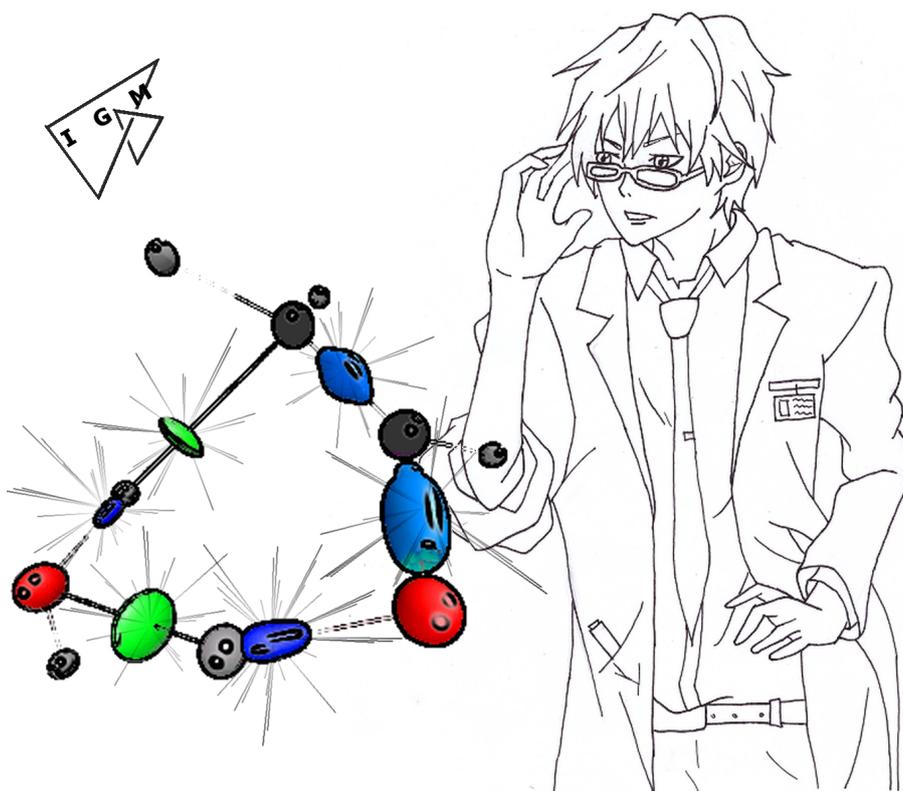


# IGMPlot Documentation

A CPU implementation of the IGM approach for detecting, quantifying and plotting molecular interactions from electron density<sup>[1]</sup>

Version 3.16 - February 2025



## Authors and Contributors

Akilan Rajamani, Corentin Lefebvre, Johanna Klein, Emmanuel Pluot, Gaëtan Rubez, Hassan Khartabil, Jean-Charles Boisson<sup>1</sup> and Eric Henon<sup>2</sup>

## License

IGMPlot is a free software. It is released under the CeCILL-c license

<http://cecill.info/index.en.html>

However, please note that a new ADF library (contact Alexei Yakovlev, SCM) introduced in this version of IGMPlot is provided under the GNU Lesser General Public License (LGPL), version 3. Please review the information provided in the source for detailed licensing information on these components.

<sup>1</sup>LICIIS, University of Reims Champagne-Ardenne, 51687 Reims, France

<sup>2</sup>Institut de Chimie Moléculaire de Reims, CNRS UMR 7312, University of Reims Champagne-Ardenne, 51687 Reims, France



### Referencing the program IGMPlot<sup>[1]</sup>

When referencing results obtained from the program "IGMPlot" in publications or other contexts, please acknowledge the software by citing the following reference:

C. Lefebvre, J. Klein, H. Khartabil, J.-C. Boisson, E. Hénon  
J. Comp. Chem 2023, 44(20), 1750-1766  
doi: <https://doi.org/10.1002/jcc.27123>

### References for citing theories and tools available in IGMPlot

#### The original IGM approach

Lefebvre C., Rubez G., Khartabil H., Boisson J.C., Contreras-García J., Hénon E.  
Phys Chem Chem Phys. 2017, 19, 17928  
doi: <http://doi.org/10.1039/c7cp02110k>

#### The quantum-mechanical theory of IGM

Lefebvre C., Khartabil H., Boisson J.C., Contreras-García J., Piquemal J.-P., Hénon E.  
Chem. Phys. Chem. 2018, 19, 1  
doi: <http://doi.org/10.1002/cphc.201701325>

#### The atomic decomposition scheme of the inter-fragment interaction

Miguel Ponce-Vargas, Corentin Lefebvre, Jean-Charles Boisson, Eric Hénon  
J. Chem. Inf. Model. 2020, 60, 1, 268  
doi: <http://doi.org/10.1021/acs.jcim.9b01016>

#### The Intrinsic Bond Strength Index (IBSI)

Johanna Klein, Hassan Khartabil, Jean-Charles Boisson, Julia Contreras-García, Jean-Philip Piquemal, Eric Hénon  
J. . Phys. Chem. A 2020, 124, 9, 1850  
doi: <http://doi.org/10.1021/acs.jpca.9b09845>

#### The Atomic Degree Of Interaction (DOI)

C. Lefebvre, Hassan Khartabil, and Eric Hénon  
Phys. Chem. Chem. Phys. 2023  
doi: <https://doi.org/10.1039/D2CP02839E>

#### The Pair Density Asymmetry (PDA)

C. Lefebvre, J. Klein, H. Khartabil, J.-C. Boisson, E. Hénon  
J. Comp. Chem 2023  
doi: <https://doi.org/10.1002/jcc.27123>

#### The descriptor qg

C. Lefebvre, J. Klein, H. Khartabil, J.-C. Boisson, E. Hénon  
J. Comp. Chem 2023  
doi: <https://doi.org/10.1002/jcc.27123>



### **Computational Resources used to develop IGMPlot**

Two centers contributed to the development of the IGMPlot code by providing High-Performance Computing environment (CPU resources, compilers, graphical interfaces, ...)

- ROMEO: the regional computational center of the University of Reims Champagne-Ardenne, <https://romeo.univ-reims.fr>
- CRIANN: the Centre Régional Informatique et d'Applications Numériques de Normandie, <https://www.criann.fr>



# Contents

<b>1</b>	<b>Introduction</b>	<b>6</b>
1.1	History . . . . .	7
1.2	General presentation of IGMPlot . . . . .	7
1.2.1	IGM principle in a nutshell . . . . .	7
1.2.2	IGM background . . . . .	8
1.2.3	Automatically separating intramolecular from intermolecular interactions . . . . .	9
1.2.4	"wave function" or "promolecular mode" . . . . .	10
1.2.5	Family of Interaction: $\delta g$ peak . . . . .	12
1.2.6	Interaction strength . . . . .	13
1.2.7	Atomic decomposition scheme . . . . .	16
1.2.8	I B S I . . . . .	17
1.2.9	Pair Density Asymmetry (PDA) . . . . .	18
1.2.10	Critical Point Analysis . . . . .	18
1.2.11	Hirshfeld Gradient Partitioning . . . . .	18
1.2.12	DOI = atomic Degree Of Interaction . . . . .	18
1.2.13	ELF . . . . .	18
<b>2</b>	<b>Installation</b>	<b>19</b>
<b>3</b>	<b>Running the program</b>	<b>20</b>
<b>4</b>	<b>Inputs - Outputs</b>	<b>20</b>
4.1	the parameter input file param.igm . . . . .	21
4.2	.xyz or wfx, wfn, rkf molecule description files . . . . .	21
4.3	Number of input files . . . . .	22
4.4	GTO or STO ? . . . . .	22
4.5	WFN/WFX/RKF remarks: . . . . .	22
<b>5</b>	<b>GRID details</b>	<b>23</b>
<b>6</b>	<b>KEYWORD documentation</b>	<b>24</b>
6.1	Optional keywords . . . . .	24
6.1.1	<b>*LIGAND</b> n r . . . . .	24
6.1.2	<b>CUBE</b> x1 y1 z1 x2 y2 z2 . . . . .	25
6.1.3	<b>RADIUS</b> x y z r . . . . .	25
6.1.4	<b>ONAME</b> name . . . . .	25
6.1.5	<b>OUTPUT</b> n . . . . .	25
6.1.6	<b>INCREMENTS</b> i1 i2 i3 . . . . .	26
6.1.7	<b>FRAG1</b> selection pattern (IGM, WFN/WFX mode) . . . . .	26
6.1.8	<b>FRAG2</b> selection pattern (IGM, WFN/WFX mode) . . . . .	26
6.1.9	<b>CUBEFRAG</b> (IGM, WFN/WFX mode) . . . . .	27
6.1.10	Fragment definition summary . . . . .	27
6.1.11	<b>IBSI</b> r (IGM, WFN/WFX mode) . . . . .	28



6.1.12	<b>*INTERMOLECULAR</b> (NCI model)	29
6.1.13	<b>CUTOFFS</b> r1 r2 (NCI + IGM models)	29
6.1.14	<b>CUTPLOT</b> r3 r4 (NCI model)	29
6.1.15	<b>CUTPLOT_IGM</b> r5 r6 (IGM model)	30
6.1.16	<b>PEAKFOCUSINTRA</b> x y (IGM model)	30
6.1.17	<b>PEAKFOCUSINTER</b> x y (IGM model)	30
6.1.18	<b>VMD_COLRANG_IGM</b> r7 r8 (IGM model)	30
6.1.19	<b>HIRSH</b> (IGMH model)	31
6.1.20	<b>CRITIC</b> (IGM model)	31
6.1.21	<b>CRITICFINE</b> (IGM model)	31
6.1.22	<b>CRITICULTRAFINE</b> (IGM model)	31
6.1.23	<b>CRITICADDSEEDS</b> (IGM model)	31
6.1.24	<b>FULLAOACC</b>	31
6.1.25	<b>ELF</b>	32
<b>7</b>	<b>General remarks</b>	<b>32</b>
7.1	Critical point search	32
7.2	Rotational invariance	33
7.3	Implementation of the IGM approach in Multiwfn	35
7.4	Hirshfeld version Gradient Based partition	36
<b>8</b>	<b>Examples</b>	<b>37</b>
8.1	Example 1 (promolecular density): to use one or two fragments ? (test1,test2)	37
8.2	Example 2 (promolecular density): vdW interactions (test3)	39
8.3	Example 3 (promolecular density): quantification of non-covalent interactions (test4 to test10)	41
8.4	Example 4 (promolecular density): Monitoring intramolecular interactions with IGM- $\delta g^{intra}$ in a peptide (test11, test12)	45
8.5	Example 5 (promolecular density): smart use of the uncoupling scheme in a trimer (test13)	47
8.6	Example 6 (promolecular density): ligand –protein (test14)	48
8.7	Example 7 (promolecular density): when to use the CUTPLOT_IGM keyword ? (test15)	50
8.8	Example 8 (promolecular density): atomic decomposition scheme of non-covalent interactions applied to host-guest assemblies (test16, test17)	51
8.9	Example 9 (QM): $H_2O \dots H_2O$ dimer, FRAG1 keyword (test18)	53
8.10	Example 10 (QM): $B_2H_6$ taken as one piece (test19)	54
8.11	Example 11 (QM): Agostic interaction in the metallic complex [Ti Cl <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub> ] <sup>+</sup> (test20)	55
8.12	Example 12 (QM): FRAG1 and FRAG2 keywords: quantifying intramolecular $\pi$ -stacking (test21)	56
8.13	Example 13 (QM): CUBEFRAG keyword (test22)	57
8.14	Example 14 (QM): IBSI keyword: a new way for probing bond strength (test23)	59
8.15	Example 15 (QM): IBSI keyword with cutoff option (test24)	61
8.16	Example 16 (QM): Bond $\delta g^{pair}$ signature and atom pair isosurface (test25)	63



8.17	ELMO possibility . . . . .	64
8.18	Example 17 (QM): DOI : the atomic Degree Of Interaction (test26, $CH_3NH_2$ ) . . . . .	65
8.19	Example 18 (QM): IGMH (test27, $CH_3OH$ ) . . . . .	66
8.20	Example 19 (QM): qg descriptor (test28, $CH_3OH$ ) . . . . .	68
8.21	Example 20 (QM): Critical point analysis (test29, benzene dimer) . . . . .	70
8.22	Example 21 (QM): Investigating inductive effect with the PDA index (test30, test31) . . . . .	72
8.23	Example 22 (QM): ADF users (test32) . . . . .	76
	8.23.1 <b>FULLAOACC</b> . . . . .	77
8.24	Example 23 (QM): ELF&IGM(test33) . . . . .	78

## 9 Future prospects and outlook 79

The IGMPlot package<sup>[1]</sup> contains the code source together with sample runs, this documentation, the software reference documentation (html), installation instructions, and Copyright.

# 1 Introduction

By using IGMPlot you can identify and quantify molecular interactions over a broad range: from non-covalent to covalent bonding. It allows you for:

- Studying molecular systems from a wave function: QM treatment from information taken in wfn, wfx or more recently in rkf (ADF) files.
- Studying large systems like ligand-protein complexes from the atomic cartesian coordinates only (promolecular ED).
- Distinguishing between intramolecular and intermolecular interactions, covalent or hydrogen-bonding or vdW interactions.
- Probing and quantifying interactions between two given fragments in a molecular system (both QM and promolecular modes).
- Determining the strength (IBSI) and asymmetry (PDA) of a given bond.
- Estimating the atomic contributions to an intermolecular interaction (QM and Promolecular modes).
- Building isosurfaces representing the regions of space where the interactions take place.
- Monitoring interactions along molecular dynamics or along a reaction path.
- Performing a critical point analysis.

From a practical perspective, an attractive feature of the IGM approach is to provide an automatic workflow delivering data that provides chemists with a visual and quantitative understanding of interactions. No topological analysis is required to study interactions and the IGM analysis can be achieved with little preparation.



## 1.1 History

IGMPlot is based on the electron density-based descriptor called  $\delta g$ . The IGM- $\delta g$  approach was initially designed to work with promolecular density (non-relaxed electron density, sum of spherically averaged neutral atomic densities).<sup>[2]</sup> But in 2018, we proposed the Gradient-Based Partitioning (GBP) that extends the IGM concept to electron density derived from a wave function (SCF).<sup>[3]</sup> Thus, thanks to the new IGM approach, detailed information can be directly obtained either on the covalent or on the non-covalent domain, for small and larger molecular systems.

In 2020, a new extension was developed able to emphasize the most relevant atomic contributions to the noncovalent interactions occurring between two fragments.<sup>[4]</sup> This development proves to be an appealing tool to shed light on the guest accommodation on a per atom basis. This possibility is available in both QM and promolecular modes. Also, in 2020 was proposed the new Intrinsic Bond Strength Index IBSI.<sup>[5]</sup> This score, (not a bond order), is obtained in an automatic workflow and is very efficient to internally probe the strength of a given atom pair, over a wide range (non-covalent to covalent, transition metal bonding, agostic interactions, ...). An IBSI scale has been proposed to range two-centre chemical bonds by their intrinsic strength.

In 2020 was proposed a new release allowing for detecting the interaction between two given sub-fragments of a single molecule, using QM electron density. This possibility is particularly attractive to assess the role of noncovalent intramolecular interactions (for instance intramolecular  $\pi - \pi$  stacking or hydrogen-bonding between two part of a single molecule along a reaction path). Also, the quantification of the detected interactions has been implemented using integration schemes and summing local quantities over the space representing the interactions. From a technical perspective, the .wfx file format is now supported by IGMPlot (in addition to .wfn). The promolecular electron density of atoms Te, I and Xe of period 5 has been implemented. In other respects, we also demonstrated the good performance of the coupling between IGM and Extremely Localized Molecular Orbitals to go beyond the promolecular ED approximation in biological systems, at low cost.<sup>[6]</sup>

In 2023 we have introduced the atomic Degree Of Interaction (DOI)<sup>[7]</sup> and provided users with the critical point analysis. The Hirshfeld-based electron density gradient partition has been included in the package.<sup>[8]</sup>

## 1.2 General presentation of IGMPlot

For more detailed information we refer the interested reader to the original papers (above-mentioned). The IGM- $\delta g$  approach is based on a new electron density (ED) reference model, the independent gradient model (IGM) really getting rid of interactions, which was presented in 2017.<sup>[2]</sup> Compared to an interacting situation, this non-interacting reference enables quantifying the net drops observed in NCI<sup>[9]</sup> 2D plots both in region of high and low ED.

### 1.2.1 IGM principle in a nutshell

The IGM- $\delta g$  approach relies on the new definition of a non-interacting system in term of ED contragradience ("ED clash" between two sources). More precisely, the new descriptor,  $\delta g$ , represents locally the difference between a virtual upper limit of the ED gradient ( $\nabla\rho^{IGM}$ )



representing a non-interacting system and the true ED gradient ( $\nabla\rho$ ). The theoretical background can be found in the papers of 2017<sup>[2]</sup> and 2018.<sup>[3]</sup> The program IGMPlot, written in C++, is a custom implementation using either promolecular density or relaxed electron density (SCF). It is designed to leverage the computational power of Multi-core CPU through parallel OpenMP programming.

