

Interacting Quantum Atoms Approach and Electrostatic Solvation Energy: Assessing Atomic and Group Solvation Contributions

*Fernando Jiménez-Grávalos, Natalia Díaz, Evelio Francisco, Ángel Martín-Pendás and Dimas Suárez**

Departamento de Química Física y Analítica. Universidad de Oviedo.

Julián Clavería 8. 33006 Oviedo (Asturias) Spain.

Keywords: Solvation • Atomic Properties • Computational Chemistry • Quantum Chemistry

Abstract: The interacting quantum atoms (IQA) method decomposes the total energy of a molecular system in terms of one- and two-center (atomic) contributions within the context of the quantum theory of atoms in molecules. Here we incorporate electrostatic continuum solvent effects into the IQA energy decomposition. To this end, the interaction between the solute electrostatic potential and the solvent screening charges as defined within the COSMO solvation model is now included in a new version of the PROMOLDEN code, allowing thus to apply IQA in combination with COSMO-quantum chemical methods as well as to partition the electrostatic solvation energy into effective atomic and group contributions. To test the robustness of this approach, we carry out COSMO-HF/aug-cc-pVTZ calculations followed by IQA calculations on more than 400 neutral and ionic solutes extracted from the MNSol database. The computational results reveal a detailed atomic mapping of the electrostatic solvation energy that is useful to assess to what extent the solvation energy can be decomposed into atomic and group contributions of various parts of a solute molecule, as generally assumed by empirical methodologies that estimate solvation energy and/or $\log P$ values.

Introduction

The Quantum Chemical Topology (QCT) methods¹ take advantage of the topological properties of scalar fields (charge density and others) in order to gain new chemical information about bonding and molecular properties. Among them, the interacting quantum atoms (IQA) approach,²⁻³ employs the first- and second-order reduced density matrices to partition the expectation values into atomic regions such as the attraction basins (Ω_A) of the gradient field of the electron density. Thus, IQA provides self-atomic energies, $E(\Omega_A)$, which tend to the free atomic energies at the limit of non-interacting atoms, and diatomic $E(\Omega_A, \Omega_B)$ energies that unambiguously discriminate between classical electrostatic and exchange-correlation energy terms. Using DFT (and HF) charge densities, IQA can be augmented with the Grimme's D3 potential,⁴⁻⁵ which yields pairwise dispersion energies $E_{disp}(A-B)$ that complement the diatomic $E(\Omega_A, \Omega_B)$ IQA terms, constituting thus an effective D3-IQA decomposition scheme⁶ applicable to medium-sized and large systems. Thus, the IQA or D3-IQA decomposition has been successfully applied to quantify many different aspects of chemical bonds and intermolecular forces. Among the various topics that have been recently addressed using IQA, we find the nature and cooperativity of H-bond interactions,⁷⁻¹⁰ halogen bonding patterns,¹¹ interactions within transition metal complexes,¹²⁻¹³ description of short-range repulsions,¹⁴ fine-tuning effects of electron correlation within covalent and non-bonded interactions,¹⁵ the categorization of non-covalent bonding and the atomic decomposition of intermolecular binding energies,⁶ etc..

Up to date all the IQA calculations have been performed considering molecular species (single molecules, dimers, clusters) in the gas-phase. However, it is clear that solvent plays a major role in determining the stability and molecular properties of organic molecules and biomolecules in solution. Moreover, since we are interested in pursuing the application of D3-IQA to quantify atomic and group energy contributions in biomolecular systems with many functional groups, the treatment of solvent effects within the IQA framework is therefore a prerequisite. To this end, the combination of implicit solvent models and quantum mechanical (QM) methods constitutes probably the best methodological

choice given that the continuum treatment of solvent focuses on the degrees of freedom of the solute(s) while it provides an accurate description of the strong, long-range electrostatic forces that dominate solvation energies in high dielectric solvents.¹⁶ In these hybrid approaches, the solute-solvent electrostatic interaction is usually described in terms of the reaction (electric) field exerted on the QM charge density of the solute by the solvent that in turn is polarized. Other empirical and semiempirical methods, which generally do not affect the charge density of the solute, have been proposed to estimate the non-electrostatic contributions to solvation energy, which are significant, especially in non-polar solvents.¹⁷

The decomposition of the QM energy in solution by IQA would render atomic and group contributions to the solute solvation energy. Nonetheless, the actual significance of this partitioning should be carefully considered. Thus, in classical Statistical Mechanics,¹⁸ the free energy of solvation of a rigid solute can be expressed as:

$$\Delta G_{solv} = -RT \ln \left\langle \exp(-V/RT) \right\rangle_{NPT}$$

where V is the solute-solvent interaction potential and $\langle \rangle$ represents the ensemble average over all possible configurations of the solvent molecules in the system. Although the total interaction energy V may be split into group/atomic contributions of the different solute atoms, the ensemble average in the above expression cannot be factorized into a product of two or more average quantities. Physically, this means that the solvation shells around the solute atoms/groups are correlated at varying degrees and, therefore, it is not feasible to achieve an exact additivity of ΔG_{solv} . Nevertheless, computational models have been developed for the fast prediction of hydration energies or partition coefficients (such as the logarithm of the octanol-water partition coefficient, which can be calculated as $\log P = (\Delta G_{solv}^{octanol} - \Delta G_{solv}^{water}) / 2.303RT$) that rely on the assumption of atomic/fragment additivity for ΔG_{solv} and derive atomic/fragment parameters using different optimization strategies.¹⁹⁻²³

In a previous work,²⁴ QM calculations have been used to heuristically determine group contributions to the free energy of solvation. However, this study is limited to a family of closely-related heterocyclic compounds, although it is concluded that group contributions are slightly affected by the chemical environment. Therefore, the IQA decomposition of the solvation free energy into *effective* atomic terms, which ultimately arise from the topological partitioning of the solute charge density, constitutes an opportunity to further assess the additivity assumption.

In the rest of the paper, we will briefly describe the theoretical details of the IQA extension to accomplish the decomposition of the QM energy of solute molecules embedded within a solvent continuum. In doing so, we will focus on the electrostatic solute-solvent interaction accounted for by the *conductor-like screening model* (COSMO). The IQA-COSMO protocol will be applied to a large set of organic molecules retrieved from the Minnesota Solvation database (MNSol),²⁵⁻²⁶ which collects experimental free energies of solvation for hundreds of solutes and QM optimized geometries for the corresponding solutes. For a subset of neutral and ionic solutes comprising 412 molecules, we perform geometry optimizations both in the gas-phase and in solution using the HF method with a triple- ζ basis set followed by full IQA calculations. Then we will assess the accuracy of the calculated solvation energies and the numerical errors in the IQA-reconstructed energies. Subsequently, we will characterize statistically the fragment-based IQA contributions to the electrostatic solvation energies. The chemical fragments comprise united atom types and functional groups that are selected using similar prescriptions to those of Meylan and Howard.²³ Finally, we will assess the goodness of the solvation energy additivity approximation by comparing between COSMO-HF energies and additive energies for an additional set of MNSol structures not considered in the IQA calculations.

Theory

IQA in the gas-phase

Starting with the atomic basins (Ω_A), which stem from the topological properties of the charge distribution $\rho(\mathbf{r})$, the IQA approach^{2,27} needs two scalar fields derived from the QM wavefunction, the

first order reduced density matrix $\rho_1(\mathbf{r}_1, \mathbf{r}_1')$ and the pair density, $\rho_2(\mathbf{r}_1, \mathbf{r}_2)$. Then IQA decomposes the total energy of a molecular system in the gas-phase (E^{gas}) as

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

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 nucleus-electron (ne) attractions and electron-electron repulsions (ee) within Ω_A . 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). Note that in the IQA terminology, *interaction* energies are the diatomic contributions to the *absolute* energy of a molecule. By grouping half the interaction energy terms involving atom Ω_A and its net energy, we define its additive energy,

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

so that the sum of all the E_{add}^A terms reproduces the total energy E^{gas} .

Implicit solvent methods: COSMO

Several excellent reviews have been published^{16-17, 28-29} that examine the various approximations underlying the implicit solvent methods. Herein, we briefly review the most basic concepts and some details of the COSMO method that are required to understand the IQA-COSMO protocol. Thus, the definition of the molecular cavity and the description of electrostatic solute-solvent interaction are the basic elements of a continuum solvent model.²⁸ The shape and size of the cavity are typically defined by a solvent excluded surface (SES), which encloses the volume in which the solvent molecules cannot penetrate. Thus, the SES, which is the boundary of the molecular cavities, can be computed using different sets of atomic radii and numerical algorithms depending on the continuum model. Among the various techniques for solving the electrostatic problem, the apparent surface charges (ASC) method²⁸ allows a direct implementation of continuum solvent effects within IQA. In this approach, the reaction field potential generated by the polarization of the dielectric medium is expressed in terms

of a set of point charges q_k assigned to small surface segments (tesserae) located at positions \mathbf{s}_k . The values of q_k are determined by imposing the proper boundary conditions on the SES that, in the COSMO model,³⁰⁻³¹ correspond to the vanishing potential on and within a grounded conductor. The COSMO model mimics solvents with finite dielectric constant by scaling down the q_k values by a factor $f(\epsilon)=(\epsilon-1)/(\epsilon+x)$ with $x=0.5$ and 0.0 for neutral and ionic molecules, respectively. The so-called outlying charge, which arises from the tail of the solute electron density that lies outside the molecular cavity, can affect negatively the results of the continuum models. In the COSMO methodology, the outlying charge correction (*occ*) is an heuristic approximation that corrects both the apparent surface charges q_k and the solute electrostatic potential $\Phi(\mathbf{s}_k)$.³² However, we found that the COSMO hydration energies with and without the *occ* term correlate with experimental data very similarly (see below) and, therefore, we decided not to include the *occ* term in the IQA decomposition.

