

# Further refinements of next-generation force fields — Nonempirical localization of off-centered points in molecules

Robin Chaudret, Nohad Gresh, G. Andrés Cisneros, Anthony Scemama, and Jean-Philip Piquemal

**Abstract:** In the present work, we investigate different possibilities for the nonempirical localization of nonatomic centers within the context of the design of new generation polarizable force fields. To do so, the positions of electron localization function (ELF) and electron pair localization function (EPLF) attractors and of Boys orbital centroids are determined for a set of thirteen saturated and conjugated biologically relevant molecules. We consider the similarities and differences in the representations of localized lone pairs and  $\pi$  delocalized systems by these approaches, as well as the effects of the basis sets and of the level of quantum chemistry (QC). All three methods give consistent results upon dealing with the localized lone pairs. Concerning aromatic systems, whereas ELF and EPLF methods give mutually consistent results, the Boys scheme breaks the symmetry by alternating the electron distributions along the C–C bonds providing a different distribution of off-centered points. We then investigate the influence of lone pair localization in the water model of the Gaussian electrostatic model/sum of interactions between fragments *ab initio* computed (GEM/SIBFA) polarizable force field, which embodies an explicit representation of the lone pairs. This is done, in a series of mono- and poly-hydrated Zn(II) complexes, by comparing the overlap-dependent repulsion ( $E_{\text{rep}}$ ) and charge-transfer ( $E_{\text{ct}}$ ) contributions to their QC counterparts. We resort to either the current GEM/SIBFA water lone pair internal coordinates or to the ones derived from the previously mentioned localization procedures. The latter enables closer reproductions by  $E_{\text{rep}}$  and  $E_{\text{ct}}$  (GEM/SIBFA) of their exchange repulsion ( $E_{\text{exch}}$ ) and  $E_{\text{ct}}$  QC counterparts. The present preliminary results show that QC localization procedures can be used to derive accurate, nonempirical positions for off-centered points intervening in the formulation of the overlap-dependent contributions of next-generation polarizable force fields.

**Key words:** electron localization function (ELF), electron pair localization function (EPLF), sum of interactions between fragments *ab initio* computed (SIBFA), Gaussian electrostatic model (GEM), Boys, extra points.

**Résumé :** Dans la présente étude, nous examinons les différentes localisations non empiriques possibles des centres non atomiques dans le contexte de la conception de champs de forces polarisables de nouvelle génération. À cette fin, nous déterminons les positions des attracteurs ELF (*Electron Localization Function*, ou fonction de localisation électronique) et des attracteurs EPLF (*Electron Pair Localization Function*, ou fonction de localisation de paire d'électrons), ainsi que des centroïdes des orbitales de Boys pour un jeu de treize molécules saturées et conjuguées ayant une importance biologique. Nous considérons les similarités et les différences entre les représentations des paires libres et des systèmes  $\pi$  délocalisés obtenues par ces approches, ainsi que les effets des bases et du niveau de chimie quantique (QC). Les trois méthodes donnent des résultats cohérents lorsque l'on s'intéresse aux paires libres localisées. En ce qui concerne les systèmes aromatiques, alors que les méthodes ELF et EPLF donnent des résultats mutuellement cohérents, la méthode de Boys rompt la symétrie en alternant la répartition des électrons le long des liaisons C–C, ce qui donne une distribution différente des points excentrés. Nous étudions ensuite l'influence de la localisation des paires libres dans le modèle de l'eau dans le champ de forces polarisable GEM/SIBFA, qui figure une représentation explicite des paires libres. Sur une série de complexes de Zn(II) mono et polyhydratés, nous comparons les contributions des énergies de répulsion ( $E_{\text{rep}}$ ) et de transfert de charge ( $E_{\text{ct}}$ ) dépendantes de la superposition à leurs équivalents dans les calculs de chimie quantique (QC). Nous utilisons les coordonnées internes des paires libres de l'eau de la méthode GEM/SIBFA courante ou celles dérivées des procédures de localisation susmentionnées. Les secondes permettent que  $E_{\text{rep}}$  and  $E_{\text{ct}}$  (GEM/SIBFA) donnent des reproductions plus proches de leurs équivalents QC  $E_{\text{exch}}$  et  $E_{\text{ct}}$ . Les résultats préliminaires présentés ici montrent que des procédures de localisation basées sur la chimie quantique peuvent être utilisées pour calculer les positions exactes, non empiriques, pour les points excentrés intervenant dans la formulation des contributions dépendantes du recouvrement des champs de forces polarisables de nouvelle génération. [Traduit par la Rédaction]

**Mots-clés :** ELF, EPLF, SIBFA, GEM, Boys, points extra-atomiques.

Received 31 December 2012. Accepted 14 March 2013.

**R. Chaudret and J.P. Piquemal.** UPMC Univ Paris 06, UMR 7616, Laboratoire de Chimie Théorique, case courrier 137, 4 place Jussieu, F-75005, Paris, France; CNRS, UMR 7616, Laboratoire de Chimie Théorique, case courrier 137, 4 place Jussieu, F-75005, Paris, France.

**N. Gresh.** Laboratoire de Chimie et Biochimie Pharmacologiques et Toxicologiques, UMR 8601 CNRS, UFR Biomédicale, Université Paris Descartes, 45 rue des Saint-Pères, 75006 Paris, France.

**G.A. Cisneros.** Department of Chemistry, Wayne State University, 5101 Cass Ave., Detroit, MI 48202, USA.

**A. Scemama.** Laboratoire de Chimie et Physique Quantiques, CNRS-IRSAMC, Université de Toulouse, France.

**Corresponding authors:** Nohad Gresh (e-mail: nohad.gresh@parisdescartes.fr) and Jean-Philip Piquemal (e-mail: jpp@lct.jussieu.fr).

This article is part of a Special Issue dedicated to Professor Dennis Salahub in recognition of his contributions to theoretical and computational chemistry.

Two of the authors (N.G. and J.-P. P.) thank Professor Dennis R. Salahub for a long-standing and enriching collaboration.

