CrossMark

ORIGINAL PAPER

On the accuracy of population analyses based on fitted densities[#]

Aurélien de la Lande[1] · Carine Clavaguéra[1] · Andreas Köster[2]

Abstract Population analyses are part of the theoretical chemist's toolbox. They provide means to extract information about the repartition of the electronic density among molecules or solids. The values of atomic multipoles in a molecule can shed light on its electrostatic properties and may help to predict how different molecules could interact or to rationalize chemical reactivity for instance. Not being physical observables to which a quantum mechanical operator can be associated, atomic charges and higher order atomic multipoles cannot be defined unambiguously in a molecule, and therefore, several population schemes (PS) have been devised in the last decades. In the context of density functional theory (DFT), PS based on the electron density seem to be best grounded. In particular, some groups have proposed various iterative schemes the outcomes of which are very encouraging. Modern implementations of DFT that are for example based on density fitting techniques permit the investigation of molecular systems comprising of hundreds of atoms. However, population analyses following iterative schemes may become very CPU time consuming for such large systems. In this article, we investigate if the computationally less expensive analyses of the variationally fitted electronic densities can be safely carried out instead of the Kohn-Sham density. It is shown that as long as flexible auxiliary function sets including f and g functions are used, the multipoles extracted from the fitted densities are extremely close to those obtained from the KS density. We further assess if the multipoles obtained through the Hirshfeld's approach, in its standard or iterative form, can be a useful approach to calculate interaction energies in non-covalent complexes. Relative energies computed with the AMOEBA polarizable forced field combined to iterative Hirshfeld multipoles are encouraging.

Introduction

Electronic population analyses (PA) represent a powerful means to connect the outputs of quantum chemistry computations to the chemical knowledge. Atomic charges may be useful to rationalize chemical reactivity, for example to investigate substituent effects on an energy profile. Atomic charges and higher order multipoles are also key quantities for molecular mechanics force fields. The possibility to extract reliable multipoles from quantum chemistry computations is actually a topic of high interest for the development of second or third generation force fields and hybrid QM/MM schemes [1–3]. Various quantum chemistry methodologies also rely on atomic charges. For the sake of illustration let us mention constrained density functional theory (DFT) whereby atomic charges are imposed to some molecular fragments in order to define diabatic states at the DFT level [4, 5]. Such diabatic

[#]Dedicated to Henry Chermette for his contributions to Density Functional Theory

This paper belongs to Topical Collection Festschrift in Honor of Henry Chermette

✉ Aurélien de la Lande
 aurelien.de-la-lande@u-psud.fr

✉ Andreas Köster
 akoster@cinvestav.mx

[1] Laboratoire de Chimie Physique, UMR 8000 CNRS/Univ. Paris-Sud, Univ. Paris-Saclay, 91405 Orsay, France

[2] Departamento de Química, CINVESTAV, Avenida Instituto Politécnico Nacional 2508, México, D.F., Mexico

 Springer

states proved to be useful for modeling electron transfer processes or charge transfers within non-covalent complexes [6].

It is actually well-known that there is no unique way to define an atom in a polyatomic molecule, and hence to define atomic charges. This is due to the fact that atomic charges are not physical observables. Despite this ambiguity of atomic charges, atom centered multipole expansions are rather reliable if the problem of atomic multipole invariance is properly addressed. A convenient way to do so is the definition of cumulative atomic multipole moments [7, 8]. These moments are built up from atomic charges, which are invariant to coordinate transformation. Because the convergence of the atom centered multipole expansion depends on the quality of the underlying atomic charges [9], population analyses are important ingredients for the development of atom centered multipole expansions with only a limited number of expansion terms. A critical point to any population scheme is to divide the real or function space in order to distribute the electron density over the atoms. It is customary to classify population schemes along various categories. The Mulliken [10, 11] and Löwdin [12] approaches as well as the more elaborated natural population analysis [13] or natural bonding orbital (NBO) [14] analysis define atomic charges by division of the function space spanned by the atomic or molecular orbitals. Another family encompasses the Becke [15], Hirshfeld [16], and Voronoi deformation density [17] approaches. These schemes are based on a real space partitioning of the electron density itself. On-going efforts are deployed by several groups to improve the quality of the charges produced by these approaches [18–20]. For example iterative schemes have been proposed such as the Hirshfeld-I [21], Hirshfeld-λI [22], the fractional occupation Hirshfeld-I [23] or the Stockholder-I methods [24]. We also mention the methods where the charges are obtained through the integration of the electronic density over topological basins of well suited functions [25]. These functions may be the electronic density (the Bader's atoms-in-molecules theory [26]) or the electron localization function [25]. For the sake of completeness, we finally mention methods like the Merz-Singh-Kollman scheme that consists in the fit of the atomic charges so as to reproduce the electrostatic potential created by the molecule [27–29]. These methods have been abundantly employed to develop molecular mechanics force fields like AMBER [30] or CHARMM [31].

In this paper, we focus on methods based on the real space integration of the density in the context of DFT. These calculations have become the most used quantum chemistry approaches because they provide an excellent cost/quality ratio. Indeed, modeling systems containing up to a few hundreds of atoms are nowadays accessible by DFT. These performances have become possible thanks to the development of ingenious algorithms for solving the Kohn-Sham equations. In particular, methods resorting on fitted densities besides the Kohn-Sham density (e.g., variational density fitting [32, 33] and Cholesky decomposition of the density matrix [34]) permit the elimination of the cumbersome evaluation of four-center electron repulsion integrals.

Performing population analyses on optimized electronic densities is usually not a CPU demanding task compared to the self-consistent-field (SCF) procedure. However, if iterative schemes are employed this task can be a computational bottleneck. This limitation can become critical for systems comprising tens or hundreds of atoms, as commonly investigated by nowadays DFT approaches. One possible strategy to overcome these potential limitations is to develop more efficient parallelization and/or grid techniques. An alternative, although not exclusive, approach is to perform analyses based on auxiliary function densities instead of Kohn-Sham densities. Even though this procedure seems rather straightforward, it must be taken into account that auxiliary densities, as obtained in density fitting approaches, are not designed to mimic the Kohn-Sham orbital density, but to provide a density from which a mathematically simpler electron-electron repulsion energy term will be calculated, avoiding explicit four-center-integrals. There is no guarantee that auxiliary densities can be used *in lieu* of Kohn-Sham densities in population analyses.

The objectives of this paper are twofold. First, we wish to assess whether auxiliary densities are suited for population analyses. By this, we mean extracting not only monopoles (atomic charges) but also atomic dipoles and quadrupoles. We have considered large sets of organic molecules in our tests. Second, we wish to investigate the capabilities of various populations schemes, including some of the most advanced to produce electrostatic multipoles that may be used for interaction energy calculations. Reliable schemes would then be valuable for second or third generation force fields or for accurate hybrid DFT/MM (molecular mechanics) approaches. As an example, relative energies of different structures between a tryptamine molecule, a water molecule and a sodium cation are computed with the AMOEBA [35, 36] force field using multipoles extracted from DFT population analyses.

The article is organized in three parts. We first present the new population schemes we have introduced in deMon2k. We then report extensive benchmark calculations on sets of organic molecules. We finally report the performances of the population schemes to produce atomic multipoles that can be used in the AMOEBA force field to reproduce structures and non-covalent energies.

