



**HAL**  
open science

## Accurately extracting the signature of intermolecular interactions present in the NCI plot of the reduced density gradient versus electron density

Corentin Lefebvre, Gaëtan Rubez, Hassan Khartabil, Jean-Charles Boisson, Julia Contreras-García, Eric Henon

### ► To cite this version:

Corentin Lefebvre, Gaëtan Rubez, Hassan Khartabil, Jean-Charles Boisson, Julia Contreras-García, et al.. Accurately extracting the signature of intermolecular interactions present in the NCI plot of the reduced density gradient versus electron density. *Physical Chemistry Chemical Physics*, 2017, 19 (27), pp.17928-17936. 10.1039/C7CP02110K . hal-02505160

**HAL Id: hal-02505160**

**<https://hal.univ-reims.fr/hal-02505160v1>**

Submitted on 14 Oct 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Accurately extracting the signature of intermolecular interactions present in the NCI plot of the reduced density gradient versus electron density †

Corentin Lefebvre,<sup>a</sup> Gaëtan Rubez,<sup>a,b,c</sup> Hassan Khartabil,<sup>a,d</sup> Jean-Charles Boisson,<sup>b</sup> Julia Contreras-García,<sup>e,f</sup> and Eric Hénon<sup>\*a</sup>

An electron density (ED)-based methodology is developed for the automatic identification of intermolecular interactions using pro-molecular density. The expression of the ED gradient in terms of atomic components furnishes the basis for the Independent Gradient Model (IGM). This model leads to a density reference for non interacting atoms/fragments where the atomic densities are added whilst their interaction turns off. Founded on this ED reference function that features an exponential decay also in interference regions, IGM model provides a way to identify and quantify the net ED gradient attenuation due to interactions. Using an intra/inter uncoupling scheme, a descriptor ( $\delta g^{inter}$ ) is then derived that uniquely defines intermolecular interaction regions. An attractive feature of the IGM methodology is to provide a workflow that automatically generates data composed solely of intermolecular interactions for drawing the corresponding 3D isosurface representations.

## 1 Introduction

Non-covalent interactions (NCI) are responsible for many properties of condensed phases, including for instance the 3-dimensional arrangement that the biological polymers adopt (DNA double helix, proteins). They also play a key role in ligand-protein bio-molecular recognition in the field of drug-design. This class of interactions spans a wide variety of attraction and repulsion forces between atoms or molecules, including van der Waals (vdW) interactions and hydrogen-bonds. Beyond the standard existing topological analysis tools of the chemical bond (AIM<sup>1,2</sup> and ELF<sup>3</sup>), in 2010, Johnson et al. presented a new approach, the so-called NCI analysis.<sup>4</sup> Also based on the electron density (ED) topology, this method enables the identification and visualization of regions of weak interactions in the 3D real space by providing chemically intuitive iso-surfaces of the reduced density gradient,  $s$  (also referred to as RDG):

$$s = \frac{1}{2(3\pi^2)^{1/3}} \frac{|\nabla\rho(r)|}{(\rho(r))^{4/3}} \quad (1)$$

where  $\rho$  represents the ED and  $|\nabla\rho(r)|$  stands for the norm of the ED gradient vector. Initially introduced in DFT developments,  $s(\rho)$  has long been used to incorporate corrections to the homogeneous electron gas in the expression of exchange correlation functionals. The needed input data for this calculation is the electron density  $\rho$  and its gradient, collected at each point of a grid encompassing the two interacting molecules. In order to differentiate repulsive from attractive interactions the sign of the second eigenvalue  $\lambda_2$  of the ED Hessian matrix is required.<sup>4</sup>

In the absence of interactions,  $s(\rho)$  shows an overall  $a\rho^{-1/3}$  shape. In an isolated atom, far from the nucleus, both  $\rho^{4/3}$  (denominator) and its gradient (numerator) are very small. But because the former approaches zero more rapidly than the latter, their ratio grows exponentially in the atomic tail, so that it tends to infinity at very low density values. Similarly, in high-density regions, towards the nuclei,  $\rho^{4/3}$  dominates over the gradient and accordingly,  $s(\rho)$  steadily decreases. In molecular systems, the 2D plot of  $s(\rho)$  exhibits new features. Deviations from the exponential decay can be observed. Mapping them back to real space, these points highlight 3D regions associated with molecular interactions by chemists. Covalent bonds are identified by troughs of RDG at large densities while spikes at low densities reveal non-covalent interactions, such as vdW contacts or hydrogen-bonds.

Until the recent work of Boto et al.<sup>5</sup>, the reduced density gradient was perceived as a DFT parameter measuring the inho-

<sup>a</sup> Institut de Chimie Moléculaire de Reims, CNRS UMR 7312, University of Reims Champagne-Ardenne, BP 1039, 51687 Reims, France. Fax: 33 (0)326913166; Tel: 33 (0)326918497; E-mail: eric.henon@univ-reims.fr

<sup>b</sup> CReSTIC EA 3804, University of Reims Champagne-Ardenne, 51687 Reims, France.

<sup>c</sup> ATOS Company, 1 rue de Provence, 38130 Echirolles, France.

<sup>d</sup> Campus Universitaire des Ardennes, 4 bd Jean Delautre, 08000 Charleville-Mézières, France.

<sup>e</sup> Sorbonne Universités, UPMC Univ Paris 06, UMR CNRS 7616, Laboratoire de Chimie Théorique, Paris, France.

<sup>f</sup> CNRS, UMR 7616, Laboratoire de Chimie Théorique, Paris, France.

† Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/b000000x/

mogeneity of the system (a correction to uniform electron gas). R. Boto et. al. recently presented an interpretation of  $s(\rho)$  in terms of the bosonic kinetic energy density since  $s^2(\rho)$  is proportional to the ratio of two kinetic energies :  $\tau_w/\tau_{TF}$  ( $\tau_w$  is the von Weizsäcker kinetic energy density,  $\tau_{TF}$  is the homogeneous electron gas kinetic energy density). More precisely,  $\tau_w(\rho) = 1/8|\nabla\rho(r)|^2/\rho(r)$  is the kinetic energy density of a many-electron system behaving as bosons (without the constraint of the Pauli exclusion principle ).<sup>6</sup> It is a non homogeneous lower boundary for the total kinetic energy density scaled by  $\tau_{TF}(r) = C_F\rho(r)^{5/3}$ ,  $C_F = 3/10(3\pi^2)^{2/3}$  to turn off the first order electron density dependence of the kinetic energy density. The new interpretation given by Boto et al. makes an interesting connection between  $s(\rho)$  and chemical bonding through the kinetic energy density concept known to play a critical role upon bond formation.