## Introduction

The modeling of large molecular complexes is generally performed in the context of “classical” all-atom force fields such as assisted model building and energy refinement (AMBER),<sup>1</sup> chemistry at Harvard molecular mechanics (CHARMM),<sup>2</sup> or Groningen molecular simulation package (GROMOS).<sup>3</sup> The intermolecular interaction energies and their gradients are obtained as sums of atom–atom contributions devoid of extra-atomic centers. Such approaches have been successful in the representation of a diversity of systems (see ref. 4 and refs. cited therein). Quantum chemistry (QC)-grounded topological analyses, such as electron localization function (ELF<sup>5–7</sup>) or electron pair localization function (EPLF<sup>8,9</sup>), have, however, shown that a significant part of the electronic density is localized on bonds and lone pairs rather than on atoms. Extending the representation of the density with off-centered points should therefore be beneficial. This then raises the issue of how to best localize the extra points, which represent the centroids of the lone pairs and of the chemical bonds.

For this purpose, we compare in the first part of this study the results from three different methods that enable extra-atomic point localization. These are the Boys procedure<sup>10</sup> and the ELF and EPLF topological analyses. Their essential features are as follows. The Boys procedure computes localized molecular orbitals (LMOs) by maximizing the distances between the centroids of all couples of orbitals. As a result, the LMOs are localized on atoms (core electrons), bonds, and lone pairs (valence electrons).

An alternative to the Boys localization is found in the context of the topological analysis approaches. In a topological analysis, a partition of the molecular space is achieved using the theory of gradient dynamical systems. This partitioning gives a set of molecular volumes or domains denoted as “basins” localized around maxima (attractors) of the vector field of a scalar (local) function. Using the right choice of scalar function, such attractors, can be related to the Boys centroids. In the “quantum theory of atoms in molecules” (QTAIM) due to Bader,<sup>11</sup> such a scalar function is the electron density itself and the basins are associated to the atoms only. To access localization going beyond atoms, one should resort to the use of another local function. For example, the ELF function can be used and is interpreted as a signature of the electronic-pair distribution. The relationship of the ELF function to pair functions has been demonstrated but ELF is defined to have values restricted to between 0 and 1. Once computed on a 3D grid, the ELF function can be partitioned into an intuitive chemical scheme compatible with the spirit of Boys localization. Indeed, core regions can be determined (if the atomic number,  $Z$ , is  $> 2$ ) for any atom  $A$ , namely the core basin labeled  $C(A)$ . Regions associated to lone pairs are labeled  $V(A)$  and bonding regions denoting chemical bonds are noted  $V(A,B)$ , in contrast to the QTAIM theory where the basins are localized on atoms only. These ELF regions closely match the domains of the valence shell electron pair repulsion (VSEPR) model. Recently, another local function providing the same information as the ELF has been proposed and is denoted as EPLF. Initially introduced within the framework of quantum Monte Carlo (QMC) approaches, EPLF is an indicator giving the average distances between an electron of a given spin located at point  $r$  and the closest spin-like and spin-unlike electrons. Its initial numerical definition has been recently modified to be analytically computable for any type of standard wave function and density functional theory (DFT) based methods. For closed-shell systems, the EPLF results are shown to closely match the ELF ones.

In this work, their similarities and differences will be evaluated on thirteen different molecular fragments frequently encountered in biological systems. These fragments differ in terms of the nature of their lone pairs and of their chemical bonds, whether single, conjugated, or double. The influence on the Gaussian basis sets and level of QC treatment will also be considered.

Water is a central molecule in condensed and biomolecular systems. Accurately reproducing its properties is therefore critical for their modeling. There exist several classical models that include nonatomic centers, whether nonpolarizable<sup>12,13</sup> or polarizable,<sup>14–16</sup> that are used in simulations of liquid water and water clusters. However, it was shown that an explicit representation of the water lone pairs is critical to reliably account for the nonadditivity of polarization and charge transfer in water clusters.<sup>14–16</sup> Subsequent to the QC analyses, we will investigate the influence of the lone pair localization on the quality of the water model<sup>14</sup> of the sum of interactions between fragments ab initio computed (SIBFA)<sup>4</sup> polarizable force field. Along with the use of distributed QC multipoles and polarizabilities to compute the electrostatic and polarization contributions, a unique feature of SIBFA is the explicit representation of lone-pair hybrids to represent the overlap-dependent, short-range contributions of  $\Delta E$ , namely repulsion,  $E_{\text{rep}}$ , in first order and charge transfer,  $E_{\text{ct}}$ , in second order. These features were essential to afford a successful reproduction of both in-plane and out-of-plane angular variations of their respective QC counterparts, exchange repulsion ( $E_{\text{exch}}$ ) and  $E_{\text{ct}}$ . This was illustrated for the first time in the case of the linear water dimer,<sup>17</sup> and thereafter in a series of cation–ligand complexes,<sup>18,19</sup> as well as complexes of ligands with a water probe.<sup>14a</sup> Toward further refining  $E_{\text{rep}}$  and  $E_{\text{ct}}$ , we will reinvestigate some previously studied<sup>19</sup> mono- and poly-hydrated Zn(II) complexes and compare their values using the current values of lone pair internal coordinates to those derived from the previously mentioned QC localization procedures. Resorting to internal coordinates derived from ELF, EPLF, or Boys analyses on isolated molecules is fully justified regarding  $E_{\text{rep}}$ , which is a first-order contribution, involving the frozen (unperturbed) functions of the interacting partners. Regarding the two second-order contributions, polarization ( $E_{\text{pol}}$ ) and  $E_{\text{ct}}$ , the possible displacements of the lone pairs due to the field of the incoming cation are accounted for implicitly in the initial calibration step, which had resorted to a divalent cation as a probe. It is also noted that for  $E_{\text{ct}}$  an additional modulation of the effective radius of the electron donor was introduced in ref. 18 as a function of the magnitude of the field generated by the electron acceptor.