Population schemes

The population schemes described thereafter have been implemented in a new version of the program deMon2k [37]. They can be classified into two categories. The first category refers to population analyses that define atomic charges from the number of electrons belonging

to each atom and its nuclear charge. The charge on atom A reads:

$$Q_A = Z_A - N_A \qquad (1)$$

where N_A represents the number of electrons on atom A and Z_A its nuclear charge. In the second category the charge is defined from a deformation density between the converged SCF density and a reference density.

$$Q_A = N_A - N_A^{ref} \qquad (2)$$

Typically N_A and N_A^{ref} refer to the number of electrons of atom A from the SCF electronic density and the so-called promolecular density which is the superposition of non-interacting atomic densities. The latter scheme will be called deformation density analyses. In both approaches, N_A is obtained by numerical integration of the electronic density over a grid of points

$$N_A = \sum_i \rho(r_i)\omega_q(r_i)\omega_A(r_i) \qquad (3)$$

where the index i loops over grid points. We have chosen Lebedev grids for the angular integration in combination with an Euler-MacLaurin radial quadrature scheme. The here used grids are identical to the default fixed grids in deMon2k for coarse, medium, and fine integration accuracy. In the above formula ω_q collects all quadrature weights, angular and radial ones, whereas ω_A is an atomic weight function for the real space partition into atomic cells. Five variants have been implemented in deMon2k. These are the Voronoi (V), Becke (B) [15], Hirshfeld (H) [16], iterative Hirshfeld (IH) [21] and, finally, the iterative Hirshfeld with fractional occupations numbers (IHFO) [23] partition schemes. The Voronoi cell of atom A is defined by all grid points that are closer to nucleus A than from any other atom. Therefore, the ω_A function takes the form:

$$\omega_A^V(r_i) = 1 \quad \text{if } |r_i - r_A| < |r_i - r_X| \ \forall \ X \neq A \qquad (4)$$

$$\omega_A^V(r_i) = 0 \quad |r_i - r_A| > |r_i - r_X| \ \forall \ X \neq A \qquad (5)$$

r_i, r_A and r_X are the positions of grid point i and of nuclei A and X, respectively. The Voronoi scheme renders a space division

by non-overlapping polyhedrons. The Becke atomic cells are defined from the Voronoi cells by making them slightly overlapping [15]. This is achieved by introducing a smoothing function to define fuzzy borders of cells.

$$\omega_A^B(r_i) = \frac{P_A(r_i)}{\sum_X P_X(r_i)} \qquad (6)$$

The cell functions $P_A(r_i)$ and $P_X(r_i)$ are defined by:

$$P_A(r_i) = \prod_{B \neq A} s(\mu_{AB}) \qquad (7)$$

The "soft" step function $s(\mu_{AB})$ is obtained by a threefold iteration of the polynom $p(\mu_{AB}) = \frac{3}{2}\mu_{AB} - \frac{1}{2}\mu_{AB}^3$. The here appearing elliptic coordinate,

$$\mu_{AB} = \frac{r_A - r_B}{R_{AB}} \ ,$$

is defined in the local coordinate system of the atom pair A and B as depicted in Fig. 1

For atom A ω_A^B equals unity close to the nuclei but it rapidly drops to zero when approaching the border of the Voronoi cell of the atom. Both the Voronoi and Becke schemes are based on geometrical considerations only. The chemical nature of the atoms composing the molecule of interest never enters into the definition of the atom cells, and then into the definition of the atomic charges. As a consequence, these population schemes may produce atomic charges that are not satisfactory from a chemical point of view. For example, charges on hydrogen atoms typically take values around -0.5, simply because the Voronoi/Becke cells of hydrogen atoms expand to half the length of the bonds in which they are engaged. The H, IH, and IHFO schemes constitute improvements in that regard. For these three schemes, the integration weights are functions of atomic reference densities, ρ_A^{ref}.

$$\omega_A^H(r_i) = \frac{\rho_A^{ref}(r_i)}{\sum_X \rho_X^{ref}(r_i)} \qquad (8)$$

The denominator in Eq. (8), $\sum_X \rho_X^{ref}$, defines the so-called promolecular density. There is some liberty to define the ρ_X^{ref}

Fig. 1 Elliptic coordinate definition for Becke weight calculation
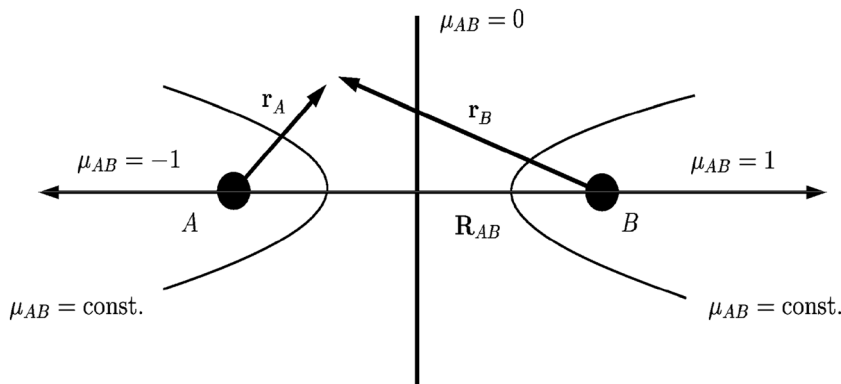
functions. In the standard Hirshfeld scheme, ρ_X^{ref} are the densities of neutral atoms. Note that other choices are acceptable, for example ρ_X^{ref} may be the densities of isolated ions. The non-uniqueness of reference density is actually a drawback of the standard Hirshfeld scheme. In deMon2k, ρ_X^{ref} are obtained by performing SCF calculations of spherically averaged neutral atoms. The ω_A^H will be close to unity near atom A but will progressively decay to zero when approaching other nuclei. Note that another drawback of the standard Hirshfeld partition is that atomic charges are generally close to zero.

To alleviate the inconvenience of the standard scheme, iterative variants have been proposed [21]. In the IH scheme, one chooses ρ_X^{ref} to be the density of an isolated atom X holding the same number of electrons N_X as the atom in the molecule (thereafter denoted ρ_X^{ref,N_X}). In other words the ω_A^{IH} function, hence the Hirshfeld cell of atom A, is adjusted iteratively so that both the reference atom and the corresponding atom in the molecule have the same number of electrons. This procedure has been shown to minimize the loss of information when defining an atom in a molecule according to the Shanon theory of information [21]. In the original article of the IH scheme, the authors proposed to define ρ_X^{ref,N_X} by interpolation between electronic densities of isolated ions, the electron numbers of which bracket N_X.

$$\rho_X^{ref,N_X} = \rho_X^{fint(N_x)}[cint(N_x)-N_x] + \rho_X^{cint(N_x)}[N_x - fint(N_x)] \quad (9)$$

In this expression, $fint(N_x)$ (resp. $cint(N_x)$) is the largest (resp. smallest) integer less (resp. greater) than or equal to N_x. Alternatively ρ_X^{ref,N_X} can be obtained by running a SCF calculation for an ion holding N_x electrons. Note that N_x is usually a non-integer number. Both variants have been tested in deMon2k and showed to give very similar atomic charges. We finally only kept the second variant based on atomic SCF calculations with non-integer electron numbers because of its simple straightforward definition.

The IHFO scheme represents an extension of the IH scheme in which both alpha (ρ^α) and beta (ρ^β) densities are integrated separately [23]. Accordingly, N_A^σ, ω_A^σ and ρ_A^{ref,N_A^σ} become spin-specific. The reference ionic densities are obtained as for the IH scheme by running SCF calculations in which the number of both alpha and beta numbers of electrons are imposed. Now the reference atom and the corresponding atom in the molecule have the same charges and spin charges. For closed-shell molecules, the IH and IHFO schemes obviously are identical but they should produce different charges for open-shell systems. Another alternative for defining ρ_A^{ref} in Eq. (7) is the IHDO-D scheme where the atomic dipoles are further imposed in the iterative procedure. We leave the introduction of the IHDO-D scheme in deMon2k for future work. Finally, we mention that ρ_A^{ref} may also be built from the densities of reference molecular fragments, as shown for example

in [17]. We have already described such an implementation in deMon2k [6].

Once atomic cells have been defined according to any of the partition schemes defined above, atomic charges are easily computed with Eqs. (1) or (2). Now, higher order moments can be defined based on the atomic cells. For example, the components of the "intrinsic" atom dipoles (μ) and quadrupoles (Θ) can be calculated by:

$$\mu_\alpha^A = \sum_i (r_{i,\alpha}-r_{A,\alpha})\rho(r_i)\omega_q(r_i)\omega_A(r_i) \quad (10)$$

$$\Theta_{\alpha\beta}^A = \sum_i (r_{i,\alpha}-r_{A,\alpha})(r_{i,\beta}-r_{A,\beta})\rho(r_i)\omega_q(r_i)\omega_A(r_i) \quad (11)$$

where $r_{A,\alpha}$ are the components of the position vector of nucleus A.