Knowing the values of the apparent surface charges q_k , the solute-solvent electrostatic energy (V_{solv}) is

$$V_{solv} = V_{solv,n} + V_{solv,e} = \sum_A \sum_k \frac{Z_A q_k}{|\mathbf{R}_A - \mathbf{s}_k|} + \sum_k \int \frac{\rho(\mathbf{r}) q_k}{|\mathbf{r} - \mathbf{s}_k|} d\mathbf{r} \quad (3)$$

where $\rho(\mathbf{r})$ is the electron charge density of the solute and Z_A is the nuclear charge of the atom A located at \mathbf{R}_A . This expression can be rewritten as

$$V_{solv} = \sum_k \Phi(\mathbf{s}_k) q_k \quad (4)$$

where $\Phi(\mathbf{s}_k)$ is the total electrostatic potential created by the solute acting on the tesserae \mathbf{s}_k . This solute-solvent interaction energy accounts for the electric work (free energy) needed to transfer the unperturbed solute from the gas-phase to the solvent cavity in the presence of the q_k charges located at the \mathbf{s}_k positions, but does not describe the polarization of the solvent continuum. Assuming a linear response regime, it can be shown that the free energy due to the building of the solvent polarization is equal to $-V_{solv}/2$.²⁹ Therefore, the total free energy gain associated to the solvation process would be $V_{solv}/2$.

To solve the QM problem of the solute embedded in the continuum, the implicit solvent methods construct effective Hamiltonians that include both the solute-solvent interaction and the solvent polarization. Mutual solute-solvent polarization effects are considered through a self-consistent reaction field (SCRF) iterative process. For HF/DFT methods, the wavefunction/charge density in solution is iteratively obtained so that the molecular/Kohn-Sham orbitals and the reaction field potential (*i.e.*, q_k values in COSMO method) are updated at each self-consistent-field cycle. After convergence, the QM energy of a molecular system and the dielectric continuum is obtained by adding the $V_{solv}/2$ term to the rest of kinetic and potential energy terms associated to the electronic and nuclear degrees of freedom of the solute.

IQA partitioning of the solute-solvent interaction energy

The IQA partitioning of the total QM energy in solution derived from the COSMO method relies on the monoelectronic character of the solute electrostatic potential, Φ . Its decomposition into atomic contributions is straightforward,

$$\Phi(\mathbf{s}_k) = \sum_A \Phi^A(\mathbf{s}_k) \quad (5)$$

so that $\Phi^A(\mathbf{s}_k)$ is the electrostatic potential created by the nuclear charge and electron density confined within the atomic basin Ω_A . This quantity is readily computable within the IQA framework, yielding thus the atomic contribution to V_{solv} ,

$$V_{solv}^A = \sum_k \Phi^A(\mathbf{s}_k) q_k \quad (6)$$

Once that the SCRF process is converged and the corresponding density matrices become available, IQA decomposes the QM energy in solution as,

$$E^{sol} = \sum_A \left(T^A + V_{ne}^A + V_{ee}^A + \frac{1}{2} V_{solv}^A \right) + \sum_{A>B} \left(V_{nm}^{AB} + V_{ne}^{AB} + V_{ne}^{BA} + V_{ee}^{AB} \right) \quad (7)$$

Solvation free energy

Ignoring thermal corrections to the free energy associated with the solute degrees of freedom, it turns out ²⁹ that the electrostatic solvation free energy ΔG_{solv} is the difference between the QM energies in solution and in the gas-phase

$$\Delta G_{solv} = E^{sol} - E^{gas} \quad (8)$$

Two separate contributions to ΔG_{solv} can be defined: the Coulomb term ΔG_{solv}^{Coul} due to the electrostatic interaction between the unperturbed (*i.e.*, not polarized) solute and the solvent continuum and the polarization or induction ΔG_{solv}^{pol} energy gained upon mutual polarization. By computing E^{sol} through a single SCRF cycle with the unperturbed gas-phase charge density, ΔG_{solv}^{Coul} can be known and thereby $\Delta G_{solv}^{pol} = \Delta G_{solv} - \Delta G_{solv}^{Coul}$. Inserting then the corresponding IQA additive energies, ΔG_{solv}^A , ΔG_{solv}^{Coul} and ΔG_{solv}^{pol} can be decomposed into effective atomic solvation energies. For example,

$$\Delta G_{solv} = E^{sol} - E^{gas} = \sum_A (E_{add}^{sol,A} - E_{add}^{gas,A}) = \sum_A \Delta G_{solv}^A \quad (9)$$

Let us stress that this is indeed an *effective* partitioning given that each ΔG_{solv}^A term collects the mutual solute-solvent polarization effects due to the charge density $\rho(\mathbf{r})$ of the solute within the basin Ω_A and *all* the apparent surface charges distributed over the molecular surface.

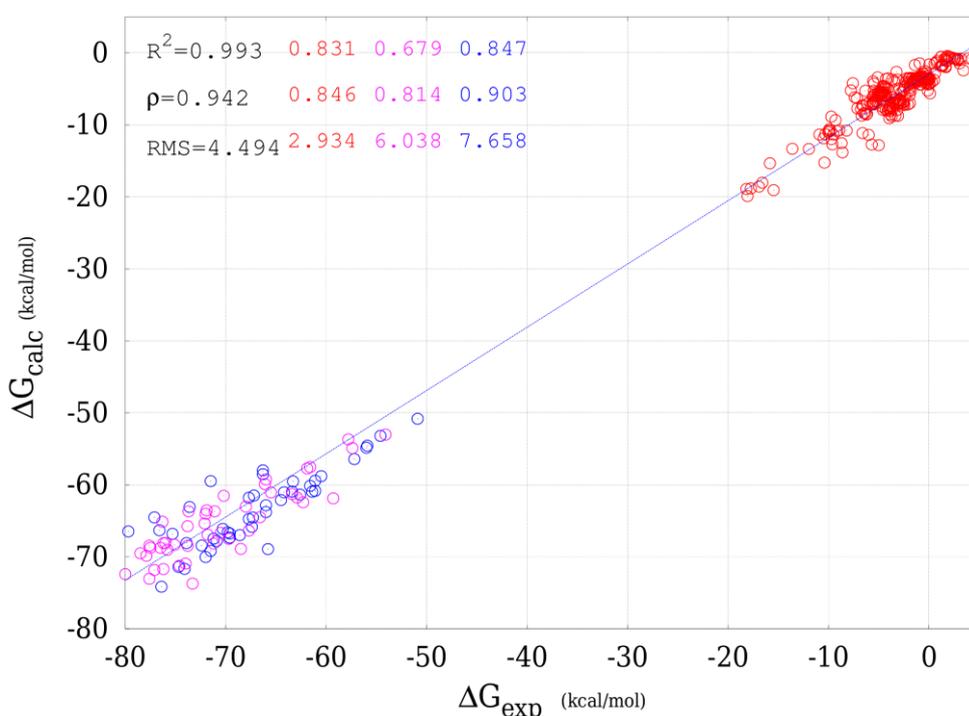
Results and Discussion

Solvation energy calculations

The MNSol database contains the Cartesian coordinates of 533 molecules solvated by pure water and of other 106 molecules in mixed aqueous organic solvent (*e.g.*, water-octanol). From these MNSol data, we selected 412 molecules to carry out the QM and IQA calculations on the basis of appropriate molecular size. Thus, small molecules containing one or two heavy atoms (*e.g.*, water, ammonia, acetylene, hydroxide anion, etc.) were not considered as the emphasis is placed on the analysis of functional group contributions. Relatively big molecules containing more than 25 atoms were not selected neither in order to keep the computational cost of the expensive IQA calculations within reasonable bounds. In the final set, 57 anionic and 49 cationic species were included. The molecular

geometries of all the molecules were fully optimized at the HF/aug-cc-pVTZ level both in the gas-phase and in the solvent continuum.

Figure 1. Comparison between the COSMO-HF/aug-cc-pVTZ calculated hydration energies (ΔG_{calc} in kcal/mol) and the corresponding experimental values (ΔG_{exp}) of the selected MNSol structures in aqueous solvent. The determination coefficient (R^2), the Spearman correlation coefficient (ρ) and the root mean square (RMS) error in kcal/mol are also indicated for the whole data set (in black) and for the neutral molecules (in red), anionic (in blue) and cationic (in magenta) categories. The blue dashed line is the least squared fit line between the calculated and the reference data.

