

Application of the Interacting Quantum Atoms Approach to the S66 and Ionic-HB Datasets for Noncovalent Interactions

D. Suárez, N. Díaz, E. Francisco, A. Martín Pendás.*

Departamento de Química Física y Analítica

Universidad de Oviedo

Julián Clavería 8. 33006 Oviedo (Asturias) Spain

E-mail: dimas@uniovi.es

Keywords: Ab initio calculations • Bond theory • Computational Chemistry • Noncovalent interactions • Quantum Chemistry

ABSTRACT The interacting quantum atoms (IQA) method can assess, systematically and in great detail, the strength and physics of both covalent and noncovalent interactions. The lack of a pair density in density functional theory (DFT), which precludes the direct IQA decomposition of the characteristic exchange-correlation energy, has been recently overcome by means of a scaling technique, what can largely expand the applicability of the method. To better assess the utility of the augmented IQA methodology in order to derive quantum chemical decompositions at the atomic and molecular levels, we report the results of HF and DFT calculations on the complexes included in the S66 and the ionic H-bond databases of benchmark geometry and binding energies. For all the structures, we perform single-point and geometry optimizations using HF and selected DFT methods with triple- ζ basis sets followed by full IQA calculations. Pairwise dispersion energies are accounted for by the D3 method. We analyze the goodness of the HF-D3 and DFT-D3 binding energies, the magnitude of numerical errors, the fragment and atomic distribution of formation energies, etc. It is shown that fragment-based IQA decomposes the formation energies in comparable terms to those of perturbative approaches and that the atomic IQA energies hold the promise of rigorously quantifying atomic and group energy contributions in larger biomolecular systems.

Introduction

The development of the Quantum Theory of Atoms in Molecules (QTAIM) has paved the way towards rigorous concepts in the theory of chemical bonding that rely on the analysis of the topology of scalar fields containing chemical information.^[1-2] In this way, a broad array of Quantum Chemical Topology (QCT) methods^[3] has been developed that provide a wealth of quantum mechanical (QM) molecular descriptors (*e.g.*, bond-critical-point properties, atomic and inter-atomic properties from which similarity measurements can be computed). Furthermore, insightful visualization and characterization of both covalent and non-covalent interactions has been greatly enhanced by analyzing the topological properties of the electron density and its reduced gradient.^[4] New quantitative structure-activity-relationships (QSAR) have been formulated using QCT properties like the localization/delocalization matrix, whose diagonal and off-diagonal elements encapsulate one- and two-electron information about the charge distribution.^[5]

Among the various QCT methods, the interacting quantum atoms (IQA) approach^[6-7] partitions the first- and second-order reduced density matrices into the atomic regions that naturally arise in the QTAIM as the attraction basins (Ω_A) of the gradient field of the electron density. Such partitioning expresses the total energy of a molecular system as the sum of one- and two-center atomic contributions with precise, physical meaning. Thus the one-center terms (self-atomic energies, $E(\Omega_A)$) tend to the free atomic energies when the interacting atoms are far apart, while the pairwise terms ($E(\Omega_A, \Omega_B)$) include a classical electrostatic term, and an exchange-correlation energy term, which accounts for all quantum mechanical effects. This energy decomposition, which is orbital invariant, provides results that do not depend on the underlying electronic structure method and treats both intra- and intermolecular interactions, has been applied in many computational studies to quantify different aspects of chemical bonds and intermolecular forces (*e.g.*, the nature and cooperativity of H-bond interactions,^[8-11]

halogen bonding patterns,^[12] interactions within transition metal complexes^[13-14], description of short-range repulsions,^[15] fine-tuning effects of electron correlation within covalent and non-bonded interactions,^[16] etc.).

In principle IQA calculations require both the first- and second-order density matrices defined within wavefunction methods. Thus, the implementation of IQA in the PROMOLDEN code^[17] is able to process wavefunctions derived from several electronic-structure methods: canonical Hartree Fock (HF), multiconfigurational complete active space (CAS),^[7, 18] coupled-cluster (CC),^[19] and full configuration interaction (FCI). The Kohn-Sham formalism of density functional theory (DFT) does not yield well-defined second-order density matrices, what prevents a direct IQA application within DFT. One- and two-atom partitions of the DFT energies have already been proposed in the framework defined by the Hilbert space of atom-centered basis functions^[20-21] and the concept of fuzzy atoms.^[22] Within the IQA approach, it has also been realized that the lack of second-order DFT density matrices can be circumvented by defining *ad hoc* additive exchange-correlation (xc) energies and use them to scale the one- and two-atom terms of the Kohn–Sham xc energy such that the total DFT energy is exactly recovered.^[23] In this way, the applicability range of IQA is significantly expanded as modern DFT methods are the default choice for describing many classes of chemical systems. Furthermore, IQA analysis can be enhanced using DFT (and HF) Hamiltonians augmented with the popular Grimme’s D3 potential,^[24-25] which yields pairwise dispersion energies $E_{disp}(A-B)$ that can be readily used to complement the interaction $E(\Omega_A, \Omega_B)$ IQA contributions, setting up an effective IQA-D3 decomposition scheme.

A robust and computationally efficient IQA-D3 implementation would offer the possibility of gaining new insight and energetic understanding from QM studies of large non-covalent complexes and biomolecules, for example, by outlining the energy impact of specific inter- and intramolecular interactions, formulating and providing data for new quantitative structure-

activity relationships (QSAR)^[5], etc. However, these and other possible applications to larger systems still face two major challenges. First, IQA is computationally very expensive due to six-dimensional numerical integrations over the irregularly-shaped atomic basins that involve the first and second order density matrices. Secondly, numerical errors, which affect the accuracy of the total QM energy reconstructed by IQA, arise in the construction of the interatomic surfaces delimiting the atomic basins and in the radial and angular numerical quadratures within those basins. Owing to the computational cost and, to a lesser extent, the numerical errors, application of IQA to non-covalent complexes and biomolecules has been scarce and, therefore, further computational work is still needed to assess its strengths and weaknesses in this area.

As a first step in the pathway to routine IQA applications to the study of intermolecular complexes, we report in this work the IQA analysis of a large family of non-covalent complexes. In particular, we considered the popular S66 and S66x8 databases of benchmark geometry and binding energies constructed by Hobza and co-workers.^[26-27] This database provides reference data for 66 model complexes formed between 14 neutral monomers, which bear typical polar and non-polar functional groups characteristic of biomolecules and/or constitute biomolecular building blocks. Interactions between neutral and charged fragments are also examined using the ionic H-bond database,^[28] which contains high level reference data for 15 dimers that are described following similar prescriptions to those used in the S66 set. For all the structures, we perform single-point and geometry optimizations using HF and selected DFT-D3 methods with a triple- ζ basis set followed by full IQA calculations. Focusing on relative trends, we analyze first the goodness of the HF-D3 and DFT-D3 formation energies of the dimers as compared with the high level reference data. Then we examine the magnitude of numerical errors in the IQA-reconstructed formation energies, the relative weight of intra- and intermolecular contributions to formation energies, the imprint of the mode of binding on

the IQA-D3 descriptors, their similarities and differences depending on the level of theory, the partitioning of formation and deformation energies, and the variation of the fragment-based IQA-D3 terms along the dissociation curves provided by the S66x8 data set for selected complexes. Altogether these results enlarge the application domain of IQA with respect to former applications both in terms of the nature of the examined systems and the levels of theory used, allow us to make recommendations of model chemistries and IQA settings for similar systems, and could help guide the design and validation of linear scaling IQA-D3 solutions that will be required to study larger non-covalent complexes and biomolecular systems.

Methods

IQA energy decomposition

The QTAIM atomic basins (Ω_A), which stem from the topological properties of the charge distribution $\rho(\mathbf{r})$, constitute proper open subsystems for which expectation values of QM operators can be computed.^[1] Particularly, the sum of the one-domain atomic energies,

$$E = \sum_A E_{\Omega_A} ,$$

yields the total energy for the complete system, but does not allow an appropriate consideration of interatomic or intermolecular interactions. This can be achieved with the IQA approach,^[7] which relies on the availability of two scalar fields derived from the wavefunction, the first order reduced density matrix $\rho_1(\mathbf{r}_1, \mathbf{r}_1')$ and the pair density, $\rho_2(\mathbf{r}_1, \mathbf{r}_2)$, as well as on the exhaustive partition of the real space (*e.g.*, the partitioning defined by QTAIM). Then IQA decomposes the total energy of a molecular system within the Born-Oppenheimer approximation as

$$E = \sum_A E_{net}^A + \sum_{A>B} E_{int}^{AB} = \sum_A (T^A + V_{ne}^{AA} + V_{ee}^{AA}) + \sum_{A>B} (V_{nn}^{AB} + V_{ne}^{AB} + V_{ne}^{BA} + V_{ee}^{AB}) \quad (1)$$

where $E_{net}^A \equiv E_{net}(\Omega_A)$ is the net electronic energy of atom A that includes the kinetic energy T^A and the potential energy due to nuclei-electron (ne) attractions and electron-electron repulsions (ee) within Ω_A . On the other hand, the interaction energy $E_{int}^{AB} = E_{int}(\Omega_A, \Omega_B)$ between atoms A and B in the molecular system collects various potential energy terms (nn , en , ne and ee). In addition, the two-electron energy (V_{ee}^{AA}, V_{ee}^{AB}) can be divided into classical and non-classical contributions by expressing the reduced pair density matrix as the sum of the Coulombic and exchange-correlation densities (i.e. $\rho_2(\mathbf{r}_1, \mathbf{r}_2) = \rho_1(\mathbf{r}_1)\rho_1(\mathbf{r}_2) + \rho_2^{xc}(\mathbf{r}_1, \mathbf{r}_2)$). This allows to define a purely classical (electrostatic) component of the interaction energy, $V_{class}^{AB} = V_{nn}^{AB} + V_{ne}^{AB} + V_{ne}^{BA} + V_{ee, Coul}^{AB}$, along with a quantum (exchange-correlation) contribution such that $E_{int}^{AB} = V_{class}^{AB} + V_{xc}^{AB}$. This decomposition can be used to better assess the role of electrostatic and quantum effects in intermolecular contacts and/or in chemical bonds.

The calculation of the kinetic and potential energy terms in eq.(1) involves the integration of the first- and second-order reduced density matrices over the atomic basins Ω_A and Ω_B as described in detail elsewhere.^[18, 29] It is important to note that the calculation of the ee terms is rather expensive because it requires the evaluation of six-dimensional integrals of the reduced pair density,

$$V_{ee}^{AB} = \int_{\Omega_A} d\mathbf{r}_1 \int_{\Omega_B} d\mathbf{r}_2 r_{12}^{-1} \rho_2(\mathbf{r}_1, \mathbf{r}_2)$$

More specifically, the PROMOLDEN code tackles the two-electron integrations by using separate radial-angular grids and expanding the interelectronic coordinate r_{12}^{-1} in terms of one and two-center multipolar series that lead to a N^4 integration algorithm for HF methods with a large prefactor associated to grid operations (N being the number of electrons).