All the electronic densities that have been introduced above are obtained from the Kohn-Sham molecular orbitals (MOs). In deMon2k, these MOs are expanded within the LCGTO approximation (linear combination of Gaussian-type orbitals). The corresponding density is given as:

$$\rho(r) = \sum_{\mu,\nu} P_{\mu\nu}\mu(r)\nu(r) \quad (12)$$

where $P_{\mu\nu}$ is an element of the density matrix and μ, ν represent GTOs. Greek letters are used as indexes and also to label the GTOs. In deMon2k auxiliary densities (denoted by ρ) are also introduced to reduce the scaling of the calculation of the Coulomb interaction. The auxiliary density, ρ, is expanded as a linear combination of auxiliary functions $\bar{k}$:

$$\rho(r) = \sum_{\bar{k}} x_{\bar{k}}\bar{k}(r) \quad (13)$$

where $x_{\bar{k}}$ are the so-called Coulomb fitting coefficients. In deMon2k, the $\bar{k}$ are primitive Hermite GTOs [38]. The coefficients $x_{\bar{k}}$ are obtained from the variational fitting of the Coulomb potential as proposed by Dunlap [32, 33]. Because we have at hand such auxiliary densities in DFT calculation in deMon2k, we may expect them to be valuable to perform population analysis *in lieu* of the Kohn-Sham density. The fitted density may be used to calculate the number of electrons for each atom by replacing ρ with ρ in Eqs. (1) or (2). It can also be used to calculate the integration weights involved in the Hirshfeld schemes. In both cases, one should expect an important saving of computer time since the Kohn-Sham density is expressed as a sum of products of atomic orbitals whereas the fitted density is a simple linear combination of auxiliary functions. Note that the number of atomic orbital products greatly exceeds the number of auxiliary functions which are typically 3 to 5 times the number of basis functions. Thus, significant computational savings can be expected. Approximating the Hirshfeld weights using the fitted density is certainly less dramatic than

integrating this density itself instead of the Kohn-Sham density. In our implementation the Hirshfeld weights are *always* calculated with $\overset{\sim}{\rho}$ while the liberty is left to the user to integrate either the Kohn-Sham, ρ, or fitted, ρ, densities. To conclude this section, we stress that for deformation density analyses, although the fitted reference densities, ρ_A^{ref}, are used to calculate the integration weights, the Kohn-Sham reference densities, ρ_A^{ref}, are used to calculate N_A^{ref}.

Accuracy of population analyses

In this section, we assess the accuracy of population analysis performed from the auxiliary function density within the Becke, Hirshfeld (standard and iterative variant), and Voronoi deformation density (VDD). To this end we consider two sets of molecules. The first one is an ensemble of 66 organic molecules relevant to biological structures (thereafter referred as S66). It contains C, H, N, and O atoms. The S66 set of molecules has been reported recently by Řezác *et al.* in the context of benchmarking computations of interaction energies by quantum chemistry methodologies [39, 40]. Although the present paper is not devoted to this topic, the S66 set still provides a valuable ensemble of organic molecules to test our population analysis implementation. The second set contains 40 halogenated organic molecules, also provided by Řezác and Hobza [41]. In total 105 organic molecules are considered. These test sets encompass 987 H, 584 C, 70 N, 96 O, 2 S, and 74 halogen (X) atoms. We used the DZVP-

GGA (double zeta with valence polarization functions, calibrated for generalized-gradient-approximation functionals) [42] basis set and the PBE exchange correlation functional [43]. The XC energy and potential have been integrated numerically on an adaptive grid of medium accuracy [44]. The auxiliary density has been used to compute both the classical Coulomb and XC potential following the so-called auxiliary DFT (ADFT) framework [45]. Various auxiliary basis sets have been considered. Auxiliary basis sets are generated by an automatic procedure implemented in deMon2k that depends on the atomic orbital basis set. The GEN-An auxiliary function sets contain groups of auxiliary functions with *s* and *spd* angular momenta. The index n determines the number of auxiliary function sets, *i.e.*, the number of these sets increase with increasing n [42]. We have considered the GEN-A1, GEN-A2, and GEN-A3 auxiliary function sets, as well as the GEN-A2* that is supplemented by *f* and *g* auxiliary functions. Numerical integrations involved in population analysis have been carried out with fixed grids of medium accuracy. For the iterative schemes the iterations were pursued until the root-mean-square error was below 10^{-5}.

We first consider atomic charges obtained by four population schemes. We report the mean unsigned error (MUE) and the maximum error (MAXERR) between atomic charges obtained by analyzing the Kohn-Sham and the auxiliary density in Figs. 2 and 3, respectively. For simplicity, we will refer to them as the KS (BASIS) and auxiliary (AUXIS) charges. The calculations are repeated for four sets of auxiliary functions.

Fig. 2 Mean unsigned error (MUE) between atomic charges (in atomic units) of the fitted density with respect to KS density analysis for four populations schemes. The symbol X refers to F, Cl, Br, and I atoms. Four population schemes are investigated (Becke, VDD, standard Hirshfeld, and iterative Hirshfeld) in combination with four auxiliary basis sets (GEN-A1, GEN-A2, GENA3, GEN-A2*). Please note that different scales are used for the graphs
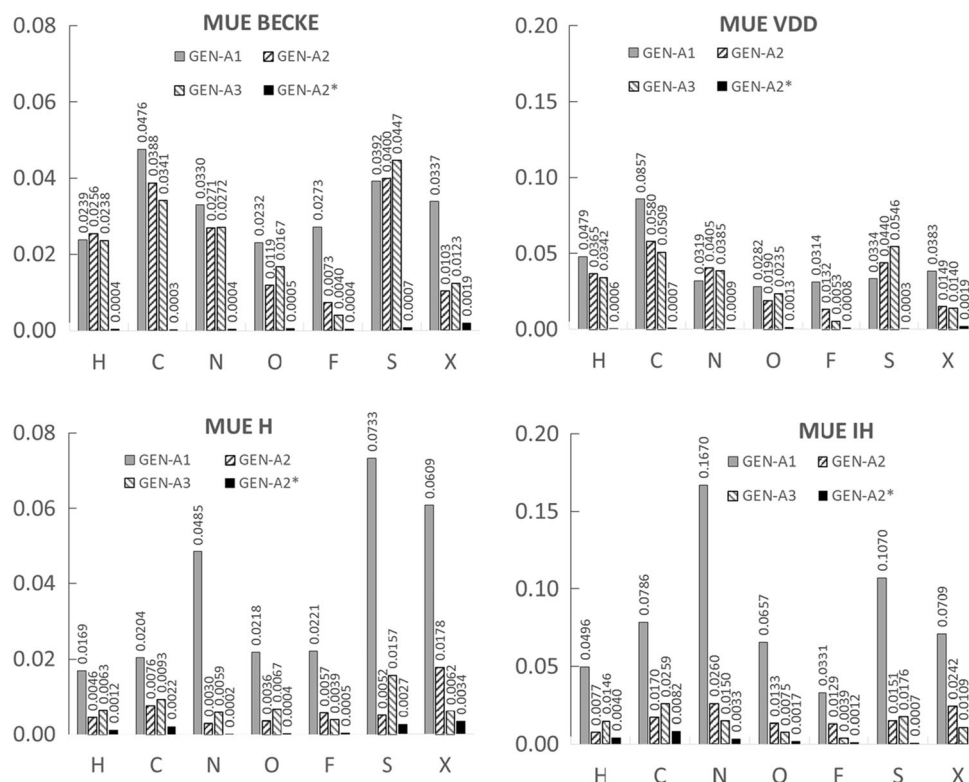
With the Hirshfeld schemes the differences between the KS and auxiliary charges decrease when going from GEN-A1 to GEN-A2. Passing to GEN-A3 does not guarantee a better convergence. For the Becke and VDD schemes none of the GEN-A1, -A2 or -A3 auxiliary function sets allow to match the atomic charges obtained by integration of the KS density. Similar conclusions can be drawn for the maximum errors (Fig. 3). Note that the maximum errors with the GEN-A1 auxiliary function set can be quite large (0.3 e$^-$). For any of the four population schemes investigated here, it is the addition of angular flexibility in the auxiliary function set (i.e., GEN-A2*) that enables a significant decrease of the MUE and of the maximum error. In conclusion the GEN-A2* auxiliary function set seems to offer an excellent accuracy close to 0.01 e$^-$ of the fitted density analysis compared to the KS density. On the other hand, auxiliary function sets comprising only s and spd sets (GEN-An, n = 1, 2 or 3) should only be used in population analyses of the fitted electronic densities if qualitative results are aimed.