### 1.2.2 IGM background

The local descriptor  $\delta g$  assesses the mutual penetration of two electron density (ED) sources (two atoms, or two intramolecular fragments, or two molecules) by the deviation of the ED gradient from a non-interacting reference. For more details, we refer the reader to IGM fundamentals in the original papers.<sup>[2,3]</sup>

Given a geometry, the electron density (ED)  $\rho$  is computed at each node of a grid encompassing the molecular system. It can be realized either using the promolecular approximation or from a wave function obtained after QM calculations or after the transfer of ELMOs.<sup>[6]</sup> Its gradient  $\nabla\rho$  is also computed. Afterwards, the fragment components of the true ED derivative ( $\frac{\partial\rho_A}{\partial x}$ ,  $\frac{\partial\rho_B}{\partial x}$ , ...) are determined. Note that we are not limited to a fragmentation scheme involving only two fragments, but we could consider 3, 4, ... fragments. In the most simple application of IGM, the ED gradient is fragmented in  $n$  contributions,  $n$  being the number of atoms. This requires using a partition scheme of the ED gradient (not of the ED). Within the promolecular approach, this is straightforward since the ED is constructed as the sum of the ED of free atoms, so the promolecular ED gradient is also the sum of atomic ED gradient contributions. On the other hand, when the system is described using a wave-function, we use the Gradient Based Partition GBP able to assign each atom a ED gradient contribution.<sup>[3]</sup> Say that we are interested in a system made of two interacting fragments A and B. Between A and B, the two individual derivative contributions:  $\frac{\partial\rho_A}{\partial x}$ ,  $\frac{\partial\rho_B}{\partial x}$ , have opposite signs, which means that the fragment EDs mix when the two fragments approach each other. This overlap between ED clouds results in an attenuation of the total ED gradient in the region between fragments. To quantify this attenuation, we simultaneously consider two molecular models: the interacting (real) system, and a non-interacting reference (i.e., the so-called Independent Gradient Model (IGM)). At a given point of the grid, for the real system featuring the interaction, the true ED derivative is in absolute value:

$$\left| \frac{\partial\rho}{\partial x} \right| = \left| \frac{\partial\rho_A}{\partial x} + \frac{\partial\rho_B}{\partial x} \right| \quad (1)$$

The previous expression holds for the three directions  $x$ ,  $y$  and  $z$ . For the virtual non-interacting reference (having the same ED as the true molecular system), the ED derivative, in absolute value, is obtained by adding the absolute values of the individual fragment components:

$$\left| \frac{\partial\rho^{IGM}}{\partial x} \right| = \left| \frac{\partial\rho_A}{\partial x} \right| + \left| \frac{\partial\rho_B}{\partial x} \right| \quad (2)$$

This partitioning procedure cancels the derivative attenuation, which is expected from the addition of two interpenetrating fragment derivatives having opposite signs in the region between the two ED sources. The norm  $|\nabla\rho^{IGM}|$  resulting from equation (2) represents an upper limit



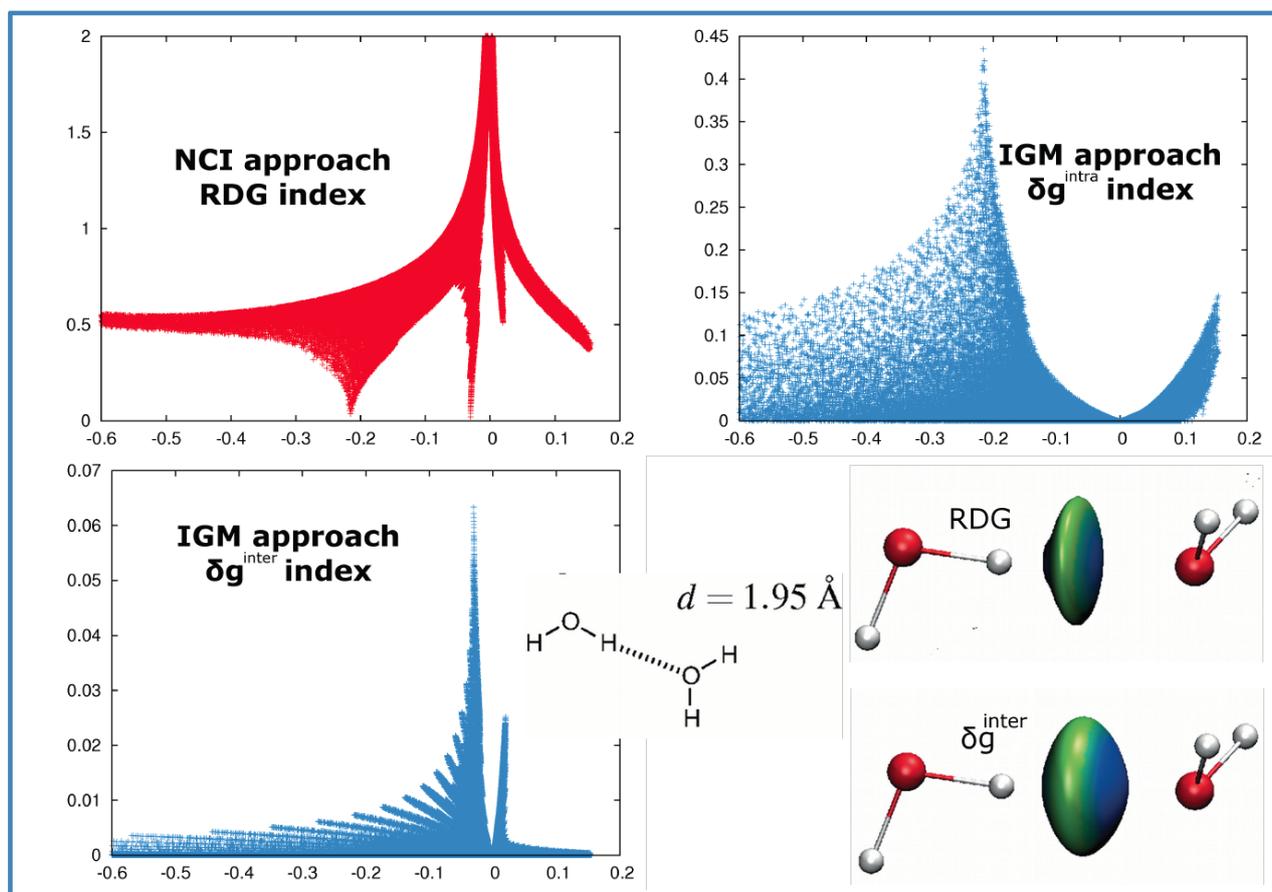
of the norm  $|\nabla\rho|$  of the true ED gradient derived from ED derivatives given by equation (1). Thereby, we calculate the local  $\delta g$  descriptor as follows:

$$\delta g = |\nabla\rho^{IGM}| - |\nabla\rho| \quad (3)$$

which quantifies the contragradience between the two fragment ED sources. Non-zero values of  $\delta g$  exclusively correspond to interaction situations: the larger  $\delta g$ , the stronger the interaction.

### 1.2.3 Automatically separating intramolecular from intermolecular interactions

The IGM approach is able to reveal all the interactions between atoms in the molecule but an attractive feature of the IGM methodology is to provide an uncoupling "intra/inter" scheme between two user-defined fragments that automatically extracts the signature of intra- ( $\delta g^{intra}$ ) (inside the fragments) and inter- ( $\delta g^{inter}$ ) (between the fragments) molecular interactions (see Fig. 1). It allows then for drawing the corresponding 3D iso-surface representations in real space with software like VMD. This automatic intra/inter separation can be carried out either using promolecular ED or ED coming from QM calculations (wave function mode). The user has only to supply the definition of two fragments in terms of atom indexes. The sign of the second eigenvalue of the ED hessian matrix serves to differentiate non-bonding ( $\lambda_2 > 0$ ) from attractive ( $\lambda_2 < 0$ ) interactions. The nature of the interaction is color-coded: blue for attractive interactions, green for weakly non-bonding or attractive interactions or red for non-bonding situations. The use of promolecular ED is particularly suited to describe ligand-protein interactions while relaxed ED (SCF) obtained from wave function calculations extends the range of applicability to organic chemistry, inorganic, chemical reactivity, ... Of course, the study of strong interactions (like covalent bonding) requires the use of relaxed ED obtained by means of QM calculations. So far, promolecular electron density computation is limited to atoms of periods 1,2,3,4 and Te, I, Xe atoms of period 5. Concerning the wave-function mode, the needed wave function file can be obtained from a QM program like Gaussian, NWChem, ORCA, ... IGMPlot supports both the .wfn and .wfx input file formats.



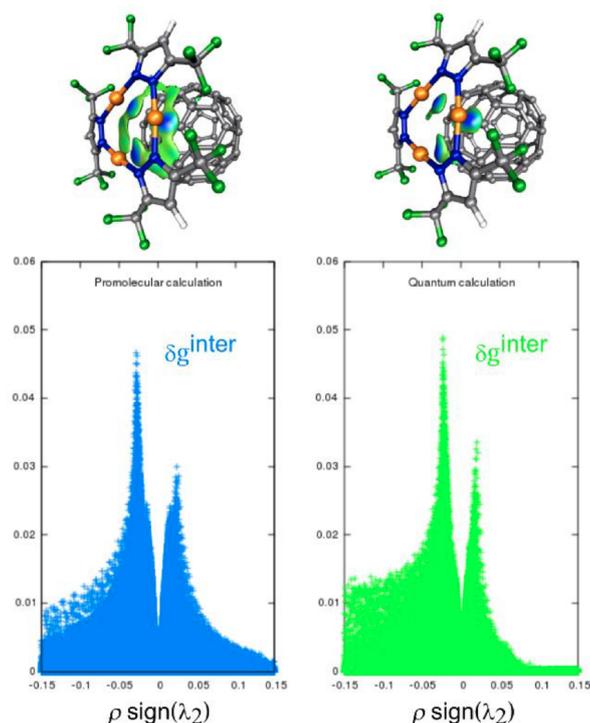
**Figure 1.** Water dimer treated with promolecular electron density (dedicated to weak interactions); NCI plot on top left panel (global signature); intramolecular IGM- $\delta g^{intra}$  signature on top right panel; automatically extracted hydrogen-bonding IGM- $\delta g^{inter}$  signature at bottom left ; isosurfaces bottom right.

The detection of intermolecular interaction regions takes place automatically from the moment the two fragments definition (atom indexes) has been supplied. Since the intra/inter interactions are well separated by the IGM- $\delta g$  approach, the program is then able to deliver intermolecular interactions (generally weak interactions) deprived of intramolecular interactions (covalent bonding, but also weak intramolecular interactions like intramol. H-bond or ring closure interactions) without any specific manipulation in the input parameters. Both in the "wave function" and "promolecular" modes, it is not necessary to choose appropriate ED cutoffs in a way that produces well separated strong/weak interaction domains. But this option is still available.

#### 1.2.4 "wave function" or "promolecular mode"

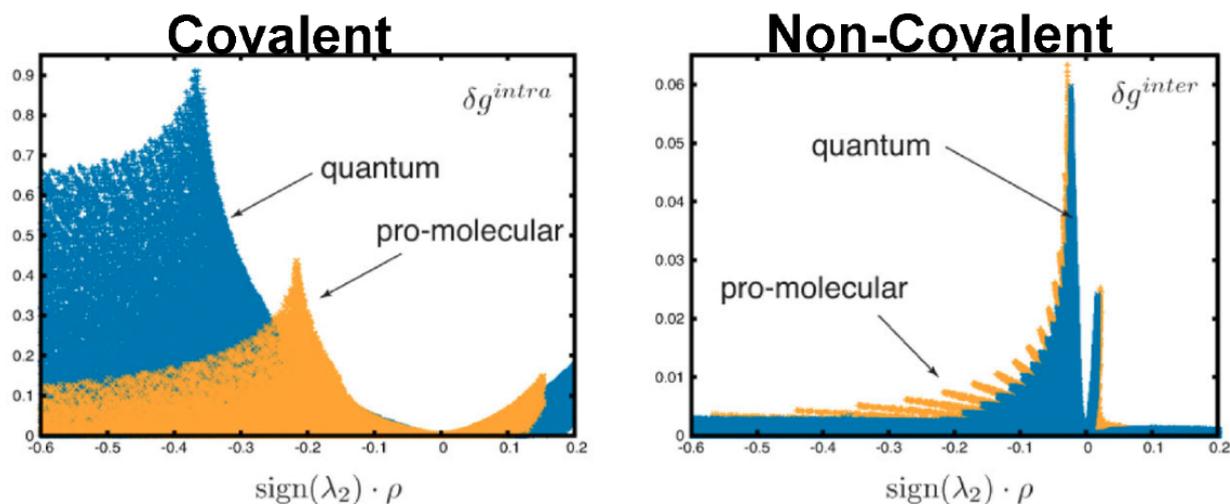
What is the difference between the "promolecular" and "wave function" modes? The promolecular electron density (ED) is the sum of simple exponential atomic functions. These exponential functions are tabulated and stored in the IGMPlot program for atoms of periods 1-4 and for atoms Te, I and Xe of period 5. It lacks ED relaxation, but this approximate density has

however shown to provide similar results to the relaxed one as long as computations remain in the noncovalent domain. Thus, promolecular densities are useful to study large systems, typically ligand–receptor interactions, because its calculation is very fast, and it only requires the geometry as input (the ED is calculated by IGMPlot). Similar topological features can be obtained from both promolecular and quantum EDs, as can be observed from Fig. 2 showing the interaction between a trinuclear copper complex and a fullerene molecule.



**Figure 2.** Trinuclear copper complex; comparison of the "promolecular" and "wave function" (M06-2X/6-311G(d)) modes to obtain the intermolecular interaction signature.

Of course, the promolecular ED cannot be employed to study strong interactions (covalent bonding, metal coordination). As can be seen on Fig. 3 for the water dimer, the promolecular ED clearly underestimates the covalent O-H interaction (left panel) :  $\delta g^{intra}$  (promol.) = 0.4 versus  $\delta g^{intra}$  (QM) = 0.9 for the two covalent peaks. Conversely, both levels of theory give  $\delta g^{inter}$  = 0.06 for the intermolecular signature peak height (on the right).



**Figure 3.** Water dimer; comparison of the "promolecular" and "wave function" (M06-2X/6-311G(d)) modes for intra- and intermolecular interactions in the dimer.

The promolecular ED is limited to the 36 first elements of the periodic table (H to Kr) and to atoms Te, I and Xe of period 5. In the promolecular mode, only the atomic coordinates have to be supplied. The promolecular density is obtained in IGMPlot from a sum over three simple atomic exponential functions. It lacks SCF relaxation. The parameterized exponents and coefficients of sphericalized atomic densities for atoms of periods 1, 2 and 3 have been taken from the program NCIPlot. These densities are developed over a sum of 3 exponential atomic pieces. For H and He, only the first coefficient has a non-zero value. For the atoms of period 2 the third coefficient is set to 0. In order to provide such an analytical function for atoms of period 4 and atoms Te, I and Xe of period 5, without any additional CPU cost, an expression of only 3 exponential atomic pieces has been tailored to match the neutral averaged spherical atomic ED (DFT B3LYP/6-311+G\*, B3LYP/Def2TZVPD for period 5). For periods 4 and 5, only the long-range ED beyond the vdW radius is well described. This was done to describe atoms of period 4 and 5 (such as Br and I) that can be involved in non-covalent interactions within ligand-protein complexes for instance. **Let's recall that such promolecular density is able to predict low-density only; i.e., to describe the non-covalent domain. It cannot be used to describe covalent (nor metal coordination) situations.**

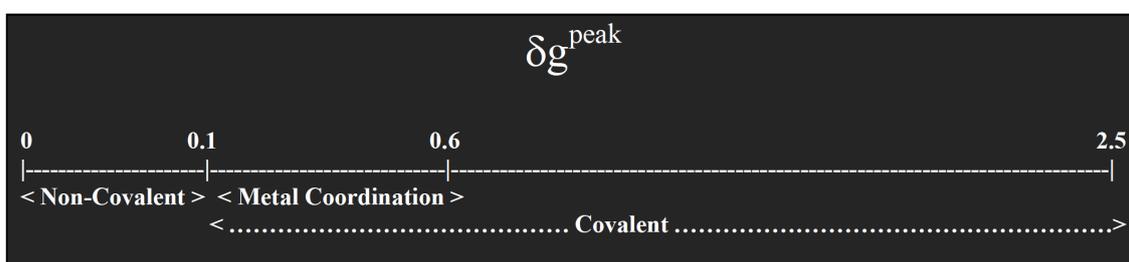
### 1.2.5 Family of Interaction: $\delta g$ peak

The  $\delta g$  descriptor is not dimensionless, linking its value to the familiar concept of interaction intensity. Hence, the  $\delta g$  peak height allows for ranking the interactions in categories. Based on our experience :

- weak non-covalent interactions never (or hardly) exceed  $\delta g$  peak heights of 0.1 a.u. (QM or promolecular treatments)
- vdW interactions do not extend beyond 0.02-0.03  $\delta g$  peak heights (QM treatment)

- H-bonding generally may extend up to a maximum of  $\delta g$  peak = 0.1 a.u. ( $\delta g$  peak for H-bond in water is 0.06 a.u. with a QM treatment)
- pure covalent bonding  $\delta g$  peak ranges from 0.2 up to around generally 1.0 in "common" molecules ( $\delta g$  peak for covalent O-H bond in water is 0.9), but it can be as large as 2.5 for exotic species like  $O_2^{2+}$
- metal coordination  $\delta g$  peak range between 0.1 and 0.6 a.u. (QM treatment)

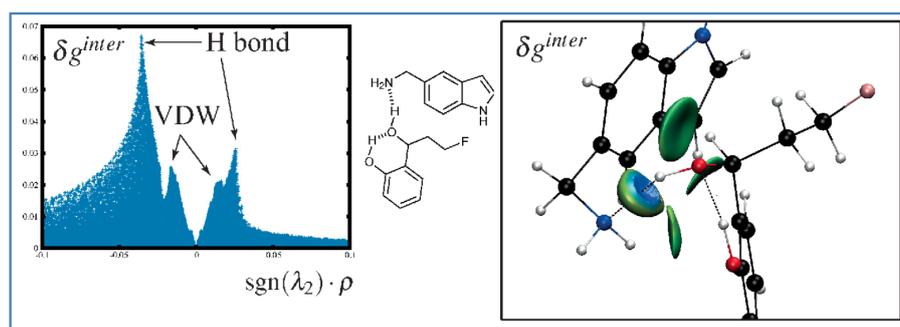
Based on these values, an indicative  $\delta g$  peak scale has been established (see Fig. 4). Just remind that this scale has been derived within the conventional IGM approach using the Gradient-Based partition of the ED gradient, not the Hirshfeld partition of the gradient (IGMH).