The NCI methodology was implemented by one of the authors,<sup>7</sup> allowing to plot the reduced density gradient  $s(\rho)$  versus ED from either quantum mechanical ED or from pro-molecular densities. The latter densities are the sum of atomic densities, without the relaxation induced by the molecular potential.<sup>8,9</sup> This non-relaxed ED is calculated at a given point,  $(x, y, z)$ , by summing the spherically averaged neutral atomic densities  $\rho_i$  centered at the atomic positions  $(x_i, y_i, z_i)$ :  $\rho(x, y, z) = \sum_{i=1}^N \rho_i(r_i)$ . In the program NCIPLT-1.0, for the first 18 elements of the periodic table, the isolated atomic density  $\rho_i$  is obtained from a linear combination of simple exponential functions:  $\rho_i(r_i) = \sum_{j=1}^{n_s} a_{i,j} e^{-b_{i,j} r_i}$  with  $a_{i,j}$  and  $b_{i,j}$  (both having positive values) adjusted to closely fit spherically averaged *ab initio* ED for atom  $i$  for each atomic shell  $j$ . The ED  $x$  component of the gradient follows from this definition (similar equations apply to components  $y$  and  $z$ ):

$$\frac{\partial\rho}{\partial x} = \sum_{i=1}^N \frac{\partial\rho_i}{\partial x} = \sum_{i=1}^N \frac{x_i - x}{r_i} \sum_{j=1}^{n_s} a_{i,j} b_{i,j} e^{-b_{i,j} r_i} \quad (2)$$

where  $r_i$  refers to the radial distance from atom  $i$ . However,  $s(\rho)$  does not allow an *ab-initio* separation in intramolecular and intermolecular interactions, which is of crucial importance in the analysis of big systems. For instance, when studying the interactions between a ligand and a protein, it is obvious that the protein itself involves internal hydrogen-bonds and van der Waals interactions that mix with the ligand-protein interactions in the NCI plot. The initial NCI approach does not provide an "absolute" way to separate these contributions. In the original code, this problem is addressed by means of a density threshold value used to discard the grid nodes for which more than a fraction (default value is 0.95) of the total pro-molecular density comes from only one molecule. This limit is however arbitrary.

Another difficulty concerns the integration of quantities over NCI regions. Actually, recent applications of NCI tackle the problem of finding a relationship between NCI regions and energetic properties.<sup>10,11</sup> These approaches involve the integral of local quantities over the space representing the interactions. To identify this region, J. Contreras-García proposes a multiple-stage approach.<sup>10,12</sup> First, both the monomer and dimer NCI 2D plots are determined. Next, the lower edge of the monomer plot is splined. Then, only the points of the

dimer plot lying below this upper-limit are considered for the integration. However, this approach has a major drawback. In the general case, a heterodimer is considered leading to two reference splines. As a result, the border delimiting the NCI region in the plot of the interacting system is no longer well defined. Another approach to find the relationship between stabilization energy and ED topological features was developed by G. Saleh et al.<sup>11</sup> They performed the integration of the kinetic energy density over NCI grid points having a RDG value below a given threshold ( $s=0.5$ ). However, the choice of isosurface is arbitrary and ignores some grid points in the interaction region.

In this study, we propose a way to solve these issues for pro-molecular densities by accurately defining the intermolecular interaction region in the 2D NCI plot  $s(\rho)$ . Starting on a very simple diatomic case, we focus on the origin of the trough associated with the interaction in the NCI plot. This preliminary approach leads to a deeper understanding of the concept of interaction involved in the NCI methodology. The effect on several local indexes for intermolecular interactions is analysed. Consequently, we develop a new pro-molecular approach able to identify and isolate the interactions between molecules or, more generally, between user-defined fragments.

## 2 Methodological approach

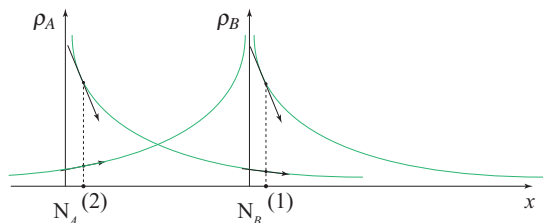
### 2.1 The Independent Gradient Model

The function  $s$  was initially chosen over other gradient functionals due to its better enhancement of non-covalent features (see S.I. in Ref. 4). The gradient collapse in the  $s(\rho)$  plots plays a key role in the success of identifying interactions. It is then important to look closer at the origin of this feature. The emergence of opposed sign in the sum of (eqn (2)) will determine or not the apparition of  $s(\rho)$  drops. However, talking about a "drop" implicitly involves that there is a reference clear to the human eye. In this contribution, we undertake the search of such reference.

In a molecular system the atoms represent electron density sources. In regions between two sources there is a crucial attenuation in the total gradient such that the numerator of  $s(\rho)$  suddenly reaches lower values than those found in the corresponding non-interacting system having the same electron density. And this is so even at the pro-molecular level. This non-interacting system corresponds to a virtual reference in which the individual terms involved in the expression of  $\nabla\rho$  are added up in absolute value (eqn (3)), yielding  $|\nabla\rho^{IGM}|$  as an upper limit to  $\nabla\rho$  with *IGM* standing for the Independent Gradient Model.

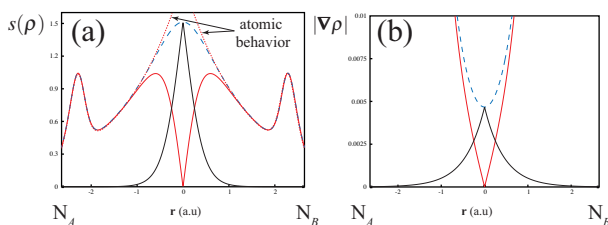
$$\left(\frac{\partial\rho}{\partial x}\right)^{IGM} = \sum_{i=1}^N \left|\frac{\partial\rho_i}{\partial x}\right| = \sum_{i=1}^N \frac{|x_i - x|}{r_i} \sum_{j=1}^{n_s} a_{i,j} b_{i,j} e^{-b_{i,j} r_i} \quad (3)$$

This is made apparent in Fig 1 for a molecular system made of two nitrogen atoms. While the two atomic contributions  $\frac{\partial\rho_A}{\partial x}$  and  $\frac{\partial\rho_B}{\partial x}$  add up at point (1) in (eqn (2)) for atoms A and B, respectively, they have opposite signs at point (2). A similar concept had been developed by Wilson and Goddard<sup>13</sup> for orbital pairs having opposite gradients leading to the principle of orbital contra-gradience needed for a chemical bond to be formed. When con-



**Fig. 1** schematic representation of pro-molecular atomic electron densities  $\rho_A$  and  $\rho_B$  for inter-nuclear regions and outside the atomic contact zone.

sidering the sum of all individual atomic gradients at one point, this interference phenomenon weakens the total gradient. This explains the fact that bond critical points appear already at the promolecular level, with quantum interference usually just modulating their final position and properties. The gradient collapse can be assessed with respect to the gradient vector  $\nabla\rho^{IGM}$  obtained without interference, i.e. by adding the absolute values of every atomic contribution at point 2 while considering the same total ED  $\rho$  in equation (3). The molecular system associated with the IGM model is thus virtual. Each atom preserves its initial distance  $r_i$  to the grid node during the calculation but the destructive interference of gradients is turned off.



**Fig. 2** (a)  $s^{IGM}(\rho)$  (dashed blue line),  $s(\rho)$  (solid red line, atomic behavior with dotted red line), and deviation  $\delta_s$  between them (black line) (b)  $|\nabla\rho^{IGM}|$  (blue dashed),  $|\nabla\rho|$  (red), and corresponding difference  $|\nabla\rho^{IGM}| - |\nabla\rho|$  (black). Properties calculated along the inter-nuclear axis for  $N_2$  with the zero set at the BCP.