We now turn to the analysis of the intrinsic dipole moments. In Fig. 4, we report the root-mean-square-deviation between the norms of the intrinsic dipole moments obtained with the AUXIS and BASIS approaches. In Fig. 5, we report the angles between the dipoles obtained with the two approaches. The atomic dipoles obtained with GEN-A1 are clearly not reliable. In particular, the orientation of the dipoles obtained from the integration of ρ or of ρ can be extremely different (see Fig. 5). The situation is largely improved with GEN-A2 or GEN-A3 both in term of the norms and orientations of the dipole moments. When using GEN-A2* the comparison is, as for atomic charges, much more satisfactory. In most cases the RMSD between the dipoles obtained with both approaches is below 0.01 D, while the orientation of the AUXIS dipoles is below 1° from the BASIS dipoles.

Computational performances

In this section, we report the efficiency of our iterative Hirshfeld population analysis implementation, which is the most time consuming partition scheme here discussed. To this end, we optimized the insulin molecule at the PBE/DZVP/GEN-A2 level of theory employing ADFT. This molecule contains 784 atoms (H, C, N, O, and S) and a total of 3078 electrons. The optimized geometry is depicted in Fig. 6.

For the optimized insulin structure, we performed IH analyses of the KS (BASIS) and fitted (AUXIS) densities employing a varying number of compute cores. The same level of theory as for the structure optimization was used, i.e., PBE/DZVP/GEN-A2. The resulting timings are depicted in Fig. 7 as a function of the number of cores. All calculations are performed with Intel® Xeon™ E5-2650v2 (2.6 GHz) 8 core CPUs with 4 GB RAM per core. To guide the eye the individual data points in Fig. 7 are connected. As expected the analysis of the fitted density is always significantly faster

Fig. 3 Maximum error between atomic charges obtained through the analyses of the auxiliary density and the Kohn-Sham density. The symbol X refers to F, Cl, Br, and I atoms. Four population schemes are investigated (Becke, VDD, standard Hirshfeld, and iterative Hirshfeld) in combination with four auxiliary basis sets (GEN-A1, GEN-A2, GENA3, GEN-A2*). Please note that different scales are used for the graphs
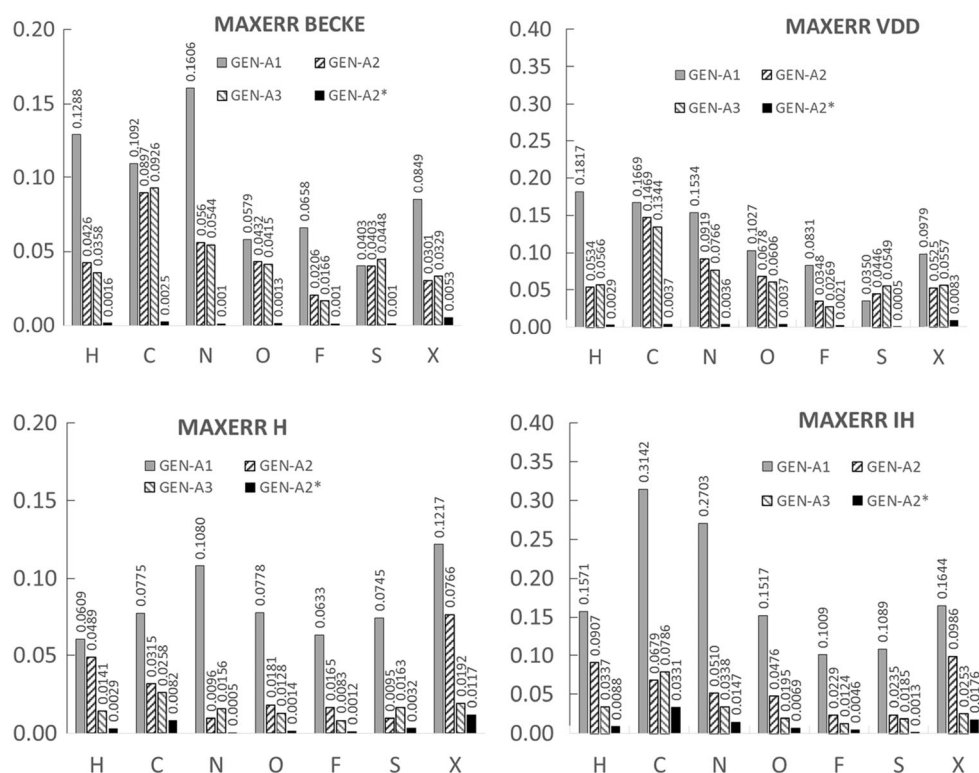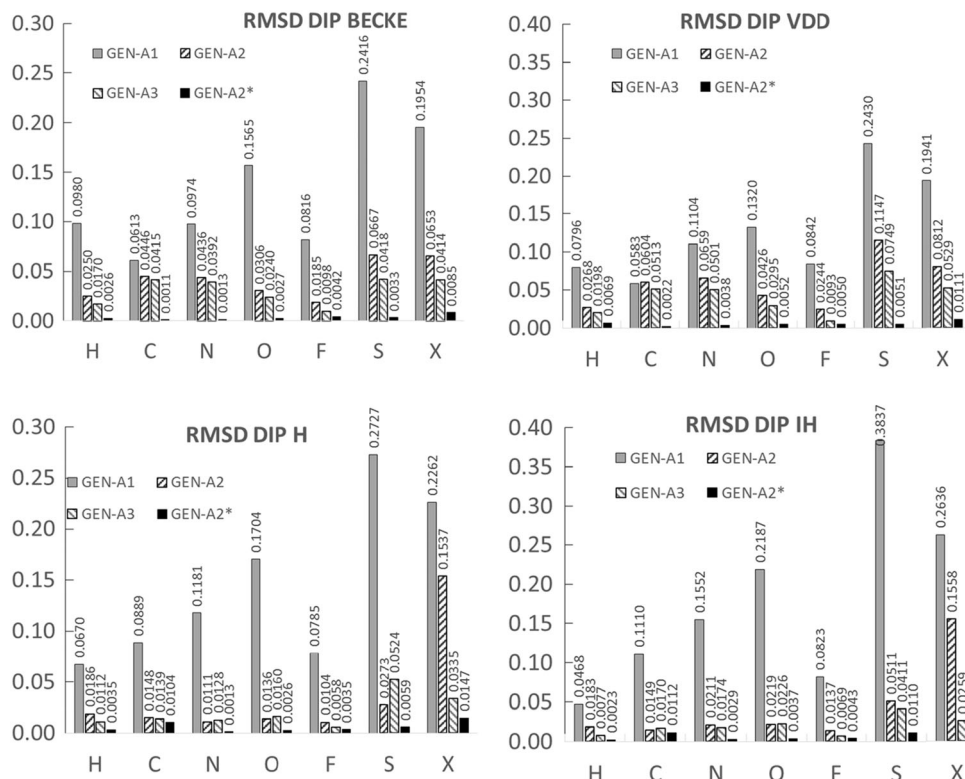
Fig. 4 Root mean square deviation between the norms of the intrinsic dipole moments (in D) obtained with the AUXIS and BASIS approaches. The symbol X refers to F, Cl, Br and I atoms. Four population schemes are investigated (Becke, VDD, standard Hirshfeld and iterative Hirshfeld) in combination with four auxiliary function sets (GEN-A1, GEN-A2, GEN-A3, GEN-A2*). Please note that different scales are used for the graphs

(factor of around 2) than the KS density. The scaling with respect to the number of cores is rather satisfying. As Fig. 7 shows computational savings are still gained when passing from 96 to 128 cores.