**Figure 4.**  $\delta g$  peak scale for indicative purposes

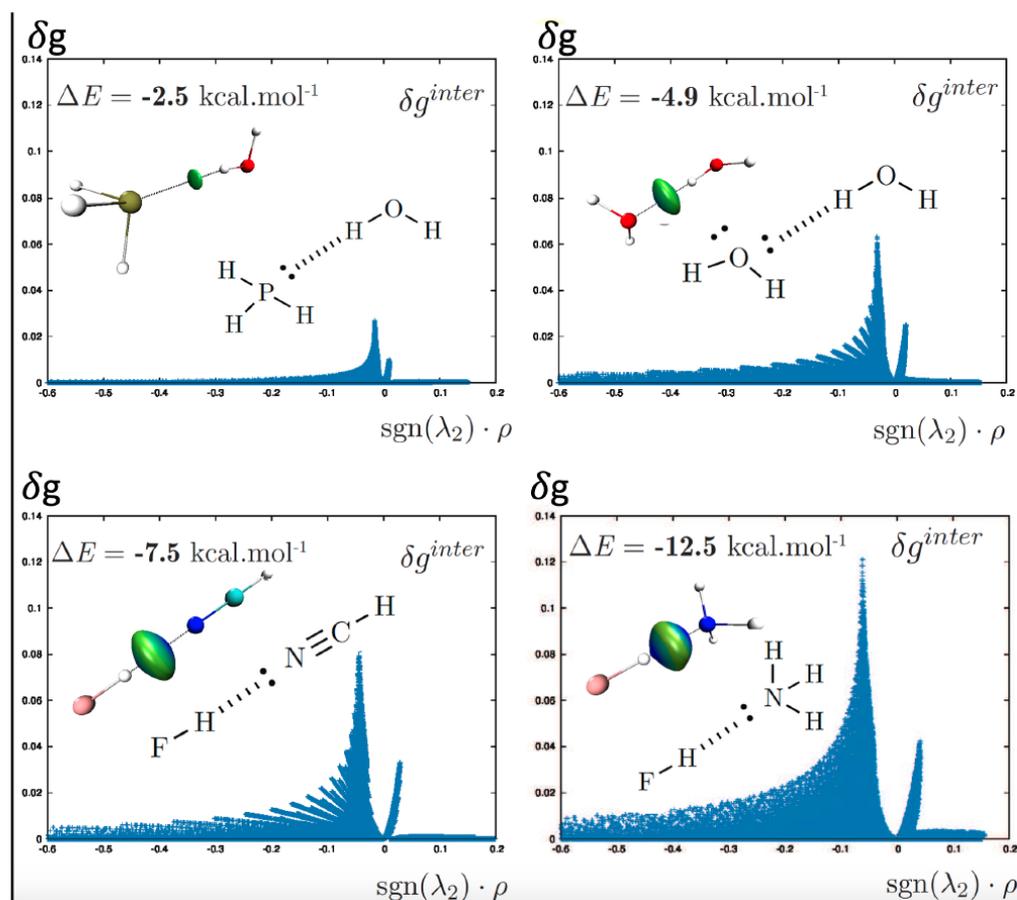
### 1.2.6 Interaction strength

As illustrated on Fig. 5. The hydrogen-bond spike emerges at a  $\delta g^{inter}$  value (around 0.065 a.u.) significantly greater than for van der Waals contacts (0.025 a.u.).



**Figure 5.** H-bond and vdW interactions; promolecular mode

In the series of 4 dimers (Fig. 6), we observe that the peak heights are related to the dimer stabilization energies obtained from CCSD(T) ab initio calculations (see Table below). The values fit relatively well to a linear correlation for  $\delta g^{inter}$ .



Complex	$d$ (Å)	$\rho$	$\delta g^{inter}$	$\Delta E$
$\text{PH}_3 \cdots \text{H}_2\text{O}$	2.65	0.016	0.027	$-2.5^a$
$\text{H}_2\text{O} \cdots \text{H}_2\text{O}$	1.93	0.032	0.060	$-5.0^b$
$\text{HCN} \cdots \text{HF}$	1.84	0.044	0.090	$-7.5^b$
$\text{HF} \cdots \text{NH}_3$	1.68	0.063	0.121	$-12.6^b$

<sup>a</sup> CCSD(T)-F12a/VQZ-F12. <sup>b</sup> CCSD(T)/CBS + REL + CV.

**Figure 6.** H-bond and vdW interactions; promolecular mode

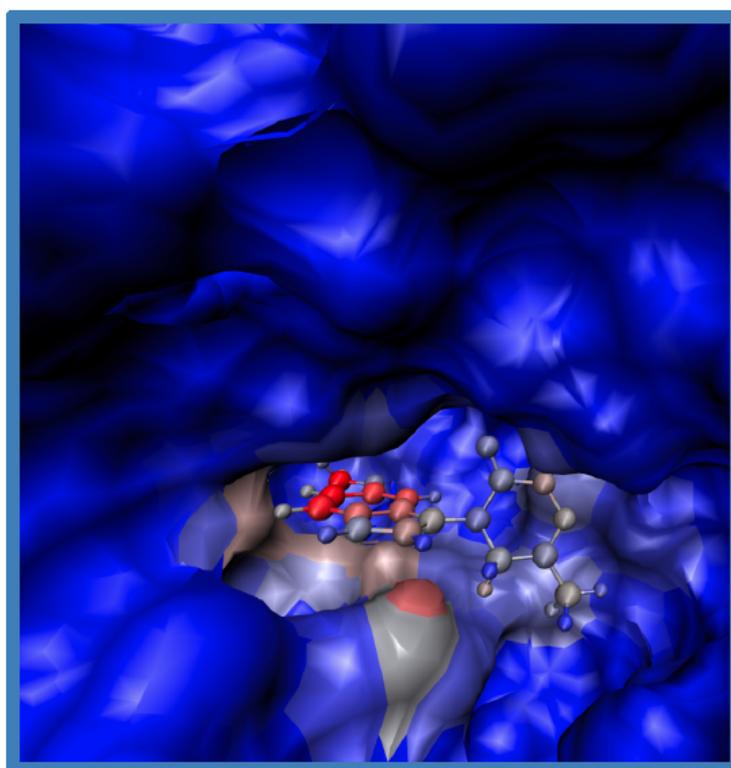
The IGM reference gives a unique definition of the interaction region: non-zero values of  $\delta g^{inter}$  exclusively correspond to interaction situations. As a consequence, the IGM approach overcomes difficulties to define regions within integration procedures to quantify interactions. An integration scheme has been proposed for the first time in 2020 to quantify intermolecular interactions (see the IGMPlot output on Fig. 7). It is explained in more details in the example section below.



The development of an even more accurate energy function is under progress, using machine learning possibilities. Also, the integration of  $\delta g^{intra}$  representing the intramolecular interactions present in fragment(s) is reported in the IGMPLOT output (not shown here).

### 1.2.7 Atomic decomposition scheme

In the both QM and promolecular modes, when two interacting fragments are defined, a compelling feature of the IGMPLOT program is to carry out an atomic decomposition scheme<sup>[4]</sup> in order to estimate the influence of a given atom in the intermolecular region between two user-selected fragments. A new  $\delta g^{inter/At}$  index has been derived from the IGM approach. Its value quantifies locally the contribution of atom X to the interaction between the two fragments A and B. It can be integrated over the IGM grid to get the associated relative atomic contribution, noted  $\Delta g^{inter/At}$ . Atoms are then colored according to their  $\Delta g^{inter/At}$  score (expressed as a percentage) to obtain a coupled analysis: isosurfaces/atomic participation, providing a rapid and rich picture of their role in the formation of the host-guest assembly. A VMD visualization state session AtContribInter.vmd input file is automatically prepared to run vmd program and to get these pictures.



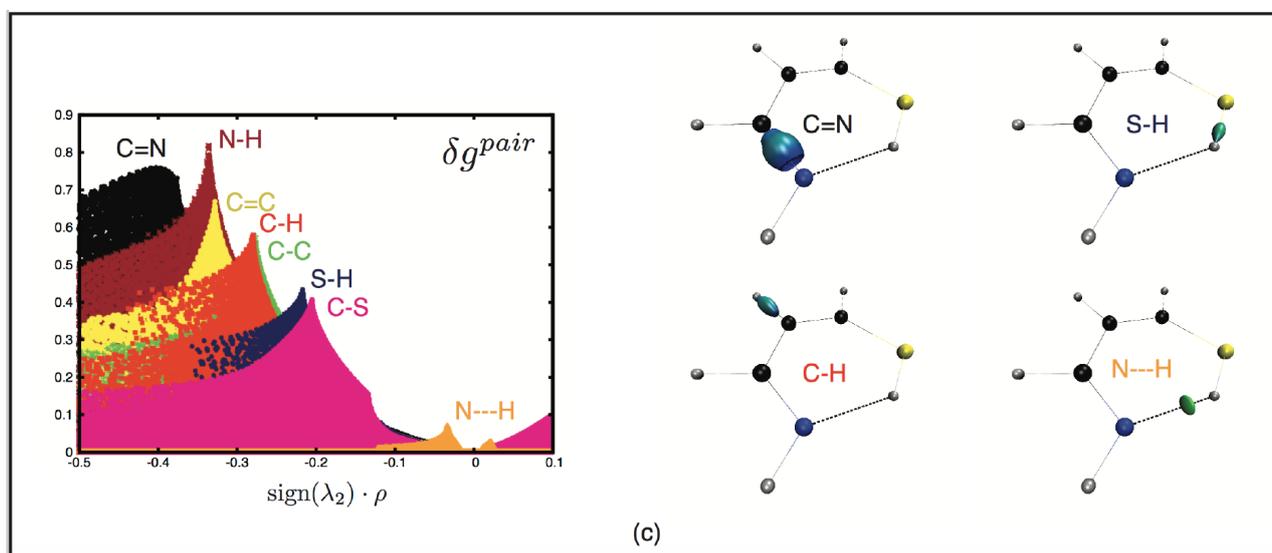
**Figure 8.** Ligand – Protein example; atoms colored according to their contribution to the ligand-protein intermolecular interaction

This complementary analysis brings an additional quantitative side to the initial IGM analysis. This possibility is illustrated on Fig. 8 where we have reported an analysis of non-covalent

interactions between a kinase protein and a ligand. The atoms of the complex are colored according to their contribution to the iso-surfaces (summing  $\delta g^{inter/At}$  over the grid and using a Blue-to-Red color scheme, with blue: no contribution to the interaction, red: largest relative contribution to the interaction).

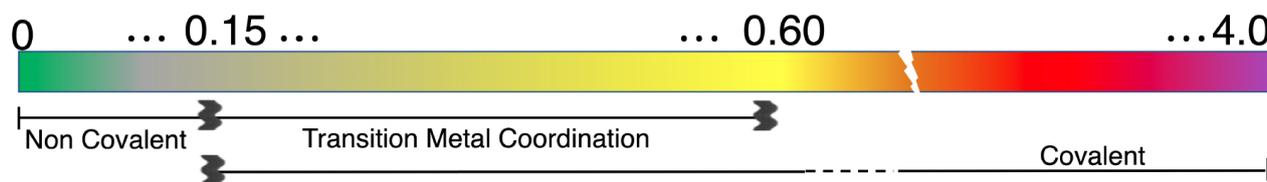
### 1.2.8 IBSI

The IBSI<sup>[5]</sup> (Intrinsic Bond Strength Index) is a score very efficient to internally probe the strength of a given atom pair in molecular situation. Closely related to local bond stretching force constant, it is not affected by negative transition state curvature nor ring constraint. It stems from the integration of the IGM- $\delta g^{pair}$  signature obtained for a given atom pair using the new bond descriptor  $\delta g^{pair}$  (see examples of such signatures on Fig. 9).



**Figure 9.** IGM is able to focus on a selected atom pair

Based on a large set of compounds (235 species, 677 atom pairs), we suggest to distinguish between weak interactions, coordinate covalent and covalent bonds according to this scale (the limits being only indicative on Fig. 10):



**Figure 10.** IBSI scale for indicative purposes

The IBSI does not belong to the conventional class of bond order (like Mulliken, Wiberg or Mayer, giving the number of electron pairs shared between two atoms), but rather assesses the



intrinsic bond strength. The IBSI is defined in a parameter-free manner and its implementation is very easy with IGMPlot: it only requires a wave function file as input. It has been shown little dependent on the level of theory and basis set. In addition to this score, the spatial representation of  $\delta g^{pair}$  isosurfaces reflect the common chemical concept of bond in a clear and intuitive manner, being distinctly different for single or double bonds. It paves the way for probing specific interactions like agostic interactions in transition metal compounds, three-centers two electrons bond, but also for targeted mechanistic exploration of reactions by monitoring selected covalent pair-interactions.

### 1.2.9 Pair Density Asymmetry (PDA)

In order to provide users with a simple tool to assess inductive effects on specific bonds in molecules, the new Pair Density Asymmetry (PDA) index has been devised.<sup>[1]</sup> The PDA index gives a measure of the ED asymmetry in between the two atoms and the direction of the asymmetry.

### 1.2.10 Critical Point Analysis

A critical point analysis is now made available within IGMPlot.<sup>[1]</sup> Quantities like the laplacian, kinetic energy density, local energy density, ... are extracted and reported in a separate file. Also, critical point positions can be viewed from a vmd session file automatically generated (see examples in the following).

### 1.2.11 Hirshfeld Gradient Partitioning

The Hirshfeld-based partition (HBP) of the electron density gradient is now made available to perform IGMH calculations.<sup>[8]</sup> It can be used as an alternative of the Gradient-based partition (GBP) to perform all the calculations and generate 2D-plots and 3D isosurfaces.

### 1.2.12 DOI = atomic Degree Of Interaction

We have introduced in 2023 the atomic Degree Of Interaction (DOI),<sup>[7]</sup> a new concept rooted in the electron density-based Independent Gradient Model (IGM). Capturing any manifestation of electron density sharing around an atom, including covalent and non-covalent situations, this index reflects the attachment strength of an atom to its molecular neighbourhood (see examples below for more details).

### 1.2.13 ELF

We have introduced in 2025 the novel approach ELF<sup>&</sup>IGM, which combines Electron Localization Function (ELF) and Independent Gradient Model (IGM) for intuitive electron structure analysis. Its speed and reliability make it valuable both for initial explorations and as a complement to conventional ELF analysis. The ELF<sup>&</sup>IGM framework serves as an accessible entry point for researchers interested in purely visual ELF analysis, enabling quick exploration of challenging systems ranging from transition metal complexes to large biological molecules such as proteins and DNA-protein assemblies.



## 2 Installation

IGMPLOT is written in C++. It has been installed and tested on several platforms: computational centers (linux), MacOS, Windows10, and several compilers and versions (GNU, Intel, PGI). It is a standalone program. Default is to use the OpenMP programming interface. OpenMP has to be installed prior to IGMPLOT for OpenMP compilation.

- General recommendations (Linux, MacOS, Windows):
  1. For OpenMP compilation, first, make sure having openMP installed on your machine (<http://www.openmp.org>).
  2. Uncompress the IGMPLOT archive: `'tar -xvjf IGMPLOT-x.y.tbz2'`. Change to directory `'IGMPLOT-x.y/source'`.
  3. Edit the Makefile and select:
    - the family of your C++ compiler (CppCompilerFamily): three choices are proposed (GNU, INTEL, PORTLAND)
    - the version of your C++ compiler (CppCompilerVersion)
    - if you want to use OpenMP or not: the use of openMP can be disabled in the file 'Makefile' by (1) commenting out the line `'OpenMP=YES'` while (2) uncommenting `'#OpenMP=NO'`.
    - the path to your compiler (including its name)
- Make sure having the appropriate compiler environment. Then type `'make'` in a terminal. The executable name is `'IGMPLOT'`. Make sure that this new command can be reached (adjust your PATH variable).
- On MacOS machines, a sequential version of IGMPLOT can be obtained with the Clang compiler. In the Makefile choose the options:
  1. CppCompilerFamily=GNU
  2. CppCompilerVersion=5\_and\_above
  3. OpenMP=NO
  4. CC=g++
- On MacOS machines, to leverage OpenMP multicore execution, you must install a gcc (g++) version different from the one provided within the compiler front end "Clang" which until now has not built-in support for OpenMP. You might install gcc with the command: `'brew install gcc'` (or `'brew upgrade gcc'`). Prior to that, make sure having installed the latest Xcode and Command Line Tools for Xcode. This way, the g++ compiler will be installed somewhere like `/usr/local/Cellar/gcc/7.1.0/bin/g++-9`. In this example, make sure the `g++-9` command be available with your PATH and adjust the IGMPLOT makefile accordingly :
  1. CppCompilerFamily=GNU



2. CppCompilerVersion=5\_and\_above
  3. OpenMP=YES
  4. changing the g++ command with g++-9 (CC=g++-9 for instance).
- Windows: only the Windows10 version has been successfully tested. The linux bash shell on Windows10 has been employed. Indeed, a full Ubuntu-based Bash shell that can run Linux software directly on Windows is now available on Windows 10. It isn't a virtual machine. It allows you to run the Bash shell and the exact same binaries you would normally run on Ubuntu Linux. After having enabled the "Windows Subsystem for Linux (Beta)", and installed a gcc developping environment (gcc, makefile), just follow the above-mentioned recommendations (and finally, type 'make').

IGMPlot is supplemented with the small shell script "vmdpath". This shell converts .vmd visualization states produced by IGMPlot (igm.vmd, nci.vmd, atContr.vmd) to a new local.vmd file by automatically adjusting file paths to current path (were cube files have really been stored). This is useful since, in the visualization state files (.vmd), absolute file paths are used (which can differ from the real ones after downloading the result from a distant computational center for instance).