Thus, we can define a new index referred to as  $s^{IGM}$  based on  $|\nabla\rho^{IGM}|$ , the norm of the ED gradient calculated within the Independent Gradient Model. Fig 2 illustrates this new function (in blue color, dashed line) for the nitrogen system. Since the interference between  $N_A$  and  $N_B$  is disabled in this independent gradient model,  $s^{IGM}$  is always greater or equal to  $s$ , the largest deviation occurring at the bond critical point (BCP). In the atomic region both functions are superposed whereas a deviation is observed in the contact zone. An accurate definition of the interaction region results when plotting the difference,  $\delta_s$  (black line on Fig 2):

$$\delta_s = s^{IGM} - s \quad (4)$$

By plotting the numerators of  $s^{IGM}$  and  $s$  (right panel of Fig 2), although both decrease when moving towards the BCP,  $|\nabla\rho|$  decays faster than  $|\nabla\rho^{IGM}|$ . This is the reason why  $\delta_s$  and  $\delta g = |\nabla\rho^{IGM}| - |\nabla\rho|$  functions have the same shape. It emerges from this short analysis that troughs in the 2D plot of  $s$  are caused at the pro-molecular level by ED gradient collapses that can also

be detected and quantified by computing the new descriptor  $\delta g$ :

$$\delta g = |\nabla\rho^{IGM}| - |\nabla\rho| \quad (5)$$

$\delta g$  relies on the independent gradient model which provides an upper limit corresponding to a situation where the gradient interference is always constructive. In the following, for the sake of simplicity, this reference will be referred to as "non-interacting" reference. This ensures that non-zero values of  $\delta g$  exclusively correspond to molecular interaction situations (intra- or intermolecular). Within this approach, the presence or not of opposite signs terms in the components expression giving  $|\nabla\rho|$  serve to define the nature of the region. Thus, points where all terms forming component  $\partial\rho/\partial x$  have the same sign are outside the interaction region. In contrast, points inside the interaction region give rise to partial or total cancellation in  $|\nabla\rho|$ . This cancellation will depend on the relative position of atoms (their densities) and their chemical hardness (exponential exponents).

A deeper understanding of the physical meaning of  $\delta g$  can be achieved from a simple exponential model. Assuming two interacting atoms  $A$  and  $B$  (an extension of the situation described in Fig 1), whose tails can be described by an exponential with characteristic exponents  $\alpha$  and  $\beta$ , respectively, their densities,  $\rho_i(x)$  along the bonding direction,  $x$ , will be given by:

$$\rho_A = e^{-\alpha(x-x_A)} \quad (6)$$

$$\rho_B = e^{-\beta(x_B-x)} \quad (7)$$

where we have assumed that atom  $A$  is at  $x_A$  and atom  $B$  at  $x_B$  ( $x_A < x_B$ ) and only the bonding direction is taken into consideration for the sake of clarity. Hence, the density gradient is given by  $\nabla\rho = -\alpha\rho_A + \beta\rho_B$ , so that the position of the bond critical point is given by:  $\alpha\rho_A = \beta\rho_B$ . The analytic expression for  $\delta g$  depends on the location of the point studied:

- before the BCP (i.e. within  $A$ 's Bader basin):  $\alpha\rho_A > \beta\rho_B$  and  $\delta g = 2\beta\rho_B = 2\nabla\rho_B$
- after the BCP (i.e. within  $B$ 's Bader basin):  $\alpha\rho_A < \beta\rho_B$  and  $\delta g = 2\alpha\rho_A = 2\nabla\rho_A$

In other words, comparing density gradients with and without interference gives rise to a quantity,  $\delta g$ , which highlights in each Bader atom the penetration of the density from the neighboring one.

## 2.2 The descriptor's dimension

The starting point and rationale for this descriptor is the idea that interaction zones are regions of space where atomic electron densities interfere.<sup>14,15</sup> From what we have seen in Section 2.1, these regions can be defined comparing the sign of the different pro-molecular atomic ED gradient components which contribute to the resulting gradient  $|\nabla\rho|$ .

Several descriptors have been introduced in the literature in which the atomic overlapping region is detected thanks to the deviations of the ED from the single-exponential behavior normally observed in monoatomic regions (e.g. NCI peaks).<sup>4,16,17</sup>

This ability to detect the interacting region descends from the  $\nabla\rho$  term, although different dimensions are obtained upon division by  $\rho^n$  (e.g.  $n=0$  in  $\nabla\rho$ ;  $n=1$  in LED<sup>16</sup>;  $n=4/3$  in NCI<sup>4</sup>).

The exponent  $n$  determines the features obtained in the analysis. In NCI, the Thomas-Fermi scaling gets rid of the electron density dependence, which is convenient to observe all interactions on the same grounds. Thus, a very low value of  $s$  will enable to visualize all interactions around the BCPs. However, it makes the  $s(\rho)$  absolute value not characteristic of the molecular regions. Indeed, the  $s$  value will not enable to distinguish strong from weak interactions. Moreover, if high density cutoffs are used, undesirable atomic features will also appear<sup>18</sup>. Hence, within the NCI framework, a preliminary visual inspection of the 2D  $s(\rho)$  plot is necessary to check whether a point deviates from the exponential reference behavior.

Other non-dimensionless quantities have been proposed, e.g. LED,<sup>16,19</sup> which also identify deviation from exponential behavior. These provide dimension dependent quantities, which thus depend on the electron density. Hence, on the one hand they give very different magnitudes in the core and the valence, but on the other hand, this dependence enables to identify the region. The  $\delta g$  descriptor belongs to this family.

Along this contribution we will mainly focus on  $s$ ,  $g$  and  $\tau_w$  (introduced above). In the latter case, we define  $\delta k$  as

$$\delta k = |\tau_w^{IGM}| - |\tau_w| = \frac{1}{8} \frac{(\nabla\rho^{IGM})^2 - (\nabla\rho)^2}{\rho} \quad (8)$$

### 2.3 Automatic detection of intermolecular interactions

Calculating the  $\delta g$  gradient-based descriptor for two interacting species A and B will detect and isolate all gradient interferences irrespective of whether they are associated with intramolecular (or even interatomic, in between different shells) or intermolecular interactions. However, the independent gradient model can be adapted in such a way that only the interactions between the two molecules cancel while intramolecular gradient features are preserved by splitting the sum over atoms in (eqn (3)) in two intramolecular building-blocks:

$$\left(\frac{\partial\rho}{\partial x}\right)^{IGM,inter} = \left|\sum_{i=1}^{N_A} \frac{\partial\rho_i}{\partial x}\right| + \left|\sum_{i=1}^{N_B} \frac{\partial\rho_i}{\partial x}\right| \quad (9)$$