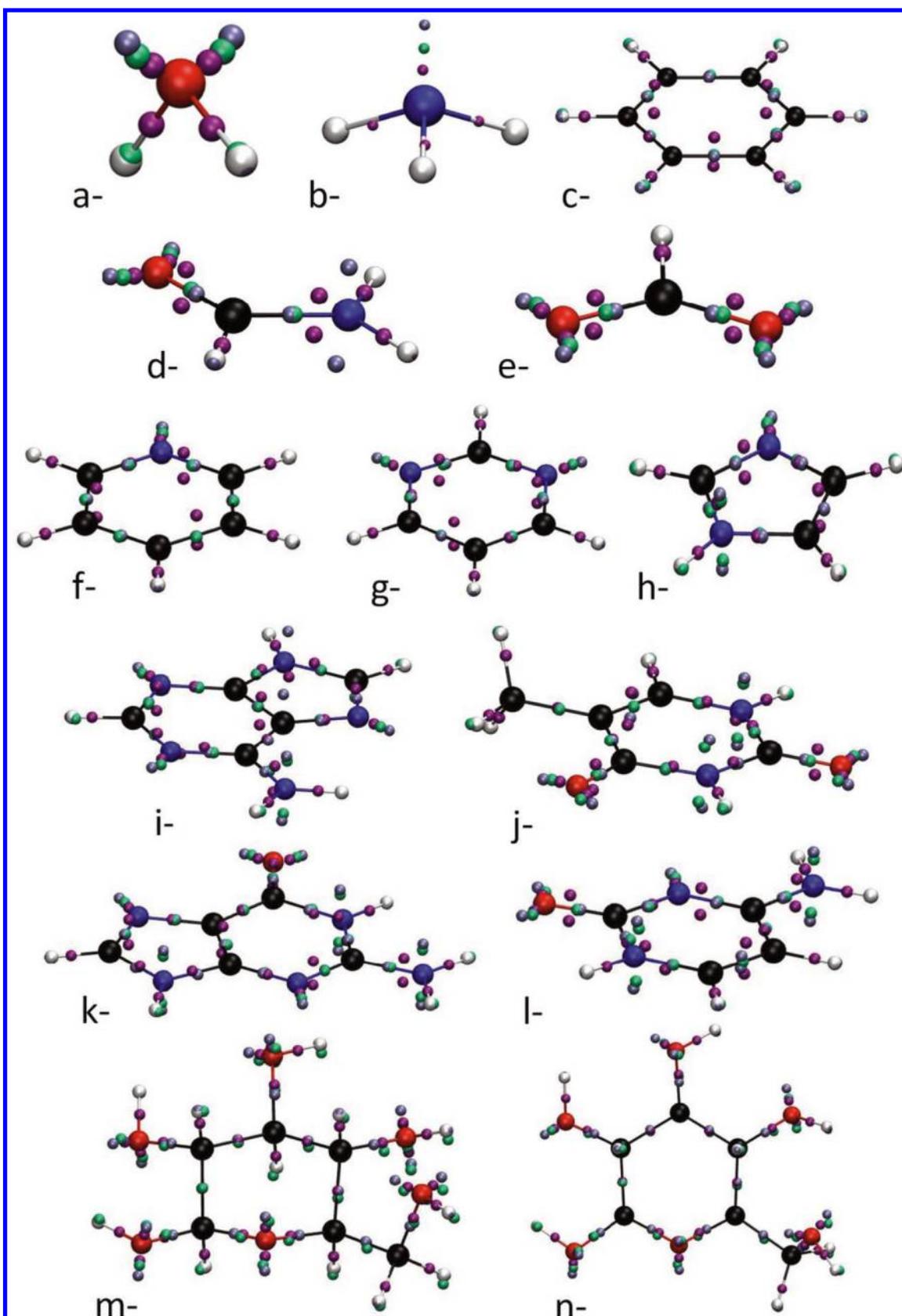
## Quantum chemistry

### Computational chemistry

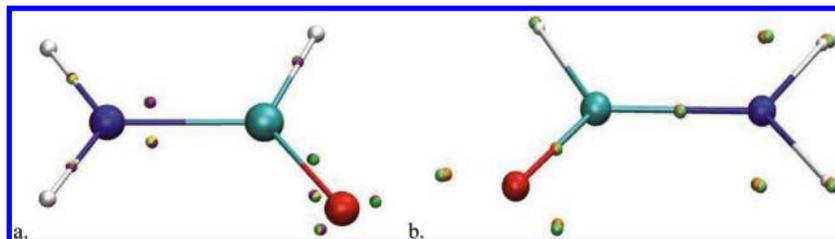
The geometries of the thirteen molecules considered were optimized using the B3LYP<sup>20</sup> DFT functional. At their energy minima, a vibration analysis was performed showing the absence of imaginary frequencies. The off-atomic centers, denoted as extra points hereafter, were localized using Boys procedure as well as ELF and EPLF topological analyses. For ELF, the TopMod program was used to derive the positions of attractors.<sup>21</sup> Additional details about the ELF topological analysis and its applications to biological systems can be found in a recent review paper.<sup>7</sup> For EPLF, upon using the original program<sup>9</sup> an attractor localization procedure was developed for closed-shell molecules, which localizes in the three-dimensional molecular space the different minima of the EPLF. The EPLF attractors were defined as the local maxima of the EPLF as in the case of the ELF attractors.

For formamide, different basis sets were used to locate the Boys centroids and the ELF attractors, namely 6-31G,<sup>22</sup> 6-31++G, 6-31G\*\*, 6-31++G\*\*, 6-311G,<sup>23</sup> 6-311++G\*\*, and CEP 4-31G(3d).<sup>24</sup> For all three localization procedures, the QC computations were performed at the HF, BLYP, B3LYP, and mPWPW91 level.<sup>25</sup> MP2 and CCSD density were also derived to compute ELF and EPLF attractors and test the influence of the post-HF effects on the attractor positions. The energy decompositions on the Zn(II) complexes were done with the reduced variational space analysis (RVS) procedure<sup>26</sup> at the HF/CEP 4-31G(2d) level.

