential.

Table 3: Statistical measurements of the correlation between either the zeroth-order intermolecular electrostatic energy $E_{ele}^{0,AB}$ and the electrostatic interaction energies between monomers approximated by the QTAIM multipoles (truncated at $l = 2$), AMOEBA multipoles (also $l_{max} = 2$) and RESP point charges.

Multipolar approximation	Complex type	R^2	ρ	RMS
QTAIM	Global	0.970	0.958	1.0
	H-bond	0.956	0.904	1.4
	Mixed	0.644	0.768	0.9
	Dispersion	0.955	0.795	0.5
AMOEBA	Global	0.953	0.972	1.3
	H-bond	0.904	0.841	2.0
	Mixed	0.800	0.845	0.7
	Dispersion	0.939	0.893	0.4
RESP	Global	0.974	0.962	0.8
	H-bond	0.981	0.918	0.7
	Mixed	0.456	0.687	1.1
	Dispersion	0.948	0.831	0.3

We see in Table 3 that either the QTAIM or the AMOEBA multipoles show an $R^2 > 0.9$ for the full S66 set and for the H-bond/dispersion subsets. Their performance is less satisfactory for the case of the mixed complexes ($R^2 > 0.6 - 0.8$). The RMS errors, in turn, become higher for H-bonded systems having stronger interactions. Nevertheless, the RMS errors are, in general, close to 1 kcal mol^{-1}. The RESP charges (monopoles) give again electrostatic energies that are quite close to the reference $E_{ele}^{0,AB}$ values for the H-bond/dispersion structures, albeit they exhibit worse correlation for the mixed complexes ($R^2 = 0.4$).

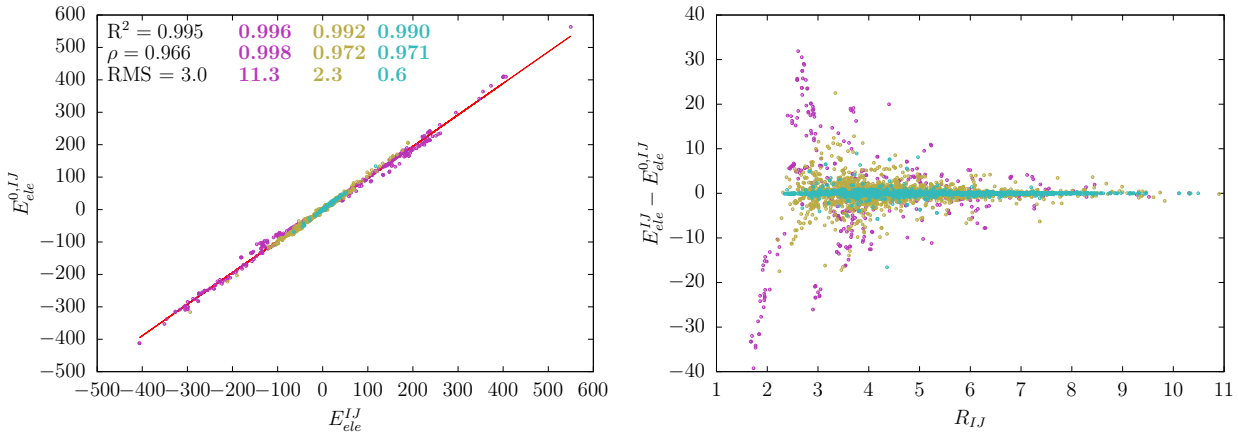
With respect to the set of RESP charges, our calculations indicate that the multipolar potentials tend to moderately improve the electrostatic description of the mixed complexes, but not for the more abundant H-bond and dispersion-dominated complexes. The good agreement between multipolar and RESP energies is probably due to the fact that the RESP charges incorporate in an effective way higher order effects due to the the parameterization procedure. The QTAIM multipoles, which are derived directly from the charge density within each atomic basin without further assumptions, allow us to assess in detail the performance of strict multipolar electrostatics. In this respect, it is interesting to note that the QTAIM and AMOEBA multipoles result in a quite similar performance, showing thus that the use of pure QTAIM multipoles in the construction of electrostatic potentials —an idea already considered in the FFLUX force field— yields accurate electrostatics free from the inclusion of other effects that can be introduced when fitting the set of the DMA multipoles to the molecular electrostatic potential. In fact, as shown by Kosov and Popelier,[36] these multipoles readily reproduce the ESP without the need of any constraint.

4.5 Electrostatic interactions at the atomic level

IQA permits an unambiguous decomposition of the intermolecular interaction energy into a sum of interatomic terms that enables a thorough analysis of the global molecular properties and an assessment of their atomic origins.

Concerning the error sources and the level of coincidence between the zeroth-order energies $E_{ele}^{0,IJ}$ with those calculated with the fully-relaxed wave function E_{ele}^{IJ}, Figure 4 reveals a high agreement between them at the atomic level. Only in the case of polar contacts the RMS error is relatively high (>10 kcal mol^{-1}), but the overall RMS remains about 3 kcal mol^{-1} due to the larger abundance of weak contacts. Figure 4 also displays the error plot of each $E_{ele}^{0,IJ}$ with respect to its corresponding counterpart E_{ele}^{IJ} as a function of the interatomic distance. In accordance with prior results showing higher errors in H-bonded complexes, the IQA partitioning emphasizes which specific polar contacts (e.g., short O-H···O H-bonds) are the major source of error. In fact, the atom-atom interactions are sometimes a few orders of magnitude higher than the inter-fragment energies, what highlights the well-known relevance of error cancellation in computing a global quantity. In this respect, Tables S3-S5 contain the atomic decomposition of E_{ele}^{AB} in the case of the acetic acid dimer to further illustrate this effect.

Figure 4: Left: Correlation between the exact atomic interactions E_{ele}^{IJ} and those subject to the zeroth-order approximation $E_{ele}^{0,IJ}$ displaying also the statistical measurements. Right: Deviation of $E_{ele}^{0,IJ}$ with respect to E_{ele}^{IJ} as a function of the interatomic distance R_{IJ}. Energies are given in kcal mol^{-1} and distances in Å.



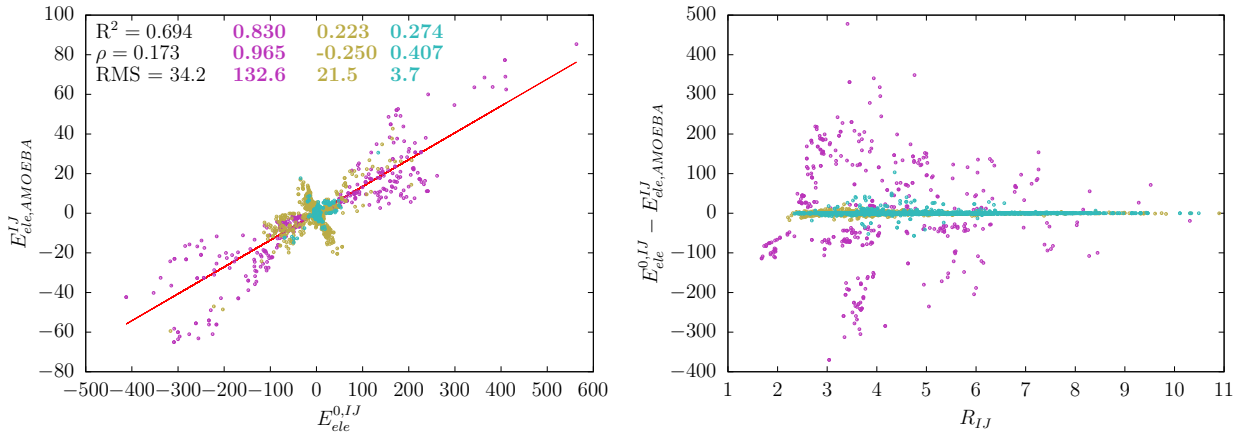
Although the molecular quality of the QTAIM-based multipolar energies is similar to those of the AMOEBA-multipoles/RESP-charges, we also examined their pairwise contributions in order to analyze the convenience of using one or another scheme to decipher the atomic origins of the global electrostatic effects.

As expected, the IQA interatomic terms correlate almost perfectly with the QTAIM multipolar ones $E^{0,IJ}_{ele,mp}$ ($R^2 = 1.000$ and $RMS = 0.6$ kcal/mol; see Figure S4), confirming thus that these are the natural heir of the parent $E^{0,IJ}_{ele}$. On the other hand, however, the AMOEBA and RESP energies are much less correlated (R^2 of 0.7 and 0.4, respectively) with the $E^{0,IJ}_{ele}$ values as shown in Figure 5, where the correlation plot between the AMOEBA atomic-pair energies $E^{IJ}_{ele,AMOEBA}$ and the corresponding $E^{0,IJ}_{ele}$ value (left), along with the distance-dependence of the associated differences (right) are depicted. The RMS errors of the energy differences are quite high ($\simeq$ 34 kcal mol^{-1}) and some error values become dramatic for strong polar contacts while others are reduced for weak interactions at the expense of loosing any evidence of correlation. Considering again the example of the acetic acid dimer, the largest differences between $E^{0,IJ}_{ele}$ and the QTAIM-multipole $E^{0,IJ}_{ele,mp}$ (about 6 kcal/mol) arise in the atoms involved in the OH···O bonds, the rest of pair interactions having much lower differences (< 0.5 kcal mol^{-1}; see Tables S6-S8). When comparing $E^{0,IJ}_{ele}$ and $E^{IJ}_{ele,AMOEBA}$ (or $E^{IJ}_{ele,RESP}$), the largest discrepancies amount to hundreds of kcal mol^{-1} and they involve not only the atom pairs in the OH···O moiety, but the methyl C atoms too.