Again, the associated molecular system is virtual. Each interacting partner preserves its initial geometry and atoms interact within each molecule but no grid node lie between these two virtual fragments. In this "IGM,inter" model, only the intermolecular interactions are canceled. Therefore, uncoupling the intra- and intermolecular contributions leads to:

$$\delta g^{inter} = |\nabla\rho^{IGM,inter}| - |\nabla\rho| \quad (10)$$

$$\delta s^{inter} = s^{IGM,inter} - s \quad (11)$$

$$\delta k^{inter} = k^{IGM,inter} - k \quad (12)$$

so that

$$\delta g^{intra} = |\nabla\rho^{IGM}| - |\nabla\rho^{IGM,inter}| \quad (13)$$

$$\delta s^{intra} = s^{IGM} - s^{IGM,inter} \quad (14)$$

$$\delta k^{intra} = k^{IGM} - k^{IGM,inter} \quad (15)$$

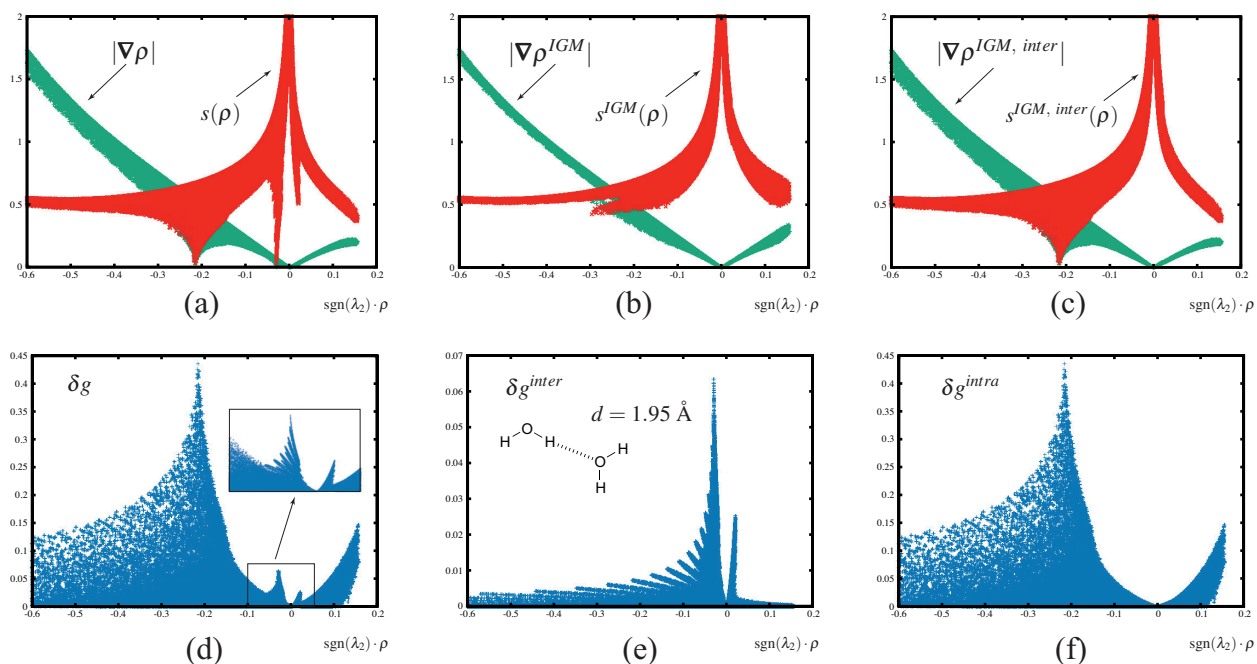
with  $\delta g = \delta g^{intra} + \delta g^{inter}$  and similar for  $s$  and  $k$ .

All computations were carried out with the NCIgpu code for pro-molecular ED.<sup>20</sup> Full geometry optimizations were performed at the B3LYP/6-311++G\*\* level of theory in gas phase using the Gaussian 09 package.<sup>21</sup> All minima were carefully characterized through harmonic frequency analysis (stationary points with no imaginary frequencies).

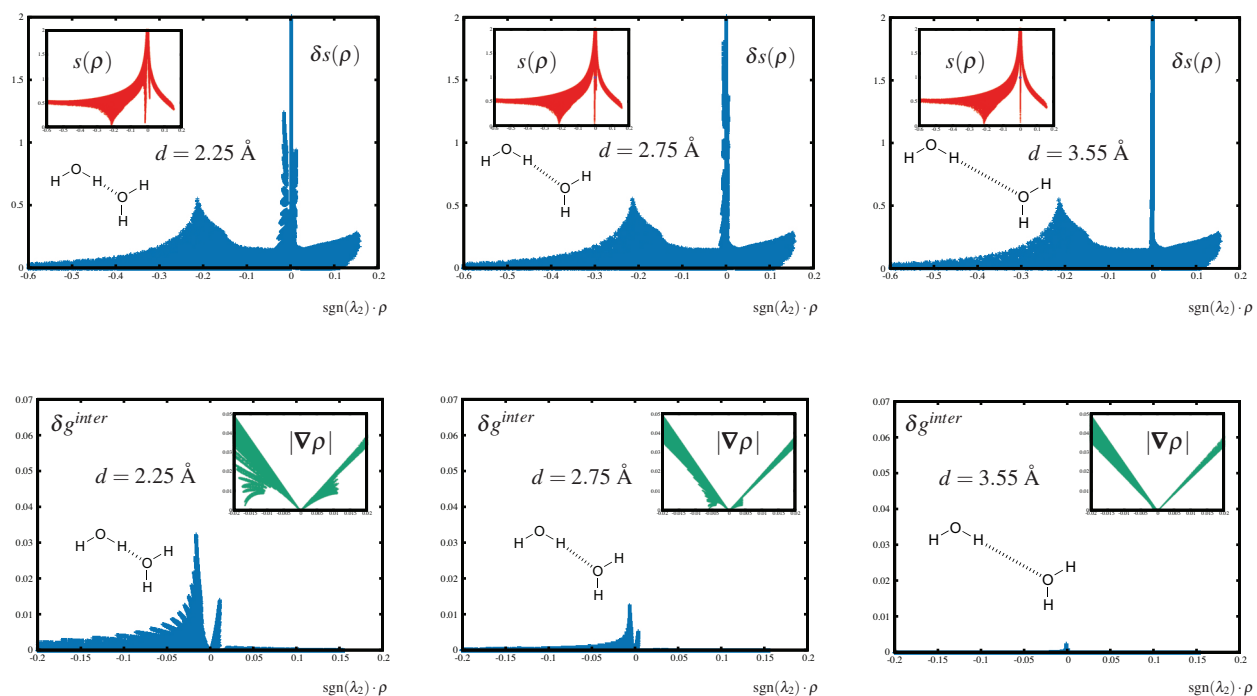
## 3 Results and discussion

To explore the possibilities provided by the various descriptors, we first analyze the water dimer. As can be seen in Fig 3a, superposition of  $s(\rho)$  and  $|\nabla\rho|$  on the same graph clearly shows that the gradient carries the information on the interaction, the division by  $\rho^n$  ( $n \geq 1$ ) amplifying the signal in NCI and LED. Compared to the standard NCI plot, using the  $|\nabla\rho|^{IGM}$  gradient norm instead of  $|\nabla\rho|$  in the calculation of  $s(\rho)$ , but using the same value of  $\rho$ , leads to a new NCI plot ( $s^{IGM}(\rho)$ , panel (b)) without any spike: this is our interaction-less reference. This 2D plot has the overall shape characteristic of a system without any interaction in the NCI index representation, nor covalent interactions neither weak interactions. This supports the idea that, given a grid node, taking the real ED but considering the independent gradient model provides suitable gradient reference for all kind of situations in the NCI plot representation within the pro-molecular approximation: a non-interacting reference which moreover enables to identify intermolecular or intramolecular cases. Furthermore, the uncoupling scheme we propose in order to specifically remove the intermolecular troughs in  $s(\rho)$  delivers the expected result in Fig 3c where the sharp spike associated with the hydrogen-bond in the standard NCI plot of the water dimer is now eliminated whereas the intramolecular bond remains.