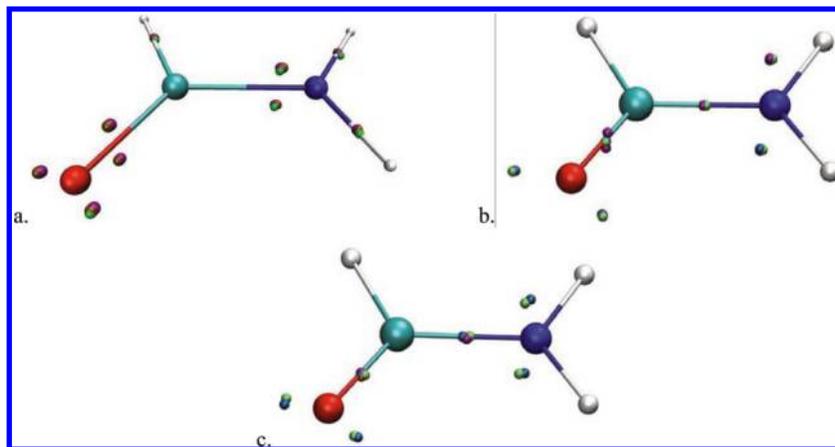
**Fig. 1.** ELF (grey) and EPLF (green) attractors and Boys (purple) centroid localizations for (a) water, (b) ammonia, (c) benzene, (d) formamide, (e) formate, (f) pyridine, (g) pyrimidine, (h) imidazole, (i) adenine, (j) thymine, (k) guanine, (l) cytosine, and (m and n) glucose.



**Fig. 2.** Positions of the (a) Boys centroids and (b) ELF attractors for different basis sets: CEP 4-31G (2d) (orange, only for ELF), 3-31G (purple), 6-31++G (blue), 6-31G\*\* (yellow), 3-31++G\*\* (green), 6-311G (pink), and 6-311++G\*\* (magenta). The hydrogen atoms appear in white, the carbons in cyan, the oxygens in red, and the nitrogens in blue.



**Fig. 3.** Positions of the (a) Boys centroids, (b) ELF, and (c) EPLF attractors for different calculation methods: HF (purple), BLYP (brown), B3LYP (green), and mPWP91 (clear blue), and also only for ELF and ELF MP2 (grey) and CCSD (dark blue). The hydrogen atoms appear in white, the carbons in cyan, the oxygens in red, and the nitrogens in blue.



## Results and discussion

Figure 1 represents the different molecules studied and their ELF and EPLF attractors in grey and green colors, respectively, as well as the centroids of the Boys localized orbitals shown in purple. These are water, ammonia, benzene, formamide, formate, pyridine, pyrimidine, imidazole, adenine, thymine, guanine, cytosine, and glucose. The coordinates of the molecules and of the centroids are given in the Supplementary data. We give separately in the Supplementary data the locations of the centroids currently used in SIBFA for several representative molecules: water, ammonia, formamide, formate, imidazole, and benzene. These are given with internal coordinates. For single bonds, the three methodologies give similar results and localize the extra point at the bond midpoint. Some differences arise, however, upon dealing with chemical bonds with a H atom. For these, both ELF and EPLF localize the attractor on the hydrogen atom while the Boys centroids are located along the bond, at distances about 0.4 Å from the hydrogen.

In the following section, delocalized electron doublets included in  $\pi$  systems will be denoted as  $\pi$  doublets to differentiate them from the “localized” doublets, such as lone pairs, that will be called  $\sigma$  doublets. As expected,<sup>27</sup> the three methodologies give similar results and localize the extra point at the bond midpoint. Some differences arise, however, upon dealing with chemical bonds with a H atom. For these, both ELF and EPLF localize the attractor on the hydrogen atom while the Boys centroids are located along the bond, at distances about 0.4 Å from the hydrogen.

In the following section, delocalized electron doublets included in  $\pi$  systems will be denoted as  $\pi$  doublets to differentiate them from the “localized” doublets, such as lone pairs, that will be called  $\sigma$  doublets. As expected,<sup>27</sup> the three methodologies give similar results and localize the extra point at the bond midpoint. Some differences arise, however, upon dealing with chemical bonds with a H atom. For these, both ELF and EPLF localize the attractor on the hydrogen atom while the Boys centroids are located along the bond, at distances about 0.4 Å from the hydrogen.

In the following section, delocalized electron doublets included in  $\pi$  systems will be denoted as  $\pi$  doublets to differentiate them from the “localized” doublets, such as lone pairs, that will be called  $\sigma$  doublets. As expected,<sup>27</sup> the three methodologies give similar results and localize the extra point at the bond midpoint. Some differences arise, however, upon dealing with chemical bonds with a H atom. For these, both ELF and EPLF localize the attractor on the hydrogen atom while the Boys centroids are located along the bond, at distances about 0.4 Å from the hydrogen.

There are more important differences between these methods with delocalized systems having double bonds or  $\pi$  doublets. For double bonds, Boys localization does not manage to conserve the planar symmetry of the system and will sometimes locate centroids out of the plane of the bond. In benzene, for example, using this procedure one C–C out of two has an in-plane centroid while the vicinal C–C bond has one centroid above and the other below the plane (see Fig. 1c). In the case of imidazole and formamide, ELF and EPLF locate the  $\pi$  doublets above and below the plane, close to the perpendicular to the N atom. The Boys procedure does not localize the  $\pi$  doublets on the N atoms but localizes them instead along the C–N bonds, on both sides of the plane, as for “banana bonds”. This highlights the shortcoming of the Boys procedure for delocalized systems because of the constraint that each LMO must have an integer number of electrons (1 or 2).