Clearly, the dissimilarity between the IQA $E^{0,IJ}_{ele,mp}$ energies and the $E^{IJ}_{ele,AMOEBA}/E^{IJ}_{ele,RESP}$ values was not entirely unexpected given that the RESP charges are derived from the molecular ESP and the AMOEBA multipoles are obtained by the DMA protocol. In fact, a difference of one order of magnitude between the atom-atom electrostatic interactions from IQA and MM potentials has also been noticed previously.[69] The data in Figure 5 and in Tables S6-S8 show in further detail the actual discrepancies between the various atomic representations of electrostatic effects and suggest that, although the diverse atomic multipoles employed in classical potentials may yield molecular electrostatic energies that resemble to each other to a large extent, the atomic decomposition of those energies may be more questionable, what, in turn, can negatively affect the interpretation of local electrostatic interactions and/or result in artefacts while dealing with QM and MM short-range electrostatics in hybrid QM/MM methodologies. In this respect, a comparison of IQA-like $E^{0}_{ele,mp}$

energies with those produced by classical pairwise potentials may be useful to guide the construction of new potentials more consistent with QM atomic energy contributions and/or to select a particular potential.

Figure 5: Comparison between the AMOEBA atomic electrostatic interactions $E^{IJ}_{ele,AMOEBA}$ and the IQA unrelaxed atomic terms $E^{0,IJ}_{ele}$ (all in kcal mol^{-1}). On the left is the correlation plot, and on the right, the differences as a function of the distance (Å).

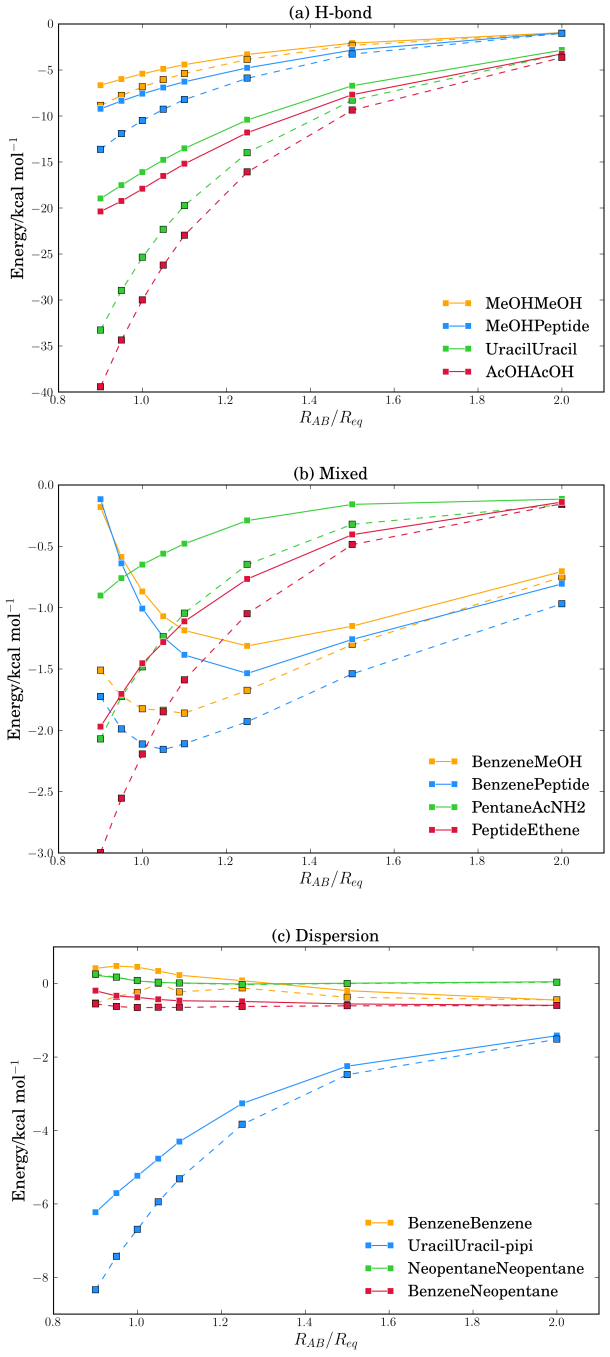


4.6 Assessment of short-range electrostatics at varying distances

A subset of complexes from the S66x8 dataset was selected to reassess at varying intermolecular separations the validity of the previous results, which were obtained at the equilibrium geometries. In particular, we reexamined the comparative performance of the zeroth-order approximation with regard to the fully-relaxed QM electrostatic energies as well as the various multipolar descriptions with respect to the zeroth-order energies.

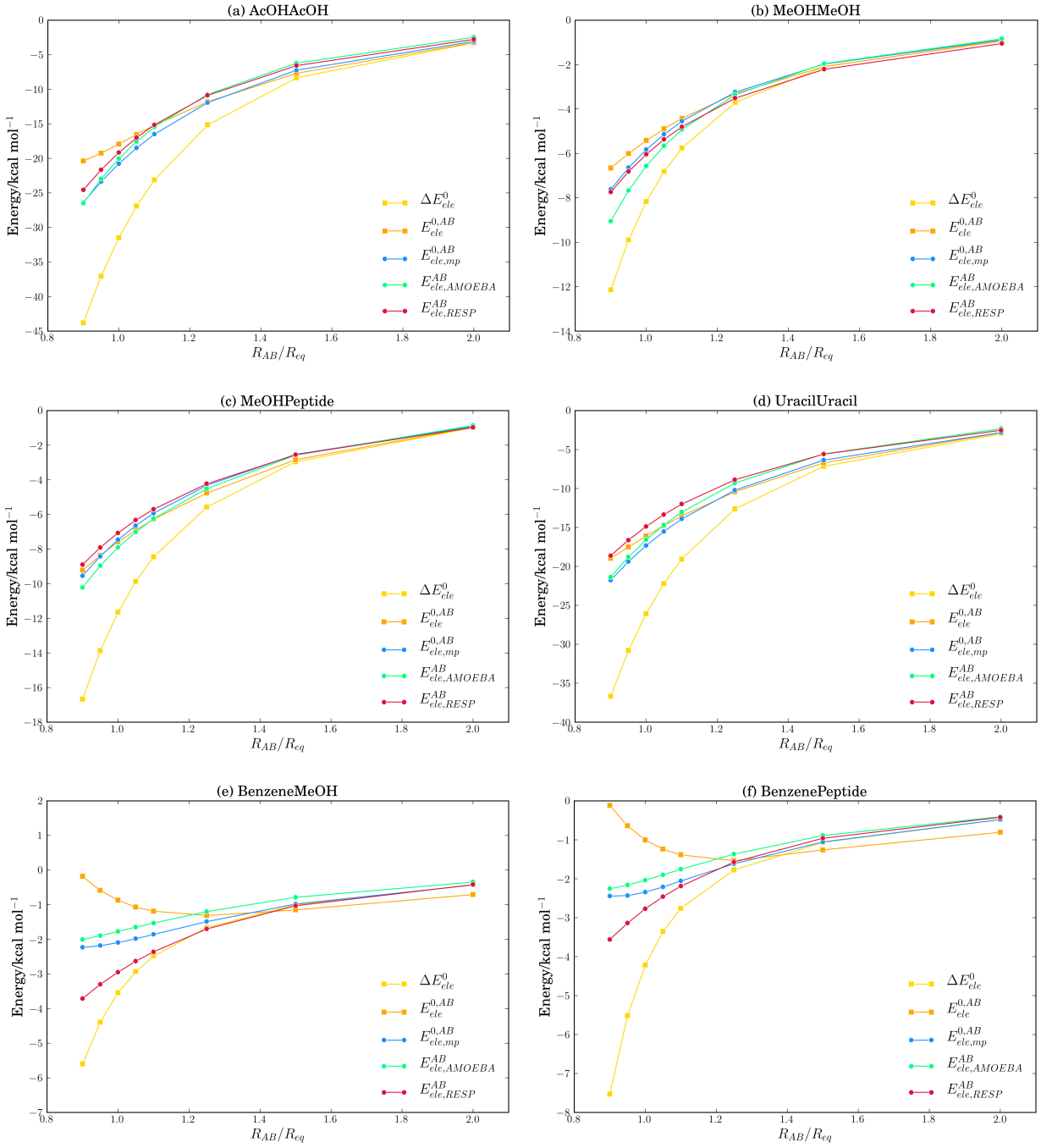
Figures 6 a-c show that the unrelaxed interfragment energies $E^{0,AB}_{ele}$ follow the same trends than the relaxed ones E^{AB}_{ele}. As expected, they start diverging as the intermolecular separation is decreased, most likely due to the strengthening of charge polarization, charge-penetration and charge-transfer effects that attenuate the pairwise electrostatic forces. The magnitude of these effects is clearly system-dependent, being almost negligible in the case of the non-polar, dispersion complexes like the benzene or neopentane dimers. The shape and slope of the E^{AB}_{ele} and $E^{0,AB}_{ele}$ curves is also indicative of the type of complexes. Thus, the

Figure 6: Intermolecular electrostatic interactions computed either exactly (E_{ele}^{AB}; dashed lines) or from frozen isolated molecular wave functions ($E_{ele}^{0,AB}$; solid lines) as a function of the intermolecular distance relative to the equilibrium one (R_{AB}/R_{eq}). The systems comprise H-bond complexes, mixed species and dispersion-dominated complexes.