Let's now compare the results of the IGM reference on the different descriptors. The  $\delta g$  descriptor is plotted as a function of the signed ED ( $sign(\lambda_2)\rho$ ) in panel (d) of Fig 3. It displays a broad spike in the high-density, high- $\delta g$  region associated with covalent bonds whilst the hydrogen-bond is revealed at low-density, low- $\delta g$  values by two small and sharp spikes known to correspond to attractive and repulsive contributions on the  $s(\rho)$  plot. This makes sense since larger gradient interferences are expected in covalent regions compared to non-covalent ones. Due to the dimension of  $\nabla\rho$  analyzed in the previous section,  $\delta g$  has different values at the various BCPs, hence the magnitude of  $\delta g$  will be larger the more  $|\nabla\rho|$  collapses (i.e. in stronger bonds). This means that the value of  $\delta g$  itself can be used to distinguish bond types, just like in the case of LED. Moreover, thanks to the IGM,inter splitting scheme, an automatic detection of inter- and intramolecular interactions can be processed (panels (e) and (f), respectively). This approach has its drawback, since the pro-molecular approach fails for covalent bonds. However, as a first order approximation, it enables to



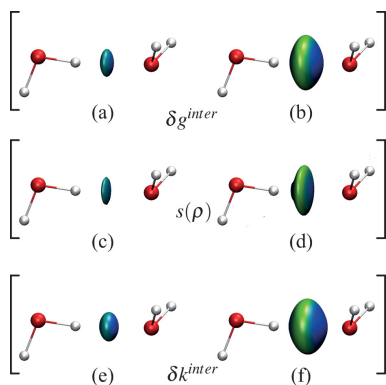
**Fig. 3** RDG (red,  $s(\rho)$  and  $s^{IGM}(\rho)$ ) are computed using the same value of  $\rho$ , ED gradient norm (green),  $\delta s(\rho)$  and  $\delta g(\rho)$  (blue) for water dimer; (a) and (d): all interactions are preserved, (b) all interactions cancel, (c) and (f): only the interactions between the two molecules are canceled, (e) intermolecular interactions.



**Fig. 4** Evolution of  $(s, \delta s)$  or  $(\nabla \rho, \delta g)$  along the reaction path of the water dimer dissociation.

distinguish intra from intermolecular bonds in big systems, which is the main aim of this work. We will dwell on how to adapt this model to relaxed densities in future contributions.

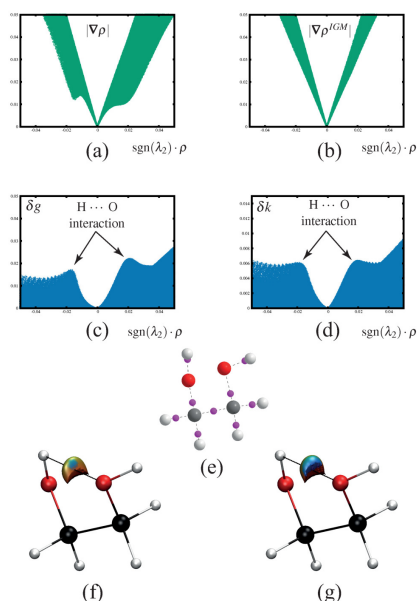
In Fig 4, we examine the behavior of the  $s$  function subjected to the IGM model reference as we stretch the hydrogen bond. A well-delimited covalent region is observed in the resulting  $\delta s$  function. At low density, a peak associated with the hydrogen-bond is observed. But in addition, an extremely sharp and very large spike invariably appears near  $\rho = 0$ , irrespective of the dimer separation distance. This undesired feature results from the mathematical form of  $\delta s$  and is explained in more details in ESI. In contrast, the  $\delta g^{inter}$  representation results in well delimited spikes at low ED regions (bottom of Fig 4) with peaks gradually vanishing as the H $\cdots$ O dimer distance increases. This descriptor is visually more appealing for the IGM approach, so we will continue our analysis of intermolecular interactions within the IGM model with  $\delta g^{inter}$ .



**Fig. 5** Comparison between isosurfaces (isovalues in a.u.) for the water dimer;  $\delta g^{inter}(\rho)$  at (a)  $\delta g^{inter} = 0.034$ , (b)  $\delta g^{inter} = 0.012$  and  $s(\rho)$  at (c)  $s = 0.4$ , (d)  $s = 0.74$  and  $\delta k^{inter}(\rho)$  at (e)  $\delta k^{inter} = 0.012$ , (f)  $\delta k^{inter} = 0.005$ ; color coding in the ED range  $-0.05 < \text{sign}(\lambda_2)\rho < +0.05$  a.u.

The next logical step is to use the data from the 2D plot of  $\delta g^{inter}(\rho)$  (Fig 3e) to build 3D plots depicting isosurfaces that represent intermolecular interactions (Fig 5a and 5b obtained for two different isocontour values). By definition, these surfaces enclose the BCP.  $\delta g^{inter}$  surfaces can be colored according to the standard NCI approach. That is, the sign of the second eigenvalue of the ED hessian matrix is used to differentiate between attractive and repulsive situations in regions of the  $\delta g^{inter}(\rho)$  surface where the local electronic charge concentrate ( $\lambda_2 < 0$ ) or is depleted ( $\lambda_2 > 0$ ) perpendicular to the interatomic line. As in the case of  $s(\rho)$ ,  $\delta g^{inter}$  is centered around the BCP. It gives rise to a coherent picture with stabilizing interactions mainly accumulated in the center of the envelop (i.e., the BCP shows the greatest density and  $\lambda_2 < 0$ , although not directly apparent in Fig 5a and b showing isosurfaces enclosing that BCP). The shape of  $\delta g^{inter}$  differs however from that of  $s(\rho)$  displayed in panels c and d. Indeed, not dividing by  $\rho^{4/3}$  in  $\delta g^{inter}$  alters the shape of the peaks leading to broader and slightly convex isosurfaces. This is not due to the IGM model. Indeed,  $\delta s^{inter}$  ( $s$  subjected to the IGM model) is shown to lead to a narrow isosurface in the interaction region like  $s(\rho)$  does (as explained in ESI, see Fig S2b).

Finally, the  $\delta k^{inter}$  isosurface of the water dimer is displayed in Fig 5e and 5f for two different isovalues.  $\delta k^{inter}$  provides the release in kinetic energy density  $\tau_w$ , within the bosonic model, in going from our IGM model to the real system. We observe here that the most significant drops in kinetic energy  $\tau_w$  with respect to the IGM model expand over a greater volume along the H $\cdots$ O hydrogen bond due to the division by  $\rho$  in  $\tau_w$ . The  $\delta k^{inter}$  2D plot employed to draw the associated isosurface for the water dimer is given in Electronic Supplementary Information (Fig S1). It is worth noting that, though a division by  $\rho$  is performed in the calculation of  $\delta k$ , this descriptor does not lead to an artifact as described earlier in this paper for the  $\delta s$  descriptor near  $\rho = 0$ . The reason is that the numerator of  $\delta k$  approaches zero more rapidly than  $\rho$  does in that case.