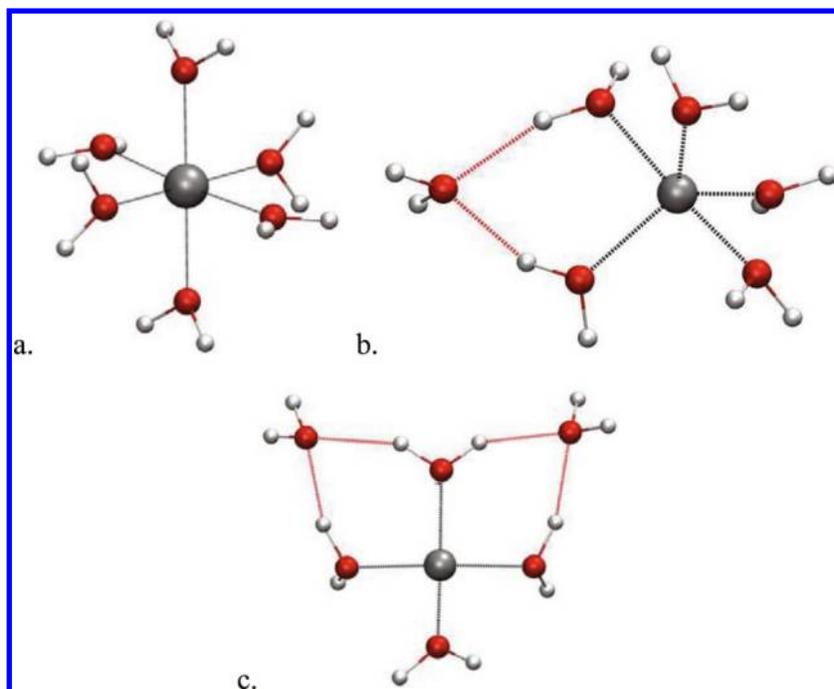
The localization procedure appears to be mostly independent of the level of theory used. One exception is found for formamide, as observed in Figs. 2 and 3, which give the localization of the extra points for the different methods. Thus, at the HF level ELF localizes two attractors along the C–O bond, instead of above and below the plane. This could be due to the absence of correlation.

## Molecular mechanics

### Methodology

In this section, we will consider a test set of  $\text{Zn}^{2+}$  mono- and poly-hydrated complexes, namely  $[\text{Zn}(\text{H}_2\text{O})^{2+}]$ ,  $[\text{Zn}(\text{H}_2\text{O})_6^{2+}]$ ,  $[\text{Zn}(\text{H}_2\text{O})_{5/1}^{2+}]$ , and  $[\text{Zn}(\text{H}_2\text{O})_{4/2}^{2+}]$ . We will evaluate the impact of using coordinates for the O lone pairs derived from the ELF, EPLF, or Boys localization procedures instead of those derived from their standard internal variables in SIBFA or Gaussian electrostatic model (GEM)/SIBFA.<sup>28</sup> The  $[\text{Zn}(\text{H}_2\text{O})_6^{2+}]$ ,  $[\text{Zn}(\text{H}_2\text{O})_{5/1}^{2+}]$ , and

**Fig. 4.** Representation of the (a)  $[\text{Zn}(\text{H}_2\text{O})_6]^{2+}$ , (b)  $[\text{Zn}(\text{H}_2\text{O})_{5/1}]^{2+}$ , and (c)  $[\text{Zn}(\text{H}_2\text{O})_{4/2}]^{2+}$  complexes. The hydrogen atoms appear in white, the carbons in cyan, the oxygens in red, and the zinc ion is in grey.



$[\text{Zn}(\text{H}_2\text{O})_{4/2}]^{2+}$  complexes have, respectively, six, five, and four water molecules in the first hydration shell and zero, one, and two in the second hydration shell (Fig. 4).

To do so, we will use the GEM/SIBFA framework<sup>28</sup> that combines the SIBFA water model<sup>14</sup> to a Gaussian electrostatic model (GEM)<sup>15,28</sup> density representation of the metal cation. The SIBFA molecular mechanics procedure aims at reproducing each separate contribution of the QC interaction energy, namely Coulomb ( $E_c$ ), exchange repulsion ( $E_{\text{exch}}$ ), polarization ( $E_{\text{pol}}$ ), and charge transfer ( $E_{\text{ct}}$ ) at the HF level and correlation/dispersion at the post-HF level. In SIBFA,  $E_{\text{rep}}$  and  $E_{\text{ct}}$  are computed as approximations of the overlaps between the localized orbitals of the interacting molecules. It uses off-centered points to represent the centroids of the localized lone pair hybrids. Therefore, using alternative values for the internal coordinates than the default ones in current use can impact the magnitudes of these two contributions. It is noted that the off-centered points are not fictitious atoms with partial atomic charges and (or) van der Waals radii, as done in some variants of classical force fields<sup>29</sup> and as was the case in the earliest version of SIBFA.<sup>30</sup> With the lone-pair positions computed using the coded internal coordinates,  $E_{\text{rep}}$  and  $E_{\text{ct}}$  are calibrated by fitting the van der Waals radii of the bearer atoms to reproduce the radial dependencies of the QC  $E_{\text{exch}}$  and  $E_{\text{ct}}$ , respectively, using the RVS analyses. Resorting to QC-derived internal coordinates for the lone-pair centroids, reflecting their “true” values within the molecular structure of the ligand rather than the standard values, could accordingly deliver more reliable values for the van der Waals radii.

SIBFA was recently interfaced<sup>28</sup> with the GEM procedure resulting in a further increase of accuracy upon computing cation-ligand complexes. We have accordingly resorted to the GEM/SIBFA methodology<sup>28</sup> in the present study. Presently, the Zn(II) cation is represented with its frozen density expressed in terms of Gaussian Hermite<sup>31</sup> while the ligand is represented with its distributed multipoles. The O van der Waals radii were recalibrated to be consistent with the set of internal lone-pair coordinates given by the different localization procedures. This was done by fitting the van der Waals radii of O to calibrate  $E_{\text{rep}}$  and  $E_{\text{ct}}$  on the

**Table 1.** Exchange-repulsion and charge-transfer energies for ab initio, SIBFA, and modified SIBFA versions so that the lone pair position correspond to Boys, ELF, or EPLF in (a)  $[\text{Zn}(\text{H}_2\text{O})_6]^{2+}$ , (b)  $[\text{Zn}(\text{H}_2\text{O})_{5/1}]^{2+}$  and (c)  $[\text{Zn}(\text{H}_2\text{O})_{4/2}]^{2+}$  complexes. The differences between ab initio and classical results are also computed.