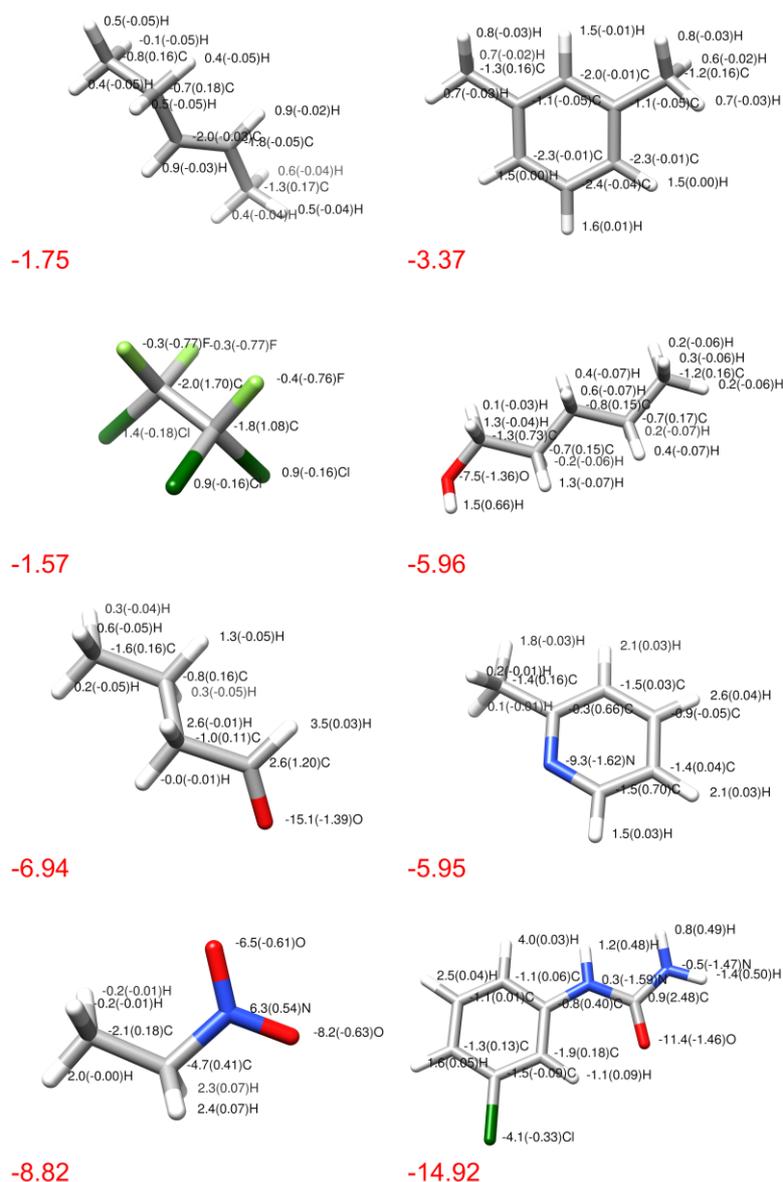


In Figure 1 the experimental hydration energies of 364 molecules are compared with the COSMO-HF/aug-cc-pVTZ ΔG_{solv} values. The global correlation between calculated and experimental data is strong as the determination coefficient R^2 amounts to 0.993. The root mean squared (RMS) error is significant, 4.5 kcal/mol, which is not entirely unexpected due to the lack of electron correlation effects, non-polar solvation contributions and conformational sampling. The performance of the COSMO-HF/aug-cc-pVTZ level depends on the charge state of the solute molecules. Thus, the

solvation energies for neutral and anionic molecules show a good correlation with experimental values ($R^2=0.83$ and 0.85 , respectively) whereas the prediction capacity for the cationic species turns out to be lower ($R^2=0.68$). We also note in passing that the COSMO-HF/aug-cc-pVTZ ΔG_{solv} energies including the outlying charge correction have very similar statistical metrics ($R^2= 0.992, 0.846, 0.844$ and 0.688 for the full set, neutral, anionic and cationic solutes, respectively).

The reliability of the COSMO-HF ΔG_{solv} values is not far from that of more sophisticated QM solvent models like the embedded cluster reference interaction site model (EC-RISM) integral equation theory coupled with the MP2/6-311+G(d,p) ab initio method.³³ This optimized EC-RISM protocol, which incorporates conformational sampling, yields a global $R^2=0.99$ and RMS error of 2.4 kcal/mol for the MNSol structures in water. The statistical measurements of the EC-RISM data are also less satisfactory if neutral and charged species are analyzed separately ($R^2=0.89, 0.88$ and 0.85 for neutrals, anions and cations, respectively). Hence, we conclude that the HF/aug-cc-pVTZ level, which is particularly suitable for carrying out the IQA calculations, captures reasonably well the trends exhibited by the hydration energies of neutral and anionic molecules, the case of cations being somewhat less satisfactory.

Figure 2. Atomic distribution of the IQA ΔG_{solv}^A energies (in kcal/mol) and Bader atomic charges (in parentheses) for some neutral MNSol molecules: *E*-2-pentene, *m*-xylene, 1,1,2-trichloro-1,2,2-trifluoroethane, pentanol, butanal, 2-methylpyridine, nitroethane, and 3-chlorophenylurea. Total ΔG_{solv} values in kcal/mol are also indicated (in red).

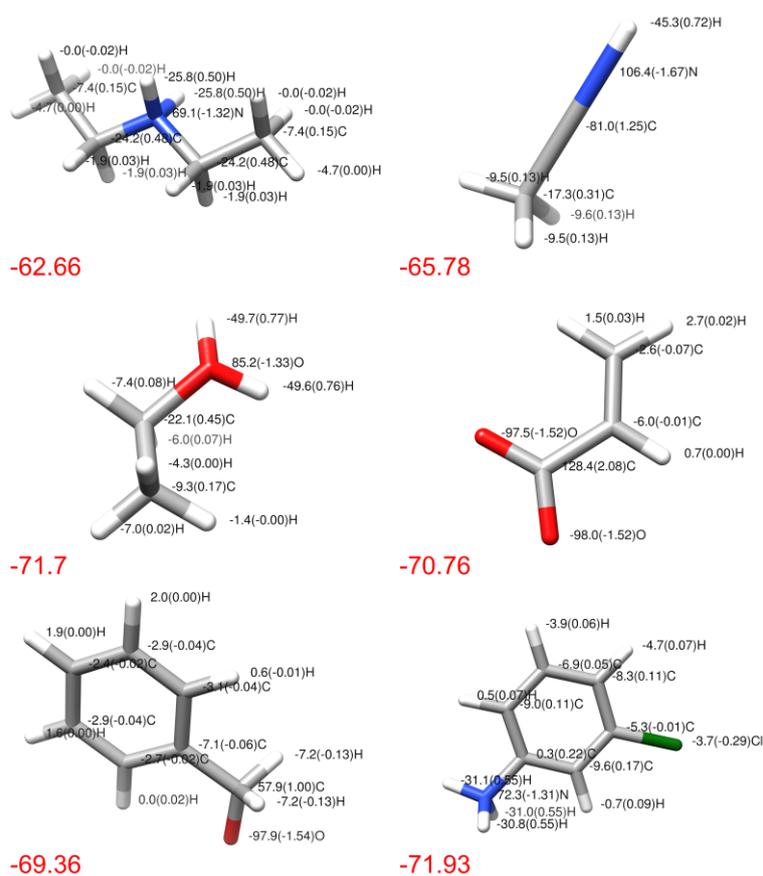


IQA decomposition of solvation energy

The calculation of the IQA energy terms involves six-dimensional numerical integration over the atomic basins, which is computationally expensive and introduces some numerical error.³⁴⁻³⁵ To estimate this error, we compared the solvation energies ΔG_{solv} derived from the gas-phase and COSMO HF calculations with their counterpart values obtained from the IQA-reconstructed energies ($\Delta G_{\text{solv}}^{\text{IQA}}$). The absolute differences $|\Delta G_{\text{solv}} - \Delta G_{\text{solv}}^{\text{IQA}}|$ can be considered as a measure of the “IQA numerical error” in the decomposition of solvation energies. Its mean value is 0.76 ± 1.36 kcal/mol, which corresponds to a average error per atom of 0.06 ± 0.10 kcal/mol. The magnitudes of these error estimates are similar to those previously found in the IQA decomposition of formation energies for non-covalent complexes.⁶ We note again that the actual interest of the IQA energy partitioning resides in the atomic and/or fragment-based IQA components and they have values ranging from ~ 0.5 to tens of kcal/mol in absolute value (see below) that are well above the mean numerical error per atom.

Figure 2 displays the stick models of the COSMO HF/aug-cc-pVTZ optimized structures for selected neutral solutes together with the $\Delta G_{\text{solv}}^{\text{A}}$ energies (in kcal/mol) and the charge corresponding to each atomic basin (Figure S1 in the supporting information shows the same data for the whole set of structures). For the non-polar hydrocarbon molecules, the effective atomic contributions to the electrostatic solvation energy are mainly negative (favorable) for the C atoms and positive for Hs, the absolute values of $\Delta G_{\text{solv}}^{\text{A}}$ being small (~ 0.5 - 2.0 kcal/mol). The atomic charges throughout these hydrophobic molecules are also quite small (0.01 - 0.05 in absolute value) and they are uncorrelated with the $\Delta G_{\text{solv}}^{\text{A}}$ values. Indeed another non-polar molecule, the halogen-substituted ethane, exhibits relatively large atomic charges (e.g., $q_{\text{A}} = -0.8, +1.7$), but $|\Delta G_{\text{solv}}^{\text{A}}|$ values < 1.6 kcal/mol (see Figure 2).