DFT-IQA

In principle, the kinetic energy, the monoelectronic potential energy and the Coulombic electron repulsion energy calculated with DFT methods can be processed by IQA following

similar prescriptions as those of the HF method. However, the lack of a DFT pair density $\rho_2(\mathbf{r}_1, \mathbf{r}_2)$ precludes the direct IQA decomposition of the characteristic DFT exchange-correlation energy E_{xc}^{DFT} . As recently proposed,^[23] a pragmatic alternative can be formulated by defining an effective atomic xc energy:

$$E_{xc}^{A,DFT} = \int_{\Omega_A} \rho(\mathbf{r}) \varepsilon_{xc}(\mathbf{r}) d\mathbf{r} + a_0 E_x^{A,KS}$$

where $\rho(\mathbf{r})$ and $\varepsilon_{xc}(\mathbf{r})$ stand for the electron density and the pure (*i.e.*, non-hybrid part) exchange-correlation functional, respectively, and a_0 is the fraction of the HF exchange in hybrid functionals. The $E_x^{A,KS}$ exchange energy is obtained from the HF-like exchange density $\rho_x^{KS}(\mathbf{r}_1, \mathbf{r}_2)$ built up with the corresponding Kohn-Sham (KS) orbitals,

$$E_x^{A,KS} = \frac{1}{2} \int_{\Omega_A} d\mathbf{r}_1 \int d\mathbf{r}_2 r_{12}^{-1} \rho_x^{KS}(\mathbf{r}_1, \mathbf{r}_2).$$

The $E_x^{A,KS}$ integrals and their inter-atomic counterparts $E_x^{AB,KS}$ can be computed using the expressions of conventional IQA for the HF exchange (V_{xc}^{AB}). For each atom A , the following scaling factor λ_A can be computed

$$\lambda_A = \frac{E_{xc}^{A,DFT}}{E_x^{A,KS}} = a_0 + \frac{1}{E_x^{A,KS}} \int_{\Omega_A} \rho(\mathbf{r}) \varepsilon_{xc}(\mathbf{r}) d\mathbf{r}.$$

These quantities are used to derive scaled intra- and interatomic xc energies:

$$\tilde{E}_{xc}^{AB} = \frac{1}{2} [\lambda_A + \lambda_B] E_x^{AB,KS}$$

either for $A=B$ or $A \neq B$. As noticed in ref. 20, the total DFT energy is exactly recovered by summing the scaled xc energies,

$$E_{xc}^{DFT} = \sum_A \tilde{E}_{xc}^{AA} + \sum_{A>B} \tilde{E}_{xc}^{AB}$$

Although this particular scaling technique is not the only way to recover the total energy of a KS-DFT calculation with IQA-like quantities,^[30-31] previous test calculations have shown that it is computationally efficient and gives results that are chemically meaningful and easy to

interpret. In this work, we will further explore its usefulness to yield IQA descriptors $E_{net}(\Omega_A)$, $E_{int}(\Omega_A, \Omega_B)$ for medium-sized systems as described by several DFT methods.

Fragment-based IQA

The application of IQA to characterize the energy change upon the *formation* of a given complex requires separate IQA calculations for the complex $\mathbf{P}\cdots\mathbf{Q}$ and the separate \mathbf{P} and \mathbf{Q} fragments followed by the collection and organization of the various energy terms. As a matter of fact, the grouping of the IQA energy components has been performed using different notations and protocols. For example, in a previous work,^[10] we have introduced the concept of *group energies* $E^{\mathcal{G}}$ as the sum of the net and interaction energies of the collection of atoms that constitute the molecular group \mathcal{G} and used them to analyze the change in energy involved in the formation of molecular clusters. Similarly, the so-called fragment-attributed-molecular system-energy-changes (FAMSEC) expressions have been defined from IQA descriptors.^[32] These and other formalisms, which can be readily interconverted into each other, can be designed or chosen depending on the particular problem at hand.

In what follows, we will introduce the arithmetic formulas of the fragment-based IQA quantities used in this work and the terminology adopted to describe them. Thus, the IQA energy terms in eq. (1) are first collected in following symmetric matrix $\boldsymbol{\eta}$:

$$\boldsymbol{\eta} \equiv \begin{bmatrix} E_{net}(\Omega_1) & E_{int}(\Omega_1, \Omega_2)/2 & \dots & E_{int}(\Omega_1, \Omega_N)/2 \\ E_{int}(\Omega_2, \Omega_1)/2 & E_{net}(\Omega_2) & & \vdots \\ \vdots & & \ddots & \\ E_{int}(\Omega_N, \Omega_1)/2 & E_{int}(\Omega_N, \Omega_2)/2 & \dots & E_{net}(\Omega_N) \end{bmatrix}$$

which is similar to the localization-delocalization matrix defined in QCT.^[5] The sum of the elements of any row or column A of the ‘‘IQA matrix’’ $\boldsymbol{\eta}$ defines the additive energy of atom Ω_A ,

$$E_{add}^A \equiv E_{add}(\Omega_A) = E_{net}(\Omega_A) + \frac{1}{2} \sum_{B \neq A} E_{int}(\Omega_A, \Omega_B),$$

so that the sum of all E_{add}^A reproduces the total energy E . At this point, it must be stressed that the IQA terminology assumes that *interaction* energies are the diatomic contributions to the *absolute* energy of a molecule or supermolecule, that is, they are not relative energies measuring the stability of a complex with respect to the separate fragments. To avoid confusions with the IQA interaction energy terms, the relative energy of a $\mathbf{P}\cdots\mathbf{Q}$ complex with respect to the separate fragments will be termed as its *formation* energy instead of interaction energy.

The study of molecular complexes using IQA can be done by partitioning the $\boldsymbol{\eta}$ matrix of the complex into submatrices associated to the molecular fragments. For example, we can define the following block matrices for two interacting monomers \mathbf{P} and \mathbf{Q} ,

$$\boldsymbol{\eta}_{\mathbf{PQ}} = \begin{bmatrix} \boldsymbol{\eta}_{\mathbf{PQ}}^{\mathbf{P}} & \boldsymbol{\eta}_{\mathbf{PQ}}^{\mathbf{PQ}} \\ \boldsymbol{\eta}_{\mathbf{PQ}}^{\mathbf{PQ}} & \boldsymbol{\eta}_{\mathbf{PQ}}^{\mathbf{Q}} \end{bmatrix}, \quad \boldsymbol{\eta}_{\mathbf{P}\cup\mathbf{Q}} = \begin{bmatrix} \boldsymbol{\eta}_{\mathbf{P}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\eta}_{\mathbf{Q}} \end{bmatrix}$$

where $\boldsymbol{\eta}_{\mathbf{PQ}}$ is the IQA matrix of the complex, and $\boldsymbol{\eta}_{\mathbf{PQ}}^{\mathbf{P}}$ and $\boldsymbol{\eta}_{\mathbf{PQ}}^{\mathbf{Q}}$ denote its submatrices associated to the \mathbf{P} and \mathbf{Q} fragments, respectively, while the non-diagonal block $\boldsymbol{\eta}_{\mathbf{PQ}}^{\mathbf{PQ}}$ comprises the atomic *interaction* energies $E_{\text{int}}(\Omega_A, \Omega_B)/2$ with ... Similarly, the matrix $\boldsymbol{\eta}_{\mathbf{P}\cup\mathbf{Q}}$ collects the IQA matrices of the isolated monomers (but keeping their internal geometries to those in the complex) such that the $\mathbf{P}\cdots\mathbf{Q}$ *formation* energy results from the sum of the elements of the subtraction matrix $\boldsymbol{\eta}_{\mathbf{PQ}} - \boldsymbol{\eta}_{\mathbf{P}\cup\mathbf{Q}}$

$$\Delta E = \sum_{AB} (\boldsymbol{\eta}_{\mathbf{PQ}} - \boldsymbol{\eta}_{\mathbf{P}\cup\mathbf{Q}})_{AB}$$

The relative importance of intra- and inter-fragment contributions to the $\mathbf{P}\cdots\mathbf{Q}$ formation energy can be assessed using the submatrices $\boldsymbol{\eta}_{\mathbf{PQ}}^{\mathbf{P}} - \boldsymbol{\eta}_{\mathbf{P}}$, $\boldsymbol{\eta}_{\mathbf{PQ}}^{\mathbf{Q}} - \boldsymbol{\eta}_{\mathbf{Q}}$ and $\boldsymbol{\eta}_{\mathbf{PQ}}^{\mathbf{PQ}}$. For example, the sum of the elements of the intra-fragment $\boldsymbol{\eta}_{\mathbf{PQ}}^{\mathbf{P}} - \boldsymbol{\eta}_{\mathbf{P}}$ matrix measures the electronic deformation energies of the respective fragments:

$$\underline{E_{def}^P = \sum_{AB} (\boldsymbol{\eta}_{PQ}^P - \boldsymbol{\eta}_P)_{AB} \quad E_{def}^Q = \sum_{AB} (\boldsymbol{\eta}_{PQ}^Q - \boldsymbol{\eta}_Q)_{AB}}$$

Hence, E_{def} is mainly associated to purely quantum mechanical effects (Pauli repulsion, charge transfer, induction-polarization, e^- delocalization) that modify the fragment electron density within its atomic basis upon complex formation,^[8] likewise the covalent bond energy of a diatomic molecule is modulated by the same effects. Note also that E_{def}^P , E_{def}^Q do not include the *distortion* of the nuclear framework unless the reference state is taken as the isolated fully geometrically relaxed fragments \mathbf{P}^{iso} and \mathbf{Q}^{iso} . On the other hand, the stabilization achieved by inter-fragment interactions is collected by the matrix elements of $\boldsymbol{\eta}_{PQ}^{PQ}$, defining thus the $\mathbf{P}\cdots\mathbf{Q}$ interaction energy within the $\mathbf{P}\cdots\mathbf{Q}$ complex,

$$E_{int}^{PQ} = \sum_{ij} (\boldsymbol{\eta}_{PQ}^{PQ})_{ij} .$$

Of course, similar fragment-based decompositions can be readily extended to other IQA components (e.g., V_{class}). Particularly, the total E_{int}^{PQ} can be easily decomposed into a classical (Coulombic) interaction energy $E_{int,class}^{PQ}$ and a purely QM correlation-exchange contribution $E_{int,XC}^{PQ}$. Finally, it is straightforward to see that the total formation or stabilization energy of the $\mathbf{P}\cdots\mathbf{Q}$ complex can be written as the sum of the deformation and interaction energy terms,

$$\Delta E = E_{def}^P + E_{def}^Q + E_{int}^{PQ} = E_{def}^P + E_{def}^Q + E_{int,class}^{PQ} + E_{int,XC}^{PQ}$$

This expression basically *defines* the fragment-based IQA decomposition of ΔE . At this point, it must be stressed that conceptual and lexical links have been established^[33] between the IQA components and those defined in other methodologies like the energy decomposition analysis (EDA),^[34-35] natural energy decomposition analysis (NEDA)^[36], or even perturbative approaches like the well-established SAPT (symmetry-adapted perturbation theory)^[37]. For instance, according to the conventional EDA prescriptions, the formation energy is calculated as the sum of a classical electrostatic interaction energy between non-deformed fragments (

V_{elstat}), a Pauli repulsion term (ΔE_{Pauli}) that incorporates the kinetic and potential energy changes due to the antisymmetrization of the fragment wavefunctions and an orbital relaxation energy (ΔE_{orb}) that arises from inter-fragment charge transfer and polarization effects,

$$\Delta E = \Delta V_{elstat} + \Delta E_{Pauli} + \Delta E_{orb}$$

When two fragments interact weakly, a first order perturbation treatment may be reasonable and the orbital relaxation is negligible. In this scenario, it has been shown that the $E_{def}^P + E_{def}^Q + E_{int,xc}^{PQ}$ sum in IQA (termed as exchange-correlation-repulsion) converges onto ΔE_{Pauli} while $E_{int,class}^{PQ}$ approaches to V_{elstat} . In a certain way IQA unravels ΔE_{Pauli} since this contribution results from the balance between intra-fragment deformation and inter-fragment exchange-correlation effects. Although the extension of e^- rearrangement is generally small in non-covalent complexes, previous computational experience^[8-9, 33] has revealed that $E_{int,xc}^{PQ}$ is a favorable contribution that can be one order of magnitude greater than the formation energy ΔE . However, this inter-fragment exchange stabilization is compensated to a great extent by the intra-fragment electronic distortion mainly associated to the antisymmetrization of the fragments wavefunction. This combination of opposite but similarly large energies result in a small formation energy that may become dominated either by the electrostatic term in polar complexes or the net exchange-correlation-repulsion term in non-polar ones.