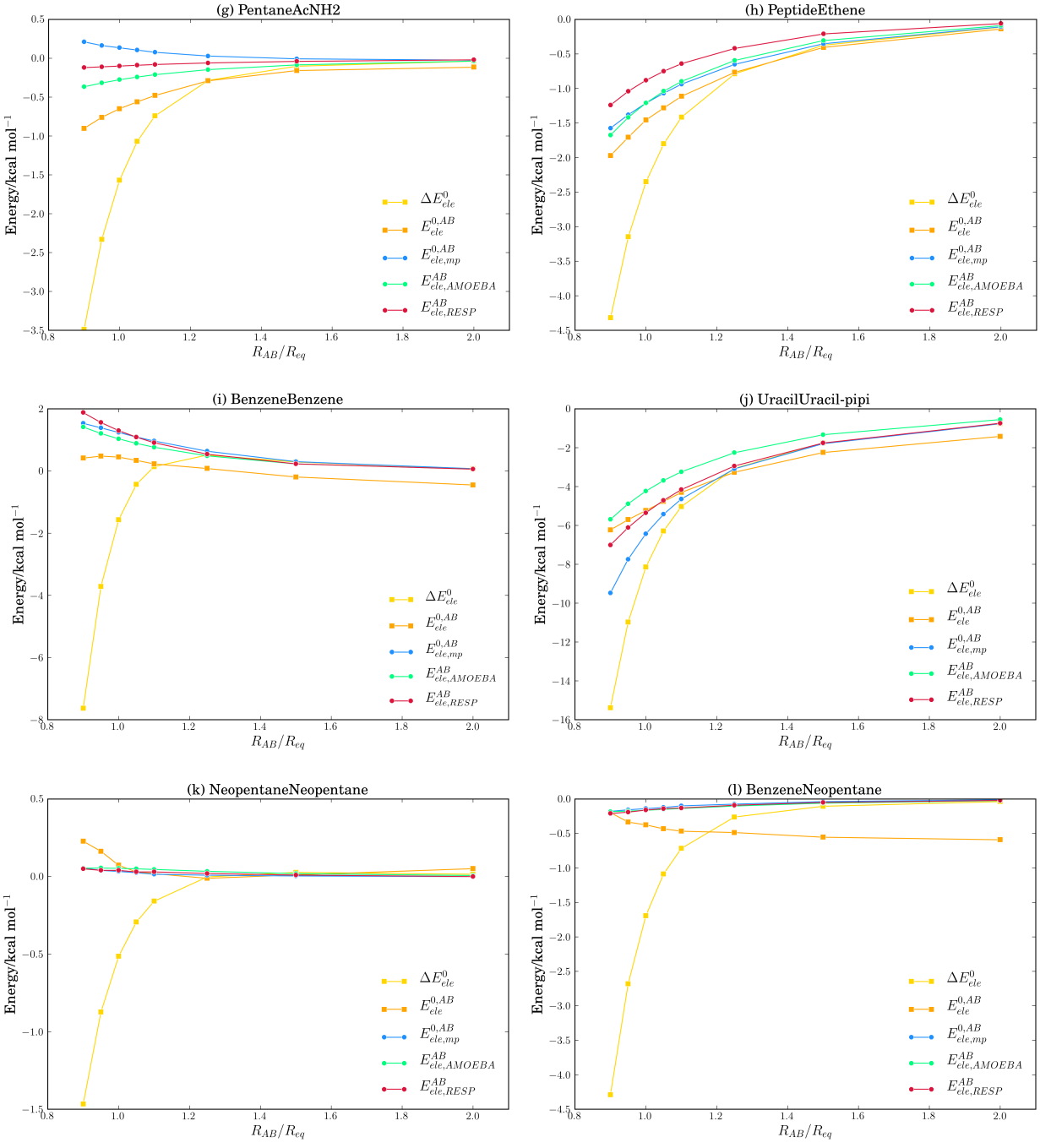


electrostatic stabilization of the four H-bond complexes and others (e.g., the π-complex of the uracil dimer) is continuously reinforced upon shortening the monomer-monomer distance,

Figure 7: Comparison between $E_{ele}^{0,AB}$, as well as ΔE_{ele}^{0}, and the multipolar energies $E_{ele,mp}^{0,AB}$, $E_{ele,AMOEBA}^{AB}$ and $E_{ele,RESP}^{AB}$ for the set of 12 S66x8 complexes selected herein. Systems a-d belong to the H-bond group, whereas e-h and i-l are part of the mixed and dispersion categories, respectively.



reflecting the major electrostatic control of these systems. In contrast, the T-shaped benzene complexes with methanol or acetamide reach an electrostatic minimum at a distance longer



than the equilibrium one while the small electrostatic energies of the dispersion dimers (i.e. +1,-1 kcal/mol) change very little along the curves (some small leaps in these curves are due to residual errors arising in the numerical integration over the atomic basins). This behavior signals the interplay of electrostatics and QM effects and the minor role of electrostatics in determining the stability of these systems.

Let us now analyze the variation with intermolecular distance of ΔE^0_{ele} and its interfragment component $E^{0,AB}_{ele}$ in comparison with the QTAIM/AMOEBA/RESP multipolar descriptions (see Figures 7). The continuous charge density electrostatics and the classical potentials coincide at the longest intermolecular distance (given by the ratio $R_{AB}/R_{eq} = 2$ between the intermolecular separation and the equilibrium one) for all the complexes. The deviation between ΔE^0_{ele} and $E^{0,AB}_{ele}$ seems particularly interesting because it reveals somehow the underlying charge penetration effects associated to the density overlap (see below). For the H-bond and some mixed complexes, the two curves decrease with lowering separation, but they split gradually from $R_{AB}/R_{eq} < 1.6$. The global ΔE^0_{ele} stabilization nearly doubles $E^{0,AB}_{ele}$ at R_{eq}, showing thus the large impact of intramolecular electrostatics as defined in the IQF framework. Besides the smaller magnitude and relevance of electrostatics, the inter- vs. intramolecular balance is differently modulated in the dispersion π complexes and in the T-shaped ones, in which the deviation between the global and interfragment electrostatics becomes significant only at very close distances, the global ΔE^0_{ele} energy being suddenly reinforced with density overlap whereas the interfragment component remains nearly constant or becomes slightly repulsive.

The plots of the classical energies (QTAIM/AMOEBA/RESP) with R_{AB}/R_{eq} show that their values and distance-dependence are basically those of the interfragment $E^{0,AB}_{ele}$ energy. A closer inspection reveals that the QTAIM/AMOEBA/RESP energies tend to overestimate/underestimate the $E^{0,AB}_{ele}$ for the H-bond/dispersion dimers, respectively. Nevertheless, their overall similarity further supports the identification of $E^{0,AB}_{ele}$ as the reference IQF descriptor for analyzing and assessing the performance of the various pairwise potentials. Furthermore, Figure 7 shows again that the QTAIM/AMOEBA multipolar terms behave in a similar fashion, no matter the scheme employed for developing the atomic multipoles. It is also remarkable that the use of RESP charges yields similar intermolecular electrostatics even at short distances.

5 Charge penetration under the QTAIM scrutiny

Following the prescriptions introduced in Theory and Methods, the zeroth-order electrostatic formation energy ΔE^0_{ele} of each S66 complex was decomposed by combining its real-space decomposition into non-overlapping atomic basins with the charge density approximation (i.e., $\rho^0_{AB} = \rho^0_A + \rho^0_B$). This strategy leads to the IQF-based charge penetration energies, $E^{IQF}_{ele,pen}$, composed by the mutual interaction between densities inside the same basin ($E^A_{ele}(\rho^0_A, \rho^0_B)$ and $E^B_{ele}(\rho^0_A, \rho^0_B)$) and an inter-basin interaction $E^{BA}_{ele}(\rho^0_A, \rho^0_B)$ between ρ^0_A lying in basin B and the opposite ρ^0_B inside basin A, as described in Eq. 18. This constitutes an *effective* penetration energy in the sense that the molecular identity between two overlapping fragments becomes necessarily blurred so that fragment properties are dependent upon the scheme followed to dissect the global charge density into its constituents. Nevertheless, the topological analysis of ρ^0_{AB} yields a consistent identification of molecular fragments so that the associated charge-penetration analysis can give useful insight into the electrostatics of non-covalent complexes.