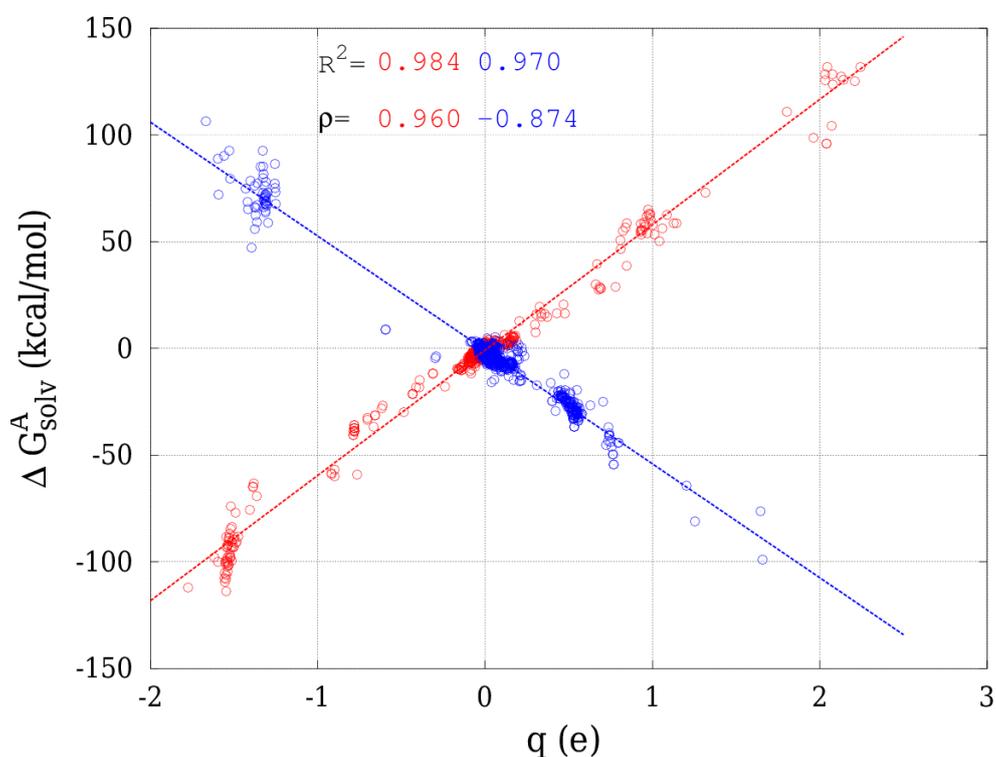
Figure 3. Atomic distribution of the IQA ΔG_{solv}^A energies (in kcal/mol) and Bader atomic charges (in parentheses) for some ionic MNSol molecules: *N*-ethylethanamine (+), acetonitrile (+), ethanol (+), acrylic acid (-), phenylmethanol (-), 3-chloroaniline (+). Total ΔG_{solv} values in kcal/mol are also indicated (in red).



The stronger hydration induced by the polar groups (alcohol, aldehyde, amino, nitro, etc.) into the neutral solutes, whose total ΔG_{solv} values are around -6, -8 kcal/mol, arises from relatively large ΔG_{solv}^A contributions associated to the atomic basins of the polar functional groups. For example, the three atoms of the aldehyde $\text{HC}=\text{O}$ group in butanal result in a $\Delta G_{\text{solv}}^{\text{CHO}}$ term of -9.0 kcal/mol, the molecular ΔG_{solv} value being -6.9 kcal/mol. The non-polar hydrocarbon moieties in the mono-substituted compounds yield atomic ΔG_{solv}^A values (~ 0.5 - 2.0 kcal/mol in absolute value) that are similar to those in the non-polar hydrocarbon molecules. The ΔG_{solv}^A and q^A values of polar sites are more widely distributed than in non-polar molecules and exhibit R^2 values ranging between 0.4 and 0.8 depending

on the solute molecule. Thus, it seems that atomic charges significantly determine the polar $\Delta G_{\text{solv}}^{\text{A}}$ contributions.

Figure 4. Correlation plot between the Bader's charges of atomic basins (q^{A} , derived from COSMO-HF/aug-cc-pVTZ density) and the corresponding IQA $\Delta G_{\text{solv}}^{\text{A}}$ energies for ionic molecules: cationic (in blue) and anionic (in red). The dashed lines are the least squared fit lines. The determination coefficient (R^2) and the Spearman correlation coefficient (ρ) are also indicated.



Ionic solutes are characterized by large solvation energies of tens or even hundreds of kcal/mol in water. For the cationic and anionic solutes shown in Figure 3, the IQA COSMO-HF calculations indicate again that the ionic functionalities ($-\text{OH}_2^+$, $-\text{COO}^-$, $-\text{NH}_3^+$, ...) concentrate the solute-solvent interaction and present the largest $\Delta G_{\text{solv}}^{\text{A}}$ energies. For example, the carboxylate group in the acrylic acid gives a $\Delta G_{\text{solv}}^{\text{COO}^-}$ term of -67.1 kcal/mol while the calculated hydration energy is -71.7 kcal/mol. However, the global charge of the ionic species also affects the atomic $\Delta G_{\text{solv}}^{\text{A}}$ contributions of non-polar sites that tend to have values from ± 5 to ± 20 kcal/mol, the greater contributions occurring at the vicinal positions with respect to the charged groups. With respect to the neutral solutes, the $\Delta G_{\text{solv}}^{\text{A}}$

energies and the Bader charges q^A are distributed over much wider ranges so that a stronger dependency can be expected. As a matter of fact, linear regression in the $(\Delta G_{\text{solv}}^A, q^A)$ data set derived from *all* the anionic or cationic solutes results in overall R^2 coefficients of 0.98 and 0.97 for cations and anions, respectively (see Figure 4). The COSMO-HF q^A values were used in this correlation analysis, but nearly identical statistical parameters are obtained if the gas-phase HF charges are adopted instead. Therefore, we conclude that the $\Delta G_{\text{solv}}^A / q^A$ relationship would be transferable for other ions and that the atomic charge distribution closely determines the hydration of the ionic solutes. We also note that, for the slightly polar or non-polar solutes, the electrostatic ΔG_{solv}^A values would probably be controlled by other multipolar terms (dipole, quadrupole, ...) associated to the solute charge density.

Assessment of atomic/fragment contributions to solvation energy

Inspection of the ΔG_{solv}^A energies in Figures 2-3 shows that the solvation contributions of two covalently bonded atoms have opposite signs as they usually have opposite q^A charges too. In the case of *A-B* polar bonds that imply a significant charge separation, the *A* and *B* contributions can be comparable to the total ΔG_{solv} and even larger for ionized groups. For example, the ammonium group $-\text{NH}_3^+$ in 3-chloroaniline (see Figure 3) has a $\Delta G_{\text{solv}}^{\text{NH}_3^+}$ of -20.7 kcal/mol arising from the sum of the N (+72.3 kcal/mol) and H terms (-31.1, -31.1 and -30.8). Therefore, we believe that the best strategy for standardizing and analyzing the IQA decomposition of the solvation energy would consist of adopting a *united atom* approach in such a way that H contributions in $-\text{XH}_n$ fragments are merged with that of the heavy atom X. Subsequently, the resulting united atom X' groups can be useful to define a set of atom types and functional groups into which a given organic molecule can be formally decomposed.

It may be interesting to note that the *united atom* approach suggested by the IQA analysis has been used in empirical solvation and/or $\log P$ methods like the atom/fragment contribution method of Meyland and Howard,²³ which defines 130 fragments to estimate the $\log P$ of organic compounds. This

model employs multiple linear regression of thousands of compounds to estimate the $\log P$ in terms of fragment contributions (f_G) using the following equation,

$$\log P = \sum_G f_G n_G + \sum_I c_I n_I + b \quad (10)$$

where b is a regression parameter, n_G is the number of times that a group occurs in the structure and c_I are specific correction terms that apply only for a subset of fragment combinations involving aromatic ring substituents, ring strain, electronic conjugation, etc. Thus, in this scenario, the f_G terms could be somehow related to the IQA ΔG_{solv}^G terms.

Table 1 lists the 51 atom types/functional groups (G) that were selected for analyzing the fragment-based IQA contributions to the electrostatic solvation energy. The mean values and standard deviations of the corresponding ΔG_{solv}^G energies are also collected in Table 1. We decided to analyze the fragment contributions of groups that appear at least 5 times in the set of MNSol structures. For this reason, several groups appearing in the solute molecules (*e.g.*, thiol $-\text{SH}$, phosphate $-\text{PO}_4$, etc.) are not included in Table 1. Therefore, our analysis is not exhaustive and is not oriented to derive a working additive model of hydration energies for organic molecules, but to find out whether or not we can extract useful information from the IQA decomposition.