Pairwise DFT-D3 dispersion energy

Inclusion of long-range dispersion energy in QM calculations is indispensable to reproduce the energetic properties of large biomolecular systems. Therefore, several methods have been proposed to correct the well-known shortcoming of HF and semilocal KS-DFT methods in properly describing dispersion energy.^[24-25] One of the most popular methods is the third-generation dispersion (D3) correction for DFT,^[38] which is a semiclassical potential inspired on the London formula for the dispersion attraction between two atoms A and B at large

distance ($-C_{AB}/R^6$). At short interatomic distances, the AB dispersion energy tends to a constant contribution to the total correlation energy and this asymptotic behavior can be reproduced using the so-called Becke-Johnson (BJ)^[39] rational damping function. Thus, the D3(BJ) dispersion energy is given by the following formula that includes two multipolar contributions ($n=6, 8$)

$$E_{disp}^{AB} = - \sum_{n=6,8} s_n \frac{C_n^{AB}}{R^n + f(R_{AB}^o)^n}$$

where C_n^{AB} is the dispersion coefficient of the AB pair that is precomputed from accurate dynamic polarizabilities of atoms and small molecules, R is the interatomic distance, s_n is an empirical scaling factor ($s_6=1$ for most common DFT methods) and $f(R_{AB}^o) = a_1 R_{AB}^o + a_2$ is the BJ damping function with $R_{AB}^o = \sqrt{C_8^{AB}/C_6^{AB}}$ and a_1, a_2 being fitting parameters. Optimized a_1, a_2 and s_8 values are available for many DFT functionals and the HF method.^[25] For clarity reasons, the D3(BJ) correction will be dubbed simply as D3 all throughout the manuscript.

The pairwise formalism for the correction of dispersion energy can be combined with the IQA decomposition scheme in order to obtain more reliable quantitative trends and categorizations of non-covalent interactions using a D3-corrected HF or DFT method. To this end, we simply add the E_{disp}^{AB} energies to the IQA interaction energies $E_{int}^{AB} = E_{int}(\Omega_A, \Omega_B)$ such that the total D3-corrected energy results,

$$E = \sum_A E_{net}^A + \sum_{A>B} (E_{int}^{AB} + E_{disp}^{AB}) \quad (2)$$

Of course inclusion of the dispersion energy does not alter the underlying electron density used in the IQA calculations. We also note in passing that another combination of an empirical dispersion energy and a QM energy decomposition method has been recently proposed^[40] in order to replace the costly ab initio dispersion calculations in the DFT-SAPT method by the D3 formula with a reparametrized damping function.

Results and Discussion

Choice of datasets and levels of theory

The S66 database^[27] has been designed to cover typical non-covalent interactions in biomolecular systems in a balanced manner. This has been achieved by the selection of 14 monomers, ranging from small molecules with important functional groups to building blocks like uracil, which are combined to form the 66 dimers that exhibit various binding modes (*e.g.*, linear H-bond, cyclic H-bond, π - π , etc.). Three different categories are distinguished *a priori* in the dataset, the H-bond complexes (23 dimers), the dispersion-dominated or non-polar complexes (23), and a mixed set (20) exhibiting polar-nonpolar interactions (*e.g.* X-H \cdots π). On the other hand, the “ionic H-bond” set, which is consistent with the S66 one, covers the typical charged moieties found in proteins (mimicked by acetate, methylammonium, guanidinium and imidazolium) interacting with small donor/acceptor molecules (water, methanol, methylamine, and formaldehyde) through short H-bonds.

The benchmark datasets will allow us to examine in detail how the IQA approach can perform in the characterization of different types of polar and non-polar direct interactions. Keeping in mind future IQA applications to larger systems, we restricted our study to self-consistent-field (SCF) QM methods like HF or KS-DFT that are efficiently implemented as linear scaling methods in many software packages and are, therefore, routinely employed in QM and hybrid QM/MM studies of biomolecules. We chose the HF method, not only because its dispersion-corrected HF-D3 form is a valuable tool to describe the structure and energetics of biomolecules,^[41-42] but because HF lacks entirely dispersion energy and, thereby, HF-D3 with the BJ damping function can provide a straight physical partitioning of energy. Among the myriad of local DFT methods, we selected the well-known hybrid B3LYP functional, which is still widely used for the prediction of all kinds of molecular properties, and PW6B95, which is one of the Minnesota hybrid functionals that has been found as one of the most robust hybrid

functionals in combination with D3.^[43] We also employed the PBE0 method, which is the hybrid version of the nonempirical PBE functional and has been recently found to yield better electron densities as compared with other DFT methods.^[44] The three DFT methods are augmented with the D3 formula although the weight of the dispersion correction varies significantly. All these methods were combined with a triple- ζ basis sets (cc-pVTZ or aug-cc-pVTZ) in all the calculations, what constitutes a compromise choice between quality and computational cost. In this respect, it has been noticed^[45] that the cc-pVTZ basis offers a good balance in terms of its description of noncovalent interactions arising from both electrostatic and dispersion forces and introduces a small basis set superposition error.

Comparison between benchmark and calculated interaction energies

Statistical error measures of the selected levels of theory against the benchmark values in the reference datasets are collected in Table 1 (correlation plots are shown in Figures S1 and S2). In this Table, the influence of the D3 dispersion energy and that of the counterpoise (CP) correction to the basis set superposition error (BSSE) are assessed separately.

Table 1. Statistical measurements comparing the calculated formation energies with the benchmark values: The determination coefficient (R^2), the Spearman correlation coefficient (ρ), the root mean square (RMS) error and mean absolute deviation (MAD) in kcal/mol.

	R^2	ρ	RMS	MAD
HF-D3	0.991	0.976	1.322	1.128
CP HF-D3	0.989	0.972	0.881	0.625
IQA HF-D3	0.978	0.957	1.580	1.230
B3LYP-D3	0.994	0.986	1.189	1.009
CP B3LYP-D3	0.998	0.996	0.337	0.207
IQA B3LYP-D3	0.980	0.960	1.340	1.089
PW6B95-D3	0.994	0.982	0.678	0.582
CP PW6B95-D3	0.996	0.988	0.386	0.306
IQA PW6B95-D3	0.988	0.977	0.665	0.542
PBE0-D3	0.991	0.980	1.123	0.885
CP PBE0-D3	0.994	0.991	0.460	0.317
IQA PBE0-D3	0.982	0.961	1.057	0.842
Ionic H-bond complexes				
	R^2	ρ	RMS	MAD
HF-D3	0.951	0.914	1.193	0.875
CP HF-D3	0.951	0.914	1.125	0.814
IQA HF-D3	0.930	0.932	1.222	0.967
B3LYP-D3	0.995	0.982	0.822	0.688
CP B3LYP-D3	0.995	0.982	0.706	0.545
IQA B3LYP-D3	0.992	0.982	0.817	0.669
PW6B95-D3	0.997	0.989	0.273	0.221
CP PW6B95-D3	0.997	0.986	0.344	0.296
IQA PW6B95-D3	0.991	0.982	0.508	0.404
PBE0-D3	0.993	0.989	1.177	0.978
CP PBE0-D3	0.993	0.989	1.074	0.848
IQA PBE0-D3	0.988	0.954	1.123	0.814

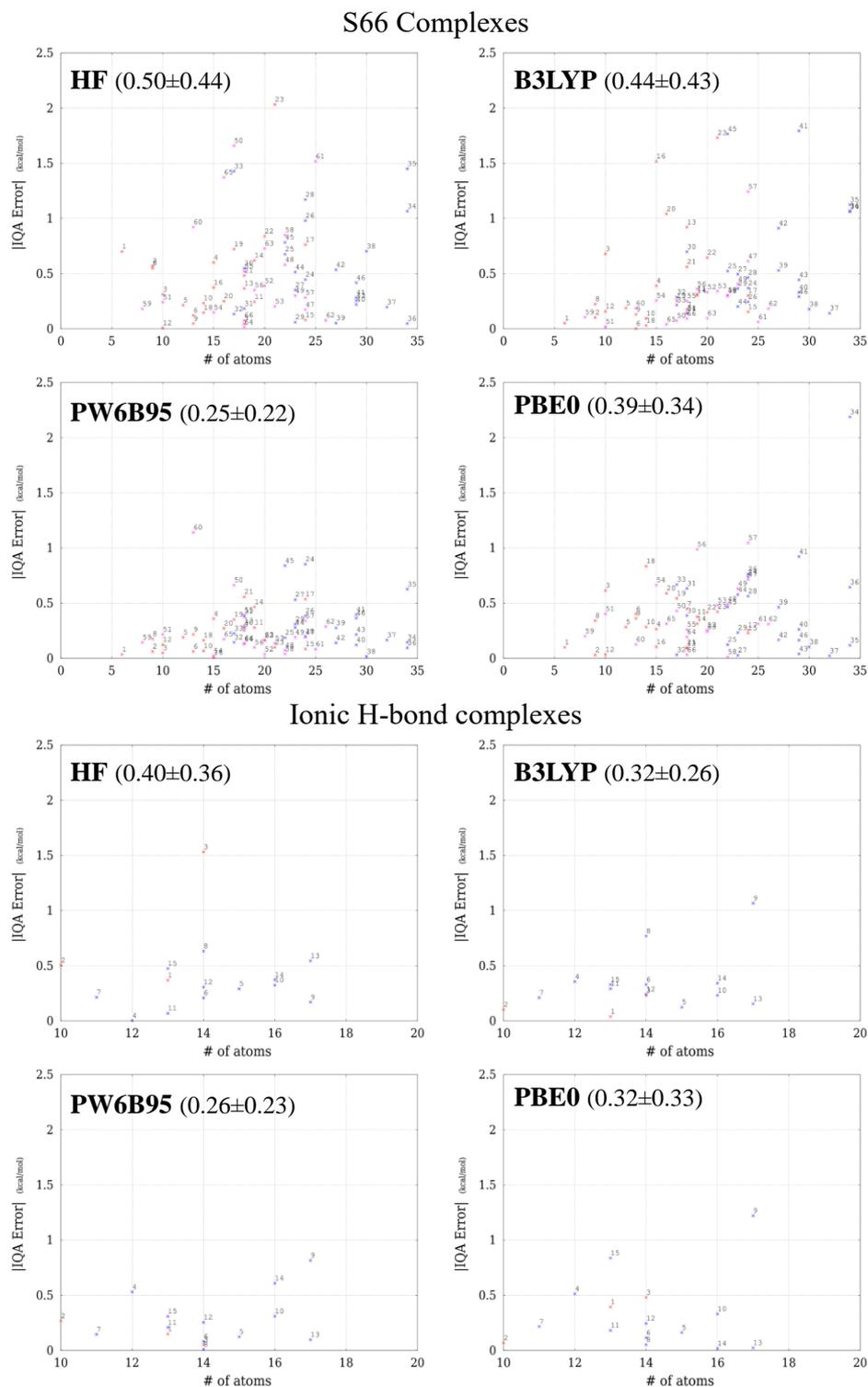
Let us discuss first the energy calculations on the S66 complexes. The four selected levels of theory (HF/cc-pVTZ, B3LYP/cc-pVTZ, PW6B95/cc-pVTZ and PBE0/cc-pVTZ) in combination with the D3 formula produce formation energies that are satisfactorily correlated to the reference data ($R^2=0.991-0.994$). The four methods perform better for complexes in the H-bond and “others” categories ($R^2\sim 0.997-0.998$) than for dispersion complexes ($R^2\sim 0.886-0.938$). In terms of the mean absolute deviations (*MAD*), the PW6B95-D3 formation energies are the most accurate (*MAD*=0.58 kcal/mol). The relatively good accuracy of the PW6B95-based data is not unexpected as this semiempirical functional has been optimized for thermochemistry applications, but the good performance of the other methods including HF-D3 is equally remarkable (*MAD*=0.9-1.1 kcal/mol). These errors are partly due to the basis set limitations as shown in a previous study that has assessed DFT-D3 methods ^[46] with a quadruple- ζ basis set (def2-QTZVP) against the S66 data set, finding *MAD* values of 0.18 kcal/mol and 0.28 kcal/mol for the PW6B95 and B3LYP methods, respectively.