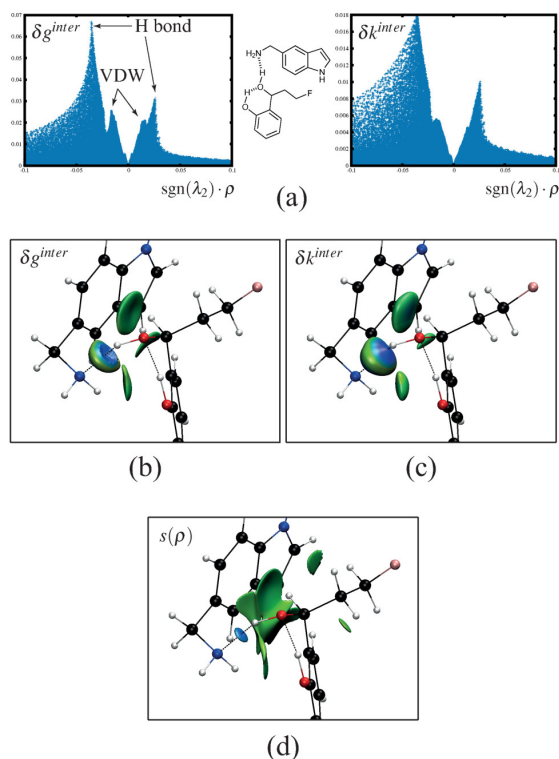
**Fig. 6** Intramolecular weak interaction in ethane-1,2-diol; (a)  $|\nabla\rho|$  2D plot; (b)  $|\nabla\rho|^{IGM}$  (c)  $\delta g$  (d)  $\delta k$ , (e) AIM bond critical points (purple)<sup>22</sup>, (f)  $\delta g(\rho) = 0.016$  a.u. isosurface, (g)  $\delta k(\rho) = 0.005$  a.u. isosurface; color coding in the ED range  $-0.03 < \text{sign}(\lambda_2)\rho < +0.03$  a.u.

Similarly to the reduced density gradient used in the NCI approach,  $\delta g$  can also reveal interactions beyond the presence of critical points.<sup>23</sup> This fact is illustrated in the case of the ethane-1,2-diol on Fig 6. The BCPs in the molecule are reported on Fig 6e (small purple spheres) and do not reveal any critical point between the hydroxyl groups. This is confirmed by the plot of  $|\nabla\rho|$  which never falls to zero in the low ED region around  $\rho = 0.02$  a.u. (Fig 6a). However, a small peak shouldering in the main covalent spike occurs at low  $\delta g$  ( $\delta k$ ) low ED values, revealing this interaction on the 2D plot (Fig 6c and 6d). The corresponding isosurface appear between the hydroxyl groups demonstrating the ability of the  $\delta g$  and  $\delta k$  descriptors to evidence weak interactions in the absence of critical point. It is to be noticed that moving from the  $\delta g$  to the  $\delta k$  descriptor slightly alters the 2D plot representation, leading to peaks less well-characterized. The  $\delta g^{inter}$  ( $\delta k^{inter}$ ) decomposition scheme was not used here (ethanediol) to directly extract this weak interaction in an automatic way because this is



an intramolecular contact and we wanted to avoid the definition of arbitrary fragments. This interaction was not found within AIM (analysis of  $\nabla\rho$ ) when only looking at BCPs. However, with IGM,  $\nabla\rho$  (the basis of AIM) allows for recovering very weak intramolecular interactions if we consider  $\nabla\rho$  deformations.

These new descriptors can also describe van der Waals (vdW) interactions. We now extend the analysis to the heterodimer displayed in Fig 7 which bears different kinds of weak interactions (intra- and intermolecular). Two pairs of spikes appear in the 2D plot of  $\delta g^{inter}$  for this complex, associated with hydrogen-bond and vdW contacts, both in the attractive and repulsive parts. The hydrogen-bond spike emerges at  $\delta g^{inter}$  value significantly

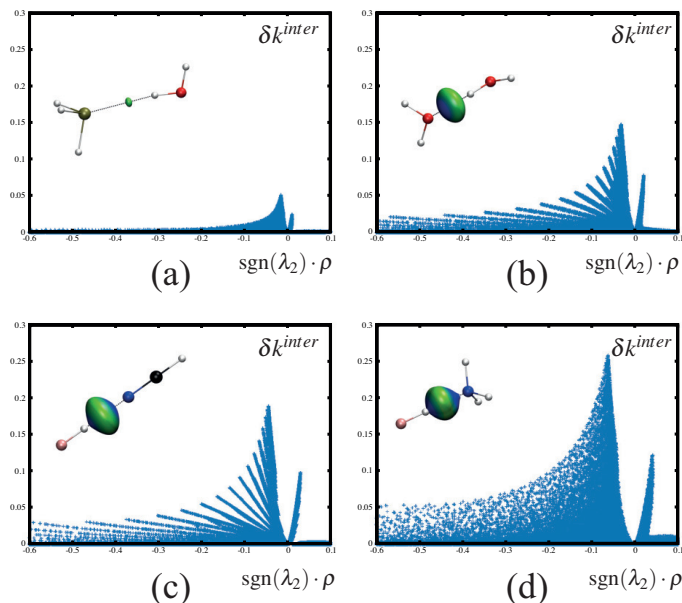


**Fig. 7** Intermolecular hydrogen-bonding and van der Waals contacts in a complex formed by two molecules bearing hydroxyl and amino groups; (a)  $\delta g^{inter}/\delta k^{inter}$  2D plots, (b)  $\delta g^{inter}(\rho) = 0.015$  isosurface, (c)  $\delta k^{inter}(\rho) = 0.004$  isosurface, (d)  $s(\rho) = 0.3$  isosurface; color coding in the ED range  $-0.05 < \text{sign}(\lambda_2)\rho < +0.05$  a.u.

greater than for van der Waals contacts, mainly for the attractive part, linking these results to the familiar concept of interaction strength. As already indicated above, the  $\delta k^{inter}$  2D plot is less informative than the  $\delta g^{inter}$  because spikes are separated to a lesser extend (panel (a) of Fig 7). In particular, van der Waals interactions appear as small shoulderings drowned out by the prominent hydrogen-bond peak in the  $\delta k^{inter}$  representation. 3D isosurfaces obtained with either  $\delta g^{inter}$  or  $\delta k^{inter}$  are shown in Fig 7 (panels b and c, respectively). As previously observed for the water dimer, the  $\delta g^{inter}$  and  $\delta k^{inter}$  isosurfaces have comparable shape. Similarly to the water dimer example, they denote regions resembling those obtained within the parent standard NCI approach (panel d). However,  $\delta g^{inter}$  and  $\delta k^{inter}$  display vdW isocontours which are more pronounced and hence well-delimited (les pixelated)

than  $s$  isosurfaces. As above-explained (and detailed in Fig S2b of ESI),  $s(\rho)$  produces narrower isosurfaces due its mathematical form which strengthens points of low ED compared to  $\delta g$  (or even  $\delta k$ ).