Table 1. Selected atom/functional groups (G) for the analysis of electrostatic solvation. The number of ΔG_{solv}^G values (n), their mean value (μ in kcal/mol), standard deviation (σ), skewness (skw) and excess kurtosis (kurt) are given. The symbol of groups belonging to anionic/cationic compounds is augmented by (-)/(+), respectively. Charged groups are denoted by placing their sign into square brackets ([+]/[-]).

Atom/ Group	Description	n	μ	σ	Skw	Kurt
-Br	bromine	29	-4.7	3.0	0.0	-1.1
=C<	sp ² C	10	0.9	1.4	0.2	-1.7
>C<	sp ³ C	39	-2.2	1.5	0.0	-1.2
C(Ar)	aromatic C	150	-0.8	1.0	0.0	0.4
C(Ar)(-)	aromatic C	25	25.5	27.9	-0.1	-1.8
C(Ar)(+)	aromatic C	24	-3.5	4.3	-1.4	1.7
=CH-	CH (sp ² C)	25	0.7	2.0	1.5	1.8
>CH-	CH (sp ³ C)	28	0.2	1.2	0.1	-0.3
=CH ₂	terminal CH ₂ (sp ² C)	13	-0.7	1.6	1.1	0.2
-CH ₂ -	methylene	316	0.6	1.4	0.8	0.9
=CH ₂ (-)	terminal CH ₂ (sp ² C)	6	-8.8	9.4	-0.3	-2.0
-CH ₂ (-)	methylene	33	22.1	20.9	-0.2	-1.7
-CH ₂ (+)	methylene	53	-22.8	11.8	0.3	-0.9
-CH ₃	methyl	283	0.4	1.1	1.1	1.0
-CH ₃ (-)	methyl	35	-4.0	4.1	0.4	-0.2
-CH ₃ (+)	methyl	52	-24.0	15.2	-0.6	-1.0
CH(Ar)	aromatic CH	379	0.2	1.5	0.4	1.5
CH(Ar)(-)	aromatic CH	66	-2.6	3.4	0.5	-0.8
CH(Ar)(+)	aromatic CH	73	-8.9	5.0	-0.6	1.1
-CHO	aldehyde	7	-9.4	2.7	-0.4	-0.9
>CH(-)	CH (sp ³ c)	7	24.9	31.7	-0.6	-1.4
-Cl	chlorine	89	-2.1	2.4	-0.4	-0.7
-Cl(-)	chlorine	8	-17.4	4.1	0.3	-1.8
-CN	cyanide	8	-12.5	1.5	-0.3	-1.0
-CO-	carbonyl	44	-11.3	2.0	0.0	0.4
-CONH-	amide	5	-12.0	1.9	0.5	-1.9

-CONH2	terminal amide	7	-12.3	2.6	0.1	-1.7
-COO-	ester	9	-9.3	0.3	0.2	-1.4
-COO[-]	carboxylate	8	-62.2	8.4	0.2	-1.9
-COOH	carboxyl	8	-10.1	1.2	0.6	-0.7
>C<(-)	sp3 c	7	95.0	5.5	0.3	-1.4
-F	fluorine	92	-0.3	1.6	-1.1	0.5
-F(-)	fluorine	21	-36.8	2.2	0.0	-1.1
>N-	tertiary amine	7	-2.3	1.3	1.0	-0.7
N(Ar)	aromatic n	30	-6.8	4.7	1.0	0.0
>NH-[+]	tertiary ammonium	5	46.6	4.1	0.7	-1.4
-NH-	secondary amine	11	-3.4	0.9	0.8	-0.1
-NH2	primary amine	17	-3.5	0.7	0.4	-0.7
-NH2-[+]	secondary ammonium	12	17.5	1.1	-0.5	-0.1
-NH3[+]	primary ammonium	14	-20.5	1.3	-1.1	0.5
NH(Ar)	aromatic NH	25	3.3	2.5	-0.1	-1.2
-NHCONH2	carbamide	5	-11.0	1.9	-0.3	-2.2
-NO2	nitro	13	-8.1	1.4	-0.3	-0.4
-O-	ether	40	-2.1	2.9	0.6	-0.7
-O[-]	alkoxide	27	-96.4	9.1	0.0	-0.9
-OH	alcohol	26	-4.3	1.9	0.5	-0.2
-O(H)[-]	H-bonded alkoxide	10	-89.5	5.2	0.6	-0.4
-ONO2	nitrate	5	-3.8	2.2	-0.2	-2.0
-S-	sulfide	5	-6.8	2.1	-0.1	-2.2
-SS-	disulfide	11	1.7	7.1	-0.5	-1.4

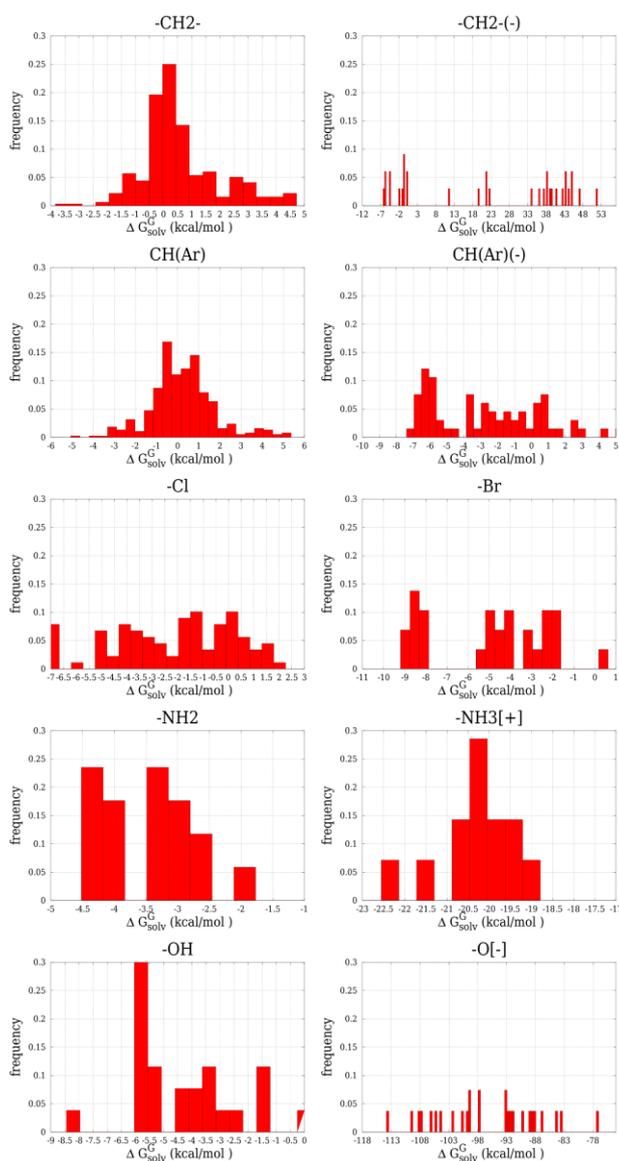
The selection and notation of the groups listed in Table 1 is partially based on the fragments defined by Meyland and Howard²³ in their empirical *logP* method. For example, the symbol “>C<” stands for a tetra-substituted *sp*³ C atom, -CH₂- for a methylene group, CH(Ar) for an aromatic CH fragment, -NH₂ for a neutral primary amine group, -NH₂-[+] for a protonated secondary amine, and so on. A particular feature of our fragment selection is that, for the majority of neutral fragments *G*, we also distinguish two ionic versions labelled as *G*(-) and *G*(+) depending on the global charge of the molecule

bearing the corresponding fragment. For instance, $-\text{CH}_2(+)$ and $-\text{CH}_2(-)$ stand for the methylene groups in cationic and anionic molecules, respectively. This distinction is due to the significant variation in the ΔG_{solv}^G values due to the global charge of the molecule and the relative positioning of the G group with respect to the ionic groups (see below).

Besides the mean values (μ) of ΔG_{solv}^G and their standard deviations (σ), Table 1 contains the skewness (skw) and excess kurtosis indexes that measure, respectively, the asymmetry and the shape of the peak and tails of the underlying distribution with respect to the normal distribution ($skw=0$ and $kurt=0$). The ΔG_{solv}^G data corresponding to fragments in neutral molecules result in more or less narrow distributions ($\sigma \sim 2.0$ kcal/mol) that are moderately asymmetrical, albeit with varying shapes as the kurtosis coefficient can be positive or negative. In the case of ionic solutes, both the ionic functionalities and the neutral fragments have ΔG_{solv}^G distributions that are quite wide (e.g., $\sigma \sim 9$ kcal/mol and above in anionic systems) and predominantly flat (*i.e.*, negative excess kurtosis).