Preliminary calculations for the ionic H-bond complexes (not shown here) were carried out with the same cc-pVTZ basis set used in the S66 structures, but the resulting formation energies were less accurate than the S66 ones, having larger *MADs* around 2-3 kcal/mol. We decided then to add diffuse Gaussian type functions into the basis set, which are usually required for anionic and highly polarizable systems. The augmented basis sets improve the accuracy of the formation energies, the *MAD* values lying in the [0.2-1.0] kcal/mol interval. The R^2 values and the scattering plots in Figure S2 reveal that the performance of the HF-D3 method for the ionic complexes is not as good as those of the DFT methods (e.g., R^2 of HF-D3 data is 0.951 vs 0.997 at PW6B95-D3), PW6B95-D3 being also rather accurate with an *MAD* value of 0.22 kcal/mol.

The influence of the BSSE in the calculated formation energies is tested by means of the CP procedure. The CP-corrected energies hardly affect the correlation measurements of the S66 complexes (all the R^2 values remain nearly constant). The CP-corrected formation energies

exhibit smaller MAD values between 0.62 and 0.22 kcal/mol for HF-D3 and B3LYP-D3, respectively, in the S66 complexes, showing thus that the cc-pVTZ formation energies perform similarly as the def2-QTZVP ones after removal of the BSSE. For the ionic H-bond complexes, a larger basis set, aug-cc-pVTZ, is used, the estimated BSSE have smaller values and the impact of the CP corrections on the *RMS* errors is more erratic (slight reductions or increases). Therefore, since the BSSE does not alter the relative trends in the formation energies derived from the cc-pVTZ or aug-cc-pVTZ basis sets and the average CP-correction is little or moderate, we did not consider BSSE effects in the subsequent IQA analysis of interaction energies. This choice is also necessary to avoid important problems in the QTAIM space partitioning (*e.g.*, spurious distortions in the interatomic surfaces, abundance of nonnuclear maxima on the charge density)^[8] associated to the ghost functions in the IQA calculations.

Figure 1. Numerical errors (in kcal/mol) of the IQA-reconstructed formation energies at the various levels of theory. Mean values and standard deviations are given in parentheses. The coloring of the S66 data points corresponds to H-bond (in red), non-polar (in blue) and mixed (in magenta) categories. For the ionic H-bond errors, red coloring indicates anionic complexes and blue is for cationic complexes.



Assessment of numerical errors in the IQA-calculations

The PROMOLDEN settings used in the IQA analysis of the S66 complexes were adopted after having verified their satisfactory performance in reproducing the total energy of a set of 22 small molecules (see Table S1) with a mean error below 0.001 hartrees (~ 0.6 kcal/mol) for the four theoretical methods. However, such numerical errors typically increase with the number of atoms and/or the presence of intermolecular contacts that may reshape atomic basins. Hence, to better assess the robustness of the IQA approach to partitioning formation energies (ΔE), we compare herein the formation energy ΔE values derived from the total HF or DFT energies (E) with their counterpart differences obtained from the IQA-reconstructed energy (referred to thereafter as E^{IQA}) of the complex and separate fragments. Thus, we define the typical “IQA numerical error” as the absolute difference $|\Delta E - \Delta E^{IQA}|$ and the results are plotted against the atomic size for all the complexes in Figure 1.

We see first in Figure 1 that most of the error values are below ~ 1 kcal/mol for the HF, B3LYP and PBE0 energies or ~ 0.5 kcal/mol for PW6B95. For a few S66 complexes, the IQA integration of their charge densities results in larger $|\Delta E - \Delta E^{IQA}|$ errors close to 2.0 kcal/mol, especially with the HF and B3LYP methods. The integration of the PWB96B5 density results in average IQA numerical errors that are quite small, 0.25 ± 0.22 and 0.26 ± 0.23 kcal/mol for the S66 and the ionic H-bond complexes. Furthermore, these error values are completely uncorrelated with the atomic size of the complexes (see Figure 1), what suggests that the ΔE^{IQA} values may benefit from error compensation and anticipates also the utility of IQA for larger systems. The numerical errors seem also randomly distributed given that the comparison between the benchmark ΔE energies and the ΔE^{IQA} values at the different levels of theory yields both R^2 determination coefficients and MAD values that remain quite similar to those obtained with the original HF/DFT ΔE data (see Table 1).

From data in Figure 1, we conclude that the “IQA numerical errors” remain within reasonable bounds (~ 0.5 kcal/mol on average). Certainly, for the weakly-bound non-polar complexes, these small absolute errors translate into large percentage relative errors in the ΔE^{IQA} values. Nevertheless, the actual interest of the IQA energy partitioning resides in the atomic and/or fragment-based IQA components and their changes upon complex formation, which are generally one or two orders of magnitude larger than the complex formation energies (see below). Thus, the IQA numerical errors illustrated in Figure 1 can be taken as indicative of the uncertainty in the IQA descriptors and, consequently we expect that any structure-energy relationships formulated with them would not be affected by numerical inaccuracies provided that the corresponding energy variations are above ~ 0.5 - 1.0 kcal/mol.

Fragment-based IQA partitioning

As above discussed, IQA decomposes the formation energy ΔE of the binary $P \cdots Q$ complex into deformation (E_{def}^P and E_{def}^Q) and interaction energies (E_{int}^{PQ}) that, in turn, are readily computed from the atomic and inter-atomic IQA energies arranged in the matrices $\eta_{PQ}^P - \eta_P$, $\eta_{PQ}^Q - \eta_Q$ and η_{PQ}^{PQ} . Each E_{def} value reflects the change in the IQA descriptors of a given fragment within the complex and in its respective isolated state, whereas E_{int}^{PQ} collects the atomic interaction energies between the P and Q atoms while E_{disp}^{PQ} corresponds to the D3 dispersion correction. Thus, the formation energy emerges from two opposite effects, the intra-fragment distortion and the inter-fragment attraction

$$\Delta E = (E_{def}^P + E_{def}^Q) + (E_{int}^{PQ} + E_{disp}^{PQ})$$

Further insight can be gained by decomposing E_{int}^{PQ} into the classical (Coulombic) interaction energy $E_{int,class}^{PQ}$ and the purely QM correlation-exchange contribution $E_{int,xc}^{PQ}$. This enables the collection of the exchange-correlation-repulsion energies plus the dispersion correction into a

single contribution, $E_{xcr}^{PQ} = E_{def}^P + E_{def}^Q + E_{int,xc}^{PQ} + E_{disp}^{PQ}$, which, in turn, allows us to describe the total formation energy as the result of two contributions, QM and classical:

$$\Delta E = E_{xcr}^{PQ} + E_{int,class}^{PQ}$$

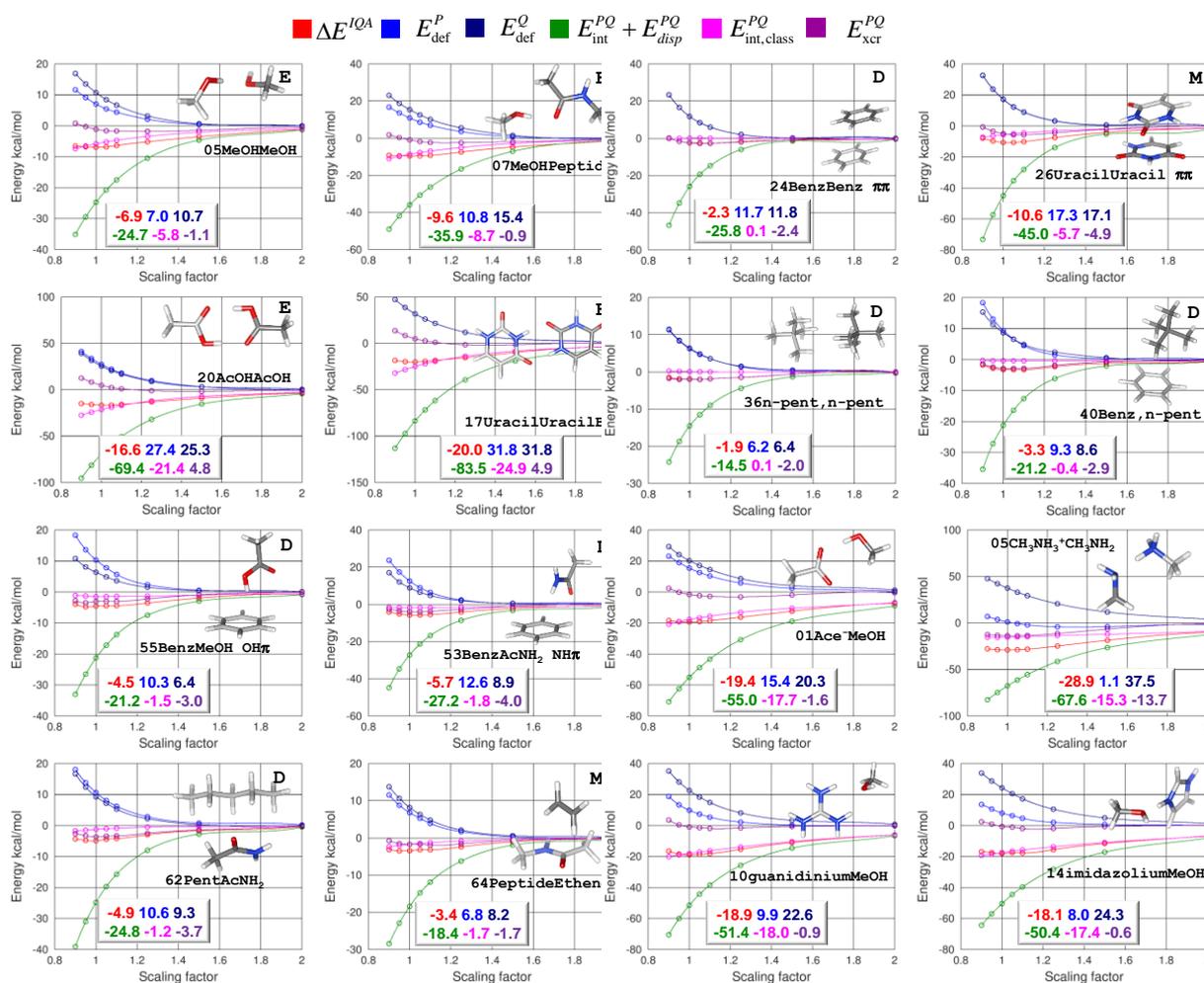
The various IQA-D3 quantities ($E_{def}^P, E_{def}^Q, E_{disp}^{PQ}, \dots$) for all the S66 complexes are collected in Tables S2-S6 in the Supporting Information. For the sake of brevity, we focus on a selection of the S66 complexes (4 H-bond, 4 dispersion and 4 “others” complexes) for which the IQA energies were calculated at eight different intermonomeric distances as defined in the S66x8 dataset. These calculations were done at the PW6B95-D3/cc-pVTZ level and the results are displayed in Figure 2 (the numerical values of the IQA contributions to ΔE for the most stable geometries are also indicated in Figure 2). We note first that the IQA data unambiguously quantify the intra-fragment distortion and the inter-fragment attraction energies, which have absolute values that are ~ 3 -4 (H-bond complexes) or ~ 5 -10 (dispersion complexes) times larger in absolute value than ΔE . Several trends can be outlined from these data: (i) the electronic distortion of the H-donor fragments in H-bond complexes is lower by a few kcal/mol than that of the H-acceptor monomer regardless of their molecular size; (ii) the two interacting moieties in asymmetrical dispersion complexes (*e.g.*, benzene $\cdots n$ -pentane) experience quite similar distortion effects; (iii) the most polar fragment in the *a priori* mixed complexes (*e.g.*, benzene \cdots methanol) tend to be less distorted.