Fig. 5 Average angle between the intrinsic dipole moments obtained with the AUXIS and BASIS approach. The symbol X refers to F, Cl, Br, and I atoms. Four population schemes are investigated (Becke, VDD, standard Hirshfeld, and iterative Hirshfeld) in combination with four auxiliary function sets (GEN-A1, GEN-A2, GEN-A3, GEN-A2*). Please note that different scales are used for the graphs
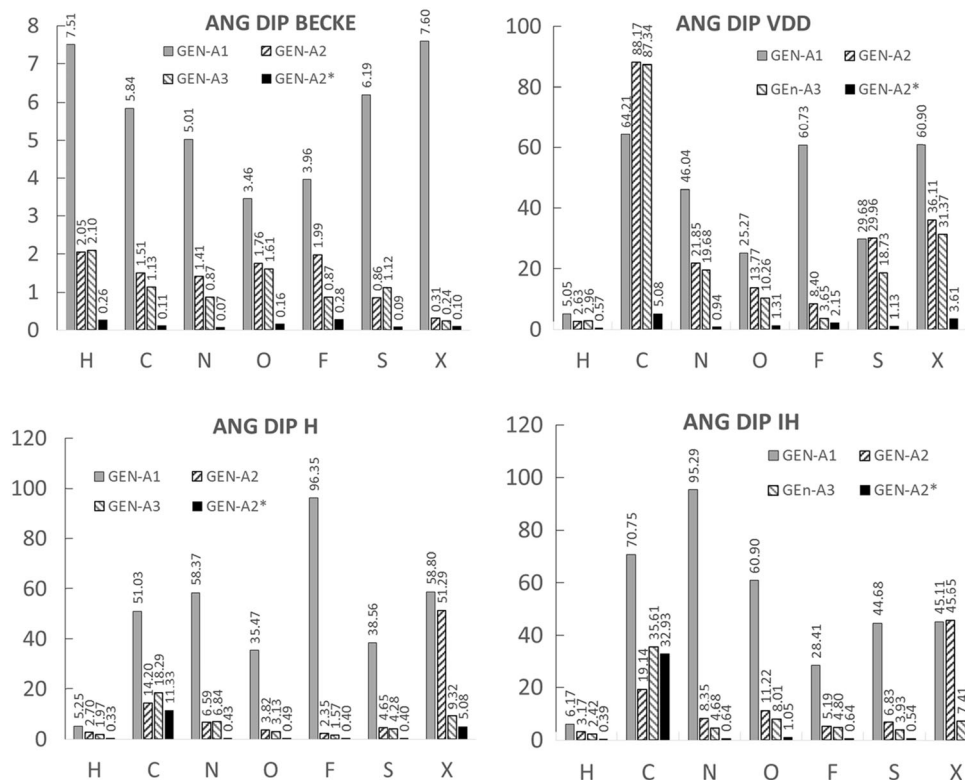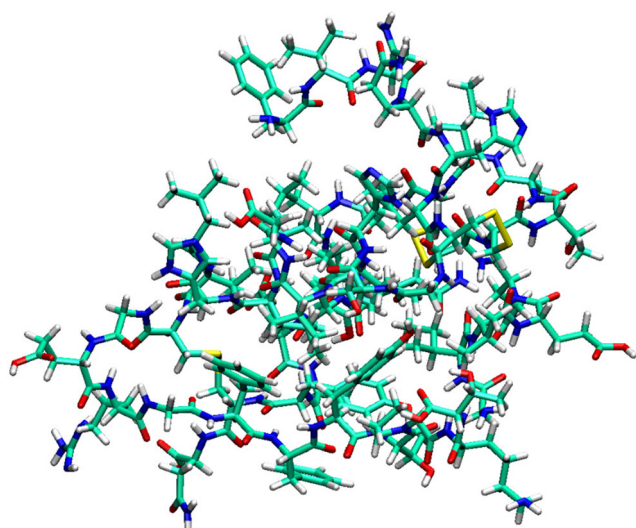
Fig. 6 3D representation of the geometrically optimized insulin molecule. Color code: H in white, C in green, N in blue, O in red, and S in yellow

To put the timings in Fig. 7 into perspective we note that the structure optimization of insulin took around 1200 optimization steps. This optimization required 3 weeks on 32 of the above specified Xeon™ cores. Thus, the here reported timings for the iterative Hirshfeld analysis are a small overhead to the structure optimization. Because they scale similar to the SCF and geometry optimization this relation holds also for larger number of cores, e.g., the above shown 128 cores. In fact, in the case of the insulin molecule the CPU time needed for the iterative Hirshfeld analysis is only 2 to 3 times larger as for a single point ADFT energy calculation. Therefore, the here discussed population analysis implementation can be used for larger systems that are of interest to biological chemistry or material science.
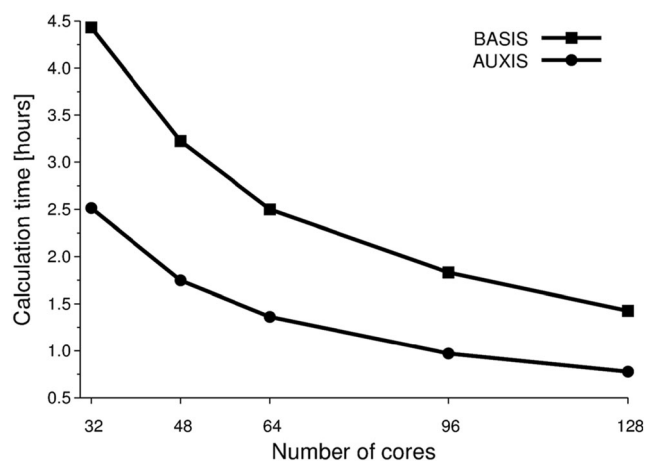

Fig. 7 Computational time for an iterative Hirshfeld analysis of insulin employing the Kohn-Sham (BASIS) or fitted (AUXIS) densities as a function of the number of compute cores

Electrostatic interaction calculations with amoeba

In this section we wish to determine if the multipole distribution obtained from the population schemes described above are suitable to calculate electrostatic interaction energies within supramolecular complexes. Previous work showed that iterative approaches give atomic charges that better reproduce the DFT electrostatic potential than non-iterative approaches [46]. We use the $Na^+(Tryp)(H_2O)$ complex to see the capability of the various population schemes to produce multipole sets in order to compute non-covalent interactions with the AMOEBA polarizable force field. The tryptamine molecule (Tryp) is derived from the tryptophan amino acid, the carboxylic acid being replaced by a hydrogen atom, and classified as a neurotransmitter. Four conformers have been selected from the previous work of Nicely and Lisy (see Fig. 1 in ref. [47] and Fig. 8). For structures A and B, the water molecule interacts both with the sodium cation and the amino group, making a hydrogen bond with the amino nitrogen. They differ mainly by the orientation of the ethylamine side chain. In structure C, the sodium cation is "sandwiched" between the water and the amino group whereas in structure D, the water molecule is acceptor of a hydrogen bond from the indole N-H.