### 3 Running the program

IGMPlot needs a single parameter input file param.igm. In addition to specify the filename describing the molecule (xyz, or wfn or wfx), this file contains specific IGMPlot keywords. The list of keywords is described hereafter. Example jobs are made available in directory 'samples'. You can run these examples individually or make use of the run.sh shell to execute them sequentially. To execute IGMPlot, in a terminal, type "IGMPLOT param.igm > igm.log" to run the job. In the OpenMP mode the default is to use the maximum of CPU cores available on the machine. This can be changed by setting the OMP\_NUM\_THREADS variable. For instance (linux, MacOS), in the terminal, type: export OMP\_NUM\_THREADS=8. Note that a "runinfo" file is generated after a while and regularly updated to display timing information (Fig. 11).

```
Current TIMING Info
TOTAL grid points n1.n2.n3: 2.60e+04 Completed: 1.27e+04 Remaining: 1.33e+04
TOTAL grid points n1.n2.n3: 100.0% Completed: 48.8% Remaining: 51.2%
Elapsed time: 56 seconds. Estimated time remaining: 58 seconds.
```

Figure 11. "runinfo" file with timing information

### 4 Inputs - Outputs

Two "modes" are made available in IGMPlot: (1) using promolecular ED or (2) using ED coming from a single wave function file. IGMPlot will detect the running mode according to the file extension of the input file(s) you provide to describe the molecule: .xyz (promolecular ED



calculated by IGMPlot) or .wfn/.wfx (text files) or .rkf (ADF binary file) (quantum calculations).

The IGM- $\delta g$  approach can be applied to a single whole system or to two user-defined fragments. In the latter case, the user must provide the definition of two fragments whose interaction is studied. Depending on the mode you will choose (QM or PROMOLECULAR), the definition of the two fragments is made differently. In this case (fragmentation scheme enabled), both intramolecular interactions and intermolecular interactions will be characterized.

Two kinds of files are needed to provide IGMPlot options (param.igm) and molecular features (xyz or wfn or wfx or rkf).

#### 4.1 the parameter input file param.igm

- The first line must contain the number  $n$  of files describing the molecular system. Currently,  $n$  is an integer in the range [1-2].  $n=1$  if you intend to use a wave function file (wfn or wfx or rkf).  $n=1$  or  $n=2$  if you intend to use promolecular density (xyz mode). According to the filename extension supplied immediately after (.xyz or .wfn or .wfx or .rkf) IGMPlot will appropriately proceed with a promolecular or a wave function calculation.
- The second line must contain the name of the first .xyz coordinate file (or of the single wfn file or single wfx file or single rkf file).
- If 2 files have to be read (promolecular mode only), the second and third lines must contain the two names of the first and second .xyz files. Example, minimal input:

```
• 2
  ligand.xyz
  protein.xyz
```

!!!! In promolecular mode, default is to build the IGM grid around the smallest molecule to save CPU time. Nevertheless a buffer is applied around this box to encompass part of the second molecule as well.

- The rest of the keywords is optional.

#### 4.2 .xyz or wfx, wfn, rkf molecule description files

- QM mode: one single file has to be supplied to specify both the geometry and the wave function (.wfn or .wfx or .rkf). In the QM mode, the definition of the two fragments will be achieved through specific keywords specified in the param.igm input file (FRAG1, FRAG2 or CUBEFRAG). Otherwise, the system will be considered as a whole ( $\delta g^{inter}=0$ ). For more details on the WFX file format specification go to the website of AIMAll program (<http://aim.tkgristmill.com/wfxformat.html>). For more details about the adf.rkf binary content see <https://www.scm.com/doc/ADF/Appendices/TAPE21.html>.
- Promolecular mode:



- Two .xyz files to specify the geometry of the two fragments: both inter- and intramolecular interactions will be characterized
- One .xyz file : the system will be considered as a whole:  $\delta g^{inter}=0$ , intramolecular interactions will be characterized

In the promolecular mode, the definition of the two fragments is achieved through the number of input .xyz files supplied.

IGMPlot proceeds with inputs and outputs in a very simple way. Once the fragment definition is provided in the param.igm input file, IGMPlot only requires a very limited set of keywords in the standard case.

### 4.3 Number of input files

- Within the QM mode: a wave function is needed to derive the ED and then two input files are needed: one wfx (or wfn or rkf) file describing the wave function (and the geometry) and in addition, a 'param.igm' parameter file containing the keywords. The definition of the two fragments to be studied can be set in param.igm through the keyword FRAG1, FRAG2 or CUBEFRAG (see below).
- With promolecular ED: in the standard case three input files are needed: two xyz file describing the cartesian coordinates of the two interacting partners; in addition, a 'param.igm' file containing the keywords is needed. It is also possible to use the program to study non-covalent interactions within one single molecule: only one single .xyz input file is required in that case instead of two. The electron density (ED) is then calculated directly by IGMPlot using the promolecular approximation.

### 4.4 GTO or STO ?

- IGMPlot is able to compute Gaussian Type Orbitals (GTO) and their derivatives from standard WFN and WFX files. Orbitals s, p, d, f and g are handled.
- IGMPlot is able to compute Slater Type Orbitals (STO) and their derivatives when a binary file adf.rkf generated by ADF is supplied. Orbitals s, p, d, and f are handled.

IGMPlot will proceed with the appropriate treatment (GTO or STO) according to the file extension: wfn, wfx for GTO and rkf for STO. However, to obtain a detailed description of the 'wfx' or 'wfn' file format for a specific quantum chemistry software, you will need to refer to the documentation or manuals of that software.

### 4.5 WFN/WFX/RKF remarks:

- IGMPlot has been devised to deal with either Gaussian Type Orbitals (GTO) by taking information in a WFN file or WFX file or with Slater Type Orbitals (STO) by taking information in a binary file generated by the program ADF.
- the WFN file generated by the ADF utility 'adf2aim' must not be used by IGMPlot. Instead, it is dedicated to the third party program Xaim (<http://www.quimica.urv.es/XAIM>).



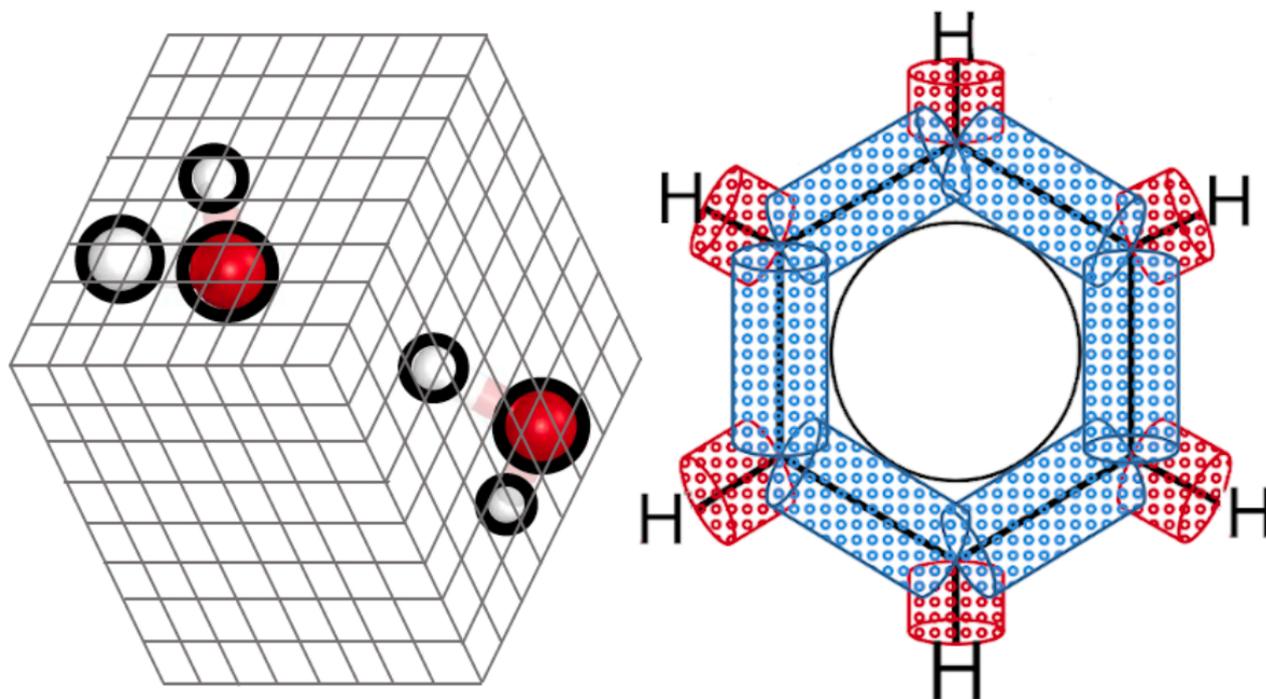
- the WFN file generated from the Gaussian program does not support ECP basis properly (the ECP data is treated as a second "basis set", and it is necessary to remove it manually from the wfn file so as to describe valence electrons only). You should prefer in that case the Ouput=WFX option in Gaussian to generate a WFX file used by IGMPlot.
- Note however that, upon using a WFX file, IGMPlot does not make use of the additional core density function data to represent the electron density of ECP-modeled core electrons (when pseudo-potentials are used). According to the AIMAll recommendations (<http://aim.tkgristmill.com>), for an atom for which a "small-core" or "medium-core" ECP is used, ignoring this core ED will not affect significantly the topology of the ED in the valence region and therefore the IGM- $\delta g$  analysis and critical point search can be performed properly. For an atom for which a "large-core" ECP is used, modelling the core ED cannot be ignored and the use of IGMPlot is not yet adapted to this situation.
- IGM calculations have been successfully carried out using much kinds of QM methods (HF, MP2, MP4, CCSD, ...) including semi-empirical methods like PM7, covering then single- and multi-determinant approaches. Actually, it is to be noticed that the IGM formalism holds within the natural orbital representation of the electron density allowing the implementation of the IGM approach to multideterminant wave functions, with fractional values of the molecular orbital occupation numbers reported in the WFN/WFX file.
- Restricted and open shell IGM calculations have been successfully tested; IGM has also been tested based on WFX files coming from TD-DFT excited state calculations.

**Outputs** Output files are generated within the IGM approach (details are given hereafter in the keyword section OUTPUT).

## 5 GRID details

IGMPlot employs a numeric grid. The default IGM grid is rectangular cuboid (Fig.12). It permits the generation of  $\delta g$  isosurfaces. For IBSI(bond strength) calculations, a cylindrical grid is employed, which only allows for IBSI and PDA score calculations.

There are three possibilities to position the IGM rectangular grid, with LIGAND, CUBE or RADIUS keywords. If none of these 3 keywords is given the default is to build this grid based on coordinates of the atoms (given in angströms). Within the promolecular mode using two .xyz coordinate files, the default is then to position the grid such as to encompass the first molecule specified in the param.igm input file (adding a 3 Å buffer around it). Within the promolecular mode using only one single .xyz file, the grid will encompass the whole molecular system. In QM mode (wfn or wfx) also, all the atoms will be enclosed by the grid. In every case, in order to ensure a correct grid in planar or linear cases, a minimum default buffer distance of +/- 3 Å is added to the axes in all directions. These defaults can be changed by using LIGAND, CUBE or RADIUS keywords to position the grid somewhere specifically.



**Figure 12.** The rectangular grid type is used for every calculation except for bond studies (IBSI) where an ultrafine cylindrical grid is automatically determined by IGMPlot for each specified bond

The IBSI calculation uses an ultrafine cylindrical grid. From version IGMPlot 3.0, the z-axis of the cylindrical grid covers the whole internuclear A-B axis (previously, the grid z axis range was  $[z_A + \epsilon : z_B - \epsilon]$ ,  $\epsilon < 0.02 \text{ \AA}$ ). Thus, very small changes might be occasionally observed on the IBSI values, but insignificant.

## 6 KEYWORD documentation

A set of options is available. The corresponding keywords are described below.

### 6.1 Optional keywords

Keywords marked with an asterisk \* are to be used only within the promolecular mode.

#### 6.1.1 \*LIGAND n r

Promolecular mode only. Only when two .xyz files are provided.

- n is the index of the xyz file used to position the grid. The order of introduction in the .igm file: 1 or 2, is used to designate the molecule (1 or 2) encompassed by the grid.
- r is a buffer distance in  $\text{\AA}$  (taken around the molecule used to define the calculation grid) Generally, within protein-ligand systems, the ligand is used as reference to locate the



grid because it is the smallest molecule. More generally, it is recommended to select the smallest molecule to lighten the calculation.

After reading the atomic positions the program determines the dimension of the grid increased by a "r" Å buffer that encompasses one of the molecules. Default xyz file index is 1 and default buffer is 3 Å.

### 6.1.2 CUBE x1 y1 z1 x2 y2 z2

- 6 real numbers (defining two points in real space)

This option is used to specifically set the IGM 3D rectangular grid by using the two extremities of a cube diagonal (x1,y1,z1) to (x2,y2,z2), all in Å.

### 6.1.3 RADIUS x y z r

- x y z determine the position around which interactions are represented for a radius r (all in Å).

The IGM 3D rectangular grid will be determined to encompass the given sphere of radius r and center (x,y,z).

### 6.1.4 ONAME name

- name stands for the prefix be passed to the output file names (default is 'mol').

### 6.1.5 OUTPUT n

- n is an integer in the range 1 – 5; default is 5.

1. will print these 2 files:

- nci.dat: 2-columns file (signed ED and RDG)
- igm.dat: 4-columns file (signed ED, RDG,  $\delta g^{intra}$ ,  $\delta g^{inter}$ )

2. will print these two .cube files (NCI mode only):

- dens.cube: signed ED
- RDG.cube: reduced density gradient

3. will print these three files (IGM mode):

- dens.cube: signed ED
- dgInter.cube:  $\delta g^{inter}$  values
- dgIntra.cube:  $\delta g^{intra}$  values

4. will print these three files:

- AtContribInter.vmd: vmd session file to display atomic contributions to inter-molecular interaction



- coord.xyz: xyz coordinates
  - AtContribInter.dat: atomic contributions to intermolecular interaction
5. will print 11 files: full output
- nci.dat
  - igm.dat
  - dens.cube
  - RDG.cube
  - nci.vmd: vmd session file to display RDG iso-surfaces
  - dgInter.cube
  - dgIntra.cube
  - igm.vmd: vmd session file to display separately  $\delta g^{inter}$  and  $\delta g^{intra}$  iso-surfaces
  - coord.xyz
- either (fragmentation scheme enabled):
- AtContribInter.vmd: vmd session file to display atom contributions to  $\delta g^{inter}$
  - AtContribInter.dat (atom contributions to  $\delta g^{inter}$ )
- or (no fragment defined):
- weakAtomContrib.vmd: vmd session file to display atom colored contributions to weak interactions only
  - AtomDOI.dat: degree of interaction of each atom in the molecule

### 6.1.6 INCREMENTS i1 i2 i3

- 3 real numbers

This option sets the increments (supplied in Å) taken along the x, y, z directions of the IGM rectangular grid. The default is set to 0.1, 0.1, 0.1 Å.

### 6.1.7 FRAG1 selection pattern (IGM, WFN/WFX mode)

QM mode only. Enables the  $\delta g^{inter}$  analysis in the QM mode.

- a selection pattern to specify the atoms belonging to fragment 1 (see examples below)

Example: FRAG1 2-5;7 → atoms 2,3,4,5,7 form the fragment 1.

### 6.1.8 FRAG2 selection pattern (IGM, WFN/WFX mode)

QM mode only. Enables the  $\delta g^{inter}$  analysis in the QM mode.

- a selection pattern to specify the atoms belonging to fragment 2 (see examples below)



Example: FRAG2 1;6;8-10 → atoms 1,6,8,9,10 form the fragment 2.

When only FRAG1 (FRAG2) is specified, the remaining atoms in the system form the fragment 2 (fragment 1).

Example, for a 10-atoms system : FRAG2 1;6;8-10 → Fragment 2 = atoms 1,6,8,9,10, Fragment 1 = atoms 2,3,4,5,7

When both FRAG1 and FRAG2 are simultaneously specified, then, only the atomic contributions of the sub-system made of FRAG1+FRAG2 are considered. This option is particularly useful to study the interaction between two small fragments of a very large molecular system. This option speeds up the calculation. Illustrative examples are given below.

### 6.1.9 CUBEFRAG (IGM, WFN/WFX mode)

QM mode only. Requires CUBE keyword to be also defined.

- Only atoms located in the rectangular grid defined by the CUBE parameters will be considered in the IGM analysis.

This option is particularly useful to study interactions within a given small region of a very large system. Then, the calculation is speeded up. Illustrative examples are given below.

### 6.1.10 Fragment definition summary

Five situations are considered: no FRAG definition, FRAG1 solely defined, FRAG2 solely defined, FRAG1+FRAG2 defined, CUBEFRAG definition supplied:

	CASE1	CASE2	CASE3	CASE4	CASE5
<b>Keyword</b>	no FRAG defined	<b>FRAG1</b>	<b>FRAG2</b>	<b>FRAG1+FRAG2</b>	<b>CUBEFRAG</b> requires <b>CUBE</b> or <b>RADIUS</b> keyword as well
<b>Fragment 1</b>	All atoms	<b>FRAG1</b> selection	remainder	<b>FRAG1</b> selection	atoms in grid
<b>Fragment 2</b>	0	remainder	<b>FRAG2</b> selection	<b>FRAG2</b> selection	0
<b>Primitives considered</b>	All	All	All	Primitives on Frag1 & Frag2 atoms	Primitives on atoms within the grid



### 6.1.11 IBSI r (IGM, WFN/WFX mode)

QM mode only. Measures the bond strength (IBSI<sup>[5]</sup>) and electron density asymmetry (PDA<sup>[1]</sup>) for given atom pairs. Requires the ENDIBSI keyword to be also defined at the end of the IBSI section.

- *r* : optional, designates a cut-off radius (in Å); primitives beyond *r* Å of the given atom pair are not considered in the IBSI calculation. *r* can be omitted (full calculation in that case).

Example:

```
IBSI 4.0
1 2
12 87
6 3
ENDIBSI
```

This IBSI example will calculate the intrinsic bond strength index (IBSI) for the atom pairs: 1-2, 12-87, 6-3, using primitives within a distance cut-off of 4.0 Å of the atom pair considered. Each IBSI index is calculated for a given A-B atom pair from equation 1 of paper:<sup>[5]</sup>

$$IBSI = \frac{\int_V \frac{\delta g^{pair}}{d^2} dV}{\int_V \frac{\delta g^{H_2}}{d_{H_2}^2} dV}$$

$\delta g^{pair}$  is the descriptor quantifying the ED contragradience between atomic sources A and B. *d* is the internuclear distance between A and B. IBSI has been normalized to 1 for  $H_2$  at the M06-2X/6-31G\*\* level of theory. IBSI has been shown to be little dependent on the level of theory, but we advise not using the HF method nor the STO-3G basis set.