Another observation derived from data in Figure 2 and in Tables S2-S6 concerns the relative weights of the classical (electrostatic) $E_{int,class}^{PQ}$ and QM (exchange-correlation-repulsion) E_{xcr}^{PQ} terms. The classical term corresponds to the Coulombic interaction energy (*e-nuclei* and *e-e*) involving the atomic basins in molecules *P* and *Q*. For the H-bond complexes, the $E_{int,class}^{PQ}$ values are close to the IQA-reconstructed formation energies ΔE^{IQA} (*e.g.*, at the PW6B95-D3/cc-pVTZ level, $\Delta E^{IQA} = -6.9$ kcal/mol and $E_{int,class}^{PQ} = -6.0$ kcal/mol for the methanol dimer,

$\Delta E^{QA} = -9.2$ and $E_{\text{int,class}}^{PQ} = -8.9$ for the peptide dimer, etc.). This indicates that exchange-correlation and polarization effects involved in the deformation and interaction energies cancel each other to a large extent in such a way that electrostatics plays a major role. We note in passing that this electrostatic fingerprint of H-bonds and other subtle details like the actual compatibility between the electrostatic image and the local covalency of H-bonds have been analyzed in detail in previous IQA applications.^[8]

For the typical dispersion complexes like the benzene dimer and the alkane dimers, the Coulombic energy $E_{\text{int,class}}^{PQ}$ is very small (~ 0.5 - 0.1 kcal/mol), signaling thus the negligible role played by classical electrostatics in the stabilization of non-polar complexes in consonance with expectations. Alternatively, the balance between fragment distortion (repulsive) and inter-fragment exchange-correlation (attractive) favors the latter contribution and, therefore, their QM term, E_{scr}^{PQ} is akin to the formation energy (*e.g.*, see the neopentane dimer in Figure 2). In this respect, we note that the D3 dispersion correction E_{disp}^{PQ} , which is conceptually related to correlation energy, is the most important contribution to E_{scr}^{PQ} , playing thus an essential role from a quantitative point of view.

Figure 2. IQA deformation and interaction energy components (in kcal/mol) for selected S66 and Ionic-HB complexes at the PW6B95-D3/cc-pVTZ level as a function of the scaling factor multiplying the intermolecular distance. Connecting curves passing through the calculated values were obtained by spline interpolation. Total formation energy (ΔE^{IQA} from IQA-reconstructed energies; red curves), fragment deformation energies (E_{def}^P and E_{def}^Q in blue and blue-navy curves, respectively), fragment interaction energies including D3 dispersion ($E_{int}^{PQ} + E_{disp}^{PQ}$, green curve); $E_{int,class}^{PQ}$ from Coulombic IQA terms (in magenta); E_{xcr}^{PQ} from deformation and exchange-correlation IQA terms and D3 dispersion (in dark magenta). The interaction type (E, Electrostatics-dominated; D, Dispersion-dominated and M, Mixed) is also indicated for the S66 complexes. The numerical values of the energy components for the most stable dimer geometry are also indicated.



Several S66 complexes become simultaneously stabilized by the classical electrostatic interaction and the QM exchange-correlation-repulsion effects. They belong to the “dispersion” and “others” categories (e.g., the π - π uracil dimer, the benzene-acetamide π -complex, etc.). These intermediate cases represent a gradual variation of the electrostatic and QM binding determinants in the formation of non-covalent complexes. As a matter of fact, the varying influence of the IQA term $E_{\text{int,class}}^{PQ}$ suggests a practical way of characterizing the S66 systems as being electrostatics dominated, dispersion dominated or mixed (dispersion/electrostatics). The same categorization has been attained in the original S66 article^[27] using the DFT-SAPT approach^[37] to calculate the ratio between the electrostatics/dispersion energy. Since E_{disp}^{PQ} accounts for dispersion energy corrections that depend on the level of theory, herein we consider the ratio between the classical interaction energy and the total formation energy (including D3 corrections) of a given complex, $f = E_{\text{int,class}}^{PQ} / \Delta E$, and propose to classify it as an electrostatic complex if $f > 2/3$. Reciprocally, if $f < 1/3$ the complex would be categorized as a dispersion one, and the mixed category would correspond to intermediate values of f between $1/3$ and $2/3$. This choice of thresholds for f gives rise to groups of complexes that agree well with the empirical classification (H-bonded, dispersion, “others”) as shown in Figure 2, which also displays the labels (E, D or M) assigned to the S66 complexes at the PW6B95/cc-pVTZ level (all the numerical values of f are reported in Tables S2-S5). Thus, the family of H-bond complexes is entirely classified as electrostatic with f ratios about 0.8-0.9 for the structures having one H-bond or $f=1.2$ -1.3 for those with two H-bonds. Similarly, the empirical category of dispersion complexes is associated to f values that are close to zero or slightly negative (e.g., alkane \cdots alkane dimers). In this set of complexes the presence of the polar nucleobase uracil increases the weight of electrostatics and some uracil complexes are labelled as mixed (M). The “others” group of heterogeneous complexes receives 4 E, 6 M and 10 D labels using the PW6BP5-D3 method (Table S4). For example, dimers formed between the small ethyne

molecule and other polar molecules (water, acetic acid) turn out to be electrostatics dominated while π -complexes between benzene and polar molecules have an intermediate mixed character. The importance of electrostatics further decreases when the polar molecules interact with alkanes or in the T-structures (TS) constituted by two aromatic rings, all these complexes being labelled as D.

The E/D/M labelling derived from the IQA partitioning practically matches that produced by DFT-SAPT, the few exceptions arising in the case of some H-bond complexes with methylamine (denoted as mixed complexes by DFT-SAPT) and in a few π -complexes. We also note that the numerical values of the f index can vary with the level of theory used in the IQA calculations. As would be expected, the HF method is associated with larger classical electrostatics effects (i.e., $f > 1$ for the majority of E complexes). Some borderline cases may change their E/D/M category depending on the HF or DFT method, but the overall categorization of the S66 complexes is quite similar in the four examined levels of theory (see Tables S2-S5).

The evolution along the intermonomeric distances of the attractive and repulsive IQA-D3 terms are displayed in Figure 2. In general, these energy contributions change smoothly when the intermolecular separation is elongated or shortened around the equilibrium value (scaling factor=1). As expected, deformation energies and the correlation-exchange contribution tend to nearly zero values at the long distance regime (scaling factor=2) whereas the classical electrostatic term $E_{\text{int,class}}^{PQ}$ retains small negative values (-1, -3 kcal/mol), especially for the polar and “others” groups of complexes. These plots show that, as the two monomers approach each other, the IQA-D3 calculations consistently depict the energetic impact of the quantum/classical and intra-/inter-molecular effects involved in the non-covalent interactions, and doing so with the same kind of descriptors and terminology that have been previously

employed to describe the electronic and energetic features of covalent or ionic bond formation.^[47]

The fragment-based IQA decomposition of selected ionic H-bond complexes is also summarized in Figure 2 (full data are reported in Table S6). These are cationic or anionic dimers whose stability is expected to be largely determined by classical Coulombic interactions. Indeed this is confirmed by the f ratios having values between 0.90 and 1.00 for most of the equilibrium structures. There are, however, two exceptions labelled as mixed complexes ($f=0.5-0.6$) that involve H-acceptor methylamine molecules and correspond to the most stable complexes according to the reference data. The dominant role of Coulombic interactions is also noticed in the relatively large classical interaction energy ($\sim -6, -8$ kcal/mol) when the intermolecular separation is doubled.

The stronger electronic distortion experienced by the H-bond acceptor fragments in the S66 set is observed again in the anionic complexes (*e.g.*, see acetate \cdots methanol in Figure 2). This asymmetry in E_{def}^P and E_{def}^Q is considerably reinforced in the cationic complexes what is in consonance with the lower polarizability of cations. Thus, the IQA partitioning points out that the methylammonium, guanidinium and imidazolium cations are moderately perturbed ($E_{def}^{cation} < 10$ kcal/mol) while they efficiently polarize their H-bond acceptor partners ($E_{def}^{neutral} > 20$ kcal/mol). This effect is maximum at the methylammonium \cdots methylamine and imidazolium \cdots methylamine complexes with $E_{def}^{cation} < 5$ and $E_{def}^{neutral} > 35$ kcal/mol, and is also accompanied by the largest fragment-interaction contributions ($E_{int}^{PQ} < -60$ kcal/mol). This means that the NH \cdots N H-bonds benefit from classical electrostatic and QM exchange-correlation effects (their $E_{int,class}^{PQ}$ and E_{xcr}^{PQ} energies are comparable; see methylammonium \cdots methylamine in Figure 2), suggesting thus a reinforced covalent character, which is already present at intermediate separations (scaling factor 1.3-1.5). Most likely, this energetic pattern is the consequence of the identical or similar basicity of the two neutral monomers

(methylamine and imidazole) and is also seen in the long bond length of the acidic N-H bond (1.10 Å).

Atomic distribution of formation energies

Clearly, a significant advantage of the IQA approach is that it provides atomic partitions of all the energetic properties and that such partitioning is intimately related with the topology of the charge density and the QM description of chemical bonding.

Among the IQA descriptors that could be considered for gaining insight at the atomic level concerning the formation energies, the change induced by fragment interactions into the additive energy of atom Ω_A (with $A \in P$),

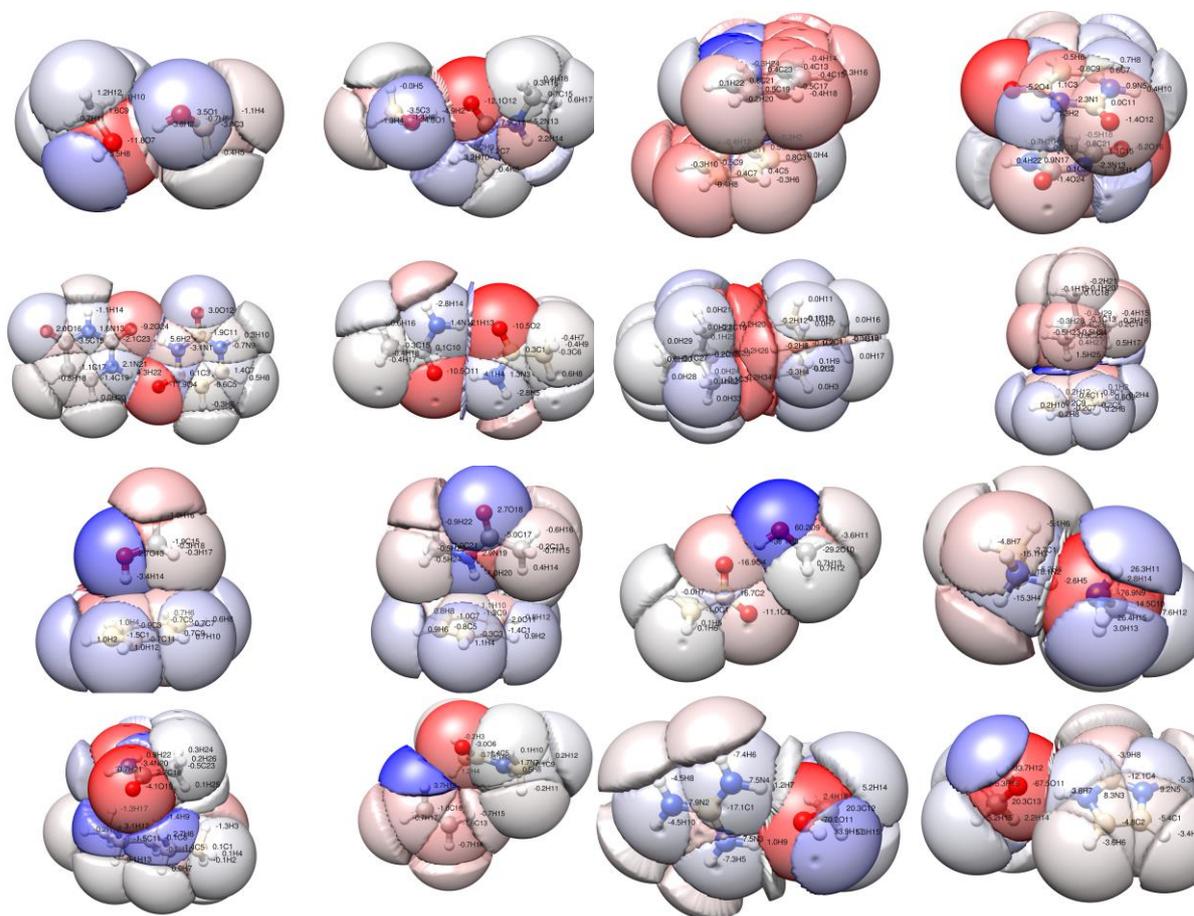
$$\Delta E_{add}^A = E_{add}^{PQ}(\Omega_A) - E_{add}^P(\Omega_A),$$

can be particularly useful to quantify the importance of individual atoms or groups of atoms in the global stabilization of the PQ complex. Of course, the sum of the ΔE_{add}^A values matches the total formation energy of the PQ complex. Hence, we calculated the ΔE_{add}^A terms including the D3 dispersion corrections for all the complexes basing on the IQA data for the binary complexes and for the separated fragments. The resulting PW6B95-D3 ΔE_{add}^A values for some representative complexes are shown in Figure 3 while Table S7 reports ΔE_{add}^A values for selected S66 complexes at the four levels of theory considered. Figures S3 and S4 in the Supporting Information show the atomic distribution of the PW6B95-D3 ΔE_{add}^A values for the S66 and the ionic H-bond data sets, respectively.