Figures 8 a-c display the various energy contributions to ΔE^0_{ele} resulting from applying Eq. 18 to each S66 complex. On the one hand, the interfragment energy $E^{AB}_{ele}(\rho^0_A, \rho^0_B)$ between the ρ^0_A in Ω_A and ρ^0_B in Ω_B, is formally not affected by charge penetration and plays a stabilizing role in all the H-bond complexes (slightly repulsive in the dispersion complexes). On the other hand, the $E^{IQF}_{ele,pen}$ term resulting from the sum of the three IQF terms ($E^A_{ele}(\rho^0_A, \rho^0_B) + E^B_{ele}(\rho^0_A, \rho^0_B) + E^{BA}_{ele}(\rho^0_A, \rho^0_B)$) turns out to be of equal importance in H-bond complexes or even more relevant in dispersion complexes for which penetration energy describes the major part of ΔE^0_{ele}. The decomposition of the penetration energy shows that it arises mainly from the mutual attraction between the ρ^0_A fraction penetrating into Ω_B and the ρ^0_B density (including the B nuclei) populating the same Ω_B basins (i.e., $E^B_{ele}(\rho^0_A, \rho^0_B)$) as well as the counterpart effect ($E^A_{ele}(\rho^0_A, \rho^0_B)$). This relevant effect arises from the penetration of hundredths e in the opposite molecular basin at the equilibrium geometries of the complexes. As shown by the integration of ρ^0_A or ρ^0_B in the corresponding basins, the

mutual charge penetration value ranges, for instance, from 0.035 e in the neopentane dimer to 0.099 e in the case of the acetic acid dimer. There is also a minor repulsive contribution owing to the purely electronic repulsion between the penetrating ρ_A^0 into Ω_B and the ρ_B^0 tail in Ω_A, which is measured by $E_{ele}^{BA}(\rho_A^0, \rho_B^0)$. Therefore, this CP energy decomposition and the conventional IQF intra- and interfragment partitioning of ΔE_{ele}^0 stress the relevant role of the intrafragment electrostatics, what may be somewhat at odds with the intuitive perception that the classical electrostatic energy of the interacting $\rho_A^0 \cdots \rho_B^0$ is dominated by interfragment terms.

Further insight can be gained by analyzing the distance dependence of the various energy terms as shown in Figures 9 for the selected complexes. The plots confirm that the three components of E_{pen}^{IQF} tend to zero when $R_{AB}/R_{eq} > 1.5$. The same plots further highlight the role of the intrafragment terms due to the mutual attraction between the nuclei associated with one fragment and the tail of the opposite electron density located in the same basin, which largely overcomes the concomitant e-e repulsion between the ρ_A^0 and ρ_B^0 electronic parts. Interestingly, the $E_{ele}^{AB}(\rho_A^0, \rho_B^0)$ energy term, formally lacking penetration effects, is modulated by the degree of the interfragment overlap so that the decreasing trend in $E_{ele}^{AB}(\rho_A^0, \rho_B^0)$ is damped out or inverted at the shorter distances. This is not entirely unexpected given that, as two initially-separated atomic basins (e.g, $\Omega_{I\in A}$ and $\Omega_{J\in B}$) approach one another, their volume, shape and electron population evolve along the R_{AB}/R_{eq} curve in response to the density overlap. We note, however, that the deviation of $E_{ele}^{AB}(\rho_A^0, \rho_B^0)$ with respect to the interfragment electrostatic energy $E_{ele}^{0,AB}$ may constitute a useful index about the specific impact of penetration effects on the pairwise electrostatics. Actually, within the QTAIM-IQF framework, it turns out that the zeroth-order interfragment $E_{ele}^{0,AB}$ energy includes a fraction of stabilizing penetration energy for $R_{AB}/R_{eq} < 1.2$ given that, for example, the loss of some electronic ρ_A^0 density from the basins of the monomer A is partially compensated by the penetration of ρ_B^0 into the same basin. The fixed multipole values in the classical potentials somehow mimic this behavior so that they remain closer to the $E_{ele}^{0,AB}$

descriptors than to $E_{ele}^{AB}(\rho_A^0, \rho_B^0)$ around the equilibrium distance.

Figure 10 compares the IQF penetration term and other relevant energetic quantities with the analogue derived from the AMOEBA+ model for the same set of 12 complexes. Thus, the AMOEBA+ CP correction[24] $E_{pen}^{AMOEBA+}$ attempts to mimic the reference SAPT's ΔE_{ele}^0 when added to the corresponding multipolar $E_{ele,AMOEBA}^{AB}$. In the present analysis, we have used previously optimized damping functions in conjunction to our derived multipoles, leading to the approximate $\Delta E_{ele}^{AMOEBA+}$ energy. As can be appreciated in the different figures, $\Delta E_{ele}^{AMOEBA+}$ presents the same trend than ΔE_{ele}^0 with very close values. Concerning the sought CP quantities, although the AMOEBA+ reference is different to that provided by the QTAIM-IQF approach, the two penetration energies exhibit a similar behavior with R_{AB}, particularly for the more stable polar complexes. According to the IQF-QTAIM partitioning, this relationship is dominated by the intramolecular terms ($E_{ele}^A(\rho_A^0, \rho_B^0)$ and $E_{ele}^B(\rho_A^0, \rho_B^0)$), pointing out again that the penetration correction incorporates intramolecular effects to a large extent.

Figure 8: Decomposition of the S66 complexes' ΔE^0_{ele} into $E^{AB}_{ele}(\rho^0_A, \rho^0_B)$ and the three IQF penetration terms $E^{A}_{ele}(\rho^0_A, \rho^0_B)$, $E^{B}_{ele}(\rho^0_A, \rho^0_B)$ and $E^{BA}_{ele}(\rho^0_A, \rho^0_B)$.

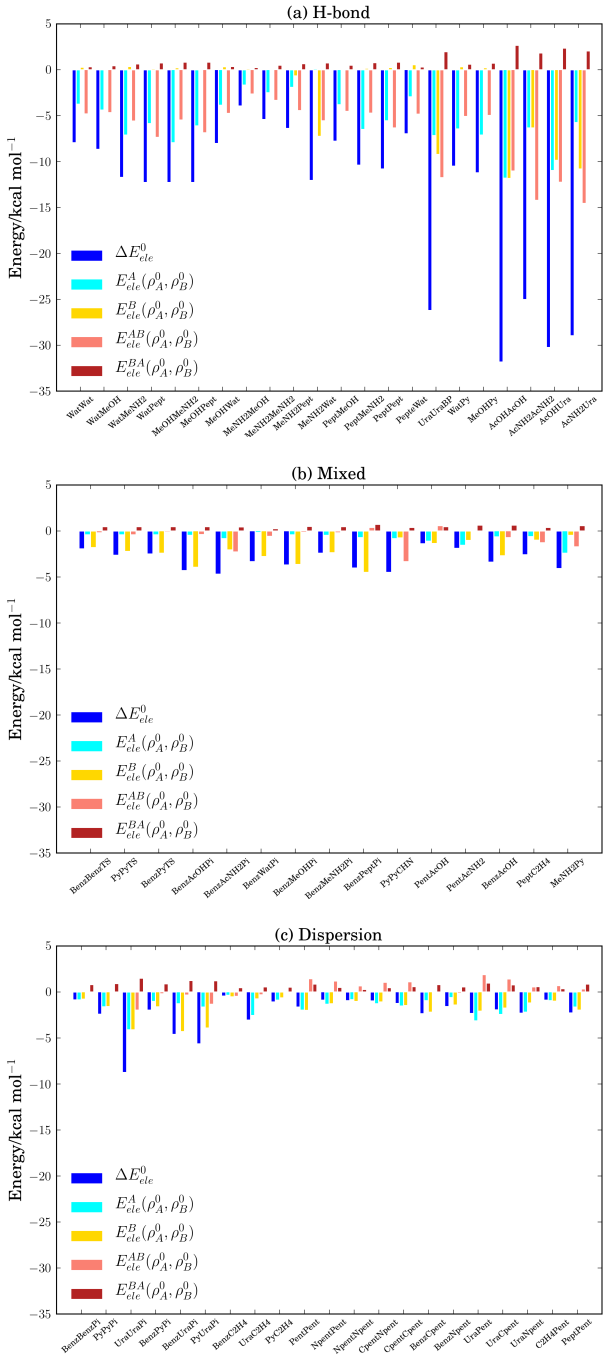
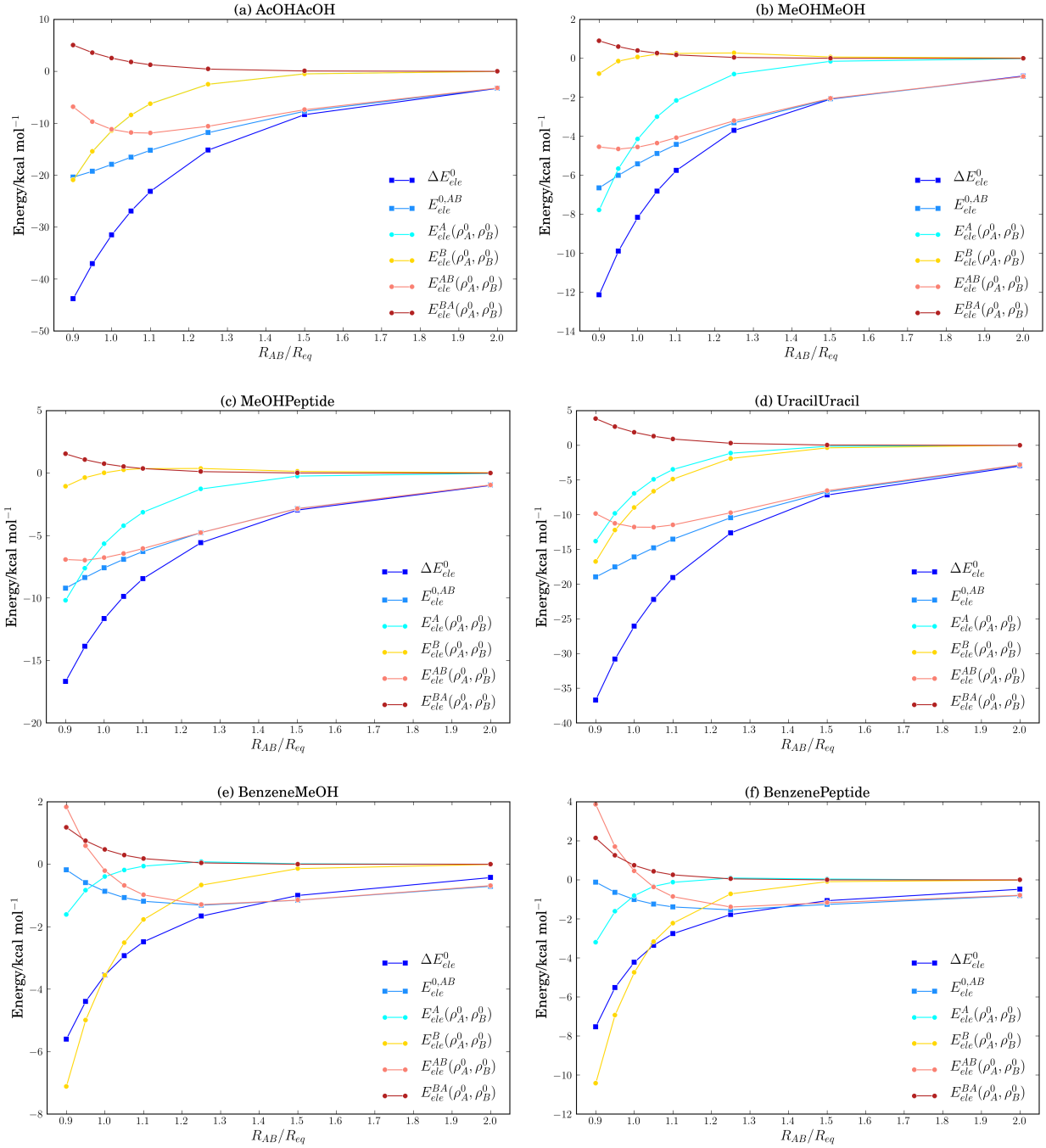
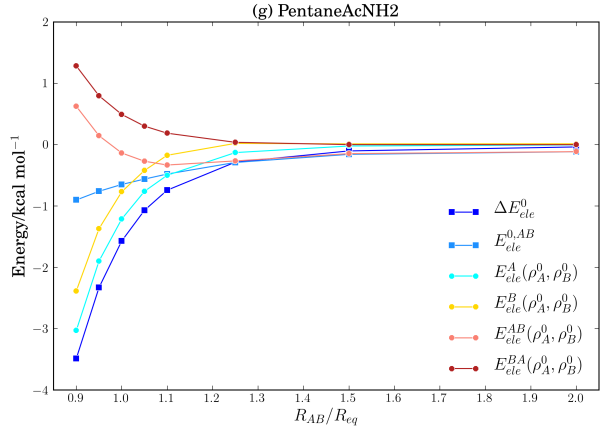
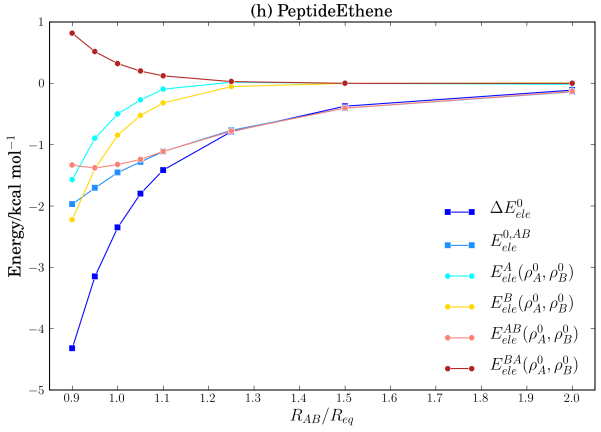
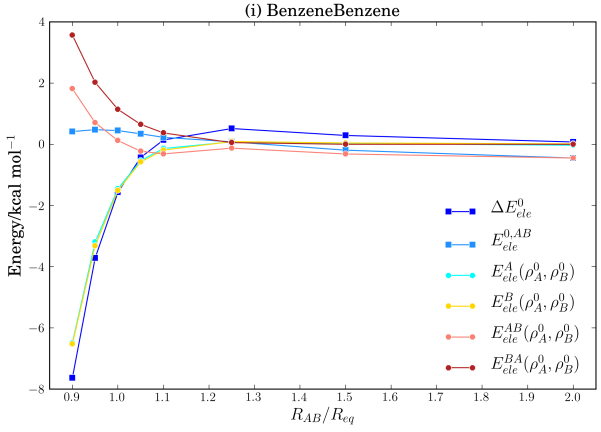