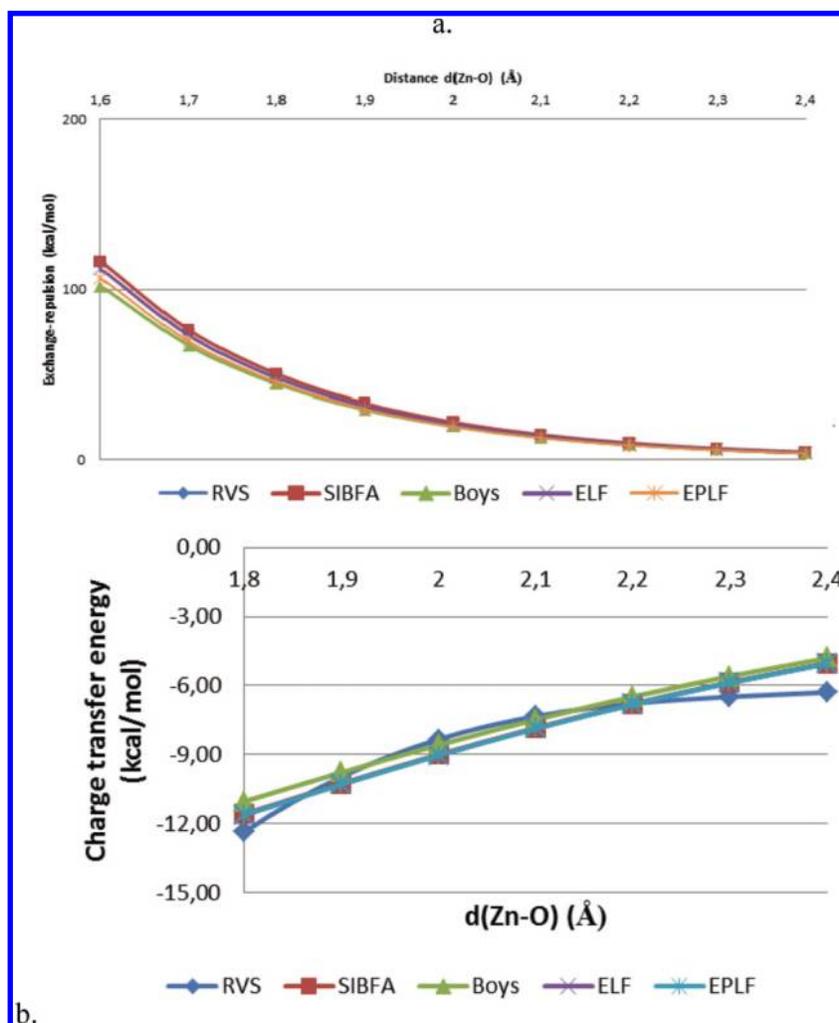
	$E_{\text{rep}}$	$\Delta E_{\text{rep}}$	$E_{\text{CT}}$	$\Delta E_{\text{CT}}$
<b>(a) <math>[\text{Zn}(\text{H}_2\text{O})_6]^{2+}</math></b>				
Ab initio	114.9	0.0	-21.3	0.0
SIBFA/current localization	120.2	5.3	-21.8	-0.5
SIBFA/Boys localization	114.4	-0.6	-20.6	0.8
SIBFA/ELF localization	116.1	1.1	-21.4	0.0
SIBFA/EPLF localization	116.3	1.4	-20.8	0.5
<b>(b) <math>[\text{Zn}(\text{H}_2\text{O})_{5/1}]^{2+}</math></b>				
Ab initio	125.4	0.0	-24.1	0.0
SIBFA/current localization	128.2	2.8	-24.6	-0.6
SIBFA/Boys localization	124.5	-0.9	-23.4	0.7
SIBFA/ELF localization	126.4	1.0	-24.1	0.0
SIBFA/EPLF localization	125.4	0.0	-23.6	0.5
<b>(c) <math>[\text{Zn}(\text{H}_2\text{O})_{4/2}]^{2+}</math></b>				
Ab initio	137.0	0.0	-26.8	0.0
SIBFA/current localization	136.8	-0.2	-27.4	-0.6
SIBFA/Boys localization	135.5	-1.5	-26.0	0.8
SIBFA/ELF localization	137.8	0.8	-26.8	-0.1
SIBFA/EPLF localization	135.7	-1.3	-26.2	0.6

Zn–H<sub>2</sub>O complexes to match the radial dependencies. They are given in Table 1.

## Results and discussion

For the mono-ligated Zn(II)–H<sub>2</sub>O complex, Fig. 5 displays the radial dependencies of  $E_{\text{rep}}$  and  $E_{\text{ct}}$  with the four sets of internal lone-pair coordinates as compared to their QC counterparts. Table 2 gives their values in the  $[\text{Zn}(\text{H}_2\text{O})_6]^{2+}$ ,  $[\text{Zn}(\text{H}_2\text{O})_{5/1}]^{2+}$ , and  $[\text{Zn}(\text{H}_2\text{O})_{4/2}]^{2+}$  hexahydrated Zn(II) complexes and the corresponding QC values. These tests constitute the first calculations aiming to evaluate the impact of three different QC-localization procedures on the ability of the short-range contributions,  $E_{\text{rep}}$  and  $E_{\text{ct}}$ ,

**Fig. 5.** Evolution of the (a) exchange-repulsion and (b) charge-transfer energies (in kcal/mol) as a function of the distance between the Zn(II) cation and the oxygen atom,  $d(\text{Zn-O})$ .



**Table 2.** Optimized van der Waals radii for the oxygen of the SIBFA water molecule for the original oxygen lone pair position and for the ab initio placed one (Boys, ELF, and EPLF).