To better characterize the width and shape of the ΔG_{solv}^G distributions, Figure 5 displays the histogram plots of a few selected groups. Closer inspection of the IQA data can reveal how the features of the various distributions are related to structural patterns. For example, electrostatic solvation of methylene ($-\text{CH}_2-$) groups is, on average, not favorable and the distribution is quite concentrated around its mean value ($\mu=0.6$, $\sigma=1.4$). The tallest distribution peak corresponds to the solvation of $-\text{CH}_2-$ fragments attached to zero or one polar groups while the shoulder at $\Delta G_{\text{solv}}^G = +3.0$ kcal/mol is due to di-substituted $\text{X}-\text{CH}_2-\text{Y}$ groups with X, Y=polar.

Figure 5. Histogram of selected fragment-based IQA ΔG_{solv}^G values.



The presence of a global negative charge dramatically changes the ΔG_{solv}^G distribution, which becomes quite wide and flat over a 40 kcal/mol interval (see $-\text{CH}_2(-)$ in Figure 5). Thus, all the $-\text{CH}_2(-)$ fragments attached to ionic groups (e.g., $-\text{NH}_3^+$) have negative contributions to ΔG_{solv} (from -40 to -20 kcal/mol). As a matter of fact, the solvation stabilization of tetrahedral ammonium groups stems from the positive ΔG_{solv}^A of the N atom and the negative (stabilizing) terms from the four surrounding groups including H atoms and $-\text{CH}_2-$ fragments (see N-ethylethanamine in Figure 3). Other methylene groups located at the β or γ positions with respect to the positively charged group have also negative ΔG_{solv}^G

values (between -12 and -3 kcal/mol), which are below the values for $-\text{CH}_2-$ in neutral molecules. The $\text{CH}(\text{Ar})$ fragments in neutral solutes give a hydrophobic contribution ($\mu=0.2$, $\sigma=1.5$) similar to that of the $-\text{CH}_2-$ groups. Most of the ΔG_{solv}^G data come from benzene ring groups although some $\text{CH}(\text{Ar})$ groups in heterocycles appear at the positive tail. The alkoxide $-\text{O}^-$ and carboxylate $-\text{COO}^-$ substituents at phenyl rings modify the hydration of the $\text{CH}(\text{Ar})$ groups, that become hydrophilic sites with negative ΔG_{solv}^G values. However, the aromatic CH groups are less perturbed by the ionic substituents than the aliphatic $-\text{CH}_2-$ fragments. The largest effect is at CH sites located at *ortho*-positions with $\Delta G_{\text{solv}}^G \sim -7/-8$ kcal/mol while those at the *meta*- and *para*- have values of $\sim -2/-3$ kcal/mol.

Other examples of fragment hydration energy distributions in Figure 5 are those of the halogen substituents (Cl and Br), which exhibit quite wide and flat histograms ($\mu=-2.1$, $\sigma=2.4$ for Cl and $\mu=-4.7$, $\sigma=3.0$ for Br). In fact a detailed examination shows that the more negative ΔG_{solv}^G values occur at monosubstituted aliphatic R-X compounds, the least favorable terms are those of polysubstituted halogenated alkanes and the values in between correspond to aromatic molecules. Neutral polar groups like $-\text{NH}_2$, $-\text{OH}$ as well as the cationic $-\text{NH}_3^+$ tend to give relatively concentrated ΔG_{solv}^G distributions so that their corresponding average values μ may be reliable estimators. In sharp contrast, anionic groups like the alkoxide $-\text{O}^-$ result in scattered data over 40 kcal/mol interval (see Figure 5). In this case the more negative ΔG_{solv}^G values for $-\text{O}^-$ are due to phenolate groups (~ -110 kcal/mol).

Additivity of fragment contributions

The analysis of the histograms in Figure 5 and other data in the Supporting Information confirms that the IQA decomposition of the COSMO-HF solvation energies can provide a detailed assessment of the fragment contributions to solvation. Although the amount of data gathered for some of the examined functionalities is limited, several structure-activity trends can be outlined regarding the constancy/dispersion of fragment contributions and their relationship with structural and substituent effects. In particular, the additivity of the mean values of the fragment ΔG_{solv}^G energies to estimate the

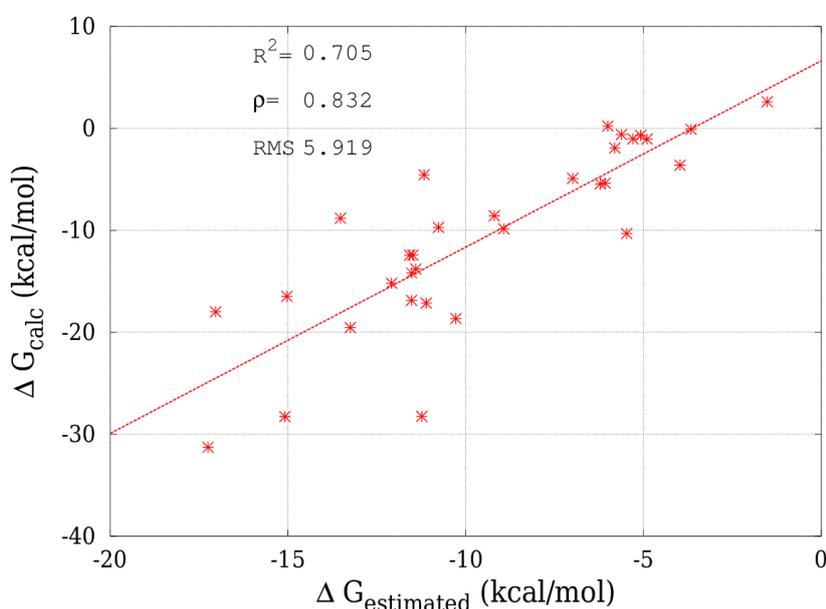
total solvation energy is not unreasonable for neutral solute molecules given that the dispersion (σ) values collected in Table 1 tend to be moderate (< 2 kcal/mol).

To find out to what extent the mean ΔG_{solv}^G values are additive, we calculated the COSMO HF/aug-cc-pVTZ solvation energy for a set of 32 MNSol molecules not considered in the former IQA calculations. All of these molecules are neutral and possess the functional groups listed in Table 1 (see Figure S2 in the Supporting Information). In Figure 6 the calculated ΔG_{solv} and the fragment-based estimations $\Delta G_{\text{estimated}} = \sum_G \langle \Delta G_{\text{solv}}^G \rangle$ are compared. The computed and estimated values show a moderate correlation ($R^2=0.70$) and the RMS error is 5.9 kcal/mol. The largest discrepancies arise in compounds that have large aliphatic or aromatic moieties (*e.g.*, octanol $\Delta G_{\text{calc}} = -6.0$, $\Delta G_{\text{estimated}} = 0.2$) whereas for molecules with 2 or more polar groups the simple additive model tends to work better (*e.g.*, 3-methylthiophenylurea $\Delta G_{\text{calc}} = -17.0$, $\Delta G_{\text{estimated}} = -18.0$).

Some empirical models developed for $\log P$ or hydration energy estimations exhibit a better performance ($R^2 \sim 0.8-0.9$).^{22-23, 36} They include, not only atomic solvation parameters, but also exposure factors and/or correcting terms that modulate the sum of atomic terms. The results shown in Figure 6 suggest that a fragment-based method including weighing parameters to be fitted against a large set of hydration energies, could be also a reasonable approach. Further improvements in the additivity of the IQA-based solvation energies could be gained by defining new atom types/functional groups as suggested by the detailed analysis of histogram data. This could be the case of the halogen atom types that may be categorized as halogen attached to either aliphatic- or aromatic-C atom. These results also indicate that the electrostatic contributions of aliphatic and aromatic sites (and eventually their non-polar terms too) should receive a special attention. Thus, correction factors could be derived to take into account the influence of ionic/polar groups on the contributions of the nearby aliphatic/aromatic fragments. Nevertheless, these and other possible alternatives are beyond the scope

of the present work, which is focused on the description of the IQA-decomposition of solvation energy rather than in the development of a fragment solvation method.

Figure 6. Comparison between the COSMO-HF/aug-cc-pVTZ calculated hydration energies (ΔG_{calc} in kcal/mol) and the estimated ones assuming the additivity of fragment contributions ($\Delta G_{estimated}$) of the selected MNSol structures in aqueous solvent. The determination coefficient (R^2), the Spearman correlation coefficient (ρ) and the root mean square (RMS) error in kcal/mol are also indicated. The dashed line is the least squared fit line between the calculated and the estimated data.



Coulomb and Polarization Effects

The Coulomb contribution ΔG_{solv}^{Coul} to the hydration energies of the MNSOL structures examined in this work was computed by means of single-point HF/aug-cc-pVTZ COSMO calculations on the gas-phase geometries and using the unpolarized (gas-phase) wavefunction. The polarization term is then obtained by subtraction, $\Delta G_{solv}^{pol} = \Delta G_{solv} - \Delta G_{solv}^{Coul}$. Herein, we briefly comment on the results (see also Supporting Information).

For the neutral solutes, the ΔG_{solv}^{Coul} values correlate with experimental data ($R^2=0.83$ and RMS error 2.3 kcal/mol) similarly as the ΔG_{solv} energies do. The cationic hydration energies are also similar. In

consonance with expectation, the lack of polarization in the ΔG_{solv}^{Coul} values of the anionic solutes results in less correlation ($R^2=0.78$) with experimental data and larger absolute errors (~ 12 kcal/mol). Concerning the stability gained by solute-solvent mutual polarization, most of the calculated ΔG_{solv}^{pol} energies have values around -1, -2 kcal/mol. Only in those molecules containing the most polarizable groups (Cl, Br, anions, ...), ΔG_{solv}^{pol} have values between -2 and -7 kcal/mol.