In Figure 3, the ΔE_{add}^A values are shown as atomic labels assigned to ball-and-stick representations of the complexes. To enhance the visual interpretation of the data, the atomic basins Ω_A are shown as colored transparent surfaces. As mentioned in Methods, the settings for the calculation of the basin surfaces in these drawings were optimized for visualization purposes, but they are useful anyway to show the relative size and shape of the atomic basins considered in the IQA calculations. For example, Figure 3 illustrate how the basins of sp^3

centers are partially hidden by the terminal basins and that the basin of hydrogen atoms involved in X-H \cdots :Y interactions is largely flattened between the X and Y basins. The numerical values of ΔE_{add}^A are mapped onto the basin surfaces (blue indicates $\Delta E_{add}^A > 0$, grey $\Delta E_{add}^A \approx 0$ and red $\Delta E_{add}^A < 0$) so that we can grasp at a glance the relevance of specific atoms for the complex stability.

Figure 3. Ball-and-stick view of selected S66 and ionic-HB complexes. Atomic labels correspond to the change of additive atomic energies (in kcal/mol) upon complex formation. Atomic basins Ω_A are shown as coloured transparent surfaces. As mentioned in Methods (blue and red colouring represents destabilized/stabilized atoms, respectively).



Some patterns about the distribution of the ΔE_{add}^A values in the S66 dataset can be found in data in Figures 3 and S3. Thus, the localized nature of H-bonds is readily seen in the energetics of structures with a single O-H \cdots :Y bond, which is largely dominated by the ΔE_{add}^A changes of the interacting atoms: the H-acceptor end (:Y) and the H atom are stabilized ($\Delta E_{add}^A \approx -5, -10$

kcal/mol) whereas the H-donor X and H-atoms attached to Y tend to be destabilized by ~5 kcal/mol. Similar variations occur in the N-H \cdots :Y bonds, albeit the two atoms in the N-H moiety are moderately destabilized. When amide groups participate in H-bond interactions, the four -O=C-NH- atoms are appreciably perturbed, but more specifically the carbonyl group atoms. The case of the uracil dimer can be of particular interest given that the ΔE_{add}^A spreading throughout the two heterocycles depicts a certain asymmetry of the two H-bond contacts (see Figure 3). Regarding the category of dispersion complexes, the ΔE_{add}^A values are quite small (< 1.0 kcal/mol for C and H atoms) and are evenly distributed across all the hydrocarbon atoms, what seems in consonance with the weak and non-specific character of the dispersion interactions. When heteroatoms (O, N) are present (*e.g.*, in pyridine, uracil), they exhibit slightly larger ΔE_{add}^A changes ($\sim \pm 2$ kcal/mol). Some trends may be also outlined as in the case of the alkane dimers in which the stabilizing effects are concentrated in the closest H-atoms between the two monomers (see the neopentane dimer in Figure 3). For the third category of *mixed* complexes, the patterns of atomic distribution of the non-polar and polar monomers resemble those in the dispersion and H-bond groups, respectively, the magnitude of the ΔE_{add}^A changes in the non-polar molecules being reinforced and those in the polar functional groups being less pronounced.

The ΔE_{add}^A values for some ionic H-bond complexes, which are also shown in Figure 3, display features akin to those of the neutral H-bond complexes in the S66 dataset. The OH/NH₂ groups belonging to the neutral monomer concentrate the more intense energetic rearrangement, irrespective of their role as H-donor (*e.g.*, acetate \cdots methanol) or H-acceptors (*e.g.*, methylammonium \cdots methylamine). As the strength of the ionic H-bond contacts is relatively large, the magnitude of the underlying ΔE_{add}^A values is also large (*e.g.*, ~60-70 kcal/mol for the O/N atoms and ~30-40 for H) and tend to be somewhat more pronounced in the cationic H-

bonds. The pictures in Figure 3 make evident again the weaker perturbation suffered by the cationic moieties with respect to the neutral or anionic fragments.

To find out to what extent the trends in the atomic distribution of the complex formation energies may depend on the level of theory, we compared the ΔE_{add}^A values as obtained with the HF-D3, B3LYP-D3, PW6B95-D3 and PBE0-D3 methods for selected S66 complexes (Table S7). It turns out that the three DFT methods provide results that are close to each other: the mean unsigned difference (MUD) of the B3LYP-D3 and PBE0-D3 ΔE_{add}^A with respect to the PW6B95-D3 ones are both 0.1 kcal/mol. The HF-D3 MUD is slightly larger, 0.3 kcal/mol. For the examined complexes, the HF-D3, B3LYP-D3 and PBE0-D3 data are well correlated against the PW6B95-D3 data (R^2 values ~ 0.95 - 0.99). The only exception arises in the neo-pentane dimer, which presents quite small ΔE_{add}^A values so that the HF-D3 data are uncorrelated with PW6B95-D3 ($R^2 = 0.50$ and 0.80 for the B3LYP-D3 and PBE0-D3). Therefore, we conclude that, in general, the IQA-D3 calculations with the examined methods give comparable and consistent results concerning the atomic energy changes. Only in the weakest interacting alkane complexes, the atomic partitioning may fluctuate with the level of theory, what is likely due to several factors (the small magnitude of the energy changes, the IQA numerical errors and the importance of correlation and exchange effects).

Effects of geometry optimization

As previously noticed, the benchmark energies in the S66 and ionic H-bond data sets have been evaluated using fixed monomer geometries. For the sake of simplicity in the comparative assessments of the formation energies, the IQA descriptors, their numerical errors and so on, those fixed geometries were employed for carrying out the HF/DFT and IQA calculations that are analyzed in the preceding subsections. Nonetheless, geometry optimization using either HF-D3 or DFT-D3 levels of theory is routinely done during the computational study of medium- and large-size molecules and, therefore, we also examined the changes in the

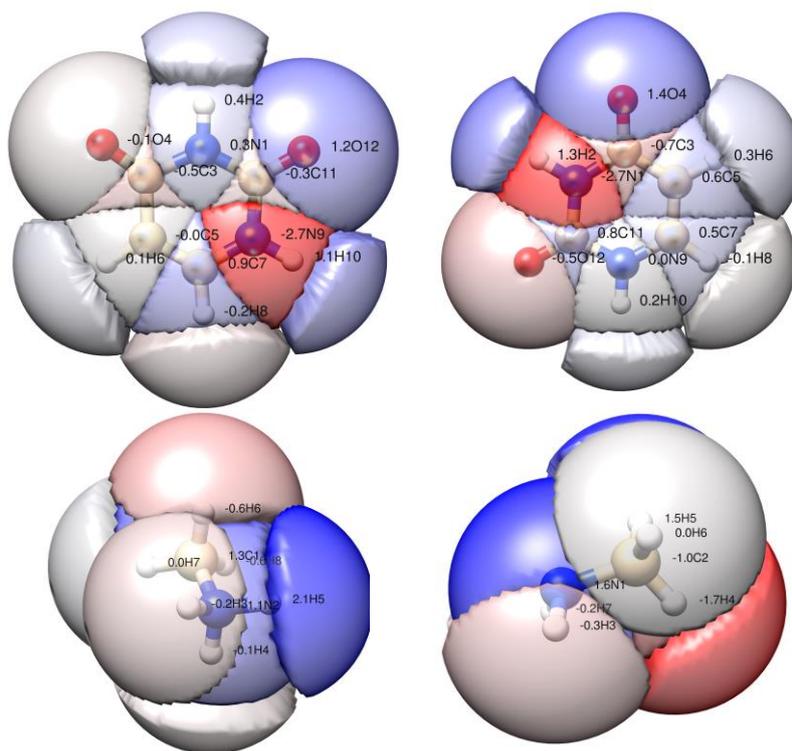
formation energies and IQA analysis after having optimized all the dimers at the corresponding levels of theory.

Concerning the equilibrium geometries of the S66 intermolecular complexes, the HF-D3 or DFT-D3 relaxed structures depart slightly from the benchmark ones. The magnitude of the deviations is generally small as characterized by the RMS deviations between Cartesian coordinates that amount to $\sim 0.04\text{-}0.08$ Å (these are typical median RMSD values; see Table S8). The B3LYP-D3 geometries are somewhat closer to the benchmark data followed by the HF-D3, PBE0-D3 and PW6B95-D3. The largest deviations arise in the π - π or T-shaped complexes with maximum RMSDs around $0.7\text{-}0.8$ Å that hint a moderate reorientation of the interacting monomers. Nevertheless, in all the cases, the nature of the interaction is well preserved and this is corroborated by the calculated formation energies ΔE (using the monomer geometries taken from the optimized dimers). Thus, the good correlation between the HF-D3 and DFT-D3 formation energies and the reference data is practically maintained when using the relaxed geometries of the dimers (see Figure S5). For example, the R^2 correlation coefficient for the HF-D3 formation energies varies from 0.991 to 0.986 Å ongoing from the fixed S66 geometries to the relaxed ones (0.994 to 0.991 for PW6B95-D3) and the RMS errors remain within reasonable bounds (1.25 and 0.73 kcal/mol for the HF-D3 and PW6B95-D3 data).

The geometry relaxation of the ionic complexes led to similar findings: the optimized structures deviate slightly from the benchmark structures (RMSDs $\sim 0.02\text{-}0.04$ Å), the most significant changes occurring at the acetate \cdots methylamine complex due to an internal rotation around the C-COO $^-$ bond and at the imidazolium \cdots methylamine complex due to small readjustments of the NH \cdots :N contact. The B3LYP-D3 geometries are on average closer to the reference structures while the HF-D3 ones show the largest RMSD scores. The impact on the formation energies is dissimilar among the levels of theory: the DFT-D3 formation energies retain fairly good correlation coefficients ($R^2=0.939\text{-}0.989$) and low RMS errors of $\sim 1.0\text{-}2.0$

kcal/mol, the B3LYP-D3 method providing the best results ($R^2=0.989$ and $RMS=1.06$ kcal/mol). The global performance of HF-D3 seems deteriorated (*e.g.* $R^2 = 0.883$) mainly because it overestimates the stability of the ionic $N-H\cdots N$ H-bonds. This greater sensitivity of the geometries and formation energies with the level of theory is probably related to the charge transfer effects in the ionic complexes that have QTAIM Δq^{PQ} values between 0.04 and 0.15 e^- whereas the corresponding Δq^{PQ} values in the neutral H-bond S66 complexes lie in the 0.01-0.04 interval.

Figure 4. Ball-and-stick views of the uracil monomers and the methylammonium and methylamine monomers in their isolated state. The atomic labels correspond to the atomic IQA partitioning (ΔE_{add} in kcal/mol) of the fragment distortion energy due to geometric rearrangement. The atomic basins are shown as transparent surfaces and colored according to the corresponding ΔE_{add} values (red means $\Delta E_{add} < 0$; blue $\Delta E_{add} > 0$).