This complex provides interesting structural pictures with various electrostatic interactions such as charge-dipole, dipole-dipole and polarization effects. The AMOEBA force field was chosen for its high-level treatment of electrostatic interactions by using a multipolar expansion up to quadrupoles on each atom and an explicit iterative polarization term. The multipoles have been computed following both IH and H schemes for the isolated tryptamine and then defined in their local atomic frame using the Orient program [48] to be used in the framework of AMOEBA. The relative energies taking
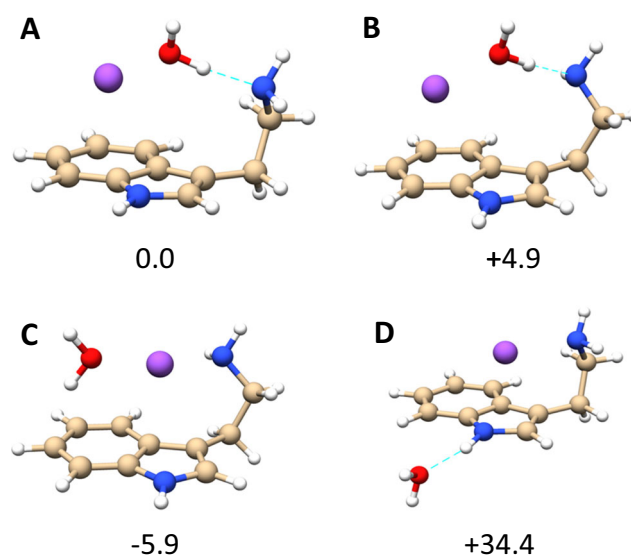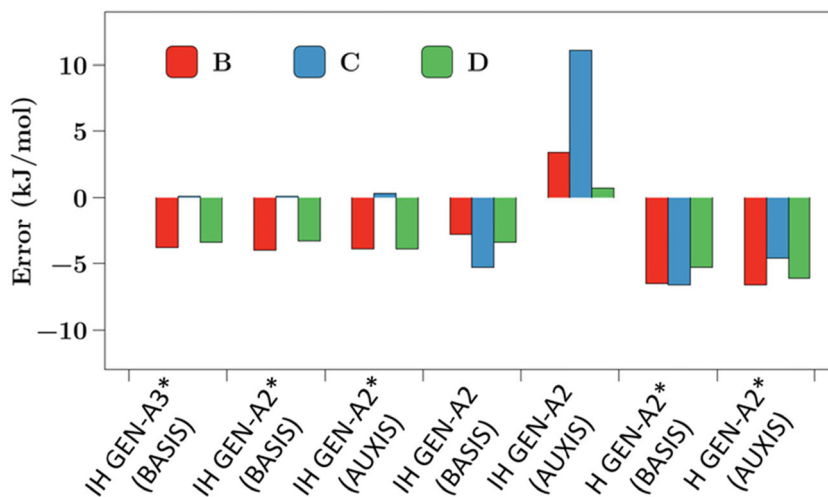

Fig. 8 Calculated structures for the $Na^+(Tryp)(H_2O)$ complex and M06-2X relative energies (kJ mol^{-1})

Fig. 9 Error (kJ mol^{-1}) on the relative energies of the four conformers between M06-2X and AMOEBA using the different multipole sets. Conformer A is taken as reference

structure A as reference, computed at the M06-2X/TZVP [49] level, are compared with the different atomic multipole sets and the energetic errors are reported in Fig. 9. When extracting

multipoles from the KS density (BASIS), the error is found to be small, in the range of the error from quantum chemistry calculations ("IH GEN-A2*/A3* (BASIS)" histograms). If the
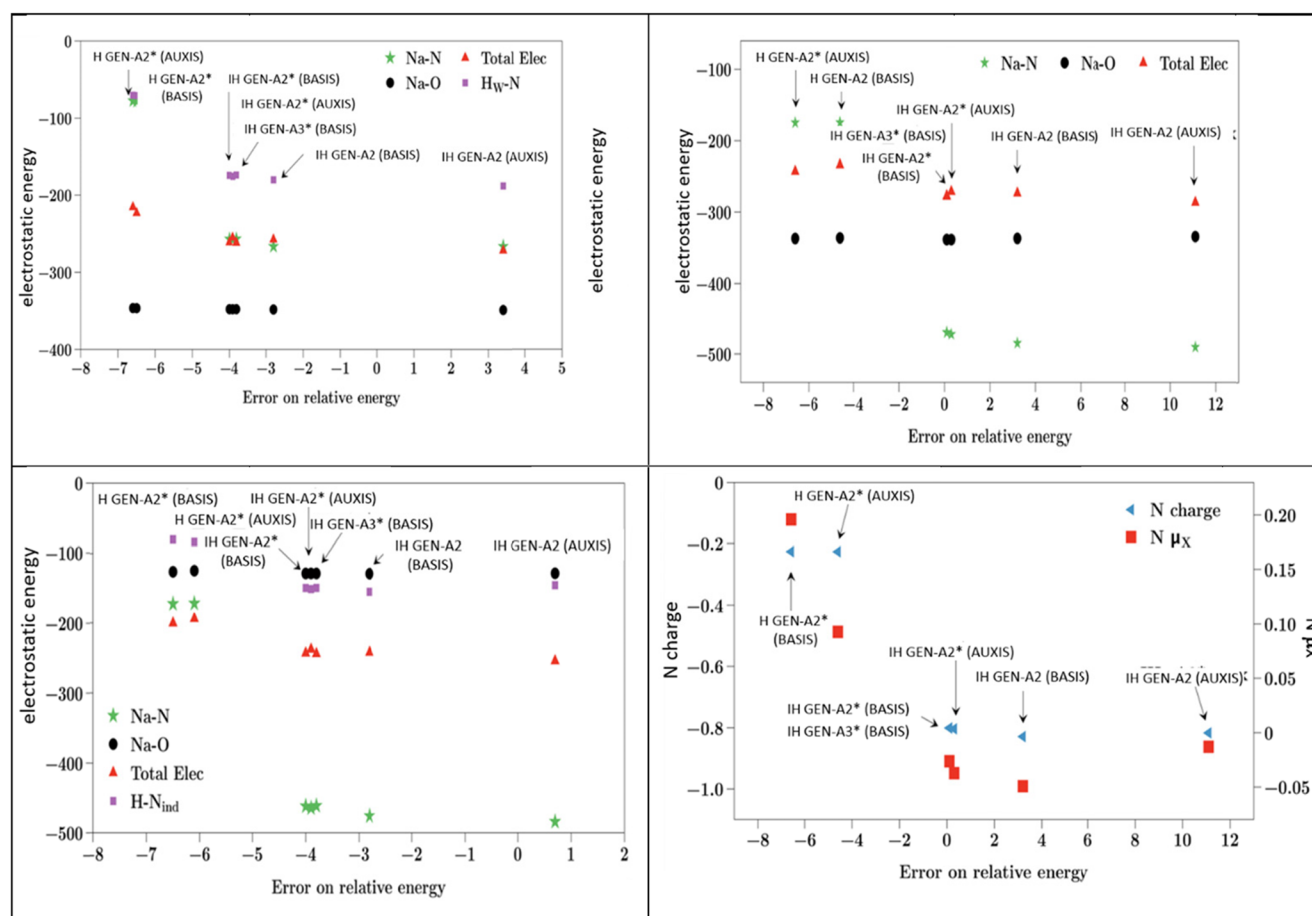

Fig. 10 Main electrostatic contribution error as a function of relative energies between M06-2X and AMOEBA using the different multipole sets (kJ mol^{-1}) for structure B (top, left), structure C (top, right), and structure D (bottom, left). See text for details. Charge and largest dipole

component on N atom in the NH$_2$ group as a function of relative energy for C (bottom, right) between M06-2X and AMOEBA using the different multipole sets (kJ mol^{-1})