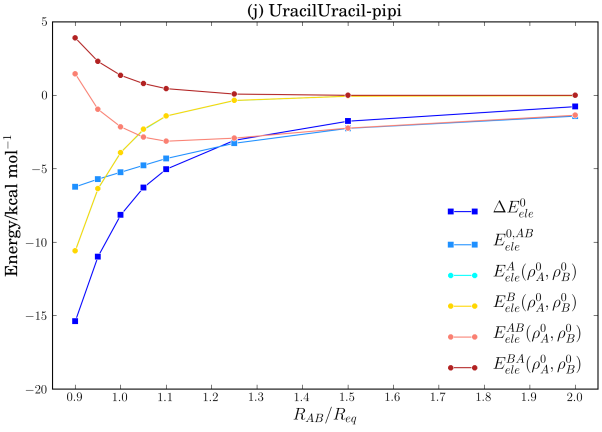
Figure 9: Evolution of the energy terms from Eq. 17, along with the $E_{ele}^{0,AB}$ IQF pair term as a function of the distance for the set of S66x8 systems chosen. Complexes a-d, e-h and i-l belong to H-bond, mixed and dispersion groups, respectively.

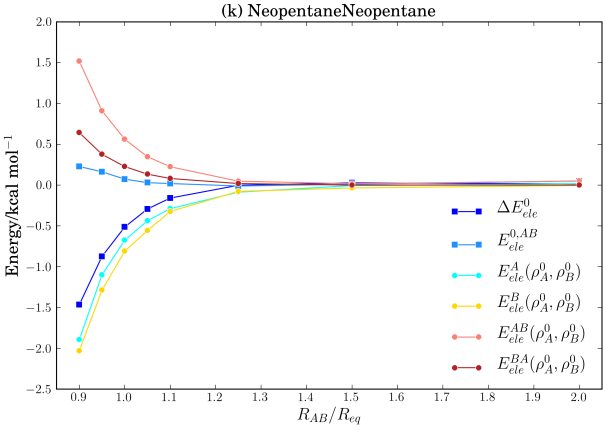



(g) PentaneAcNH2
Energy/kcal mol^{-1}
R_{AB}/R_{eq}
ΔE^0_{ele}
$E^{0,AB}_{ele}$
$E^A_{ele}(\rho^0_A,\rho^0_B)$
$E^B_{ele}(\rho^0_A,\rho^0_B)$
$E^{AB}_{ele}(\rho^0_A,\rho^0_B)$
$E^{BA}_{ele}(\rho^0_A,\rho^0_B)$


(h) PeptideEthene
Energy/kcal mol^{-1}
R_{AB}/R_{eq}
ΔE^0_{ele}
$E^{0,AB}_{ele}$
$E^A_{ele}(\rho^0_A,\rho^0_B)$
$E^B_{ele}(\rho^0_A,\rho^0_B)$
$E^{AB}_{ele}(\rho^0_A,\rho^0_B)$
$E^{BA}_{ele}(\rho^0_A,\rho^0_B)$


(i) BenzeneBenzene
Energy/kcal mol^{-1}
R_{AB}/R_{eq}
ΔE^0_{ele}
$E^{0,AB}_{ele}$
$E^A_{ele}(\rho^0_A,\rho^0_B)$
$E^B_{ele}(\rho^0_A,\rho^0_B)$
$E^{AB}_{ele}(\rho^0_A,\rho^0_B)$
$E^{BA}_{ele}(\rho^0_A,\rho^0_B)$


(j) UracilUracil-pipi
Energy/kcal mol^{-1}
R_{AB}/R_{eq}
ΔE^0_{ele}
$E^{0,AB}_{ele}$
$E^A_{ele}(\rho^0_A,\rho^0_B)$
$E^B_{ele}(\rho^0_A,\rho^0_B)$
$E^{AB}_{ele}(\rho^0_A,\rho^0_B)$
$E^{BA}_{ele}(\rho^0_A,\rho^0_B)$


(k) NeopentaneNeopentane
Energy/kcal mol^{-1}
R_{AB}/R_{eq}
ΔE^0_{ele}
$E^{0,AB}_{ele}$
$E^A_{ele}(\rho^0_A,\rho^0_B)$
$E^B_{ele}(\rho^0_A,\rho^0_B)$
$E^{AB}_{ele}(\rho^0_A,\rho^0_B)$
$E^{BA}_{ele}(\rho^0_A,\rho^0_B)$