We also checked that the numerical errors in the subsequent IQA calculations and the trends exhibited by the IQA descriptors are not substantially affected by the minor changes in the formation energies and the molecular geometries caused by geometry optimization. The

resulting mean IQA numerical errors of the S66 complexes are now 0.51 ± 0.44 , 0.37 ± 0.32 , 0.30 ± 0.25 and 0.22 ± 0.19 kcal/mol for the HF, B3LYP, PW6B95 and PBE0-based calculations, respectively, these values being indeed comparable to those reported in Figure 1. The equivalent values for the ionic H-bond complexes are 0.45 ± 0.32 , 0.25 ± 0.19 , 0.65 ± 0.86 and 0.57 ± 0.68 kcal/mol. Similarly, the decomposition of the formation energies into deformation and interaction components and the related categorization of the complexes result in analogous trends regarding the weight of intra- and inter-fragment energy changes and the E/M/D labels assigned to the complexes. Apart from minimal variations in the fragment-based or atomic IQA energies (~ 0.1 - 1.0 kcal/mol depending on their actual size) as well as a few E/M exchanges at borderline cases (see Tables S4, S6 and S9), the overall picture of formation energies outlined in the IQA calculations is well preserved in the relaxed dimers.

Geometry optimization of the complexes and the separated fragments allows us to estimate the energetic cost due to the distortion of the monomer geometries upon the formation of the complex (the so-called preparation energy in the EDA scheme). The corresponding difference between the energy of a P fragment in the dimer geometry, E^P , and in its isolated state, E^{P*} , would be amenable of IQA analysis to ascertain the localization and size of atomic distortion. However, we found that the actual values of the $E^P - E^{P*}$ differences are well below 0.5 kcal/mol for the majority of the S66 and ionic H-bond complexes. The IQA decomposition of such minor energetic effects due to geometric distortion correspond to very small changes of atomic additive energies (e.g., 0.05 , 0.1 kcal/mol) that are unsuitable for meaningful analysis. Only in the case of the doubly H-bonded complexes (uracil \cdots uracil, $\text{CH}_3\text{COOH}\cdots\text{HOOC-CH}_3$) and of the ionic complexes with short H-bonds (methylammonium \cdots methylamine, imidazolium \cdots methylamin), the internal rearrangement of the monomer geometries result in a significant energy cost of ~ 1 - 3 kcal/mol per fragment so that the IQA decomposition in terms of additive energies (ΔE_{add}^A) may render useful insight into the distortion energy. Two examples

are shown in Figure 4, which displays the atomic basins of two uracil fragments and the distribution of ΔE_{add}^A values as well as those of the methylammonium and methylamine fragments. Inspection of Figure 4 reveals the particular regions of the monomers that become strained (in blue) owing to the geometric rearrangement caused by the inter-fragment interactions. For the methylammonium and methylamine fragments, the N-H and :N H-bond ends, but also other two basins adjacent to the H-bond, become destabilized. The uracil monomers show a more complex pattern of stabilization/destabilization effects associated to their internal rearrangement (see Figure 6). Overall, although the geometric distortion in the examined complexes may have little importance, it may be interesting to note that ligand distortion can play an important role to determine the stability and selectivity of larger host-ligand complexes.

Conclusions

Certainly, the availability of standardized databases of high level ab initio energies and geometries for intermolecular complexes^[26] can be of great help to carry out computational experiments aimed to method development and validation. In this respect, the S66 and ionic H-bond datasets have been reexamined in this work using HF and DFT methods augmented with the empirical D3 dispersion corrections, but with the more specific goal of deriving fragment- and atom-based decompositions of any QM energy contributions to non-covalent binding energies using the IQA approach. The two datasets together comprise a variety of H-bond and dispersion complexes that are representative of typical intermolecular contacts in biomolecules.

The limitation to using HF and DFT methods in this study is because of the good applicability of the SCF methods to large systems and to maintain the computational cost of the IQA calculations within tractable bounds. Using triple- ζ basis sets, the examined methods predict reasonable formation energies that correlate quite well ($R^2 \sim 0.99$) with the benchmark data and have acceptable RMS errors, the PW6B95-D3 method providing the best performance. This

comparative analysis suggests that reliable energetic trends could be derived from similar QM calculations on a series of heterogeneous intermolecular complexes (*e.g.*, host-ligand systems). It may also be reasonably expected that the IQA calculations fed with HF and DFT densities could reliably measure the relative importance of functional groups, residues or specific atoms for the total stability of the complexes

The implementation of the IQA approach for the decomposition of molecular energy across the QTAIM basins relies heavily on numerical integration tasks, which, besides being expensive, introduce some numerical error into the IQA descriptors. We have shown in this work, however, that the combination of IQA settings that tradeoff accuracy and computational cost with partial cancellation of errors in the calculation of formation energies, results in IQA-reconstructed ΔE^{IQA} that have acceptable absolute errors, close to 0.5 kcal/mol on average, that can be taken as a measurement of the numerical error in the fragment- or atomic IQA components of formation energies (*e.g.*, ΔE_{add}). Such errors are generally much smaller than the values of the IQA terms, therefore, we conclude that the numerical accuracy of practical IQA calculations is good enough.

The decomposition of QM energies into chemically meaningful quantities is a topic of ongoing interest (*e.g.*, see the recent development of the variational EDA technique based on absolutely localized molecular orbitals^[48] and the interaction energy decomposition associated to the local pair natural orbital coupled cluster framework^[49]). In this respect, the usefulness of the combined IQA-D3 approach to analyze noncovalent intermolecular interactions is well illustrated in the fragment-based decomposition of the formation energies of the S66 and ionic H-bond complexes in terms of fragment-deformation (E_{def}^P), fragment- interaction (E_{int}^{PQ}) and dispersion corrections. Compared with other energy decomposition methods that normally distinguish more energy contributions (exchange, electrostatic, induction-polarization, correlation-dispersion, etc.), these IQA terms constitute a simpler partitioning that nevertheless

captures interesting features of the polar and non-polar complexes (*e.g.*, the larger electronic distortion of H-acceptor fragments in H-bond complexes). Of course additional details can be gained by looking at other energy components, among which the *classical* energy ($E_{\text{int,class}}^{PQ}$) due to Coulombic *e-e* and *e-nuclei* interactions turns out to be particularly useful to categorize the noncovalent complexes as electrostatic, dispersion or mixed. As similarly done in the original S66 reference using SAPT quantities, the ratio between $E_{\text{int,class}}^{PQ}$ and the total formation energy can be considered as a sort of noncovalent bonding index with a clear physical meaning.

Although the fragment-based IQA quantities allow us to analyze formation energies of the noncovalent dimers in comparable terms to those of perturbative approaches, the ability of IQA to split the total energy into atomic contributions is a unique feature that deserves further interest because it holds the promise of unambiguously quantifying atomic and group energy contributions in biomolecular systems with many functional groups. To this end, we have resorted to the additive atomic energies defined in IQA (although other IQA descriptors might be used). As shown graphically in Figures 4-5, the variation of these atomic properties (ΔE_{add}^A) offers a detailed picture of the *hot spots* in a given complex, which are essentially in agreement with expectations (*e.g.*, the atoms directly involved in H-bonds concentrate the energy changes while small or moderate changes are more evenly distributed in dispersion complexes). However, it is also clear that this type of IQA calculations may distinguish specific traits of functional groups behavior and/or score a series of inter- and intramolecular interactions in larger systems.

Altogether, the results presented in this work give support to the use of HF-D3 and DFT-D3 calculations followed by IQA-D3 energy decompositions to take further advantage of QM studies of non-covalent complexes in the form of atomic and residue energy contributions that may give new insight into the structural features controlling the stability of these systems. Similar computational work aimed at the study of intramolecular interactions and

conformational preferences will be also required to better assess the reliability of IQA in order to study the energetics of biomolecular complexes. These and other applications will require the inclusion of environment effects into the IQA framework (continuum solvents, MM point charges, ...) as well as the design and testing of a nearly linear scaling IQA implementation having good numerical accuracy on biomolecules comprising ~100-200 QM atoms.

Computational Details

Molecular geometries and benchmark energies

Cartesian coordinates and reference interaction energies for all the complexes were retrieved from the BEGDB (Benchmark Energy and Geometry Database) website.^[50] In particular, we accessed to the S66 dataset^[27] that collects the geometries of 66 optimized dimers at the counterpoise-corrected MP2/cc-pVTZ level. The benchmark energies correspond to formation energies derived by means of a composite approach that combines CCSD(T)/aug-cc-pVDZ and MP2/aug-cc-pVXZ (X=T,Q) energies to estimate the CCSD(T) energy at the complete basis set (CBS) limit. The formation energies are calculated using fixed monomer geometries as taken from the dimer optimizations and, therefore, no contribution due to the deformation of the monomer geometry is included. Similarly, we extracted the minimum-energy structures of 15 dimers from the “ionic hydrogen bonds” dataset,^[51] which also contains MP2/cc-pVTZ molecular geometries and CCSD(T)/CBS benchmark energies.

HF and DFT calculations

To generate the molecular orbitals required by the IQA calculations, we carried out single-point HF calculations on all the reference geometries using the cc-pVTZ (S66 complexes) and aug-cc-pVTZ (ionic H-bond complexes) basis sets.^[52] Single-point DFT calculations were also performed with three different exchange-correlation functionals: the hybrids B3LYP^[53-54] and PBE0^[55-56] functionals as well as the hybrid-meta PW6B95 functional,^[57] which combines Perdew-Wang exchange^[58] and Becke-95 correlation^[59] and includes 6 parameters optimized

for thermochemical calculations. These calculations were done with the GAMESS program^[60](HF, B3LYP, PBE0) and the ORCA 3.0.3 package^[61] (PW6B95). Single-point dispersion energies were obtained separately using the DFT-D3 program^[40] and choosing the Becke-Johnson damping function^[39, 62] that correctly reproduces the asymptotic behavior of the dispersion energy at small distances.^[63] The sensitivity of the interaction energies to the Basis Set Superposition Error (BSSE)^[64] was estimated by means of the Counterpoise (CP) correction.^[65]

To further assess the influence of the various levels of theory on the formation energies and on the IQA decomposition, we relaxed the structures of the intermolecular complexes by means of unconstrained energy minimizations that were started from the corresponding benchmark geometries. The ORCA package was employed to carry out all the energy minimizations with DFT-D3 dispersion energy and gradient corrections. Subsequently, the HF or KS molecular orbitals were computed on the optimized structures using the GAMESS program excepting for the PW6B95/cc-pVTZ ones, for which the ORCA program was used again.

IQA calculations

The IQA decomposition of the total energies to yield the net atomic energies, $E_{net}^A \equiv E_{net}(\Omega_A)$, and the interaction energies, $E_{int}^{AB} = E_{int}(\Omega_A, \Omega_B)$, was performed with the PROMOLDEN program^[17] at the considered levels of theory (*e.g.*, HF/cc-pVTZ, B3LYP/cc-pVTZ, etc.). The scaled exchange-correlation terms \tilde{E}_{xc}^{AB} associated to the DFT functionals were calculated with the LIBXC library^[66] that has been interfaced to PROMOLDEN. Pairwise dispersion energy corrections E_{disp}^{AB} were obtained separately using the DFT-D3 code.