We recommend the DFT M06-2X/6-31G\*\* level of theory for its performance/price ratio and using the same method/basis set for comparative studies. Adding diffuse functions can be relevant only for weakly bonded systems. In extreme cases, in very polar bonds like LiH, AlO, AlCl, MgC, ... a larger dependence can be observed.

Unless the radius *r* is specified, the calculation of IBSI for a given atom pair accounts for all the primitives of the molecular system. Otherwise, only primitives located within *r* Å of the atom pair will be accounted for. This option is particularly useful to speed up the calculation for large systems. Actually, an ultrafine cylindrical grid is built around each bond in this calculation, which can be quite expensive. The effect of the radius on the result accuracy is exemplified below (in the section examples below).

**Note that since version 3.04, the cylindrical axis exactly matches the distance between the two atoms, while previously, a small empty gap was left between each atom of the atom pair and the beginning of the cylinder. This can lead to small differences with the numerical IBSI values produced by the previous version of the software.**

### 6.1.12 \*INTERMOLECULAR (NCI model)

Promolecular mode only

- $r$  in the range 0-1. For good results,  $r$  should be 0.8-0.9.

In the original NCIPLOT code it is possible to discard the grid nodes for which more than a fraction (default threshold value is 0.95) of the total promolecular density comes from only one molecule (A or B). Typically, this turns off the intramolecular interactions in the resulting files for NCI calculations only. IGM result files won't be affected by this parameter.

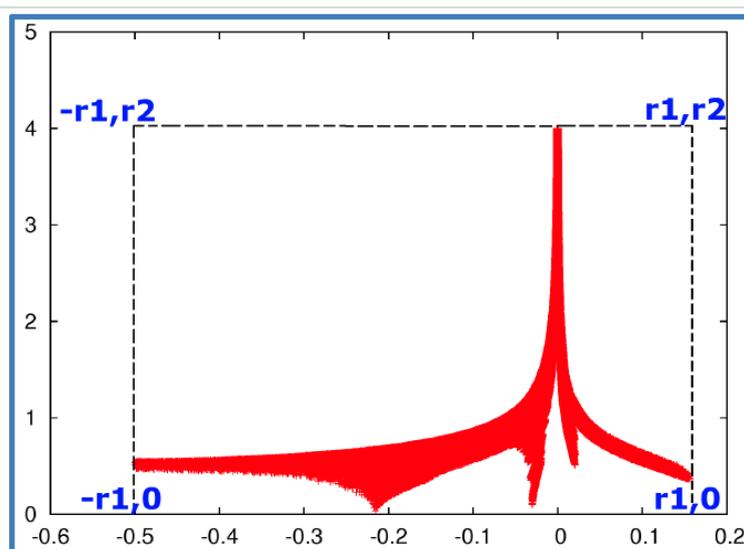
### 6.1.13 CUTOFFS $r_1$ $r_2$ (NCI + IGM models)

- 2 real numbers

Defines the density ( $r_1$ ) and RDG ( $r_2$ ) cutoffs used to print .dat files:

- $\text{sign}(\lambda_2)\rho$  range:  $[-r_1 \ r_1]$ : impacts both nci.dat and igm.dat
- RDG range:  $[0 \ r_2]$ : impacts nci.dat

$r_1$  default is 0.3 u.a.  $r_2$  default is 10.0 u.a.



### 6.1.14 CUTPLOT $r_3$ $r_4$ (NCI model)

- 2 real numbers; only concerns the NCI outputs (impacting both cube and vmd session files)

$r_3$ : density threshold, the unique criterion used to filter values written to the RDG cube file. Points with  $\text{sign}(\lambda_2)r$  in the range  $[-r_3 \ r_3]$  will be stored in the cube file.

$r_4$ : the RDG iso-value used in the nci.vmd session script generated by IGMPlot; RDG iso-surfaces will be colored according to a BGR scheme over the ED range  $-r_4 < \text{sign}(\lambda_2)\rho < r_4$  a.u.

This option might be useful to display 3D iso-surfaces coming from a specific peak-region previously identified on the 2D plot. Default values are:  $r_3=0.3$ ,  $r_4=0.3$ .



### 6.1.15 CUTPLOT\_IGM r5 r6 (IGM model)

- 2 real numbers; only concerns IGM outputs (impacting the cube files content)

r5: density threshold used to filter values written to the  $\delta g^{intra}$  cube file. Default : 1.2. Points with  $\text{sign}(\lambda_2)\rho$  in the range  $[-r5 : r5]$  will be stored in the  $\delta g^{intra}$  cube file.

r6: density threshold used to filter values written to the  $\delta g^{inter}$  cube file. Default : 0.3. Points with  $\text{sign}(\lambda_2)\rho$  in the range  $[-r6 : r6]$  will be stored in the  $\delta g^{inter}$  cube file.

This option is normally not necessary since  $\delta g^{intra}$  and  $\delta g^{inter}$  cubes files are naturally separated within the IGM model as soon as a fragmentation scheme is employed.

### 6.1.16 PEAKFOCUSINTRA x y (IGM model)

- 2 real numbers; only concerns IGM outputs (impacting the cube files content)

This keyword is followed by two numbers x and y specifying the electron density window to be considered. Only those points of the grid for which  $x \leq \text{ED} \leq y$  will be printed to the  $\delta g^{intra}$  cube. x and y can be negative or positive numbers. This means that the new ED range [x:y] takes over the range defined by the CUTPLOT\_IGM keyword in that case. That way, you can generate  $\delta g^{intra}$  isosurfaces corresponding to the peak selected in the  $\delta g^{intra}$  part of the 2D-plot signature.

A new  $\delta g^{intra}$  integration is performed specifically within this [x:y] ED window and the resulting score is reported in the igm.log. Note that the other integration schemes reported in the igm.log file are not affected by this keyword.

### 6.1.17 PEAKFOCUSINTER x y (IGM model)

- 2 real numbers; only concerns IGM outputs (impacting the cube files content)

This keyword is followed by two numbers x and y specifying the electron density window to be considered. Only those points of the grid for which  $x \leq \text{ED} \leq y$  will be printed to the  $\delta g^{inter}$  cube. x and y can be negative or positive numbers. This means that the new ED range [x:y] takes over the range defined by the CUTPLOT\_IGM keyword in that case. That way, you can generate  $\delta g^{inter}$  isosurfaces corresponding to the peak selected in the  $\delta g^{inter}$  part of the 2D-plot signature.

A new  $\delta g^{inter}$  integration is performed specifically within this [x:y] ED window and the resulting score is reported in the igm.log text output. Note that the other integration schemes reported in the igm.log file are not affected by this keyword.

### 6.1.18 VMD\_COLRANG\_IGM r7 r8 (IGM model)

- 2 real numbers; only concerns the VMD visualisation session file generated by IGMPlot (the way the isosurfaces will be colored)

r7:  $\delta g^{intra}$  iso-surfaces will be colored according to a BGR scheme over the ED range  $-r7 < \text{sign}(\lambda_2)\rho < r7$  a.u.; default = 0.3.



r8:  $\delta g^{inter}$  iso-surfaces will be colored according to a BGR scheme over the ED range – r8 <  $\text{sign}(\lambda_2)\rho$  < r8 a.u.; default = 0.08. Remark: the isovalues used to build the  $\delta g^{intra}$  and  $\delta g^{inter}$

iso-surfaces are automatically estimated by IGMPlot based on the  $\delta g^{intra}$  and  $\delta g^{inter}$  peak height (isovalues are taken as 40% of these peak heights). This can be changed into the VMD session.

### 6.1.19 HIRSH (IGMH model)

The Hirshfeld partition of the ED gradient (HBP)<sup>[8]</sup> will be used during the IGM calculations (except IBSI, which is defined using the GBP partition), instead of the default gradient-based partition (GBP).

### 6.1.20 CRITIC (IGM model)

The search for critical points of the electron density is conducted (cp, points where the gradient vanishes). The list of critical points (bond, ring and cage critical points) and properties calculated at these points are reported in the file cp.txt: coordinates, ED, ED gradient,  $\delta g$ , qg,  $\lambda_1$ ,  $\lambda_2$ ,  $\lambda_3$  (three ED hessian eigenvalues), Laplacian, kinetic energy density G, potential energy density V, local energy density H, ellipticity for (3,-1) critical points. The file cp.vmd allows for visualizing cp colored according to their type in 3D space.

### 6.1.21 CRITICFINE (IGM model)

Same as CRITIC but using a finer grid to perform the promolecular search of seeds.

### 6.1.22 CRITICULTRAFINE (IGM model)

Same as CRITIC but using an ultrafine grid to perform the promolecular search of seeds.

### 6.1.23 CRITICADDSEEDS (IGM model)

The center of atom triplets (triangle formed with 3 atoms less than 15 bohr apart) are added to the seed list formed by promolecular IGM-qg approach and midpoint centers. This solution generate a much larger seed list and is consequently much time-consuming. CRITICADDSEEDS can be used in conjunction with CRITIC or CRITICFINE or CRITICULTRAFINE.

### 6.1.24 FULLAOACC

This keyword will enable the full accuracy level for the computation of the atomic orbitals.

Indeed, particularly when evaluating local descriptors from electron density, the computational bottleneck arises from the evaluation of atomic orbitals (AOs, and their derivatives) at each point of a numerical grid that covers the entire molecule. To accelerate this process, we have implemented the computationally efficient atomic orbital pruning procedure<sup>[10,11]</sup> originally proposed by D. Kozłowski and J. Pilmé and implemented in TopChem2. It is based on a



precomputed threshold for each atomic orbital. Specifically, we calculate a radial cutoff distance  $R_{threshold}$  for each AO, defined as the distance at which the integral of the AO's radial function:

$$\int_0^{R_{threshold}} \phi(r)\phi(r)dr = 0.999999 \quad (4)$$

This threshold represents the radius within which a large amount (depending on the threshold value) of the electron density of the AO is confined. During the grid evaluation, if the distance between the grid point and the atom hosting the AO exceeds this threshold, the AO is considered negligible and is skipped in the calculation, saving significant computational time. This approach leads to a substantial performance improvement in large-scale computations.

To perform computations within the full accuracy level, you can make use of the keyword FULLAOACC in order to avoid IGMPlot using the AO pruning procedure.

Until now, this procedure is only implemented for GTO (Gaussian Type Orbitals), not yet for the STO (Slater Type Orbitals) employed by ADF.

### 6.1.25 ELF

An ELF&IGM<sup>[12]</sup> analysis will be performed.

## 7 General remarks

### 7.1 Critical point search

IGMPlot is now able to perform a critical point analysis,<sup>[1]</sup> i.e. to search for those points in space where the electron density (ED) gradient vanishes. This kind of information is of importance within the Atom In Molecule (AIM) analysis proposed by Bader. The eigenvalues of the Hessian matrix ( $\lambda_1, \lambda_2, \lambda_3$ ) of the ED evaluated at these points determine their rank (number of nonzero eigenvalues) and their signature (the algebraic sum of the sign of the eigenvalues). In molecules, most generally, the rank is 3 with four possible signatures:

- (3,-3): all curvatures of the ED are negative,  $\rho$  is a local maximum, associated with the nuclei position
- (3,-1): two curvatures are negative and one is positive, associated with so-called "bond critical points; ED is accumulated at the cp in the plan perpendicular to the internuclear axis"
- (3,+1): two curvatures are positives and one is negative, associated with so-called "ring critical points"
- (3,+3): the three curvatures of the ED are positive, associated with so-called "cage critical points"

For bond critical points, properties like: ED, its Laplacian, and other quantities, are generally calculated to characterize the bond.



Since such cp are not systematically some maxima of the ED, nor minima, finding their position requires performing algorithms like the Newton-Raphson method. But, when the point chosen as the initial point is not in the interval where the method converges, the cp search fails. Thus, one of the difficulties is then to choose appropriate starting points, called "seeds" in the following. The automatic search of critical points (characterized by  $\nabla\rho = 0$ ) performed by IGMPlot proceeds in two steps.

- seeds are collected in two different ways, before to perform a Newton-Raphson search starting from these seeds. An original feature of IGMPlot concerns the seeding step. Generally, programs performing critical search choose seed points at the midpoint between atom pairs, completed with points at the center of atom triplets (triangles), or completed by other strategies in order to avoid missing critical points, which can lead to time consuming cp search. It turns out that bond critical points, ring critical points and cage critical points results from ED contragradience between two sources located along one direction, several sources located in a two-dimensional plane, or multiple sources forming a cage framework, respectively. All these situations are naturally detected by the IGM- $\delta g$  descriptor taking maximum values at critical points. So, a fast promolecular IGM- $\delta g$  treatment is performed on a grid encompassing the molecule to find out regions of large values of  $\delta g$  serving as seeds for the next Newton-Raphson step. More precisely, qg maximum values are identified on the grid (qg is much more sensitive than  $\delta g$ ). Theses seeds are completed with only atom pair midpoints of pairs of atoms less than 15 bohr apart. This seeding strategy is very fast and turns out to be fast and to provide very good results.
- Finally, a Newton-Raphson algorithm has been implemented to search for cp starting from these seeds.

## 7.2 Rotational invariance

In the previous implementation of IGM, a small but existing dependence of the integrated  $\delta g$  score on the molecule orientation in the coordinate system was observed, which could amount to a few % to 10–15% in our studied systems. Note that however, by definition, the IBSI score was not affected by this effect since a cylindrical grid is applied in a same way to every atom pair. But other integrated score like  $\delta g$ ,  $\delta g^{intra}$ ,  $\delta g^{inter}$ , or the DOI (degree of interaction), could be influenced. The new implementations IGMPlot\_3.0 and IGMPlot\_3.08 get rid of this drawback for both QM treatments and promolecular, respectively. The idea is to work locally in the new reference frame defined by the three eigenvectors of the electron density hessian such that chemically equivalent points are strictly treated on an equal footing.

This way, all chemically equivalent atoms (i.e. related by symmetry operations) are rigorously treated equivalently. For instance, considering the benzene molecule (considered as a whole piece here), in orientation 1 or 2 (as illustrated on Fig. 13), leads to the integrated scores reported in Table 1.



Integrated score	Orientation 1	Orientation 2
$\Delta g = \int_V \delta g dV$	$1.326 \cdot 10^1$	$1.326 \cdot 10^1$
DOI for atom	orientation 1	orientation 2
C1	3.8383	3.8400
C2	3.8379	3.8390
C3	3.8386	3.8389
C4	3.8381	3.8388
C5	3.8380	3.8385
C6	3.8375	3.8391
H7	0.8444	0.8446
H8	0.8445	0.8444
H9	0.8441	0.8447
H10	0.8443	0.8445
H11	0.8444	0.8443
H12	0.8444	0.8445
IBSI for bond	orientation 1	orientation 2
C1 - C2	1.211	0.915
C2 - C3	1.211	0.915
C3 - C4	1.211	0.915
C4 - C5	1.211	0.915
C5 - C6	1.211	0.915
C6 - C1	1.211	0.915
C1 - H1	1.211	0.915
C2 - H2	1.211	0.915
C3 - H3	1.211	0.915
C4 - H4	1.211	0.915
C5 - H5	1.211	0.915
C6 - H6	1.211	0.915

**Table 1.** Integrated scores for chemical equivalent atoms and bonds in the benzene molecule addressed in two different orientations; grid steps = 0.05 Å; results obtained at the DFT M06-2X/6-31G\*\* level of theory

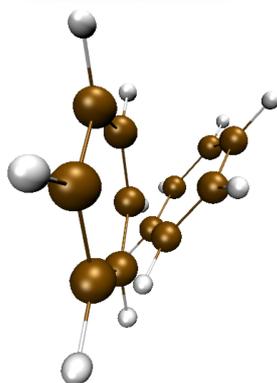
As can be seen:

- in a given orientation, chemically equivalent atoms are treated equivalently
- re-orienting the molecule does not impact atomic or bond scores
- the molecular integrated score  $\Delta g$  does not depend on the molecule orientation

So, the IGM integrated scores are rotational invariant and the symmetry of the molecule is fully respected. This treatment is now carried out both within the Quantum-Mecanical and promolecular modes.

**This may lead to small numerical differences of integrated scores compared to the previous releases of IGMPlot.** The very tiny differences which still appear in the above table result

from the use of a numerical grid with given grid increments. Very tight geometry optimization in multiple orientations and ultrafine numerical integration of two-electrons integrals in QM packages may help to reach a still better accuracy (although it is generally not necessary to derive IGM properties).



**Figure 13.** benzene molecule in two different orientations

### 7.3 Implementation of the IGM approach in Multiwfn

It is worth noticing that the promolecular version of IGM has been implemented in the Multiwfn software (<http://sobereva.com/multiwfn>)<sup>[13]</sup> in releases provided before 2021-Nov-8. After this date, the definition of the non-interacting reference and its norm have changed in Multiwfn compared to our original formula (see equation 2). Actually, in Multiwfn, the non-interacting reference for two interacting fragments A and B is now:  $|\nabla\rho^{IGM}| = |\nabla\rho_A| + |\nabla\rho_B|$ . In other words, this states that the norm  $|\nabla\rho^{IGM}|$  is the sum of the norms of the fragments  $|\nabla\rho_i|$ . This is another definition than the IGM one because it is not defined at the ED derivative level. This is an important point to us, since this is the presence of opposite signs in the real sum of the ED derivatives for each component:

$$\left| \frac{\partial\rho}{\partial x} \right| = \left| \sum_i \frac{\partial\rho_i}{\partial x} \right| \quad (5)$$

which is related to the drop in the norm of the real ED gradient  $|\nabla\rho|$  in the presence of chemical interactions, leading for instance to the drops observed in the RDG= $f(\rho)$  representation (RDG = reduced density gradient, a quantity introduced in DFT functional functionals as an inhomogeneity correction to the homogeneous electron gas). This is the idea behind the IGM methodology: quantifying these RDG drops by evaluating the presence of “contragradients” ED sources in each individual component:  $\frac{\partial\rho}{\partial x}$ ,  $\frac{\partial\rho}{\partial y}$ ,  $\frac{\partial\rho}{\partial z}$ . The original IGM definition of the non-interacting reference is carried out at the derivative level:

$$\left| \frac{\partial\rho}{\partial x} \right| = \sum_i \left| \frac{\partial\rho_i}{\partial x} \right| \quad (6)$$



and this grants that  $\delta g \neq 0$  only in interacting situation, while being exactly 0 when no opposite sign are present in the expression of ED derivatives.