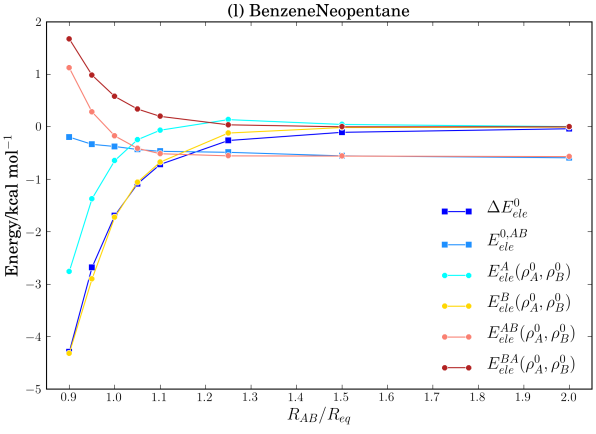
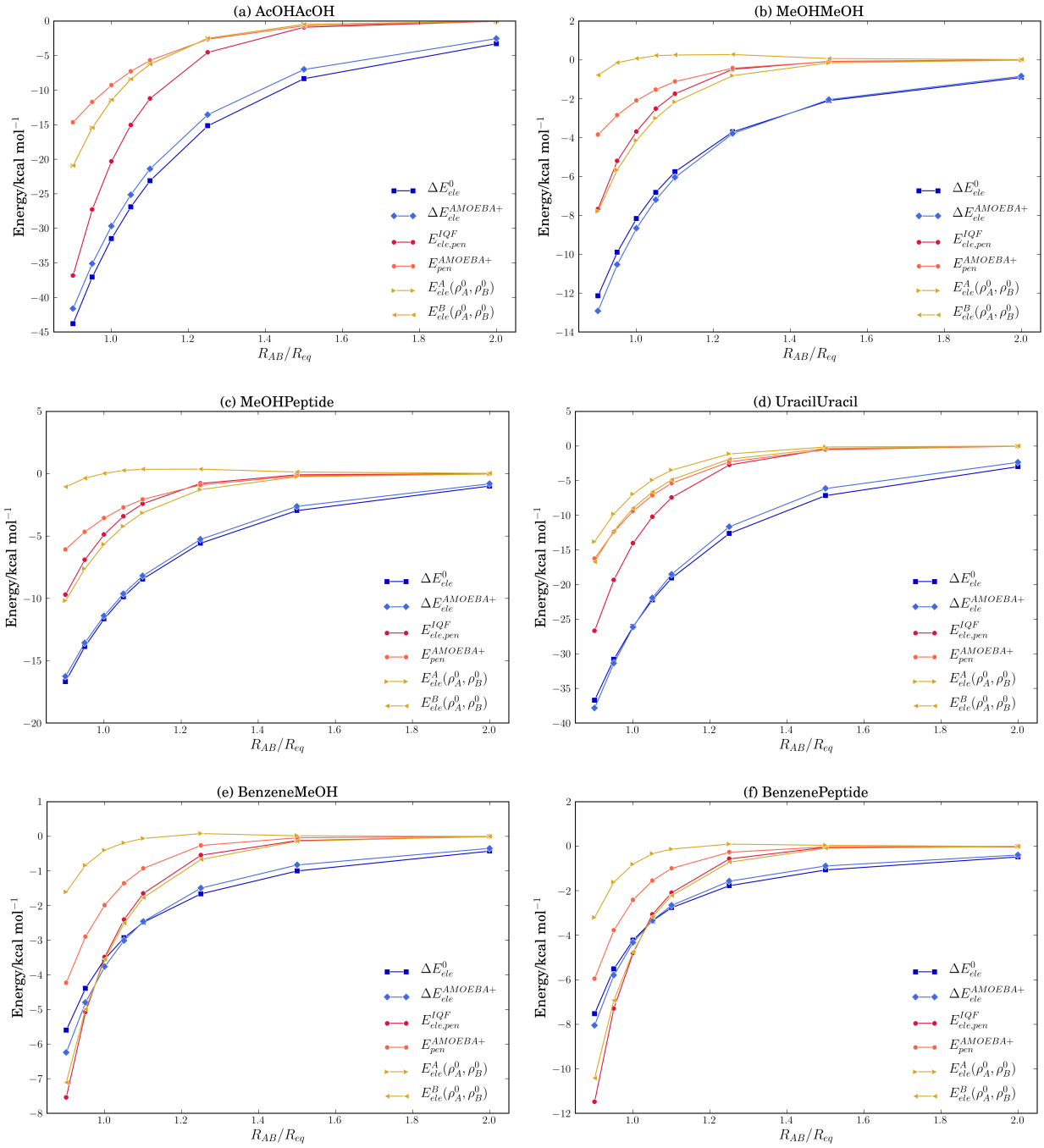
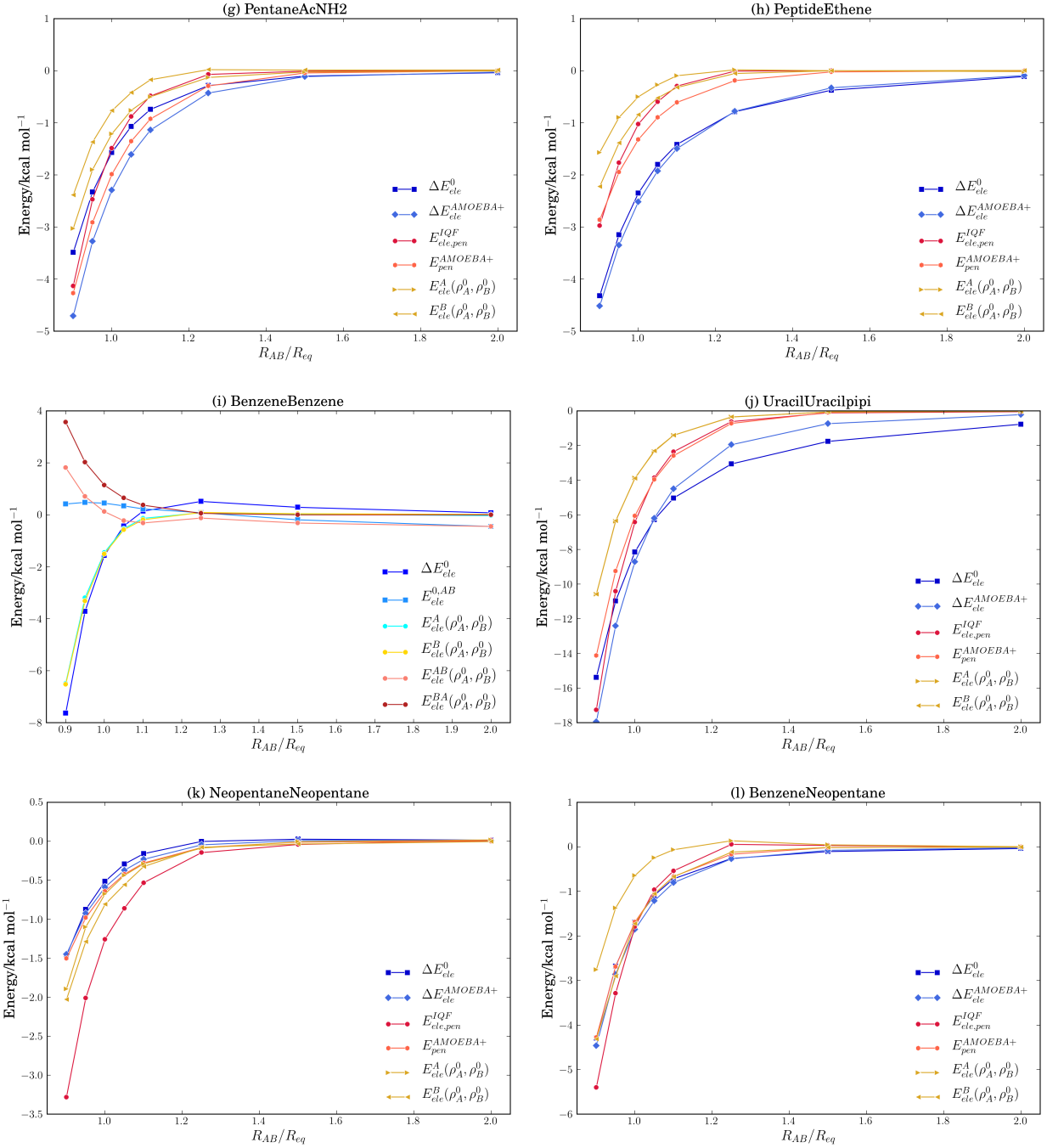

(l) BenzeneNeopentane
Energy/kcal mol^{-1}
R_{AB}/R_{eq}
ΔE^0_{ele}
$E^{0,AB}_{ele}$
$E^A_{ele}(\rho^0_A,\rho^0_B)$
$E^B_{ele}(\rho^0_A,\rho^0_B)$
$E^{AB}_{ele}(\rho^0_A,\rho^0_B)$
$E^{BA}_{ele}(\rho^0_A,\rho^0_B)$

Figure 10: Comparison between the AMOEBA+ model and the zeroth-order IQF energies.




(g) PentaneAcNH2
(h) PeptideEthene
(i) BenzeneBenzene
(j) UracilUracilpipi
(k) NeopentaneNeopentane
(l) BenzeneNeopentane
Energy/kcal mol^{-1}
R_{AB}/R_{eq}
ΔE^0_{ele}
$\Delta E^{AMOEBA+}_{ele}$
$E^{IQF}_{ele,pen}$
$E^{AMOEBA+}_{pen}$
$E^A_{ele}(\rho^0_A,\rho^0_B)$
$E^B_{ele}(\rho^0_A,\rho^0_B)$
$E^{0,AB}_{ele}$
$E^{AB}_{ele}(\rho^0_A,\rho^0_B)$
$E^{BA}_{ele}(\rho^0_A,\rho^0_B)$

6 Concluding remarks

In this work we have made extensive use of the atomic and molecular descriptors defined by the IQA/IQF energy decomposition method in order to characterize the electrostatic contributions to the formation energy of selected non-covalent complexes relevant to biomolecular systems. Besides obtaining the various inter-/intra- fragment/atomic terms, the consideration of both the fully relaxed density ρ_{AB} and the zeroth-order (unperturbed) density ρ^0_{AB}, allow us to analyze a hierarchy of approximations to the electrostatic description of those complexes, including classical potentials commonly used in MM methods.