We also examined the distribution of group contributions, $\Delta G_{solv}^{Coul, G}$, using the same group definitions and statistical indexes as in Table 1 and Figure 5. In general, the dispersion (σ) of the $\Delta G_{solv}^{Coul, G}$ data for a given group is only slightly lower than that of ΔG_{solv}^G . Thus, the Coulomb $\Delta G_{solv}^{Coul, G}$ contributions depend on the chemical environment much to the same extent as ΔG_{solv}^G . More significant variations are observed in the shape of the ΔG_{solv}^G and $\Delta G_{solv}^{Coul, G}$ distributions characterized in terms of the skewness and excess kurtosis. In particular, several non-polar groups (-CH₂-, -CH₃, CH(Ar)) in neutral molecules, which have very small group contributions to ΔG_{solv} , exhibit sharp-peaked distributions of their $\Delta G_{solv}^{Coul, G}$ values having also a large excess kurtosis. As a consequence, the corresponding polarization group components $\Delta G_{solv}^{pol, G}$ are slightly positive (+0.5, +0.8, +0.7), what is more a statistical artefact than a physical effect. The rest of functional groups show negative stabilizing $\Delta G_{solv}^{pol, G}$ values. Finally, we also examined the additivity of the mean $\Delta G_{solv}^{Coul, G}$ values, finding a worse correlation with the calculated Coulomb solvation energies than in the case of the full solvation energies. Overall, we conclude that the separate treatment of the Coulomb and polarization electrostatic effects does not lead to an improved description of the fragment contributions to solvation energies.

About the extension of IQA to other implicit solvent models

The present COSMO-IQA calculations indicate that QM energies in implicit solvent are prone to be decomposed within the IQA method. COSMO belongs to the family of ASC methods that express the electrostatic solute-solvent energy as a single sum, $V_{solv} = \sum_k \Phi(\mathbf{s}_k) q_k$, involving the solute electrostatic potential and the ASCs. As mentioned above, this feature enables IQA to absorb solvent effects into

the net atomic energies. Hence, it can be reasonably expected that other ASC methods including the original Polarizable Continuum Model (PCM)²⁸ and the integral-equation-formalism PCM (IEFPCM)³⁷ could be similarly coupled with IQA provided that q_k and Φ data are available.

IQA extensions to other QM SCRF methods as the generalized Born theory (GB) and the multipolar expansion methods would be more problematic. On one hand, the GB approach³⁸ uses a modified and further parameterized Coulombic potential and evaluates the solvation energy as the total Coulomb-like interaction over the atom pairs in the solute molecule and, therefore, it is not evident how to decompose it into meaningful atomic contributions. On the other, the multipolar expansion of the solute charge density can be extended to atom-centered expansions although there is an infinite number of manners of weighting the multipoles.³⁹⁻⁴⁰ However, the resulting solute-solvent interaction energy is written in terms of diatomic reaction potential terms that do not admit an evident atomic partitioning either.

In this work, we focus on the partitioning of the electrostatic solvation energy. However, the consideration of non-polar solvation effects within the IQA-like analysis could be feasible by means of empirical approaches. For example, cavitation free energy, solute-solvent dispersion and solvent-structural effects can be accounted for by means of the G_{CDS} empirical potential implemented in the SMD solvation method⁴¹:

$$G_{CDS} = \sum_A (\gamma_A + \gamma^M) \sigma_A$$

where γ_A and γ^M are molecular surface tension parameters and σ_A is the solvent-accessible surface of atom A. Hence, the atomic contributions to G_{CDS} could be combined with the electrostatic ΔG_{solv}^A terms to yield an atomic mapping of the total solvation energy.

Conclusions

The computational results presented in this work demonstrate that it is feasible to incorporate electrostatic solute-solvent effects into the IQA energy decomposition method. In this way the usual IQA analysis of energy differences can be carried out including continuum solvent effects, extending thus its applicability. Basing on the extensive solvation energy calculations followed by the IQA decomposition of the electrostatic solvation energy, we have also shown that IQA yields a detailed atomic mapping of solvation energies and suggests a united atom approach for considering fragment contributions. A tentative selection of fragments has been made and their solvation energies have been characterized statistically, finding that the distributions of fragment solvation energies, which may have relatively large deviation for some groups, depend on structural and substituent effects. For neutral molecules, the simple additivity assumption, commonly adopted in empirical solvation methods, leads to approximate solvation energies that exhibit only moderate correlation with reference values and have significant errors of several kcal/mol. More specific fragment-types and extra-parameters would be required to derive improved fragment solvation methods from QM SCRF and IQA calculations on a larger database of solute structures.

Computational Section

QM calculations

Cartesian coordinates and reference hydration energies for all the solute molecules were retrieved from the MNSol database. The general ab initio quantum chemistry program GAMESS⁴²⁻⁴³ was used to perform all the QM calculations. First, we relaxed all the structures by means of unconstrained energy minimizations that were started from the corresponding MNSol geometries. These calculations were carried out first in the gas-phase combining the Hartree-Fock (HF) method with the aug-cc-pVTZ basis set.⁴⁴⁻⁴⁵ The solute geometries were also optimized at the HF/aug-cc-pVTZ level in combination with the COSMO solvation model.³⁰ A dielectric $\epsilon=80$ was selected for mimicking water as solvent while a multiplicative factor of 1.2 was applied to the standard van der Waals radii for cavity

construction. Each atomic sphere that contributes to build the molecular cavity is divided into 92 tesserae. The `cosprt` module in GAMESS was locally modified to print out the charge and Cartesian coordinates of the apparent surface charges of the optimized HF-COSMO structures. The Chimera visualization system⁴⁶ was used to draw the models of the solute molecules.

IQA calculations

The IQA decomposition of the total energies was performed with a modular version of the PROMOLDEN program⁴⁷ that is being developed in our laboratory. In this version, the program reads the apparent surface charge data generated by the COSMO implementation in GAMESS in order to compute the solute-solvent interaction term V_{solv} using the same integration algorithm that is employed for computing the electron-nucleus interaction terms V_{en} .²

The IQA quantities are numerically integrated by PROMOLDEN over finite and irregular integration domains (i.e., atomic basins Ω_A) using angular and radial grids in atomic spherical quadratures that are much finer than those typically used by other QM software.^{2, 34} We employed similar integration settings to those used in previous work⁶ and that represent a compromise choice between computational cost and accuracy for small and medium-sized molecules. Thus, 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 of heavy atoms, respectively, (3890 and 590 points for hydrogen atoms). Euler-McLaurin radial quadratures were employed with 512 and 384 radial points outside and inside the β -spheres of heavy atoms, respectively (384 and 256 points for H). The largest value of the radial coordinate in the integrations was 15.0 au for heavy atoms (10.0 au for H atoms). Maximum angular moments, λ_{max} , of 10 and 6 were assigned to the Laplace and bipolar expansions of $1/r_{12}$ outside and within the β -spheres.

Acknowledgements

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

References

1. Popelier, P. L. A., Quantum Chemical Topology. In *The Chemical Bond Ii: 100 Years Old and Getting Stronger*, Mingos, D. M. P., Ed. Springer International Publishing: Cham, 2016; pp 71-117.
2. Blanco, M. A.; Martín-Pendás, A.; Francisco, E., Interacting Quantum Atoms: A Correlated Energy Decomposition Scheme Based on the Quantum Theory of Atoms in Molecules. *J. Chem. Theory Comput.* **2005**, *1*, 1096-1109.
3. Francisco, E.; Martín-Pendás, A.; Blanco, M. A., A Molecular Energy Decomposition Scheme for Atoms in Molecules. *J. Chem. Theory Comput.* **2006**, *2*, 90-102.
4. Grimme, S., Density Functional Theory with London Dispersion Corrections. *WIREs Comput Mol Sci* **2011**, *1*, 211-228.
5. Grimme, S.; Hansen, A.; Brandenburg, J. G.; Bannwarth, C., Dispersion-Corrected Mean-Field Electronic Structure Methods. *Chem. Rev.* **2016**, *116*, 5105-5154.
6. 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.
7. Martín-Pendás, A.; Blanco, M. A.; Francisco, E., The Nature of the Hydrogen Bond: A Synthesis from the Interacting Quantum Atoms Picture. *J. Chem. Phys.* **2006**, *125*.