In contrast, the Multiwfn implementation of IGM may yield  $\delta g \neq 0$  even in non-interacting physical situations. It might then lead to integrated scores of  $\delta g$  reflecting both interacting and non-interacting situations and then to misleading conclusions. The same flaw may appear in the 'QM' Multiwfn implementation of IGMH, the IGM approach using the Hirshfeld-based partition of the ED gradient instead of the Gradient-based partition.

Nonetheless, in practice, we observed only very small differences in the 2D-plot signatures and 3D-isosurfaces resulting from this 'reshaped' definition of the IGM definition. Note however that Multiwfn remains a useful tool for the calculation of other properties.

#### 7.4 Hirshfeld version Gradient Based partition

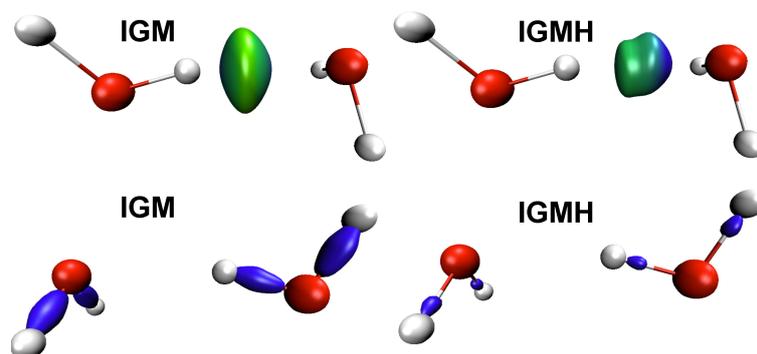
An alternative to the Gradient-Based Partition (GBP) has been proposed in the Multiwfn program to perform IGM calculation using a wave function description of molecular systems. It relies on an Hirshfeld-Based Partition of the ED gradient (HBP):<sup>[8]</sup>

$$\frac{\partial \rho_i}{\partial x} = \frac{\rho_i^{free}}{\rho^{pro}} \frac{\partial \rho}{\partial x} + \frac{\rho}{\rho^{pro}} \frac{\partial \rho_i^{free}}{\partial x} - \frac{\rho \rho_i^{free}}{(\rho^{pro})^2} \sum_j \frac{\partial \rho_j^{free}}{\partial x} \quad (7)$$

where 'free' stands for free atoms (spherically averaged ab initio electron density for the atom, not relaxed by means of an SCF procedure in molecular situation). Compared to our original Gradient-based partition, which only relies on the wave function information, this HBP alternative mixes promolecular (approximate) and wave function information. This hybrid QM/Promolecular HBP partition has been coded in IGMPlot and can be employed to perform IGM calculations ( $\delta g$ , DOI, atomic decomposition, ...).

Unfortunately, due to a bug of coding in Multiwfn, the IGMH/Multiwfn results initially presented in reference<sup>[8]</sup> are wrong. This error has been signaled in 2022 to the author of Multiwfn, who considered submitting a correction to J. Comp. Chem. journal. For more details see the erratum written by the very authors of IGMH.<sup>[14]</sup>

Using the proper Eq. 7 in IGMPlot leads to the result presented in Fig. 14 for the water dimer. Contrary to what is claimed in this paper:<sup>[8]</sup> " *The IGM maps ... have an evidently poorer graphical effect than the IGMH maps. The IGM  $\delta g^{inter}$  isosurfaces representing H-bonds and X-bond are bulgy and ugly ...*", it is quite opposite on the water dimer complex, as illustrated on Fig. 14:



**Figure 14.** GBP versus HBP ED gradient partition within IGM calculations; top :  $\delta g^{inter}$  0.02 a.u. isosurfaces; bottom:  $\delta g^{intra}$  0.5 a.u. isosurfaces

Concerning the covalent bonds O-H, isosurfaces obtained with the IGMH approach turn out to be smaller, much less "stretched" along the bond axis as illustrated. It is just like IGMH isosurfaces exhibit less "covalent" features.

In our view, the hybrid QM/Promolecular partition HBP of the ED gradient might be used to describe interactions in the covalent domain, but not employed to describe weak interactions like hydrogen-bonding. The comparison between HBP and GBP ED gradient partitions is still under investigation. But, we strongly advise using the GBP (default) ED gradient partition in IGMPLOT.

## 8 Examples

Examples given below can be found in the **sample** folder provided with this release. Note that, in addition to the IGM calculations, 'NCI' calculations (RDG descriptor) are also automatically performed within an IGMPLOT run.

### 8.1 Example 1 (promolecular density): to use one or two fragments ? (test1,test2)

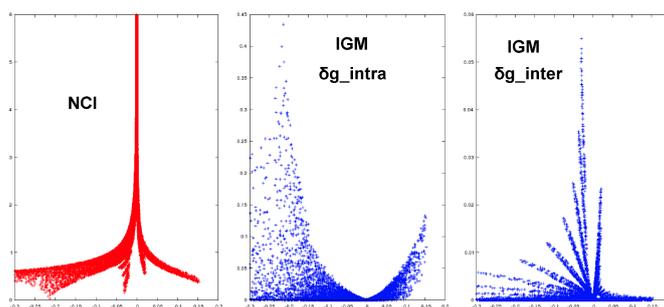
Water dimer: considered with two fragment.xyz files

The input param.igm file is reduced to 3 lines as shown below.

```
2
frag1.xyz
frag2.xyz
```

Type the following command in a terminal:  
IGMPLOT param.igm > igm.log

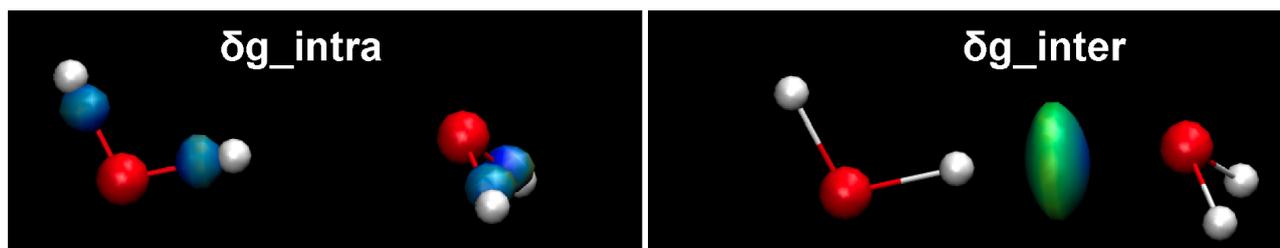
Using two fragment .xyz files in this example enables the IGM approach to automatically separate the intra/interfragment contributions. Using the generated file 'mol-igm.dat', we can construct the following 2D plots signatures (Fig.15), with gnuplot for instance: gnuplot> plot "mol-igm.dat" u 1:4 -> gives the  $\delta g^{inter}$  signature (frame of right below). Alternately, you can use the generated gnuplot input file, for instance: gnuplot mol-dgIntra.gnu.



**Figure 15.** water dimer; IGM- $\delta g^{intra}$  and  $\delta g^{inter}$  IGM 2D-signature (right, blue color); RDG signature (left, red color)

$\delta g^{inter}$  (for weak interactions) is generally one order of magnitude smaller than  $\delta g^{intra}$  (covalent). Columns 1, 3 and 4 in the mol-igm.dat output file represent the signed ED,  $\delta g^{intra}$  and  $\delta g^{inter}$ , respectively. The default increment has been used here (0.1 Å), which can be reduced to 0.05 Å for instance to get more accurate interaction signatures (using the INCREMENTS keyword). IGMPLOT has also generated the result cube files ('mol-dens.cube', 'mol-dgInter.cube' and 'mol-dgIntra.cube') to construct a colored 3D representation of these interactions using for instance the program VMD.

To make it easier the generation of these 3D representations using the VMD program, the igm.vmd session is prepared automatically for you by IGMPLOT: two representations (intra and inter) are then available in the VMD graphical interface, leading to the following Fig.16. IGMPLOT has automatically detected the maximum of  $\delta g^{intra}$  (and  $\delta g^{inter}$ ) peak values stored in the cube file such as to set the  $\delta g^{intra}$  (and  $\delta g^{inter}$ ) isovalue to 40% of this  $\delta g$  maximum value (in the generated vmd session file). Load this igm.vmd file from the VMD GUI (menu File/Load Visualization state).



**Figure 16.** Water dimer treated as two fragments;  $\delta g^{intra}$  and  $\delta g^{inter}$  are delivered separately by IGMPLOT;  $\delta g^{intra}$  0.174 a.u. with BGR color code in the range  $-0.3 < \text{sign}(\lambda_2)\rho < 0.3$  a.u. and  $\delta g^{inter}$  0.022 a.u. isosurfaces with BGR color code in the range  $-0.08 < \text{sign}(\lambda_2)\rho < 0.08$  a.u.

Don't forget to adjust manually (in this .vmd file) the absolute file path for vmd to access the data. Alternatively, you can use the "vmdpath" bash which is made available in the IGMPLOT distribution and which will transform automatically the .vmd file to adjust the absolute path to your current path before launching vmd.

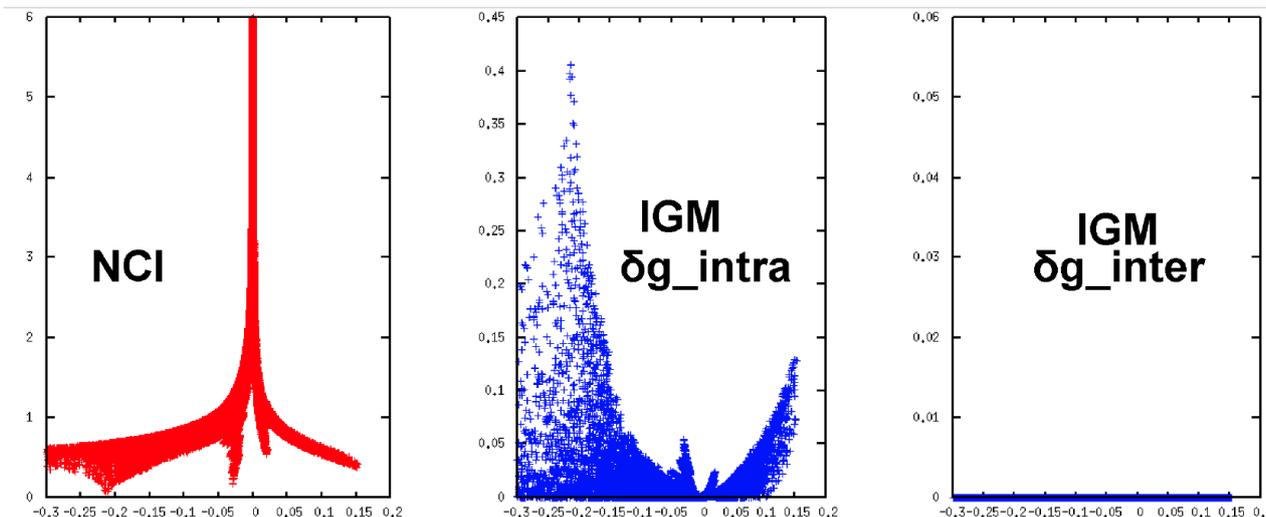
The  $\delta g$  index gives rise to a coherent picture with stabilizing interactions accumulated in the center of the envelop enclosing the bond critical point (BCP). Of course, depending of the

iso-value chosen for  $\delta g$  index, iso-surfaces will appear more or less large. The interesting thing is that using the "2-fragments mode" (by supplying two .xyz input files), the intermolecular interaction signature  $\delta g^{inter}$  is automatically extracted. Let us recall however that a promolecular ED cannot describe covalent interactions in a rigorous way. Thus, the  $\delta g^{intra}$  signature (2D or 3D isosurface) obtained with the promolecular mode is not really relevant in this case. Please, use an ED derived from wave function calculations instead to study strong interactions (see hereafter). However, in the low electron density (ED) domain, it has been shown (see for instance refs 2,3) that promolecular and quantum ED IGM signatures are very similar both in order of magnitude and in the peak height position; accordingly, the promolecular mode is very attractive for studying non-covalent interactions within large systems.

Water dimer considered with one fragment.xyz file

The input param.igm file is now reduced to 2 lines. Using the 'mol-igm.dat' file generated by IGMPlot, we can build (for instance using the gnuplot program) the following 2D plots (Fig.17). Within the IGM- $\delta g$  approach, using the single-fragment mode (supplying only one .xyz input file in promolecular mode) reveals ALL interactions on a single 2D-plot or drawing 3D-isosurfaces.

```
1
dimer.xyz
```

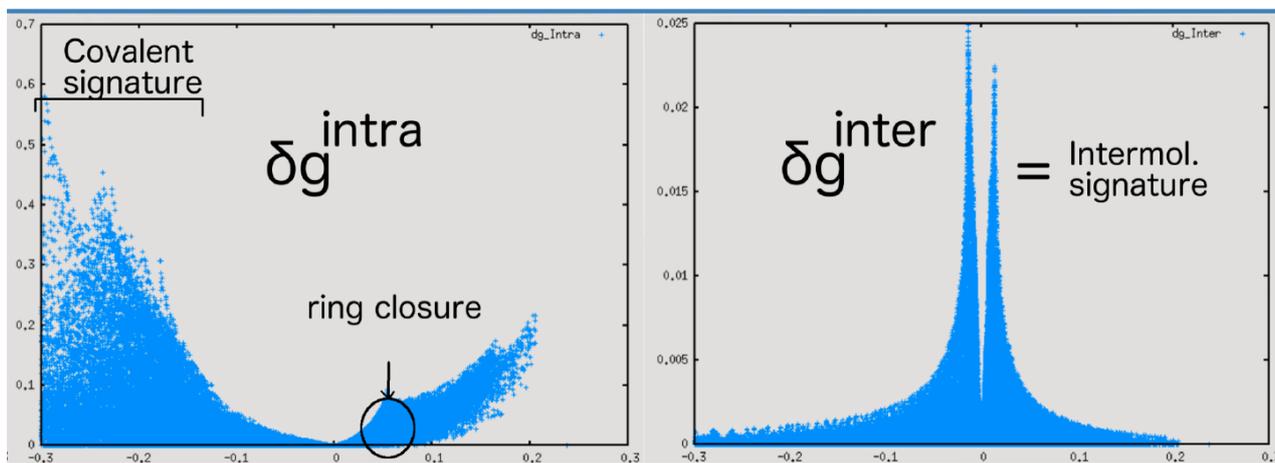


**Figure 17.** water dimer – a single .xyz file has been used;  $\delta g^{intra}$  (blue, middle) contains both intra and inter peak signatures in that case, and  $\delta g^{inter} = 0$  here (blue, right)

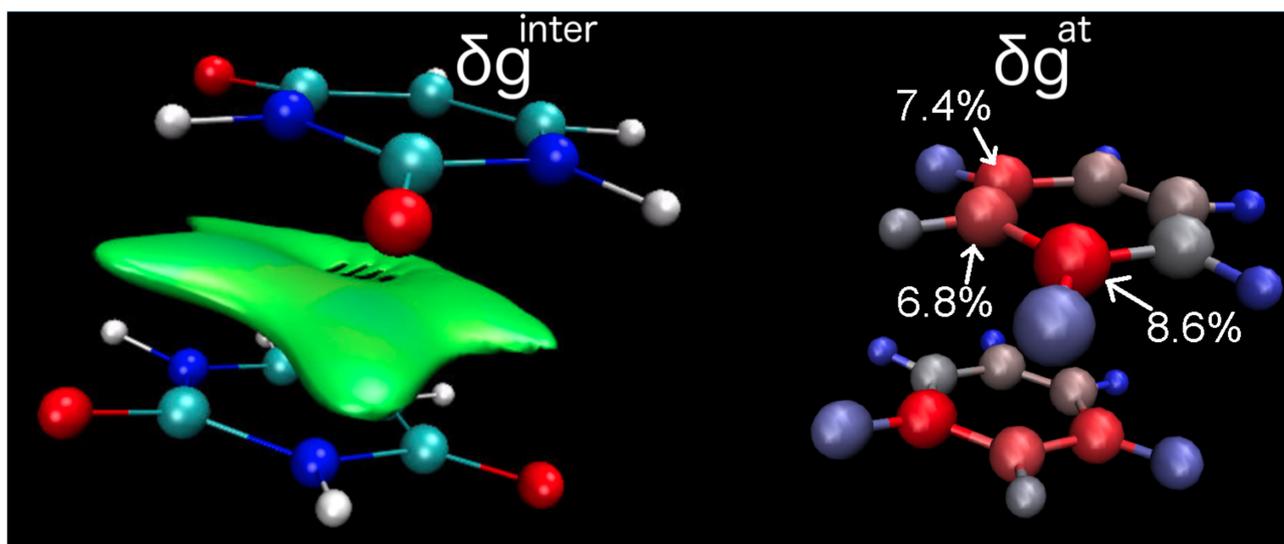
## 8.2 Example 2 (promolecular density): vdW interactions (test3)

van der Waals non-covalent interactions are also well detected by the IGM- $\delta g^{inter}$  approach. In the Uracyl dimer, thanks to the uncoupling "2-fragments" scheme and employing promolecular ED, intramolecular ring closure is naturally separated from intermolecular weak interactions occurring between the two monomers (Fig.18 and Fig.19).

```
2
frag1.xyz
frag2.xyz
```



**Figure 18.** Uracyl dimer  $\delta g^{intra}$  and  $\delta g^{inter}$  signatures



**Figure 19.** uracyl (parallel displaced) dimer;  $\delta g^{inter}$  0.010 a.u. isosurface (left) with BGR color code in the range  $-0.08 < \text{sign}(\lambda_2)\rho < 0.08$  a.u.; atomic decomposition scheme (right,  $\delta g^{inter/At}$ ) with BGryR color scale (Blue=0%, Red=8.6%)

A  $\delta g^{inter/At}$  decomposition scheme is automatically performed by IGMPlot in order to give an estimation of the contribution of each atom in the peaks appearing in  $\delta g^{inter}$  2D plots. After summing over the grid ( $\sum_{grid} \delta g^{inter/At}$  without distinction between the repulsive and attractive parts of the 2D plot), each atom is given a score (%) and coloured according to this



percentage. These data are stored to the 'AtContribInter.dat' file. This file is employed by the 'AtContribInter.vmd' VMD session file to generate the picture reported on the right of Fig.19. As can be seen, discrepancies occur between atoms showing that they contribute differently in the flat iso-surface representing the van der Waals interaction between the two monomers here. This tool can be considered as a complement to the  $\delta g^{inter}$  analysis.