The interfragment (pair interaction) IQF electrostatic energies E^{AB}_{ele} augmented with the pairwise D3 dispersion energy give overall binding energies that are close to the QM benchmark values. This observation indicates that E^{AB}_{ele} is the electrostatic descriptor that best correlates and represents non-covalent binding. Furthermore, the proper evaluation of the QM energies and the zeroth-order ones has determined that, indeed, $E^{0,AB}_{ele}$ is a good approximation of E^{AB}_{ele} at the equilibrium geometries of the complexes ($R^2 = 0.987$, $RMS = 3.124$ kcal mol^{-1}). Moreover, the agreement in the E^{AB}_{ele}-$E^{0,AB}_{ele}$ terms is also significant at shorter and longer separations. Since the interaction between unrelaxed fragments and the pairwise approximations are common assumptions made in MM methodologies, the pair interaction $E^{0,AB}_{ele}$ energy turns out to be the most appropriate descriptor to analyze and/or compare with electrostatic MM potentials. In particular, we have considered two widely used potentials relying on the RESP atomic point charges or the AMOEBA distributed atomic multipoles, respectively, as well as the multipolar potential up to the quadrupoles derived directly from the QTAIM basins. It has been found that the three MM pairwise approximations correlate very similarly with the zeroth-order IQF electrostatic energy at varying intermolecular distances and have small RMS errors ($<$ 1-2 kcal mol^{-1}), what somehow validates them and further highlights their pairwise nature.

Despite the similarity of the various pair interaction terms at the molecular level, the IQA partitioning of $E^{0,AB}_{ele}$ into diatomic contributions points out that the RESP or AMOEBA

atom-atom interactions and their zeroth-order IQA counterparts differ widely (e.g., R^2 of 0.413 for RESP and 0.690 for AMOEBA). Although this result is understandable in terms of the specific details of the RESP/AMOEBA charge/multipole derivations, it contrasts sharply with the nearly perfect match between the QTAIM atomic multipolar energies and the zeroth-order IQA reference energies (R^2 of 1.000 and an RMS of 0.6 kcal mol^{-1}). Again this is not surprising because of the direct relationship between the QTAIM multipoles and the ρ^0_{AB} density. However, taking into account that the QTAIM-multipole fragment energies perform very similarly to the RESP/AMOEBA ones, our analysis suggests that the development of MM electrostatic potentials based on the QTAIM multipoles —as the novel FFLUX force field does— may provide a more complete description of the electrostatic interactions, fully consistent with QM modeling at both the molecular and atomic levels.

We have also addressed the CP effect by means of a novel approach that does not require the adoption of an external reference to evaluate such effect and may shed light on its interpretation. Thus, the IQF-like electrostatic penetration energy is defined in terms of the interpenetration of the unperturbed ρ^0_A and ρ^0_B densities of the isolated monomers as measured over the real space partition of the whole ρ^0_{AB} density. In this way, the CP effect is dissected into different contributions that emphasize the role played by the attraction between the charge density associated with one fragment A and the fraction of the electron density of B localized in the basins delimiting the fragment A (and reciprocally between fragment B and A). Therefore, in the framework of IQA/IQF, charge penetration is essentially an intramolecular electrostatic effect.

The separation between the inter- and intrafragment electrostatics complemented with the QTAIM assessment of the CP effects illustrates a conceptual problem arising in some EDA approaches. As pointed out by Francisco and Pendás,[70] the use of interpenetrating densities in SAPT and other orbital-rooted EDAs *assumes that electrons in different fragments are distinguishable*, since those schemes differentiate between electrons described by particular orbitals centered on a particular nucleus/fragment, but that extend over the do-

mains of other systems with which they interact. In contrast, the QTAIM real space partition associates a portion of the whole electron density encompassed by well-defined boundaries to a given atomic basin, regardless of the origins of the orbital contributions that build up the final density ρ. This consideration does not pose an issue in the use of multipolar MM potentials, since as our analysis has shown, the interfragment electrostatic energy remains largely unaffected by the interpenetration effects in non-covalent complexes, but warns about the appropriateness of considering the interaction between interpenetrating densities as a target quantity in the development of MM potentials.

Finally, we note that our theoretical study may help clarify some practical issues related with the relevance of the CP corrections for MM potentials. As previously noticed, the IQF-D3 method shows that the basic pairwise approach requires only the zeroth-order interfragment electrostatic energy, which is combined with empirical dispersion terms in order to predict reasonable formation energies for non-covalent complexes. In this way, adding charge-penetration corrections to MM potentials like RESP/AMOEBA, which only require interfragment electrostatic energy, would result in an unbalanced description. This aspect, which seems to have been overlooked in previous works,[15,48] implies also that the electrostatic energy employed in popular MM force fields (AMBER, CHARMM...) cannot be compared with the global ΔE^0_{ele} energy derived from continuous charge distributions, but with its interfragment component.

On the other hand, charge penetration corrections have been derived to improve the description of the QM-MM electrostatic interactions in hybrid QM/MM methodologies. In this case, such corrections should mitigate short-range electrostatic artefacts, particularly those associated to the QM-MM covalent linkages. However, considering the highly-dissimilar interatomic electrostatic energies produced by the QM densities and the RESP/AMOEBA potentials, the usage of electrostatic parameters more akin to the QM densities at the atomic level may have a larger impact in improving the QM-MM electrostatics. Concerning the novel MM potentials inspired by the QM SAPT methodology, it is clear that the multipolar

electrostatics (interfragment) must be augmented by the CP potentials (intrafragment) if one seeks to reproduce the global electrostatics ΔE^0_{ele}. Nevertheless, the IQF/IQA approach (and other EDAs) points out that the intramolecular electrostatic energy is closely related with other energy changes induced by fragment overlap (e.g., intrafragment deformation and interfragment exchange-correlation energy), suggesting thus that the separate treatment of these effects by means of independent potential terms might be inefficient and hamper parameter development and transferability.

Acknowledgement

F.J.-G. and D.S. acknowledge the Spanish MICINN (grant PGC2018-095953-B-I00), the FICyT (grant FC-GRUPIN-IDI/2018/000177), and the European Union FEDER for financial support. F.J.-G. specially acknowledges the Spanish MICINN for predoctoral grant (BES-2016-076986). Both authors are greatly grateful to Evelio Francisco for his contribution with the development of the `mpolint` code and to Ángel Martín Pendás for his valuable suggestions and careful reading of the manuscript.

References

(1) Cisneros, G. A.; Karttunen, M.; Ren, P.; Sagui, C. Classical Electrostatics for Biomolecular Simulations. *Chemical Reviews* **2013**, *114*, 779–814.

(2) Monticelli, L.; Tieleman, D. P. *Methods in Molecular Biology*; Humana Press, 2012; pp 197–213.

(3) Leach, A. R. *Molecular Modelling: Principles and Applications*, 1st ed.; Addison Wesley Longman Limited: Essex CM20 2JE, England, 1996.

(4) Ren, P.; Chun, J.; Thomas, D. G.; Schnieders, M. J.; Marucho, M.; Zhang, J.;

Baker, N. A. Biomolecular electrostatics and solvation: a computational perspective. *Quarterly Reviews of Biophysics* **2012**, *45*, 427–491.

(5) Ewald, P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physik* **1921**, *369*, 253–287.

(6) Hockney, R. W.; Eastwood, J. W. *Computer Simulation Using Particles*; IOP Publishing Ltd, 1988.

(7) Pollock, E.; Glosli, J. Comments on P3M, FMM, and the Ewald method for large periodic Coulombic systems. *Computer Physics Communications* **1996**, *95*, 93–110.

(8) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: AnN·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **1993**, *98*, 10089–10092.

(9) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **1995**, *103*, 8577–8593.

(10) Herce, D. H.; Perera, L.; Darden, T. A.; Sagui, C. Surface solvation for an ion in a water cluster. *The Journal of Chemical Physics* **2005**, *122*, 024513.

(11) Stone, A. *The Theory of Intermolecular Forces*; Oxford University Press, 2013.

(12) Freitag, M. A.; Gordon, M. S.; Jensen, J. H.; Stevens, W. J. Evaluation of charge penetration between distributed multipolar expansions. *The Journal of Chemical Physics* **2000**, *112*, 7300–7306.

(13) Wang, B.; Truhlar, D. G. Including Charge Penetration Effects in Molecular Modeling. *Journal of Chemical Theory and Computation* **2010**, *6*, 3330–3342, PMID: 26617087.

(14) Wang, Q.; Rackers, J. A.; He, C.; Qi, R.; Narth, C.; Lagardere, L.; Gresh, N.; Ponder, J. W.; Piquemal, J.-P.; Ren, P. General Model for Treating Short-Range Electro-

static Penetration in a Molecular Mechanics Force Field. *Journal of Chemical Theory and Computation* **2015**, *11*, 2609–2618.

(15) Bojarowski, S. A.; Kumar, P.; Dominiak, P. M. A Universal and Straightforward Approach to Include Penetration Effects in Electrostatic Interaction Energy Estimation. *ChemPhysChem* **2016**, *17*, 2455–2460.

(16) Francisco, E.; Pendás, A. M. *Non-Covalent Interactions in Quantum Chemistry and Physics*; Elsevier, 2017; pp 27–64.