Unlike  $s(\rho)$  and as discussed above for LED, the  $\delta g^{inter}$  and  $\delta k^{inter}$  descriptors are not dimensionless. Hence we can directly use their values on the plot to compare their density at the BCP. Fig 8 shows 2D plots of four hydrogen bond dimers of varying strengths.



**Fig. 8**  $\delta k^{inter}$  2D plots and  $\delta k^{inter} = 0.047$  a.u. isosurfaces colored according to a BGR scheme over the range  $-0.1 < \text{sign}(\lambda_2)\rho < 0.1$  a.u., for 4 dimers: (a)  $\text{PH}_3 \cdots \text{H}_2\text{O}$ , (b) water, (c)  $\text{HF} \cdots \text{HCN}$ , (d)  $\text{HF} \cdots \text{NH}_3$ ; structures optimized at the DFT(B3LYP) 6-311++G\*\* level of theory.

The same  $\delta k^{inter}$  scale has been deliberately chosen for the sake of comparison. Besides the  $\text{PH}_3 \cdots \text{H}_2\text{O}$  complex, envelopes of comparable size are obtained for the other dimers. In that series of 4 dimers, we observe that the peak heights are related to the dimer stabilization energies obtained from CCSD(T) ab initio calculations (see Tab 1).<sup>24</sup> The values fit rather well to a linear correlation for  $\delta g^{inter}$  as well as for  $\delta k^{inter}$ . This is not surprising since  $\delta k^{inter}$  behaves similarly to  $\delta g^{inter}$ . Actually, using a simple exponential model of two interacting atoms A and B, equations 6 and 7, a linear relationship can be established at BCP between both descriptors (see ESI). Moreover, this result is consistent with the commonly admitted correlation between hydrogen bond strength and topological indicators at BCP such as the ED itself or its Laplacian ( $\nabla^2\rho$ ).<sup>25,26</sup> For these four complexes, using dispersion-corrected DFT functionals (M06-2X or wb97XD) does not alter the conclusions. Actually, using these functionals hardly changes the geometry and issued  $\delta g^{inter}$  and  $\delta k^{inter}$  values obtained with promolecular density at BCP (see Tab S1 in ESI).

To conclude, we present in Fig 9 four results obtained on a complex involving a kinase and a bis-azaindole derivative. Panel (a) represents the  $s$  function achieved with a density fraction threshold of either 90% or 95%, respectively. In the standard NCI ap-



**Table 1** H-bond length  $d$  and properties at the bond critical point (represented by the point of highest value in the 2D plot of Fig 8);  $\rho$ ,  $\delta g^{inter}$  and  $\delta k^{inter}$  in a.u.;  $\Delta E$  in kcal.mol<sup>-1</sup>

| Complex                             | $d(\text{\AA})$ | $\rho$ | $\delta g^{inter}$ | $\delta k^{inter}$ | $\Delta E$         |
|-------------------------------------|-----------------|--------|--------------------|--------------------|--------------------|
| PH <sub>3</sub> ...H <sub>2</sub> O | 2.65            | 0.016  | 0.027              | 0.054              | -2.5 <sup>a</sup>  |
| H <sub>2</sub> O...H <sub>2</sub> O | 1.93            | 0.032  | 0.060              | 0.148              | -5.0 <sup>b</sup>  |
| HCN...HF                            | 1.84            | 0.044  | 0.090              | 0.189              | -7.5 <sup>b</sup>  |
| HF...NH <sub>3</sub>                | 1.68            | 0.063  | 0.121              | 0.258              | -12.6 <sup>b</sup> |

<sup>a</sup> CCSD(T)-F12a/VQZ-F12.<sup>27</sup>

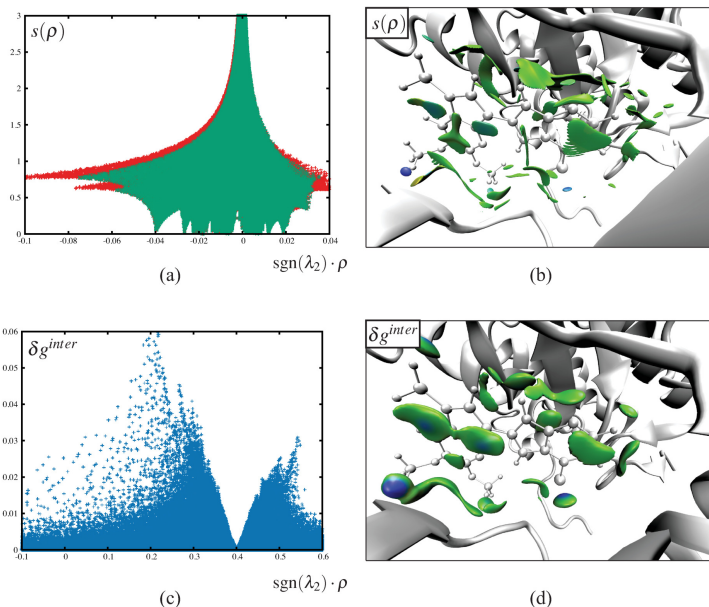
<sup>b</sup> CCSD(T)/CBS+REL+CV<sup>28</sup>

proach with pro-molecular ED, applying such a ratio threshold is the only way to attempt discarding nodes corresponding to intramolecular interaction situations (for which the density comes mostly from one single molecule). As can be seen, the same general shape of the 2D plot is obtained. Of course, depending on the specified threshold value, some differences occur on the 2D plot, however this translates into relatively small discrepancies (not reported here) onto the 3D plot in panel (b). The natural threshold (0.90 - 0.95) is what is proposed in NCIPLOT. Although choosing a threshold value remains an arbitrary decision, this threshold range is appropriate to identify non covalent interactions. In contrast, the  $\delta g^{inter}$  descriptor (2D plot in panel (c)) uniquely defines interaction isosurfaces (see panel (d)). Comparing the two descriptors  $s$  and  $\delta g^{inter}$ , we observe that (as previously underlined) the  $\delta g^{inter}$  isocontours are thicker, more pronounced and well-delimited. Again, this is not due to the IGM approach but it results from the specific mathematical form of  $s(\rho)$  compared to the  $\delta g$  one.

From a practical point of view, the IGM workflow exhibits some advantages compared to the original  $s(\rho)$  analysis. Actually, there is no need to differentiate grid points associated to intermolecular situations from other (undesired) points corresponding to intramolecular regions by using an arbitrary pro-molecular ED fraction threshold as proposed in NCIPLOT. It is worth noting that the program is able to filter weak intramolecular interactions (h-bond and ring closure interactions) from the data without any specific manipulation in the input parameters. Finally, this also clearly constitutes an advantage for integration schemes summing properties over the interaction region,<sup>10</sup> which is now well identified and isolated in real space with the  $\delta g^{inter}$  descriptor.