Localization method	O van der Waals radius (Å)
SIBFA	1.448
Boys	1.474
ELF	1.412
EPLF	1.375

to reproduce  $E_{\text{exch}}$  and  $E_{\text{ct}}$  (QC) in the context of their GEM/SIBFA formulations. They do not as such attempt to compare the relative inherent merits of the localization procedures. Some conclusions on the validity of the approach can nevertheless be drawn. The three different methods give results differing by  $<1.5$  kcal/mol in absolute values from the QC ones regarding repulsion and by  $<0.8$  kcal/mol regarding charge transfer. They bring about a significant improvement (by up to 5.3 kcal/mol out of 115) with respect to the “standard” lone pairs. The ELF lone-pair localization procedure is the one enabling the closest agreement to QC repulsion and charge transfer. One possible reason for the more uniform match of  $E_{\text{ct}}$ (SIBFA) to its QC counterpart than that of  $E_{\text{rep}}$ ,

even though both contributions are overlap-dependent, is the formulation of  $E_{\text{rep}}$  under the  $S^2/R + S^2/R^2$  form, where  $S$  is an approximation of the intermolecular interorbital overlap, and  $R$  is the distance between the centroids of the interacting overlaps. The presence of  $R$  and  $R^2$  in the denominators could confer extra sensitivity to the precise location of the centroids in the case of  $E_{\text{rep}}$ .

Beyond water, we tested this approach with other Zn(II)-ligands displayed in Fig. 1. As mentioned in the Quantum chemistry section, only few differences were visible between the different localization procedures. Therefore, their possible impact on the actual  $E_{\text{rep}}$  and  $E_{\text{ct}}$  values can be recovered by the calibration procedure (by triggering the effective radii of the atoms concerned).

## Conclusions and perspectives

The use of off-centered points is necessary to confer anisotropy to the overlap-dependent contributions of next-generation force fields.<sup>4</sup> Their exact locations can be derived upon resorting to non-empirical QC localization procedures, such as the Boys procedure or the ELF and EPLF topological analyses. These three approaches were applied to thirteen molecules encountered in biological systems and in supramolecular chemistry. ELF and EPLF gave similar outcomes. The Boys procedure had a different outcome upon handling delocalized  $\pi$  systems since it broke their symmetry localizing some bond extra points perpendicular to the conjugated system plan. Thus, for the representative case of benzene, one C–C out of two has an in-

plane centroid while the vicinal C–C bond has two centroids, one above and one below the plane. Since they resort to electron densities, the ELF and EPLF approaches can either be computed at HF, DFT, or post-HF levels. On the other hand, the Boys analysis is restricted to HF and DFT since it requires an access to the correlated orbitals that cannot be computed at the post-Hartree–Fock level. Within a given localization scheme, the location of the off-centered points was found to have little sensitivity to the size of the basis set and the level of theory.

The outputs from these procedures can be directly used to locate the nonatomic centers within force field models. We have compared the effect of the different localization procedures of the lone pairs on the GEM/SIBFA overlap-dependent contributions, exchange repulsion ( $E_{\text{rep}}$ ), and charge transfer ( $E_{\text{ct}}$ ). All three procedures enabled a closer match of these contributions to their QC counterparts (<1.5 kcal/mol out of 115 for  $E_{\text{rep}}$  and <0.8 kcal/mol out of 25 for  $E_{\text{ct}}$ ) than upon using the standard lone-pair coordinates currently used in GEM/SIBFA, for which the error could amount to up to 5 kcal/mol out of 115 in the case of  $E_{\text{rep}}$ . The ELF appeared to enable slightly improved matches to QC than the other methods. For delocalized systems, the use of the Boys procedure may be prevented or would need additional “post-Boys” regularization procedures that could lead to ambiguous results. In such cases, the ELF and EPLF could be preferable, especially as they can be used beyond the HF or DFT levels that restrict the use of the Boys approach. In contrast, the EPLF can be used for any type of post-HF wave-function method and essentially gives results equivalent to the ELF. The present preliminary results show that QC localization procedures can be used to nonempirically derive accurate positions for off-centered points intervening in the formulation of the overlap-dependent contributions of next-generation force fields. The lone-pair coordinates derived as an outcome from such procedures could be more reliable than the standard ones currently allotted to them by the GEM/SIBFA program as a function of the chemical nature of the bearer atom based on its atomic number and hybridization state. This could be particularly important when differences between atoms of the same nature arise because of the electronic structure of the fragment to which they belong. Examples are  $\text{N}_3$  and  $\text{N}_7$  atoms in guanine versus adenine, and versus the sp-like N in imidazole, or the carbonyl group in formamide as compared to each of the nucleic acid bases, etc. Such differences could also a fortiori arise between atoms of the same atomic number and hybridization state in the same molecule, such as  $\text{N}_3$  versus  $\text{N}_7$  in the purine bases or  $\text{O}_2$  versus  $\text{O}_4$  in thymine or uracil. An additional variable that can be derived is the electronic population on the doublets, which is presently two in  $\text{sp}^n$  hybrids and one in pure  $\pi$  hybrids. Finally, the information from QC localization schemes given as input to SIBFA, upon further improving the accuracy of  $E_{\text{rep}}$  and  $E_{\text{ct}}$ , could enable this procedure to better emulate both GEM and QC calculations of these contributions. It is presently coupled to automatic parametrization methods, thus further reducing the time required to develop an all-purpose methodology for peptide and protein<sup>32</sup> including correlation effects<sup>33</sup> and DNA<sup>34</sup> simulations.

### Supplementary data

Supplementary data are available with the article through the journal Web site at <http://nrcresearchpress.com/doi/suppl/10.1139/cjc-2012-0547>.

### Acknowledgements

Support from the French National Research Agency (ANR) is acknowledged (grant No. ANR-08-BLAN-0158), as well as computer time allowances by the Conseil Régional de Haute-Normandie (CRIHAN; Rouen, France, project No. 2008011) and the Grand Equipement National de Calcul Intensif (GENCI; Institut du développement et des ressources en informatique scientifique (IDRIS), Centre informatique de l'enseignement supérieur (CINES), France,

project No. x2010075027). Financial support from the Wayne State University to G.A.C. is also gratefully acknowledged.