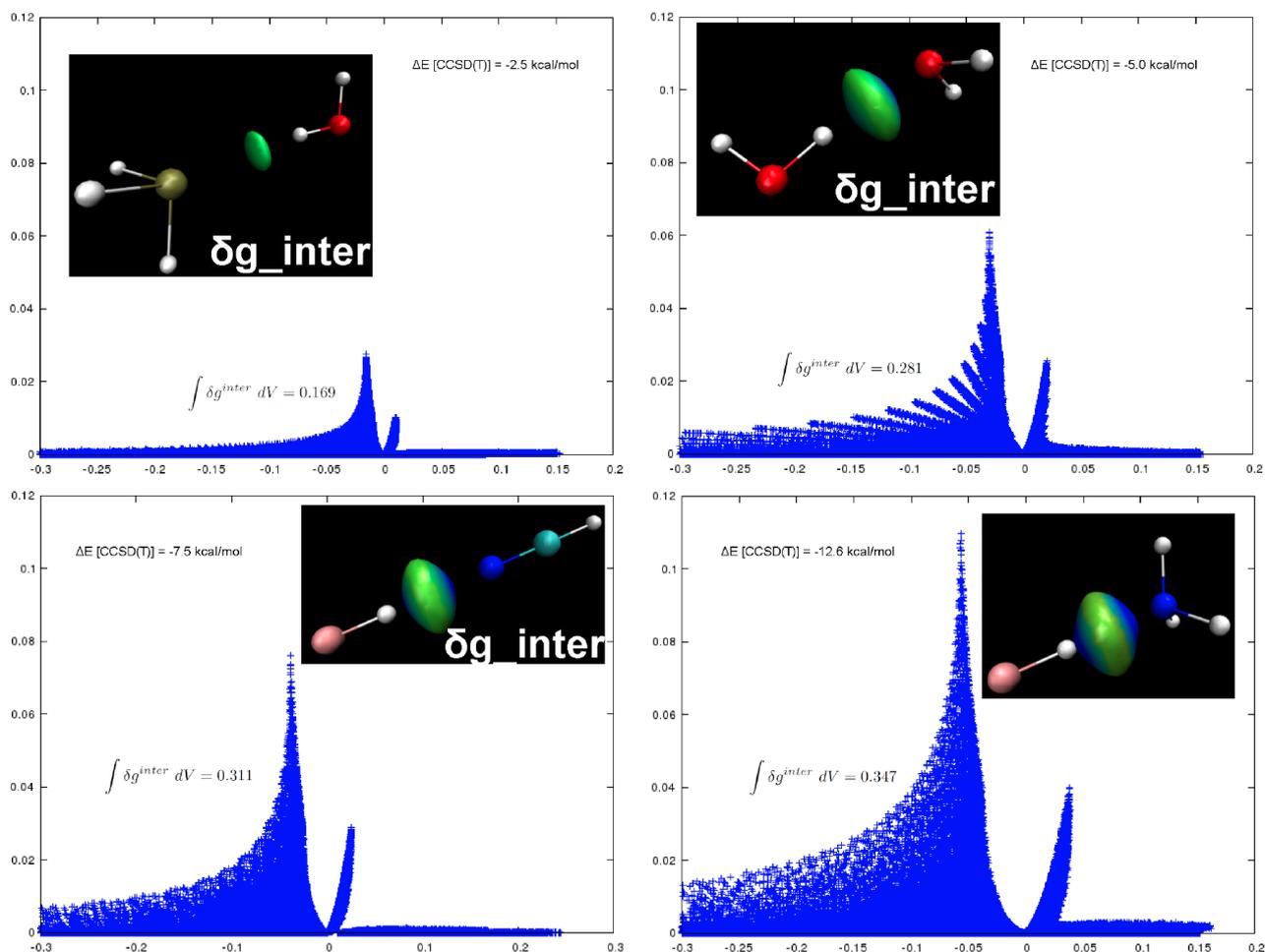
### 8.3 Example 3 (promolecular density): quantification of non-covalent interactions (test4 to test10)

Since non-zero values of  $\delta g$  exclusively corresponds to molecular interaction situations, the IGM- $\delta g$  approach is very convenient and suitable for integration schemes. Actually, within the IGM approach, the question of how defining the integration volume does not arise, one simply needs to sum  $\delta g$  over the entire grid. It is proposed in this release an integration scheme developed to relate the sum of  $\delta g^{inter}$  to the strength of interaction between fragments.

#### Link between energy stabilization and $\int \delta g^{inter} dV$ for selected dimers

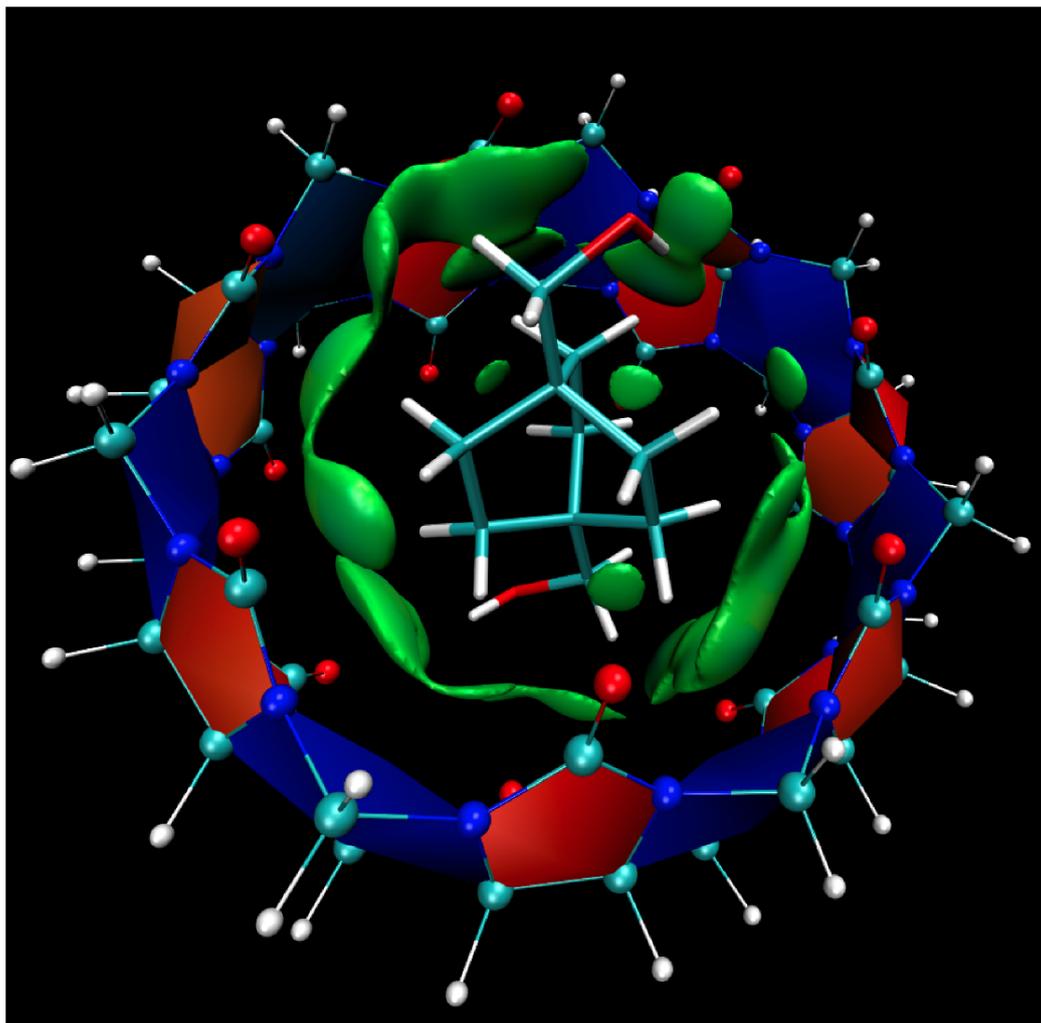
For comparison, four dimers of varying strength are described hereafter (see Fig. 18). First, we observe that the peak heights are related to the dimer stabilization energies at the CCSD(T) level of theory (see paper<sup>[2]</sup>). Furthermore, the integrated values  $\int_{entire\ grid} \delta g^{inter} dV$  fits rather well to a linear correlation for QM calculations.

Remark: Noteworthy, in the promolecular mode, some iso-surfaces might sometimes appear to be missing. For instance, in example test04, one O-H covalent bond does not appear in the igm  $\delta g^{intra}$  iso-surfaces generated with the igm.vmd script. This is because the grid employed for the analysis is built around the first fragment, surrounded by a small buffer only, not necessary encompassing the whole second fragment. This can be changed with the LIGAND keyword.



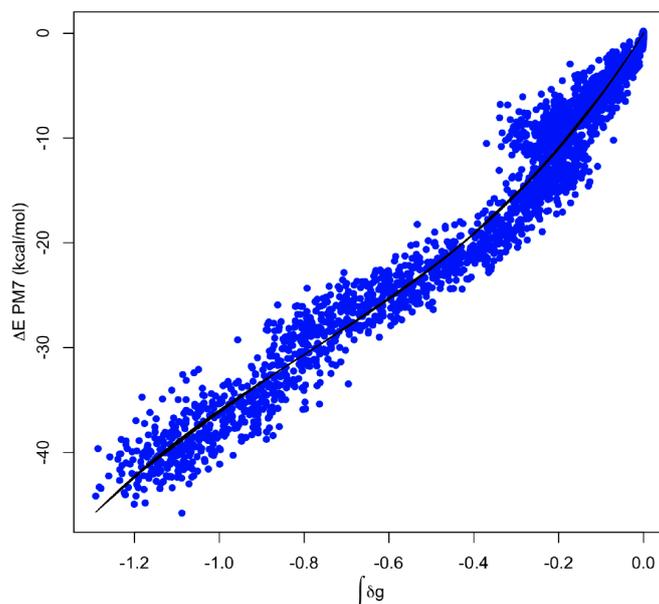
**Figure 20.** Study of 4 dimers;  $\delta g^{inter}$  0.015 a.u. isosurfaces with BGR color code in the range  $-0.08 < \text{sign}(\lambda_2)\rho < 0.08$  a.u.; the IGM calculation used a 0.05 Å grid steps

Towards the development of a scoring function from promolecular ED The idea of integrating the local descriptor  $\delta g^{inter}$  has been extended to a larger dimer along a molecular dynamic trajectory in an attempt to find out whether it was possible to establish a relationship between the integrated  $\delta g^{inter}$  and the stabilization energy assessed at the quantum mechanical level of theory. To do so, the interaction between the cucurbit[7]uril (CB7) with the guest bicyclo[2.2.2]octane (B2)<sup>[15]</sup> has been examined (Fig.21).



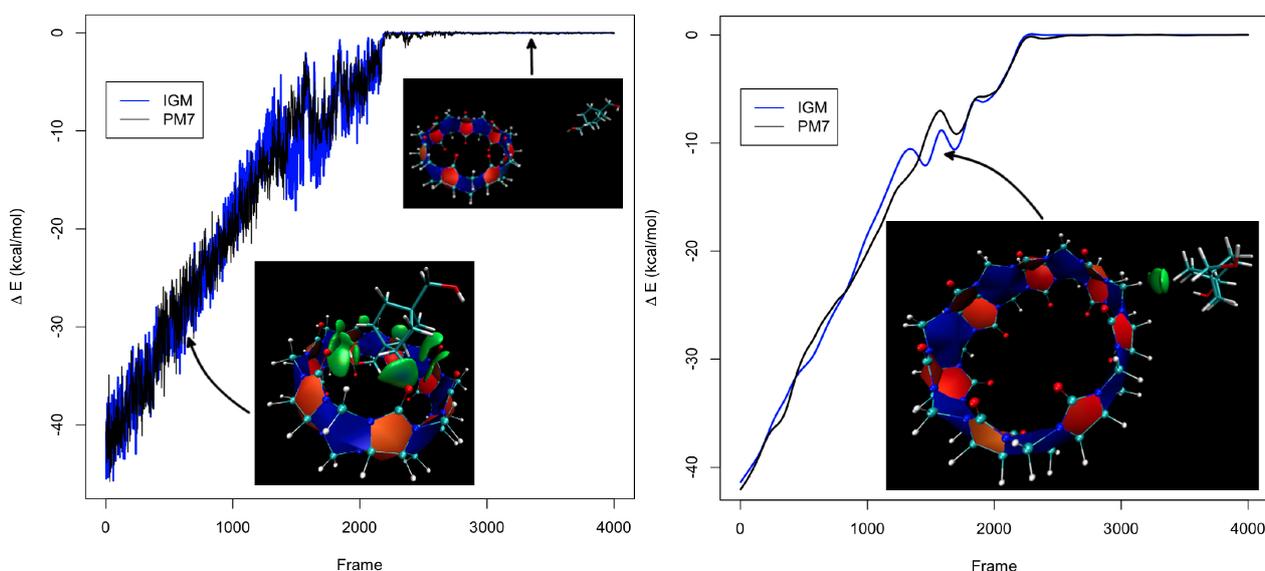
**Figure 21.** cage-guest B2...CB7 complex;  $\delta g^{inter}$  0.008 a.u. isosurface with BGR color code in the range  $-0.08 < \text{sign}(\lambda_2)\rho < 0.08$  a.u.

In order to explore an extend range of host-guest binding energies, a steered molecular dynamic was performed allowing us to observe a substrate being pulled from the binding cavity in explicit water at 300 K. From the trajectory, 4000 frames were extracted and subjected to both an IGM- $\delta g^{inter}$  calculation and a semi-empirical PM7 quantum calculation (three QM calculations are required to get the binding energy for a given dimer geometry). Several integration schemes have been considered. It turns out that by focusing on the peaks of the 2D-IGM- $\delta g^{inter}$  signature (by taking those points for which the ratio  $qg = \frac{\nabla_{rho}^{IGM}}{\nabla_{rho}}$  is greater than 1.2) and considering the grid points for which  $\lambda_2 < 0$  (attractive regions), a significant correlation is observed between  $\Delta E$  PM7 (kcal/mol) and  $\int \delta g^{inter} dV$  as illustrated on Fig. 20:



**Figure 22.** PM7 stabilization energy versus  $\int \delta g^{inter} dV$  obtained from the IGM-2-fragments approach reported for 4000 frames extracted from a steered molecular dynamics at 300K

A cubic polynomial expression has been obtained by regression that fits well with the PM7/IGM data ( $R^2=0.989$ ). This model has been used to get a simulation of the binding energy from the IGM- $\delta g^{inter}$  approach, leading to the next graph (Fig. 21) which shows the host-guest interaction energy along the studied molecular dynamic trajectory:



**Figure 23.** B2-CB7 interaction energy along the 2 ns steered MD trajectory; left panel: PM7 QM binding energy and integrated IGM- $\delta g^{inter}$  score, right panel: PM7 and IGM smoothed curves;  $\delta g^{inter}$  isosurfaces with BGR color code

As can be seen, the three stages of the substrate being pulled from the binding cavity are well recovered by the integrated IGM- $\delta g^{inter}$  score: (1) cage-guest interaction during the extraction (2) ligand-cage surface interactions before (3) the final release.

Applying this IGM- $\delta g^{inter}$  scoring function to the water dimer for instance leads to a consistent result ( $\Delta E = -5.5$  kcal/mol).

The IGM- $\delta g^{inter}$  integration scheme is not affected by the keywords CUTOFFS, CUTPLOT, CUTPLOT\_IGM (which only affect the final .dat and .cube outputfiles).

This "Inter. Energy crude Estimate" is a trial balloon, which only provides a crude estimate of the interaction energy in kcal/mol. Much must be done now to extend this estimation to other types of atoms. (extending the calibration range, accounting for steric repulsion). Also, so far, it is based on the existence of ED contragradience. Thus, long-range electrostatic interactions (which may not involve significant ED clash) cannot be taken into account (note that, in the presence of the water solvent with large dielectric constant, this issue is reduced). Also, this score is found to overestimate  $\pi$ -stacking interactions over hydrogen-bonding. This is a proof of concept, which is currently being improved using new machine learning possibilities employing a variety a different molecules to remedy these lacks and to provide a more general and still better fit with QM interaction energy. Nevertheless, the  $\delta g^{inter}$  integration reflects the strength of interaction. Of course, the IGM- $\delta g^{inter}$  approach using promolecular electron density cannot be expected to lead as accurate results as QM calculations, but it offers a fast and robust way to get an order of magnitude of interactions comparing different systems or during their evolution.

#### 8.4 Example 4 (promolecular density): Monitoring intramolecular interactions with IGM- $\delta g^{intra}$ in a peptide (test11, test12)

In addition to the IGM- $\delta g^{inter}$  local descriptor describing the interactions between two fragments A and B, the IGM approach also delivers the IGM- $\delta g^{intra}$  descriptor. It describes all the interactions present inside each fragment. Generally, these  $\delta g^{intra}$  interactions correspond to covalent bonding (Case 1 on Fig.24).