8. Guevara-Vela, J. M.; Chavez-Calvillo, R.; Garcia-Revilla, M.; Hernandez-Trujillo, J.; Christiansen, O.; Francisco, E.; Martín-Pendás, A.; Rocha-Rinza, T., Hydrogen-Bond Cooperative Effects in Small Cyclic Water Clusters as Revealed by the Interacting Quantum Atoms Approach. *Chemistry-a European Journal* **2013**, *19*, 14304-14315.
9. Guevara-Vela, J. M.; Romero-Montalvo, E.; Mora Gomez, V. A.; Chavez-Calvillo, R.; Garcia-Revilla, M.; Francisco, E.; Martín-Pendás, A.; Rocha-Rinza, T., Hydrogen Bond Cooperativity and Anticooperativity within the Water Hexamer. *Phys. Chem. Chem. Phys.* **2016**, *18*, 19557-19566.
10. Romero-Montalvo, E.; Guevara-Vela, J. M.; Costales, A.; Martín-Pendás, Á.; Rocha-Rinza, T., Cooperative and Anticooperative Effects in Resonance Assisted Hydrogen Bonds in Merged Structures of Malondialdehyde. *Phys. Chem. Chem. Phys.* **2017**, *19*, 97-107.
11. Bartashevich, E.; Troitskaya, E.; Martín-Pendás, A.; Tsirelson, V., Understanding the Bifurcated Halogen Bonding N Center Dot Center Dot Center Dot Hal Center Dot Center Dot Center Dot N in Bidentate Diazaheterocyclic Compounds. *Computational and Theoretical Chemistry* **2015**, *1053*, 229-237.
12. Tiana, D.; Francisco, E.; Macchi, P.; Sironi, A.; Martín-Pendás, A., An Interacting Quantum Atoms Analysis of the Metal-Metal Bond in M-2(Co)(8) (N) Systems. *J. Phys. Chem. A* **2015**, *119*, 2153-2160.
13. Cukrowski, I.; de Lange, J. H.; Mitoraj, M., Physical Nature of Interactions in Zn^{II} Complexes with 2,2'-Bipyridyl: Quantum Theory of Atoms in Molecules (Qtaim), Interacting Quantum Atoms (Iqa), Noncovalent Interactions (Nci), and Extended Transition State Coupled with Natural Orbitals for Chemical Valence (Ets-Nocv) Comparative Studies. *J. Phys. Chem. A* **2014**, *118*, 623-637.
14. Wilson, A. L.; Popelier, P. L., Exponential Relationships Capturing Atomistic Short-Range Repulsion from the Interacting Quantum Atoms (Iqa) Method. *J Phys Chem A* **2016**, *120*, 9647-9659.

15. McDonagh, J. L.; Silva, A. F.; Vincent, M. A.; Popelier, P. L. A., Quantifying Electron Correlation of the Chemical Bond. **2017**, *8*, 1937-1942.
16. Cramer, C. J.; Truhlar, D. G., Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chem. Rev.* **1999**, *99*, 2161-2200.
17. Tomasi, J., Thirty Years of Continuum Solvation Chemistry: A Review, and Prospects for the near Future. *Theoretical Chemistry Accounts* **2004**, *112*, 184-203.
18. Ben-Naim, A., *Statistical Thermodynamics for Chemists and Biochemists*. Springer Science Business Media: New York, 1992.
19. Park, H., Extended Solvent-Contact Model Approach to Sampl4 Blind Prediction Challenge for Hydration Free Energies. *J. Comput. Aided Mol. Des.* **2014**, *28*, 175-186.
20. Bernazzani, L.; Duce, C.; Micheli, A.; Mollica, V.; Tiné, M. R., Quantitative Structure–Property Relationship (Qspr) Prediction of Solvation Gibbs Energy of Bifunctional Compounds by Recursive Neural Networks. *Journal of Chemical & Engineering Data* **2010**, *55*, 5425-5428.
21. Kang, H.; Choi, H.; Park, H., Prediction of Molecular Solvation Free Energy Based on the Optimization of Atomic Solvation Parameters with Genetic Algorithm. *J. Chem. Inf. Model.* **2007**, *47*, 509-514.
22. Jianfeng, P.; Qi, W.; Jiaju, Z.; Luhua, L., Estimating Protein–Ligand Binding Free Energy: Atomic Solvation Parameters for Partition Coefficient and Solvation Free Energy Calculation. *Proteins* **2004**, *57*, 651-664.
23. Meylan, W. M.; Howard, P. H., Atom/Fragment Contribution Method for Estimating Octanol-Water Partition Coefficients. *J Pharm Sci* **1995**, *84*, 83-92.

24. Soteras, I.; Morreale, A.; López, J. M.; Orozco, M.; Luque, F. J., Group Contributions to the Solvation Free Energy

from Mst Continuum Calculations. *Brazilian Journal of Physics* **2004**, *34*, 48-57.

25. Marenich, A. V.; Kelly, C. P.; Thompson, J. D.; Hawkins, G. D.; Chambers, C. C.; Giesen, D. J.; Winget, P.; Cramer, C. J.; Truhlar, D. G., Minnesota Solvation Database-Version 2012. Minnesota, U. o., Ed. Minneapolis, 2012.

26. Chamberlin, A. C.; Cramer, C. J.; Truhlar, D. G., Predicting Aqueous Free Energies of Solvation as Functions of Temperature. *J. Phys. Chem. B* **2006**, *110*, 5665-5675.

27. Francisco, E.; Martín-Pendás, A., Energy Partition Analyses: Symmetry-Adapted Perturbation Theory and Other Techniques. In *Non-Covalent Interactions in Quantum Chemistry and Physics. Theory and Applications*, Otero-de-la-Roza, A.; DiLabio, G., Eds. Elsevier: 2017; pp 27-62.

28. Tomasi, J.; Mennucci, B.; Cammi, R., Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005**, *105*, 2999-3094.

29. Tomasi, J.; Persico, M., Molecular Interactions in Solution: An Overview of Methods Based on Continuous Distributions of the Solvent. *Chem. Rev.* **1994**, *94*, 2027-2094.

30. Baldrige, K.; Klamt, A., First Principles Implementation of Solvent Effects without Outlying Charge Error. *J. Chem. Phys.* **1997**, *106*, 6622-6633.

31. Klamt, A., *Cosmo-Rs from Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design* Elsevier: Amsterdam, 2005.

32. Klamt, A.; Jonas, V., Treatment of the Outlying Charge in Continuum Solvation Models. *J. Chem. Phys.* **1996**, *105*, 9972-9981.

33. Tielker, N.; Tomazic, D.; Heil, J.; Kloss, T.; Ehrhart, S.; Güssregen, S.; Schmidt, K. F.; Kast, S. M., The Sampl5 Challenge for Embedded-Cluster Integral Equation Theory: Solvation Free Energies, Aqueous pK_a , and Cyclohexane–Water Log D . *J. Comput. Aided Mol. Des.* **2016**, *30*, 1035-1044.
34. Martín-Pendás, A.; Blanco, M. A.; Francisco, E., Two-Electron Integrations in the Quantum Theory of Atoms in Molecules. *J. Chem. Phys.* **2004**, *120*, 4581-4592.
35. Martín-Pendás, A.; Francisco, E.; Blanco, M. A., Two-Electron Integrations in the Quantum Theory of Atoms in Molecules with Correlated Wave Functions. *J. Comput. Chem.* **2005**, *26*, 344-351.
36. Chung, K.-C.; Park, H., Accuracy Enhancement in the Estimation of Molecular Hydration Free Energies by Implementing the Intramolecular Hydrogen Bond Effects. *Journal of Cheminformatics* **2015**, *7*, 57.
37. Cancès, E.; Mennucci, B.; Tomasi, J., A New Integral Equation Formalism for the Polarizable Continuum Model: Theoretical Background and Applications to Isotropic and Anisotropic Dielectrics. *J. Chem. Phys.* **1997**, *107*, 3032-3041.
38. Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T., Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *JACS* **1990**, *112*, 6127-6129.
39. Rinaldi, D.; Bouchy, A.; Rivail, J.-L.; Dillet, V., A Self-Consistent Reaction Field Model of Solvation Using Distributed Multipoles. I. Energy and Energy Derivatives. *J. Chem. Phys.* **2004**, *120*, 2343-2350.
40. Rinaldi, D.; Bouchy, A.; Rivail, J.-L., A Self-Consistent Reaction Field Model of Solvation Using Distributed Multipoles. II: Second Energy Derivatives and Application to Vibrational Spectra. *Theoretical Chemistry Accounts* **2006**, *116*, 664-669.

41. Cramer, C. J.; Truhlar, D. G., A Universal Approach to Solvation Modeling. *Acc. Chem. Res.* **2008**, *41*, 760-768.
42. 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. *J. Comput. Chem.* **1993**, *14*, 1347-1363.
43. Gordon, M. S.; Schmidt, M. W., Chapter 41 - Advances in Electronic Structure Theory: Gamess a Decade Later A2 - Dykstra, Clifford E. In *Theory and Applications of Computational Chemistry*, Frenking, G.; Kim, K. S.; Scuseria, G. E., Eds. Elsevier: Amsterdam, 2005; pp 1167-1189.
44. Jr., T. H. D., Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007-1023.
45. Kendall, R. A.; Dunning Jr., T. H.; Harrison, R. J., *J. Chem. Phys.* **1992**, *96*, 6796-6802.
46. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., Ucsf Chimera--a Visualization System for Exploratory Research and Analysis. *J Comput Chem* **2004**, *25*, 1605-1612.
47. Martín-Pendás, A.; Francisco, E. *Promolden: A Qtaim /Iqa Code*, Unpublished: 2015.