### References

- Case, D. A.; Cheatham, T. E.; Darden, T. A.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668. doi:10.1002/jcc.20290.
- Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caffisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545. doi:10.1002/jcc.21287.
- Oostenbrink, C.; Villa, A.; Mark, A. E.; VanGunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656. doi:10.1002/jcc.20090.
- Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J. P. *J. Chem. Theory Comput.* **2007**, *3*, 1960. doi:10.1021/ct700134r.
- Becke, A. D.; Edgecombe, K. E. *J. Chem. Phys.* **1990**, *92*, 5397. doi:10.1063/1.458517.
- Silvi, B.; Savin, A. *Nature* **1994**, *371*, 683. doi:10.1038/371683a0.
- Piquemal, J. P.; Pilmé, J.; Parisel, O.; Gérard, H.; Fourré, I.; Bergès, J.; Gourlaouen, C.; De La Lande, A.; Van Severen, M. C.; Silvi, B. *Int. J. Quantum Chem.* **2008**, *108*, 1951. doi:10.1002/qua.21711.
- Scemama, A.; Chaquin, P.; Caffarel, M. *J. Chem. Phys.* **2004**, *121*, 1725. doi:10.1063/1.1765098.
- Scemama, A.; Caffarel, M.; Chaudret, R.; Piquemal, J. P. *J. Chem. Theory Comput.* **2011**, *7*, 618. doi:10.1021/ct1005938.
- Foster, J. M.; Boys, S. F. *Rev. Mod. Phys.* **1960**, *32*, 300. doi:10.1103/RevModPhys.32.300.
- Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; Oxford University Press, Oxford, 1990.
- Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926. doi:10.1063/1.445869.
- Mahoney, M.; Jorgensen, W. L. *J. Chem. Phys.* **2000**, *112*, 8910. doi:10.1063/1.481505.
- (a) Piquemal, J. P.; Chevreau, H.; Gresh, N. *J. Chem. Theory Comput.* **2007**, *3*, 824. doi:10.1021/ct7000182; (b) Piquemal, J. P.; Chelli, R.; Procacci, P.; Gresh, N. *J. Phys. Chem. A* **2007**, *111*, 8170. doi:10.1021/jp072687g.
- Piquemal, J. P.; Cisneros, G. A.; Reinhardt, P.; Gresh, N.; Darden, T. A. *J. Chem. Phys.* **2006**, *124*, 104101. doi:10.1063/1.2173256.
- Kumar, R.; Wang, F. F.; Jenness, G.; Jordan, K. D. *J. Chem. Phys.* **2010**, *132*, 014309. doi:10.1063/1.3276460.
- (a) Gresh, N.; Claverie, P.; Pullman, A. *Int. J. Quantum Chem.* **1982**, *22*, 199. doi:10.1002/qua.560220118; (b) Gresh, N.; Claverie, P.; Pullman, A. *Int. J. Quantum Chem.* **1986**, *29*, 101. doi:10.1002/qua.560290110.
- Gresh, N. *J. Comput. Chem.* **1995**, *16*, 856. doi:10.1002/jcc.540160705.
- Gresh, N.; Piquemal, J. P.; Krauss, M. *J. Comput. Chem.* **2005**, *26*, 1113. doi:10.1002/jcc.20244.
- (a) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648. doi:10.1063/1.464913; (b) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785. doi:10.1103/PhysRevB.37.785.
- Noury, S.; Krokidis, X.; Fuster, F.; Silvi, B. *J. Comput. Chem.* **1999**, *23*, 597. doi:10.1016/S0097-8485(99)00039-X.
- Hehre, W. J.; Stewart, R. F.; Pople, J. A. *J. Chem. Phys.* **1969**, *51*, 2657. doi:10.1063/1.1672392.
- Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650. doi:10.1063/1.438955.
- Stevens, W. J.; Basch, H.; Krauss, M. *J. Chem. Phys.* **1984**, *81*, 6026. doi:10.1063/1.447604.
- Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992**, *45*, 13244. doi:10.1103/PhysRevB.45.13244.
- Stevens, W. J.; Fink, W. H. *Chem. Phys. Letts.* **1987**, *139*, 15. doi:10.1016/0009-2614(87)80143-4.
- Savin, A. *J. Chem. Sci.* **2005**, *117*, 473. doi:10.1007/BF02708351.
- (a) Chaudret, R.; Ulmer, S.; van Severen, M. C.; Gresh, N.; Parisel, O.; Cisneros, G. A.; Darden, T. A.; Piquemal, J. P. *J. P. AIP Conf. Proc.*, **2009**; *1102*, 185. doi:10.1063/1.3108373; (b) Chaudret, R.; Gresh, N.; Darden, T. A.; Cisneros, G. A.; Piquemal, J. P. *J. Chem. Phys.* **2013**, revision submitted.
- Kolar, M.; Hobza, P. *J. Chem. Theory Comput.* **2012**, *8*, 1325. doi:10.1021/ct2008389.
- Gresh, N.; Claverie, P.; Pullman, A. *Theoret. Chem. Acc.* **1984**, *66*, 1. doi:10.1007/BF00577135.
- (a) Cisneros, G. A.; Piquemal, J. P.; Darden, T. A. *J. Chem. Phys.* **2006**, *125*, 184101. doi:10.1063/1.2363374; (b) Cisneros, G. A.; Elking, D.; Piquemal, J. P.; Darden, T. A. *J. Phys. Chem. A* **2007**, *111*, 12049. doi:10.1021/jp074817r.
- Gresh, N.; Kafafi, S. A.; Truchon, J. F.; Salahub, D. R. *J. Comput. Chem.* **2004**, *25*, 823. doi:10.1002/jcc.20012.
- Cloaguen, E. de; Courcy, B.; Piquemal, J.-P.; Pilmé, J.; Parisel, O.; Pollet, R.; Biswal, H. S.; Piuze, F.; Tardivel, M.; Broquier, M.; Mons, M. *J. Am. Chem. Soc.* **2010**, *132*, 11860. doi:10.1021/ja103996q.
- Gresh, N.; Sponer, J. E.; Spackova, N.; Leszczynski, J.; Sponer, J. *J. Phys. Chem. B* **2003**, *107*, 8669. doi:10.1021/jp022659s.