Fig. 11 AMOEBA polarization energies (kJ mol^{-1}) of the four conformers using the different multipole sets
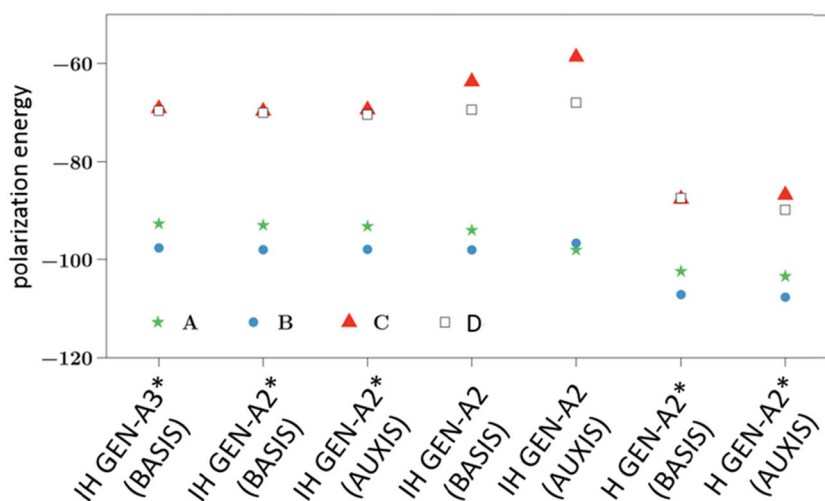
multipoles are extracted from the auxiliary density (AUXIS) the error is still very small as long as GEN-A2* is used ("IH GEN-A2* (AUXIS)" histograms). The errors associated with multipoles obtained from auxiliary density may become large when using the GEN-A2 function set ("IH GEN-A2 (AUXIS)" histograms). Finally, we find that the standard Hirshfeld scheme is less accurate to reproduce such relative energies than the iterative version in both BASIS or AUXIS approaches.

Due to the various interactions involved in the different structures, the errors to reproduce the relative energies should come from wrong specific energetic contributions. The graphs in Fig. 10 represent the values of the main components of the electrostatic energy and the total electrostatic energy for structures B, C, and D as a function of the error on the total energy relative to structure A. For the three structures, the largest error comes from the underestimation of the Na$^+$-N(amino) interaction whereas the energy for the Na$^+$-O interaction remains relatively constant. Furthermore, the water-N(amino) interaction also plays a role in structure B and the water-N(indole) one in structure D, with a less extent, respectively. Furthermore, this effect can be easily correlated to the charge and the x-component of the dipole moment of the nitrogen atom of the amino group when they are reported as a function of the energetic error for C. The large error for C in the IH scheme using the GEN-A2 auxiliary basis can be explained by a cumulative error on the charge and dipole moment on N and the small value of the polarization energy (Fig. 10).

Consequently, the polarization energy of the structures can be compared for the different multipole sets (Fig. 11). Even if the effect is smaller than for the electrostatic component, this term contributes to the non-covalent interactions. Let's focus on the IH first. When the GEN-A2 auxiliary function set is used the AUXIS and BASIS give different polarization energies that may differ by several kJ mol^{-1}. This is especially noticeable for structure C for example. On the other hand we

obtain very similar polarization energies for each structures when the GEN-A2* auxiliary function set is used (compare the "IH GEN-A2* (BASIS)" vs. "IH GEN-A2* (AUXIS)" data points). When considering the standard Hirshfeld scheme, we find that polarization energies are significantly larger. For example if we consider structure C, together with the BASIS approach (taking GEN-A2*) the polarization energy goes from 65 kJ mol^{-1} to 85 kJ mol^{-1}.

Conclusions

In this work we have been interested in the accuracy of population analyses based on fitted densities in the context of DFT. The main conclusion of our study can be summarized as follow. We found that fitted densities can actually be used instead of the Kohn-Sham density to extract electrostatic multipoles from DFT calculations. However, the quality of the auxiliary function sets to expand the fitted density has a great impact on the results and should thus be considered with care in applications. With standard sets comprising *s* and *spd* angular momentum functions only qualitative agreement between the BASIS and AUXIS atomic charges is found. Conversely, when using GEN-A2* (or GEN-A3*) the agreement between both approaches is excellent for either atomic charges or higher order multipoles. As seen in the case of insulin, the AUXIS approach offers a significant reduction of computational cost compared to the BASIS one. We finally tested the capabilities of the extracted multipoles of the tryptamine system to provide sufficiently accurate interaction energy calculation with the AMOEBA force field. In that regard the iterative Hirshfeld scheme represents a clear improvement over the traditional Hirshfeld scheme. Good results have been obtained with the IH scheme and GEN-A2* or GEN-A3* auxiliary function sets. Overall these results encourage us to pursue our ongoing efforts on the implementation of advanced

QM/MM schemes that include second and third generation force fields in deMon2k [50, 51].

References

1. Misquitta AJ, Stone AJ, Fazeli F (2014) Distributed multipoles from a robust basis-space implementation of the iterated stockholder atoms procedure. J Chem Theory Comput 10(12):5405–5418. doi:10.1021/ct5008444

2. Mei Y, Simmonett AC, Pickard FC, DiStasio RA, Brooks BR, Shao Y (2015) Numerical study on the partitioning of the molecular polarizability into fluctuating charge and induced atomic dipole contributions. J Phys Chem A 119(22):5865–5882. doi:10.1021/acs.jpca.5b03159

3. Verstraelen T, Vandenbrande S, Heidar-Zadeh F, Vanduyfhuys L, Van Speybroeck V, Waroquier M, Ayers PW (2016) Minimal basis iterative stockholder: atoms in molecules for force-field development. J Chem Theory Comput 12(8):3894–3912. doi:10.1021/acs.jctc.6b00456

4. Kaduk B, Kowalczyk T, Van Voorhis T (2012) Constrained density functional theory. Chem Rev 112(1):321–370. doi:10.1021/cr200148b

5. Voorhis TV, Kowalczyk T, Kaduk B, Wang L-P, Cheng C-L, Wu Q (2010) The diabatic picture of electron transfer, reaction barriers, and molecular dynamics. Annu Rev Phys Chem 61(1):149–170. doi:10.1146/annurev.physchem.012809.103324

6. Řezáč J, de la Lande A (2015) Robust, basis-set independent method for the evaluation of charge-transfer energy in noncovalent complexes. J Chem Theory Comput 11(2):528–537. doi:10.1021/ct501115m

7. Sokalski WA, Poirier RA (1983) Cumulative atomic multipole representation of the molecular charge distribution and its basis set dependence. Chem Phys Lett 98(1):86–92. doi:10.1016/0009-2614(83)80208-5

8. Sokalski WA, Sawaryn A (1987) Correlated molecular and cumulative atomic multipole moments. J Chem Phys 87(1):526–534. doi:10.1063/1.453600

9. Köster AM, Kölle C, Jug K (1993) Approximation of molecular electrostatic potentials. J Chem Phys 99(2):1224–1229. doi:10.1063/1.465366

10. Mulliken RS (1955) Electronic population analysis on LCAO–MO molecular wave functions. I. J Chem Phys 23(10):1833–1840. doi:10.1063/1.1740588

11. Carbó-Dorca R, Bultinck P (2004) Quantum mechanical basis for Mulliken population analysis. J Math Chem 36(3):231–239. doi:10.1023/b:jomc.0000044221.23647.20

12. Löwdin P-O (1970) On the nonorthogonality problem*. In: Per-Olov L (ed) Advances in quantum chemistry, vol 5. Academic, Cambridge, pp 185-199. doi:10.1016/S0065-3276(08)60339-1

13. Reed AE, Weinstock RB, Weinhold F (1985) Natural population analysis. J Chem Phys 83(2):735–746. doi:10.1063/1.449486

14. Reed AE, Curtiss LA, Weinhold F (1988) Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint. Chem Rev 88(6):899–926. doi:10.1021/cr00088a005

15. Becke AD (1988) A multicenter numerical integration scheme for polyatomic molecules. J Chem Phys 88(4):2547–2553. doi:10.1063/1.454033

16. Hirshfeld FL (1977) Bonded-atom fragments for describing molecular charge densities. Theoret Chim Acta 44(2):129–138. doi:10.1007/bf00549096