## 4 Summary and future outlook

In this paper, we propose a new concept to accurately extract the signature of the intermolecular interactions present in an NCI plot. This model is founded on a reference ED featuring an exponential decay characteristic of non-interacting atoms (Independent Gradient Model). We demonstrate that it allows identifying and quantifying deviations from the exponential decay region of the plot  $s(\rho)$  in the NCI approach. The model is also applied to other local descriptors such as the electron gradient, leading to  $\delta g$ . Since the gradient is density dependent,  $\delta g$  has the advantage of taking non-zero values only when ED interference occur between interacting atoms, going down to zero when the distance between fragments goes to infinity. The analysis of a simple exponential model of  $\delta g$  shows that the IGM model reveals the elec-



**Fig. 9** Biphosphorylated kinase IGF-1R (PDBID 3LVP) in complex with a bis-azaindole inhibitor; for the sake of clarity atoms and secondary structures are not colored; (a) standard NCI plot obtained with grid nodes discarded when at least 90% (green) or 95% (red) of the total pro-molecular density comes from one molecule; (b)  $s = 0.5$  a.u. isosurfaces ( $\rho_i/\rho = 0.95$ ) (c)  $\delta g^{inter}$  2D plot; (d)  $\delta g^{inter} = 0.009$  a.u.; all isosurfaces are colored according to a BGR scheme over the ED range  $-0.07 < \text{sign}(\lambda_2)\rho < 0.07$  a.u.

tron density interpenetration within AIM basins. The Weizsäcker kinetic energy density has also been subjected to the IGM model leading to the descriptor  $\delta k$  that exhibits 3D isosurfaces similar to those obtained with the  $\delta g$  descriptor.

The possibilities offered by the IGM approach and issued descriptors is illustrated on several molecular complexes involving hydrogen-bonding and van der Waals interactions. An attractive feature of the IGM methodology coupled to a given local descriptor is to provide a workflow that automatically generates data composed solely of intermolecular interactions for drawing the corresponding 2D and 3D representations. The user no longer needs to choose an iso-value (usually system dependent) in a way that produces well-separated domains. Neither is the application of ED cutoffs required. As a consequence, the IGM approach overcomes difficulties to define NCI regions within integration procedures.

In this paper, only the pro-molecular density is considered to obtain the ED. Indeed, the separation into atomic contributions is only possible in a strict sense in pro-molecular densities. However, these ED have shown to provide similar results to relaxed ones as far as non covalent interactions are concerned. Moreover, since the intra/intermolecular differentiation is done point-wise, it can be used as a first approximation in relaxed densities: each point is classified in intra/intermolecular and the relaxed RDG calculated accordingly. This option will be considered in future publications along with more accurate schemes. We hope to extend our calculations soon to an orbital-based ED gradient casting the density expression in a form that is most effective for

using the IGM model. Several ways can be considered ranging from the approximation of Mulliken to more recent sophisticated models such as the concept of quasi-atomic orbitals.<sup>29,30</sup>

## 5 Acknowledgements

We are grateful to Pr F. Bohr for helpful discussions. This work has been partially supported by the French foundation of Technological Research under grant CIFRE1782015 (PhD thesis of G. Rubez) involving an agreement between the ATOS Company and the University of Reims Champagne-Ardenne. Most of the computations reported in this study were performed on the computational centre of Champagne-Ardenne (France, <http://romeo.univ-reims.fr>). Also, we are grateful to the Multi-scale Molecular Modelling Platform (P3M, University of Reims Champagne-Ardenne, France, <http://p3m.univ-reims.fr>) for various supports and the CRIANN (<http://www.criann.fr>) for additional computing facilities.

## References

- 1 *Atoms In Molecules, A Quantum Theory*, Oxford University Press, USA, 1994.
- 2 *The Quantum Theory of Atoms In Molecules*, ed. C. F. Matta and R. J. Boyd, Wiley-VCH, Weinheim, Germany, 2007.
- 3 B. Silvi and A. Savin, *Nature*, 1994, **371**, 683–686.
- 4 E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J. Cohen and W. Yang, *J. Am. Chem. Soc.*, 2010, **132**, 6498–6506.
- 5 R. A. Boto, J. Contreras-García, J. Tierny and J. P. Piquemal, *Mol. Phys.*, 2016, **114**, 1406–1414.
- 6 C. F. von Weizsäcker, *Z. Phys.*, 1935, **96**, 431–458.
- 7 J. Contreras-García, E. R. Johnson, S. Keinan, R. Chaudret, J.-P. Piquemal, D. N. Beratan and W. Yang, *J. Chem. Theory Comput.*, 2011, **7**, 625–632.
- 8 M. A. Spackman and E. N. Maslen, *J. Phys. Chem.*, 1986, **90**, 2020–2027.
- 9 F. L. Hirshfeld and S. Rzotkiewicz, *Mol. Phys.*, 1974, **27**, 1319–1343.
- 10 J. Contreras-García, W. Yang and E. R. Johnson, *J. Phys. Chem. A*, 2011, **115**, 12983–12990.
- 11 G. Saleh, C. Gatti and L. L. Presti, *Comput. Theor. Chem.*, 2015, **1053**, 53–59.
- 12 J. Contreras-García, *Habilitation to conduct researches*, University Pierre and Marie Curie, Paris, 2015.
- 13 C. W. Wilson and W. A. Goddard, *Chem. Phys. Lett.*, 1970, **5**, 45–49.
- 14 M. W. Schmidt, J. Ivanic and K. Ruedenberg, *J. Chem. Phys.*, 2014, **140**, 204104.
- 15 *The Chemical Bond*, ed. G. Frenking and S. Shaik, Wiley-VCH, Weinheim, Germany, 2014.
- 16 H. J. Bohorquez, C. F. Matta and R. J. Boyd, *Int. J. Quant. Chem.*, 2010, **110**, 2418–2425.
- 17 P. de Silva, J. Korchowiec and T. A. Wesolowski, *Chem. Phys. Chem.*, 2012, **13** (15), 3462–3465.
- 18 J. Contreras-García, M. Calatayud, J.-P. Piquemal and J. M. Recio, *Comput. Theor. Chem.*, 2012, **998**, 193–201.
- 19 P. de Silva and C. Corminboeuf, *J. Chem. Theory Comput.*, 2014, **10** (9), 3745–3756.
- 20 G. Rubez, J.-M. Etancelin, X. Vigouroux, M. Krajecki, J.-C. Boisson and E. Hénon, *J. Comput. Chem.*, 2017, **38**, 1071–1083.
- 21 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09 Revision E.01*, Gaussian Inc. Wallingford CT 2009.
- 22 T. A. Keith, *AIMAll Version 16.08.17*, TK Gristmill Software 2016.
- 23 J. R. Lane, J. Contreras-García, J.-P. Piquemal, B. J. Miller and H. G. Kjaergaard, *J. Chem. Theor. Comp.*, 2013, **9**, 3263–3266.
- 24 N. Mohan and C. H. Suresh, *J. Phys. Chem. A*, 2014, **118**, 1697–1705.
- 25 S. J. Grabowski, *J. Phys. Org. Chem.*, 2004, **17**, 18–31.
- 26 E. Espinosa, E. Molins and C. Lecomte, *Chem. Phys. Lett.*, 1998, **285**, 170–173.
- 27 J. R. Lane and H. G. Kjaergaard, *J. Chem. Phys.*, 2009, **131**, 034307.
- 28 A. D. Boese, *Journal of Chemical Theory and Computation*, 2013, **9**, 4403–4413.
- 29 A. C. West, M. W. Schmidt, M. S. Gordon and K. Ruedenberg, *J. Chem. Phys.*, 2013, **139**, 234107.
- 30 A. C. West, M. W. Schmidt, M. S. Gordon and K. Ruedenberg, *J. Phys. Chem. A*, 2015, **119**, 10360–10367.