The IQA terms are numerically integrated by PROMOLDEN over the QTAIM basins. These constitute finite and irregular integration domains and demand the application of angular and radial grids in atomic spherical quadratures that are much finer than those typically used by QM software. Even so the previous computational experience has shown that the IQA-

reconstructed energy of small and medium-sized molecules still differs from the parent HF or DFT energies in $\sim 10^{-3}$ - 10^{-4} au. In this work, we performed a preliminary study on the accuracy of IQA calculations on 22 small molecules (CH_4 , C_2H_2 , C_2H_4 , C_2H_6 , C_6H_6 , CH_2NH , HCl , HF , CH_3Cl , CH_3F , CH_3COOH , CH_3COO^- , OH^- , NH_3 , NH_4^+ , CH_3NH_2 , CH_3NH_3^+ , CH_3OCH_3 , H_2O , CH_3OH , CO_2 , H_2CO , H_2S , CH_3SH , $\text{C}_3\text{H}_4\text{N}_2$ (imidazole) and NH_2CONH_2 (urea)). After some computational experimentation with different integration settings, the following parameters, which represented a compromise choice between computational cost and accuracy, were set to carry out the rest of PROMOLDEN integrations. First a β -sphere around each atom was considered (*i.e.*, a sphere completely contained inside the atomic basin), with a radius equal to 60 % the distance of its nucleus to the closest bond critical point in the electron density. High-quality Lebedev angular grids were used with 5810 and 974 points outside and within the β -spheres, respectively. Euler-McLaurin radial quadratures were employed with 512 and 384 radial points outside and inside the β -spheres, respectively, the largest value of the radial coordinate in the integrations being 15.0 au. Maximum angular moments, λ_{max} , of 10 and 6 were assigned to the Laplace and bipolar expansions of r_{12}^{-1} outside and within the β -spheres. As shown in the Supporting Information, these settings give a mean unsigned difference between the IQA-reconstructed and the parent energies that is equal or below 0.001 hartrees for the 22 molecule set.

As already mentioned, we adopted the space partition into atomic basins derived from QTAIM, which is the common choice in IQA calculations although other divisions are also possible. In principle, every atomic basin shows one electron density cusp around the nucleus, but non-nuclear maxima (NNM) can also appear giving rise to extra QTAIM basins. Such NNM basins are not uncommon in molecules with multiple bonds (e.g., N_2) although their appearance depends on the level of theory and the bond distance. Among the systems studied in this work, NNM were only located in the middle of the $\text{C}\equiv\text{C}$ bond of ethyne molecules.

Hence, one NNM basin was added to the atomic basins of ethyne in order to fully integrate the first- and second-order density matrices. Regardless of the unclear meaning of the individual NNM IQA energy terms, the presence of the NNM basins does not introduce any difficulty or ambiguity in the fragment-based IQA decomposition of the interaction energies.

The Chimera visualization system^[67] was used to draw the ball-and-stick models of the complexes surrounded by the translucent surfaces of the QTAIM basins as calculated by the PROMOLDEN code. For reasons of computational efficiency and better handling of computer images, the basin surfaces were generated using a (θ, ϕ) angular grid of 72 x 36. For aesthetic considerations, the largest value of the radial coordinate in these surface calculations was limited to 4.0 au.

Acknowledgements

This research was supported by the CTQ2015-65790-P (MINECO, Spain) and the GRUPIN14-049 (FICyT, Spain) grants.

REFERENCES

- [1] R. F. Bader, *Chem. Rev.* **1991**, *91*, 893–928.
- [2] *The quantum theory of atoms in molecules: From solid state to DNA and drug design*, WILEY-VCH, Weinheim, **2007**.
- [3] P. L. A. Popelier, in *The chemical bond ii: 100 years old and getting stronger* (Ed.: D. M. P. Mingos), Springer International Publishing, Cham, **2016**, pp. 71-117.
- [4] J. Contreras-García, E. R. Johnson, S. Keinan, R. Chaudret, J.-P. Piquemal, D. N. Beratan, W. Yang, *J. Chem. Theory Comput.* **2011**, *7*, 625-632.
- [5] C. F. Matta, *J. Comput. Chem.* **2014**, *35*, 1165-1198.
- [6] E. Francisco, A. M. Pendas, M. A. Blanco, *J. Chem. Theory Comput.* **2006**, *2*, 90-102.

- [7] M. A. Blanco, A. M. Pendas, E. Francisco, *J. Chem. Theory Comput.* **2005**, *1*, 1096-1109.
- [8] A. M. Pendas, M. A. Blanco, E. Francisco, *J. Chem. Phys.* **2006**, *125*.
- [9] J. M. Guevara-Vela, R. Chavez-Calvillo, M. Garcia-Revilla, J. Hernandez-Trujillo, O. Christiansen, E. Francisco, A. M. Pendas, T. Rocha-Rinza, *Chemistry-a European Journal* **2013**, *19*, 14304-14315.
- [10] J. M. Guevara-Vela, E. Romero-Montalvo, V. A. Mora Gomez, R. Chavez-Calvillo, M. Garcia-Revilla, E. Francisco, A. M. Pendas, T. Rocha-Rinza, *Phys. Chem. Chem. Phys.* **2016**, *18*, 19557-19566.
- [11] E. Romero-Montalvo, J. M. Guevara-Vela, A. Costales, Á. M. Pendás, T. Rocha-Rinza, *Phys. Chem. Chem. Phys.* **2017**, *19*, 97-107.
- [12] E. Bartashevich, E. Troitskaya, A. M. Pendas, V. Tsirelson, *Computational and Theoretical Chemistry* **2015**, *1053*, 229-237.
- [13] D. Tiana, E. Francisco, P. Macchi, A. Sironi, A. M. Pendas, *J. Phys. Chem. A* **2015**, *119*, 2153-2160.
- [14] I. Cukrowski, J. H. de Lange, M. Mitoraj, *J. Phys. Chem. A* **2014**, *118*, 623-637.
- [15] A. L. Wilson, P. L. Popelier, *J Phys Chem A* **2016**, *120*, 9647-9659.
- [16] J. L. McDonagh, A. F. Silva, M. A. Vincent, P. L. A. Popelier, **2017**, *8*, 1937-1942.
- [17] A. M. Pendas, E. Francisco, Unpublished, **2015**.
- [18] A. M. Pendas, E. Francisco, M. A. Blanco, *J. Comput. Chem.* **2005**, *26*, 344-351.

- [19] R. Chavez-Calvillo, M. Garcia-Revilla, E. Francisco, A. M. Pendas, T. Rocha-Rinza, *Computational and Theoretical Chemistry* **2015**, *1053*, 90-95.
- [20] S. F. Vyboishchikov, A. Krapp, G. Frenking, *J. Chem. Phys.* **2008**, *129*, 144111.
- [21] S. F. Vyboishchikov, P. Salvador, M. Duran, *J. Chem. Phys.* **2005**, *122*, 244110.
- [22] P. Salvador, I. Mayer, *J. Chem. Phys.* **2007**, *126*, 234113.
- [23] E. Francisco, J. L. Casals-Sainz, T. Rocha-Rinza, A. Martín Pendás, *Theoretical Chemistry Accounts* **2016**, *135*.
- [24] S. Grimme, *WIREs Comput Mol Sci* **2011**, *1*, 211-228.
- [25] S. Grimme, A. Hansen, J. G. Brandenburg, C. Bannwarth, *Chem. Rev.* **2016**, *116*, 5105-5154.
- [26] P. Hobza, *Acc. Chem. Res.* **2012**, *45*, 663-672.
- [27] J. Řezáč, K. E. Riley, P. Hobza, *J. Chem. Theory Comput.* **2011**, *7*, 2427-2438.
- [28] T. J. Hughes, S. M. Kandathil, P. L. A. Popelier, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **2015**, *136*, 32-41.
- [29] A. M. Pendas, M. A. Blanco, E. Francisco, *J. Chem. Phys.* **2004**, *120*, 4581-4592.
- [30] P. Maxwell, Á. M. Pendás, P. L. A. Popelier, *Phys. Chem. Chem. Phys.* **2016**, *18*, 20986-21000.
- [31] V. Tognetti, L. Joubert, *ChemPhysChem* **2017**, *18*, 2675-2687.
- [32] I. Cukrowski, *Computational and Theoretical Chemistry* **2015**, *1066*, 62-75.
- [33] A. M. Pendas, M. A. Blanco, E. Francisco, *J. Comput. Chem.* **2009**, *30*, 98-109.

- [34] T. Ziegler, A. Rauk, *Theoret. Chim. Acta* **1977**, *46*, 1-10.
- [35] F. M. Bickelhaupt, E. J. Baerends, in *Rev. Comput. Chem.*, John Wiley & Sons, Inc., **2007**, pp. 1-86.
- [36] E. D. Glendening, A. Streitwieser, *J. Chem. Phys.* **1994**, *100*, 2900-2909.
- [37] A. J. Misquitta, R. Podeszwa, B. Jeziorski, K. Szalewicz, *J Chem Phys* **2005**, *123*, 214103.
- [38] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.
- [39] A. D. Becke, E. R. Johnson, *J. Chem. Phys.* **2005**, *122*, 154104.
- [40] R. Sedlak, J. Řezáč, *J. Chem. Theory Comput.* **2017**, *13*, 1638-1646.
- [41] L. Goerigk, J. R. Reimers, *J. Chem. Theory Comput.* **2013**, *9*, 3240-3251.
- [42] L. Goerigk, C. A. Collyer, J. R. Reimers, *J. Phys. Chem. B* **2014**, *118*, 14612-14626.
- [43] L. Goerigk, S. Grimme, *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670-6688.
- [44] M. G. Medvedev, I. S. Bushmarinov, J. Sun, J. P. Perdew, K. A. Lyssenko, *Science* **2017**, *355*, 49-52.
- [45] K. E. Riley, P. Hobza, *J. Phys. Chem. A* **2007**, *111*, 8257-8263.
- [46] L. Goerigk, H. Kruse, S. Grimme, *ChemPhysChem* **2011**, *12*, 3421-3433.
- [47] E. Francisco, A. M. Pendas, in *Non-covalent interactions in quantum chemistry and physics. Theory and applications* (Eds.: A. Otero-de-la-Roza, G. DiLabio), Elsevier, **2017**, pp. 27-62.

- [48] P. R. Horn, Y. Mao, M. Head-Gordon, *Phys. Chem. Chem. Phys.* **2016**, *18*, 23067-23079.
- [49] W. B. Schneider, G. Bistoni, M. Sparta, M. Saitow, C. Riplinger, A. A. Auer, F. Neese, *J. Chem. Theory Comput.* **2016**, *12*, 4778-4792.
- [50] J. Rezac, P. Jurecka, K. E. Riley, J. Cerny, H. Valdes, K. Pluhackova, ., K. Berka, T. Rezac, M. Pitonak, J. Vondrasek, P. Hobza, *Collect. Czech. Chem. Commun.* **2008**, *73*, 1261.
- [51] J. Řezáč, P. Hobza, *J. Chem. Theory Comput.* **2012**, *8*, 141-151.
- [52] T. H. D. Jr., *J. Chem. Phys.* **1989**, *90*, 1007-1023.
- [53] A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098-3100.
- [54] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785-789.
- [55] J. P. Perdew, M. Ernzerhof, K. Burke, *J. Chem. Phys.* **1996**, *105*, 9982-9985.
- [56] C. Adamo, V. Barone, *J. Chem. Phys.* **1999**, *110*, 6158-6170.
- [57] Y. Zhao, D. G. Truhlar, *J. Phys. Chem. A* **2005**, *109*, 5656-5667.
- [58] J. P. Perdew, K. Burke, Y. Wang, *Phys Rev B Condens Matter* **1996**, *54*, 16533-16539.
- [59] A. D. Becke, *J. Chem. Phys.* **1996**, *104*, 1040-1046.
- [60] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, J. A. Montgomery, *J. Comput. Chem.* **1993**, *14*, 1347-1363.
- [61] F. Neese, *WIREs Comput Mol Sci* **2012**, *2*, 73-78.
- [62] S. Grimme, S. Ehrlich, L. Goerigk, *J. Comput. Chem.* **2011**, *32*, 1456-1465.

- [63] A. Koide, *Journal of Physics B: Atomic and Molecular Physics* **1976**, *9*, 3173.
- [64] N. R. Kestner, J. E. Combariza, in *Rev. Comput. Chem.*, John Wiley & Sons, Inc., **2007**, pp. 99-132.
- [65] S. F. Boys, F. Bernardi, *Mol. Phys.* **1970**, *19*, 553-566.
- [66] M. A. L. Marques, M. J. T. Oliveira, T. Burnus, *Comput. Phys. Commun.* **2012**, *183*, 2272-2281.
- [67] E. F. Pettersen, T. D. Goddard, C. C. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J Comput Chem* **2004**, *25*, 1605-1612.