(17) Zhao, L.; von Hopffgarten, M.; Andrada, D. M.; Frenking, G. Energy decomposition analysis. *WIREs Computational Molecular Science* **2017**, *8*.

(18) Patkowski, K. Recent developments in symmetry-adapted perturbation theory. *WIREs Computational Molecular Science* **2019**, *10*.

(19) Jeziorski, B.; Moszynski, R.; Szalewicz, K. *Chemical Reviews* **1994**, *94*, 1887–1930.

(20) Kitaura, K.; Morokuma, K. A new energy decomposition scheme for molecular interactions within the Hartree-Fock approximation. *International Journal of Quantum Chemistry* **1976**, *10*, 325–340.

(21) Bickelhaupt, F. M.; Baerends, E. J. *Reviews in Computational Chemistry*; John Wiley & Sons, Inc., 2007; pp 1–86.

(22) Blanco, M. A.; Pendás, A. M.; Francisco, E. Interacting Quantum Atoms: A Correlated Energy Decomposition Scheme Based on the Quantum Theory of Atoms in Molecules. *Journal of Chemical Theory and Computation* **2005**, *1*, 1096–1109.

(23) Francisco, E.; Pendás, A. M.; Blanco, M. A. A Molecular Energy Decomposition Scheme for Atoms in Molecules. *Journal of Chemical Theory and Computation* **2005**, *2*, 90–102.

(24) Liu, C.; Piquemal, J.-P.; Ren, P. AMOEBA Classical Potential for Modeling Molecular Interactions. *Journal of Chemical Theory and Computation* **2019**, *15*, 4122–4139.

(25) Vandenbrande, S.; Waroquier, M.; Speybroeck, V. V.; Verstraelen, T. The Monomer Electron Density Force Field (MEDFF): A Physically Inspired Model for Noncovalent Interactions. *Journal of Chemical Theory and Computation* **2016**, *13*, 161–179.

(26) Pendás, A. M.; Casals-Sainz, J. L.; Francisco, E. On Electrostatics, Covalency, and Chemical Dashes: Physical Interactions versus Chemical Bonds. *Chemistry – A European Journal* **2018**, *25*, 309–314.

(27) Řezáč, J.; Riley, K. E.; Hobza, P. S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *Journal of Chemical Theory and Computation* **2011**, *7*, 2427–2438.

(28) Řezáč, J.; Riley, K. E.; Hobza, P. Extensions of the S66 Data Set: More Accurate Interaction Energies and Angular-Displaced Nonequilibrium Geometries. *Journal of Chemical Theory and Computation* **2011**, *7*, 3466–3470.

(29) Bader, R. *Atoms in molecules : a quantum theory*; Clarendon Press: Oxford New York, 1990.

(30) Suárez, D.; Díaz, N.; Francisco, E.; Martín Pendás, A. Application of the Interacting Quantum Atoms Approach to the S66 and Ionic-Hydrogen-Bond Datasets for Noncovalent Interactions. *ChemPhysChem* **2018**, *19*, 973–987.

(31) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of Chemical Physics* **2010**, *132*, 154104.

(32) Becke, A. D.; Johnson, E. R. A density-functional model of the dispersion interaction. *The Journal of Chemical Physics* **2005**, *123*, 154101.

(33) Pendás, A. M.; Blanco, M. A.; Francisco, E. Two-electron integrations in the quantum theory of atoms in molecules. *The Journal of Chemical Physics* **2004**, *120*, 4581–4592.

(34) Stone, A. Distributed multipole analysis, or how to describe a molecular charge distribution. *Chemical Physics Letters* **1981**, *83*, 233–239.

(35) Stone, A.; Alderton, M. Distributed multipole analysis. *Molecular Physics* **1985**, *56*, 1047–1064.

(36) Popelier, P. L. A.; Kosov, D. S. Atom–atom partitioning of intramolecular and intermolecular Coulomb energy. *The Journal of Chemical Physics* **2001**, *114*, 6539–6547.

(37) Rafat, M.; Popelier, P. L. A. A convergent multipole expansion for 1, 3 and 1, 4 Coulomb interactions. *The Journal of Chemical Physics* **2006**, *124*, 144102.

(38) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An overview of the Amber biomolecular simulation package. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012**, *3*, 198–210.

(39) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry* **2009**, NA–NA.

(40) Oostenbrink, C.; Soares, T. A.; van der Vegt, N. F. A.; van Gunsteren, W. F. Validation of the 53A6 GROMOS force field. *European Biophysics Journal* **2005**, *34*, 273–284.

(41) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society* **1996**, *118*, 11225–11236.

(42) Engkvist, O.; Åstrand, P.-O.; Karlström, G. Accurate Intermolecular Potentials Obtained from Molecular Wave Functions: Bridging the Gap between Quantum Chemistry and Molecular Simulations. *Chemical Reviews* **2000**, *100*, 4087–4108.

(43) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current Status of the AMOEBA Polarizable Force Field. *The Journal of Physical Chemistry B* **2010**, *114*, 2549–2564.

(44) Popelier, P. L. A. QCTFF: On the construction of a novel protein force field. *International Journal of Quantum Chemistry* **2015**, *115*, 1005–1011.

(45) Konovalov, A.; Symons, B. C.; Popelier, P. L. On the many-body nature of intramolecular forces in FFLUX and its implications. *Journal of Computational Chemistry* **2020**, *42*, 107–116.

(46) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. Anisotropic, Polarizable Molecular Mechanics Studies of Inter- and Intramolecular Interactions and Ligand-Macromolecule Complexes. A Bottom-Up Strategy. *Journal of Chemical Theory and Computation* **2007**, *3*, 1960–1986.

(47) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. The Effective Fragment Potential Method: A QM-Based MM Approach to Modeling Environmental Effects in Chemistry. *The Journal of Physical Chemistry A* **2001**, *105*, 293–307.

(48) Rackers, J. A.; Wang, Q.; Liu, C.; Piquemal, J.-P.; Ren, P.; Ponder, J. W. An optimized charge penetration model for use with the AMOEBA force field. *Physical Chemistry Chemical Physics* **2017**, *19*, 276–291.

(49) Bojarowski, S. A.; Kumar, P.; Dominiak, P. M. Interplay of point multipole moments and charge penetration for intermolecular electrostatic interaction energies from the University at Buffalo pseudoatom databank model of electron density. *Acta Crystallographica Section B Structural Science, Crystal Engineering and Materials* **2017**, *73*, 598–609.

(50) 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*, 140–144.

(51) Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *The Journal of Chemical Physics* **1989**, *90*, 1007–1023.

(52) 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*, 1347–1363.

(53) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of Chemical Physics* **2010**, *132*, 154104.

(54) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *Journal of Computational Chemistry* **2011**, *32*, 1456–1465.

(55) Goerigk, L.; Collyer, C. A.; Reimers, J. R. Recommending Hartree–Fock Theory with London-Dispersion and Basis-Set-Superposition Corrections for the Optimization or Quantum Refinement of Protein Structures. *The Journal of Physical Chemistry B* **2014**, *118*, 14612–14626.

(56) Sedlak, R.; Řezáč, J. Empirical D3 Dispersion as a Replacement for ab Initio Dispersion Terms in Density Functional Theory-Based Symmetry-Adapted Perturbation Theory. *Journal of Chemical Theory and Computation* **2017**, *13*, 1638–1646.

(57) Martín Pendás, Á.; Francisco, E. PROMOLDEN. A QTAIM/IQA code (unpublished).

(58) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *Journal of Computational Chemistry* **2004**, *25*, 1157–1174.

(59) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian 03, Revision C.02. Gaussian, Inc., Wallingford, CT, 2004.

(60) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling* **2006**, *25*, 247–260.

(61) Ren, P.; Wu, C.; Ponder, J. W. Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules. *Journal of Chemical Theory and Computation* **2011**, *7*, 3143–3161.

(62) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.;

Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Jr.,; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; ; Fox, D. J. Gaussian09 Revision D.01. Gaussian Inc., Wallingford CT, 2013.

(63) Stone, A. J. Distributed Multipole Analysis: Stability for Large Basis Sets. *Journal of Chemical Theory and Computation* **2005**, *1*, 1128–1132, PMID: 26631656.

(64) Rackers, J. A.; Wang, Z.; Lu, C.; Laury, M. L.; Lagardère, L.; Schnieders, M. J.; Piquemal, J.-P.; Ren, P.; Ponder, J. W. Tinker 8: Software Tools for Molecular Design. *Journal of Chemical Theory and Computation* **2018**, *14*, 5273–5289.

(65) Francisco, E. MPOLINT. A program to calculate Coulombic interactions by means of the multipole approximation (unpublished).

(66) Eaton, J. W.; Bateman, D.; Hauberg, S.; Wehbring, R. GNU Octave version 4.4.0 manual: a high-level interactive language for numerical computations. 2018.

(67) Williams, T.; Kelley, C.; many others, Gnuplot 5.2: an Interactive Plotting Program. `http://gnuplot.sourceforge.net/`, 2019.

(68) Hunter, J. D. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering* **2007**, *9*, 90–95.

(69) Yuan, Y.; Mills, M. J. L.; Popelier, P. L. A. Multipolar electrostatics for proteins: Atom-atom electrostatic energies in crambin. *Journal of Computational Chemistry* **2013**, *35*, 343–359.

(70) Francisco, E.; Pendás, A. M. *Non-Covalent Interactions in Quantum Chemistry and Physics*; Elsevier, 2017; pp 27–64.