17. Fonseca Guerra C, Handgraaf J-W, Baerends EJ, Bickelhaupt FM (2004) Voronoi deformation density (VDD) charges: Assessment of the Mulliken, Bader, Hirshfeld, Weinhold, and VDD methods for charge analysis. J Comput Chem 25(2):189–210. doi:10.1002/jcc.10351

18. Heyndrickx W, Salvador P, Bultinck P, Solà M, Matito E (2011) Performance of 3D-space-based atoms-in-molecules methods for electronic delocalization aromaticity indices. J Comput Chem 32(3):386–395. doi:10.1002/jcc.21621

19. Matito E, Sola M, Salvador P, Duran M (2007) Electron sharing indexes at the correlated level. Application to aromaticity calculations. Faraday Discuss 135:325–345. doi:10.1039/b605086g

20. Salvador P, Ramos-Cordoba E (2013) Communication: An approximation to Bader's topological atom. J Chem Phys 139(7):071103. doi:10.1063/1.4818751

21. Bultinck P, Van Alsenoy C, Ayers PW, Carbó-Dorca R (2007) Critical analysis and extension of the Hirshfeld atoms in molecules. J Chem Phys 126(14):144111. doi:10.1063/1.2715563

22. Ghillemijn D, Bultinck P, Van Neck D, Ayers PW (2011) A self-consistent Hirshfeld method for the atom in the molecule based on minimization of information loss. J Comput Chem 32(8):1561–1567. doi:10.1002/jcc.21734

23. Geldof D, Krishtal A, Blockhuys F, Van Alsenoy C (2011) An extension of the Hirshfeld method to open shell systems using fractional occupations. J Chem Theory Comput 7(5):1328–1335. doi:10.1021/ct100743h

24. Lillestolen TC, Wheatley RJ (2009) Atomic charge densities generated using an iterative stockholder procedure. J Chem Phys 131(14):144101. doi:10.1063/1.3243863

25. Piquemal JP, Pilmé J, Parisel O, Gérard H, Fourré I, Bergès J, Gourlaouen C, De La Lande A, Van Severen MC, Silvi B (2008) What can be learnt on biologically relevant systems from the topological analysis of the electron localization function? Int J Quantum Chem 108(11):1951–1969. doi:10.1002/qua.21711

26. Bader RFW, Nguyen-Dang TT, Tal Y (1981) A topological theory of molecular structure. Rep Prog Phys 44(8):893

27. Singh UC, Kollman PA (1984) An approach to computing electrostatic charges for molecules. J Comput Chem 5(2):129–145. doi:10.1002/jcc.540050204

28. Besler BH, Merz KM, Kollman PA (1990) Atomic charges derived from semiempirical methods. J Comput Chem 11(4):431–439. doi:10.1002/jcc.540110404

29. Chirlian LE, Francl MM (1987) Atomic charges derived from electrostatic potentials: A detailed study. J Comput Chem 8(6):894–905. doi:10.1002/jcc.540080616

30. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. J Comput Chem 25(9):1157–1174. doi:10.1002/jcc.20035

31. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, Mackerell AD (2010) CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J Comput Chem 31(4):671–690. doi:10.1002/jcc.21367

32. Mintmire JW, Dunlap BI (1982) Fitting the Coulomb potential variationally in linear-combination-of-atomic-orbitals density-functional calculations. Phys Rev A 25(1):88–95

33. Dunlap BI, Rösch N, Trickey SB (2010) Variational fitting methods for electronic structure calculations. Mol Phys 108(21-23):3167–3180. doi:10.1080/00268976.2010.518982

34. Aquilante F, Pedersen TB, Lindh R (2007) Low-cost evaluation of the exchange Fock matrix from Cholesky and density fitting representations of the electron repulsion integrals. J Chem Phys 126(19):194106. doi:10.1063/1.2736701

35. Ren P, Ponder JW (2003) Polarizable atomic multipole water model for molecular mechanics simulation. J Phys Chem B 107(24):5933–5947. doi:10.1021/jp027815+

36. Ponder JW, Wu C, Ren P, Pande VS, Chodera JD, Schnieders MJ, Haque I, Mobley DL, Lambrecht DS, DiStasio RA, Head-Gordon M, Clark GNI, Johnson ME, Head-Gordon T (2010) Current status of the AMOEBA polarizable force field. J Phys Chem B 114(8): 2549–2564. doi:10.1021/jp910674d

37. Köster AM, Geudtner G, Alvarez-Ibarra A, Calaminici P, Casida ME, Carmona-Espindola J, Dominguez V, Flores-Moreno R, Gamboa GU, Goursot A, Heine T, Ipatov A, de la Lande A, Janetzko F, del Campo J-M, Mejia-Rodriguez D, Reveles J, Vasquez-Perez J, Vela A, Zuniga-Gutierrez B, Salahub DR (2016) deMon2k Version 5. Mexico City

38. Köster AM (2003) Hermite Gaussian auxiliary functions for the variational fitting of the Coulomb potential in density functional methods. J Chem Phys 118(22):9943–9951. doi:10.1063/1.1571519

39. Řezáč J, Riley KE, Hobza P (2011) S66: a well-balanced database of benchmark interaction energies relevant to biomolecular structures. J Chem Theory Comput 7(8):2427–2438. doi:10.1021/ct2002946

40. Řezáč J, Riley KE, Hobza P (2011) Extensions of the S66 data set: more accurate interaction energies and angular-displaced nonequilibrium geometries. J Chem Theory Comput 7(11):3466–3470. doi:10.1021/ct200523a

41. Řezáč J, Riley KE, Hobza P (2012) Benchmark calculations ofnoncovalent interactions of halogenated molecules. J Chem Theory Comput 8(11):4285–4292. doi:10.1021/ct300647k

42. Calaminici P, Janetzko F, Köster AM, Mejia-Olvera R, Zuniga-Gutierrez B (2007) Density functional theory optimized basis sets for gradient corrected functionals: 3d transition metal systems. J Chem Phys 126(4):044108. doi:10.1063/1.2431643

43. Perdew JP, Burke K, Ernzerhof M (1996) Generalized gradient approximation made simple. Phys Rev Lett 77(18):3865–3868

44. Köster AM, Flores-Moreno R, Reveles JU (2004) Efficient and reliable numerical integration of exchange-correlation energies and potentials. J Chem Phys 121(2):681–690. doi:10.1063/1.1759323

45. Köster AM, Reveles JU, del Campo JM (2004) Calculation of exchange-correlation potentials with auxiliary function densities. J Chem Phys 121(8):3417–3424. doi:10.1063/1.1771638

46. Van Damme S, Bultinck P, Fias S (2009) Electrostatic potentials from self-consistent Hirshfeld atomic charges. J Chem Theory Comput 5(2):334–340. doi:10.1021/ct800394q

47. Nicely AL, Lisy JM (2011) Charge and temperature effects on hydrated tryptamine cluster ions. J Phys Chem A 115(13):2669–2678. doi:10.1021/jp1059648

48. Stone AJ, Dullweber A, Engkvist O, Fraschini E, Hodges MP, Meredith AW, Nutt DR, Popelier PLA, Wales DJ (2002) Orient: a program for studying interactions between molecules. version 4.5 edn. University of Cambridge

49. Mejía-Rodríguez D, Huang X, del Campo JM, Köster AM (2015) Hybrid functionals with variationally fitted exact exchange. Adv Quantum Chem 71:41–67

50. Alvarez-Ibarra A, Köster AM, Zhang R, Salahub DR (2012) Asymptotic expansion for electrostatic embedding integrals in QM/MM calculations. J Chem Theory Comput 8(11):4232–4238. doi:10.1021/ct300609z

51. Salahub D, Noskov S, Lev B, Zhang R, Ngo V, Goursot A, Calaminici P, Köster A, Alvarez-Ibarra A, Mejía-Rodríguez D, Řezáč J, Cailliez F, de la Lande A (2015) QM/MM calculations with deMon2k. Molecules 20(3):4780