	Fragments considered	$\delta g^{intra}$	$\delta g^{inter}$
Case 1		> 0	> 0
Case 2		> 0	= 0
Case 3		> 0	= 0

**Figure 24.** Several scenarios of fragmentations within IGMPlot

When we consider the system A...B taken as one piece (one single fragment, Case 2 on Fig.24), the intra/inter uncoupling scheme automatically leads to IGM- $\delta g^{inter} = 0$  (at every grid node),

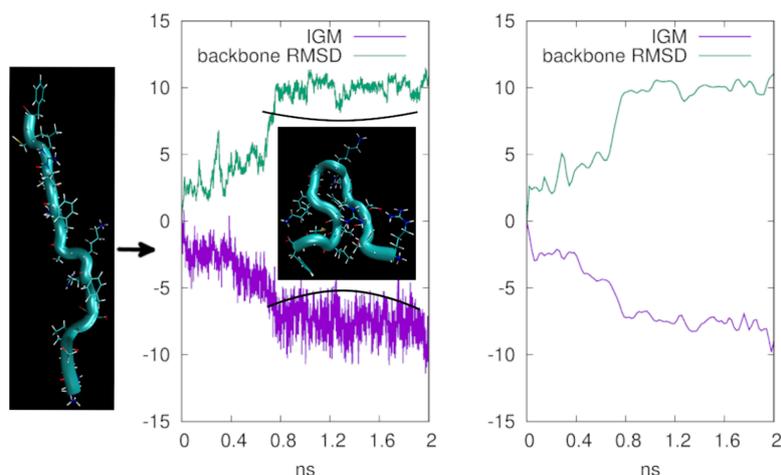


and  $\text{IGM-}\delta g^{\text{intra}}$  reveals every interaction present in the system: both covalent bonding inside A and B and weak interaction between A and B. Then, the  $\text{IGM-}\delta g^{\text{intra}}$  signature displays two domains, strong and weak interactions. Such a use (defining a single fragment) of the IGM approach does not make sense if you are interested in the intermolecular A...B interaction. In that case, it is much better to define two fragments to separate automatically the two signatures. A last case may occur when one single molecule involves both covalent and intramolecular non-covalent interactions (between C and D, Case 3). In that case, one may be interested in extracting and quantifying the intramolecular weak interaction part of the whole  $\text{IGM-}\delta g^{\text{intra}}$  signature. Indeed, for instance, it may be worthwhile to get information on folding patterns of a single peptide (like alpha helices and beta sheets), which results from intramolecular weak interactions between neighbouring amino-acids and makes up the secondary structure of a protein.

That is the reason why IGMPlot automatically provides (in the `igm.log` text outputfile) the attractive contribution to the weak interaction  $\text{IGM-}\delta g^{\text{intra}}$  signature. This way, the  $\text{IGM-}\delta g^{\text{intra}}$  can then serve to study energy changes in a single molecule which are only due to weak interactions.

This possibility (automatically extracting weak interactions in a single system considered as whole) is illustrated below during a molecular dynamic trajectory of a peptide. In this simulation, due to fluctuating weak interactions between side chains of amino acids, the non-covalent contribution to  $\text{IGM-}\delta g^{\text{intra}}$  changes. Then,  $\text{IGM-}\delta g^{\text{intra}}$  (weak interaction part) can be employed to monitor energy changes in a single molecule that results from non-covalent interactions.

The result is shown on Fig.25 which reports the back-bone RMSD of a peptide (PDB ID 1dep) and the evolution of the  $\text{IGM-}\delta g^{\text{intra}}$  (weak interaction domain) curve along a 2 ns MD trajectory at 300K. As can be seen, starting from a quasi-linear structure, the peptide adopts an organized structure (a turn) due to non-covalent interactions between amino acids. The RMSD plot reveals the presence of this organized state after 0.7 ns. Accordingly, the  $\int -\delta g^{\text{intra}} dV$  sum (weak interaction domain) decreases in the range [0: 0.7 ns] and reveals the gradual accumulation of stabilizing weak interactions, accurately reflecting the formation of this "turn" state.

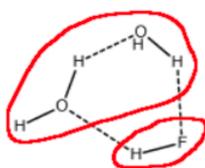


**Figure 25.** Membrane protein (PDB ID 1dep), sequence: ARG-SER-PRO-ASP-PHE-ARG-LYS-ALA-PHE-LYS-ARG-LEU-LEUCYS- PH; backbone RMSD ( $\text{\AA}$ , green) and  $\int -\delta g^{intra} dV$  (a.u., scaled by 5, weak interaction domain, purple curve) along a 2ns MD trajectory (smoothed curves on the right)

This example on a peptide proves that the integral  $\int -\delta g^{intra} dV$  obtained in the non-covalent domain and from promolecular electron density can be used to obtain meaningful chemical predictions so as to produce output data representative of the time evolution of the molecular system, with the added flexibility of focusing on a subdomain of a biomolecular system. Work is on progress to still improve the sensitivity of this approach and to provide energetics data in conventional units (kcal/mol). As already pointed out, the non-relaxed electron density obtained by means of the promolecular approach is not appropriate to describe strong interactions (covalent domain).

### 8.5 Example 5 (promolecular density): smart use of the uncoupling scheme in a trimer (test13)

Let us consider the following cyclic cluster, and say we are interested in the interaction energy between  $(H_2O)_2$  and  $HF$ :



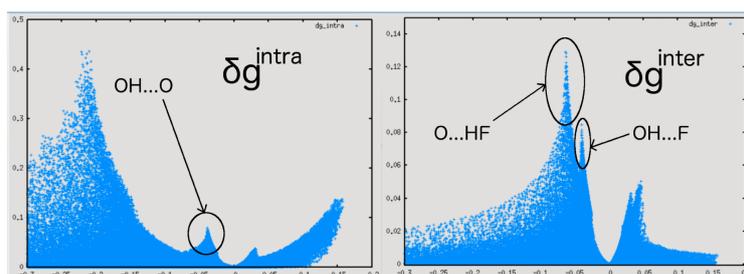
**Figure 26.** The two fragments defined for the IGM study of a cyclic cluster

```

2
frag1.xyz
frag2.xyz
INCREMENTS 0.05 0.05 0.05

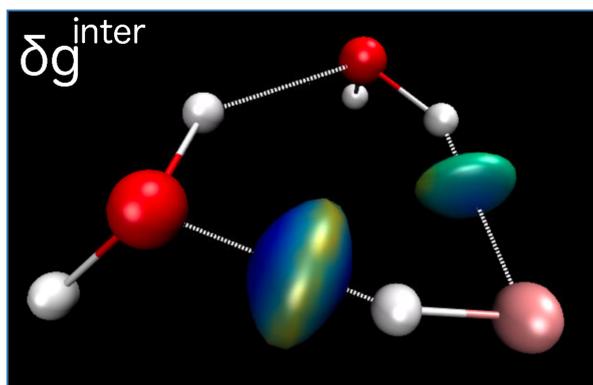
```

To access the interaction between just HF and the rest of the cluster we choose the two fragments as follow:  $(H_2O)_2 + HF$  (Fig.26). The resulting 2D plots are reported on Fig.27.



**Figure 27.**  $(H_2O)_2 \dots HF$  trimer

As can be seen, the OH...O and OH...F hydrogen bonds are automatically and naturally separated using the IGM approach. It leads to the following  $\delta g^{inter}$  3D-representation (Fig.28), not "polluted" with the water dimer (internal) hydrogen-bonding:



**Figure 28.** Customized study of the interactions in a trimer;  $\delta g^{inter}$  0.052 a.u. isosurfaces with BGR color code in the range  $-0.08 < \text{sign}(\lambda_2)\rho < 0.08$  a.u

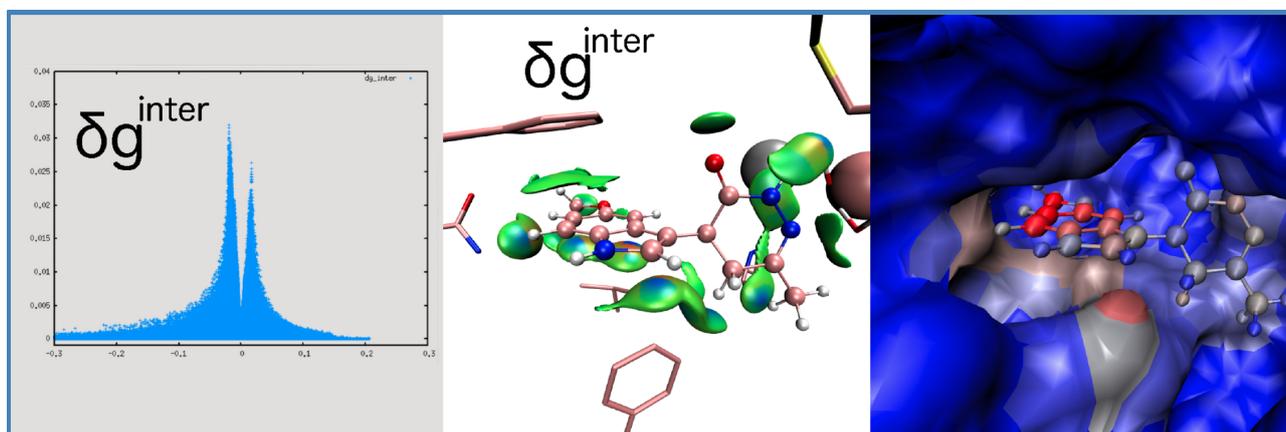
As already pointed out by the  $\delta g^{inter}$  2D-plot (peak heights on Fig.27), it is clear from this 3D-picture that the O ... HF hydrogen bond strength is larger (larger volume enclosed by iso-surface) than that of the OH...F hydrogen bond. It worth noticing that using this partition removes the ring closure (non bonding interaction) in the  $\delta g^{inter}$  representation since this feature results from the overall cluster (taken as one piece).

## 8.6 Example 6 (promolecular density): ligand –protein (test14)

For large systems, the use of the fast promolecular density (computed by IGMPlot) is advised.

From a practical point of view, studying ligand-host interactions requires choosing the ligand as the first molecule specified in the param.igm input file in order to reduce the grid size and save time. Depending on the size of the ligand and the grid step value, the computation time might be large (timing predictions are regularly updated in the runinfo file). On Fig.29 is reported the interaction between a pyridazinone derivative and an isoform of the phosphodiesterase IV (PDE4) protein.

```
2
ligand.xyz
protein.xyz
INCREMENTS 0.09 0.09 0.09
ONAME ligprot
```

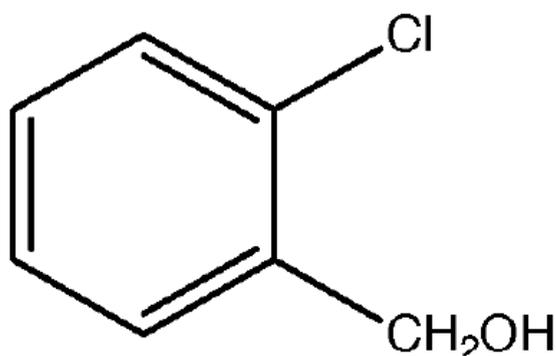


**Figure 29.** ligand – protein example; three analyses from left to right:  $\delta g^{inter}$  2D-signature,  $\delta g^{inter}$  0.013 a.u. isosurfaces with BGR color code in the range  $-0.08 < \text{sign}(\lambda_2)\rho < 0.08$  a.u., host-guest colored according to the  $\Delta g^{inter/At}$  atomic contributions to the intermolecular interactions, using a BGryR color scale (blue=0%, red=2.9%)

In tandem with  $\delta g^{inter}$  isosurfaces, the atomic scheme decomposition  $\Delta g^{inter/At}$  brings supplementary information. For instance, here, it can be seen that the indole moiety of the ligand primarily contributes to the interactions in this complex with protein "hotspots" well identified (red color). Extracting the interactions and performing the atomic contribution analysis is straightforward and parameter free.

## 8.7 Example 7 (promolecular density): when to use the CUTPLOT\_IGM keyword ? (test15)

Let us consider the single molecule on Fig.30:

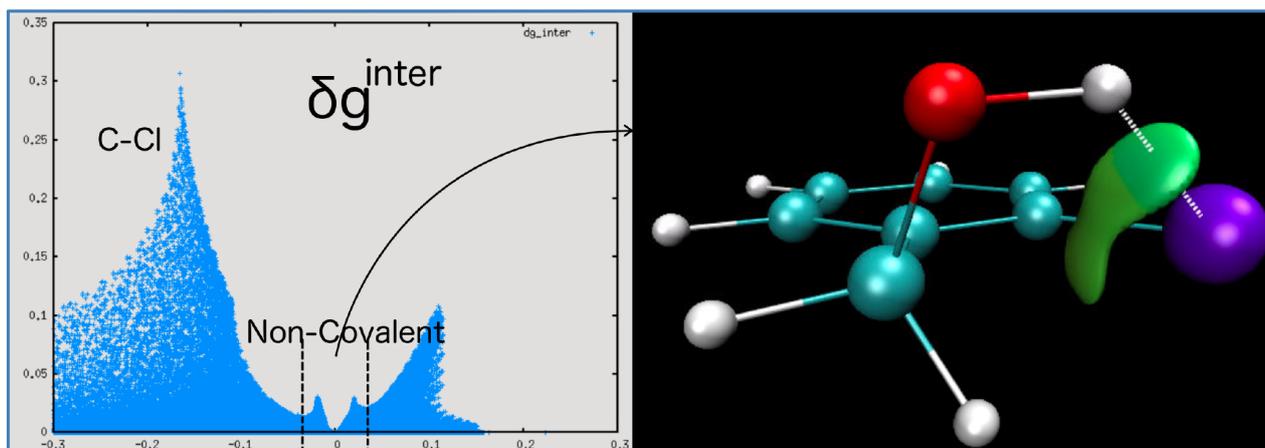


**Figure 30.** Example of molecule to probe non-covalent intramolecular interactions

We are interested in this case by non-covalent interactions located between the two substituents ( $-Cl$  and the group  $-CH_2OH$ ). To get this interaction using the promolecular ED model, the first idea is to proceed with the definition of two fragments:  $-Cl$  on one hand, and the rest of the molecule on the other hand. Unfortunately, in addition to the desired non-covalent  $Cl \dots CH_2OH$  interaction we will also get the covalent  $C-Cl$  interaction peak. Then, exceptionally, in addition, we can use the keyword `CUPLOT_IGM` (in the `param.igm` input file) to tell IGMPlot to write in the resulting cube file those points corresponding only to non-covalent interactions present on the 2Dplot reported below on Fig.31.

The second number given after the keyword `CUTPLOT_IGM` (0.03) means that only  $\delta g^{inter}$  points for which the signed ED is in the range  $[-0.03:0.03]$  will be printed to file 'dgInter.cube'. This will get rid of the C-Cl covalent bond (still present in the  $\delta g^{inter}$  representation). This ensures having the desired "intramolecular" non-covalent interaction in the resulting 3D picture reported on the right of Fig.31 :

```
2
frag1.xyz
frag2.xyz
INCREMENTS 0.05 0.05 0.05
CUTPLOT_IGM 0.3 0.03
```



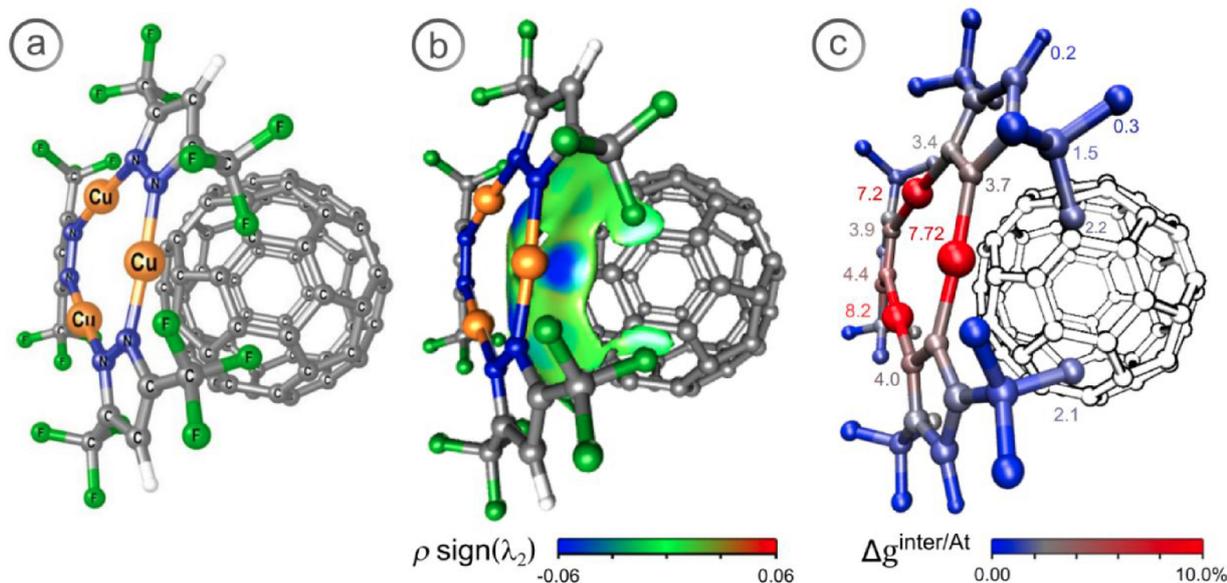
**Figure 31.** Using IGM\_CUTPLOT keyword allows for extracting and visualizing the desired weak contribution

Note that in this very special 2-fragments case, the " $\Delta E^{inter}$  Estimate" delivered by IGMplot (not affected by the CUTPLOT\_IGM keyword) includes C-Cl covalent bonding and is not meaningful (promolecular ED is not appropriate for strong interaction). In order to have an estimate of this weak intramolecular interaction between the chlorine atom and the hydroxyl group, you can opt for the 1-fragment IGM scheme and get the integrated  $\delta g^{intra/weak}$  score. Note that the PEAKFOCUSINTRA and PEAKFOCUSINTER more recent keywords make it even more flexible the choice of the ED windows to be studied (taking over the CUTPLOT\_IGM keyword in that case).

### 8.8 Example 8 (promolecular density): atomic decomposition scheme of non-covalent interactions applied to host-guest assemblies (test16, test17)

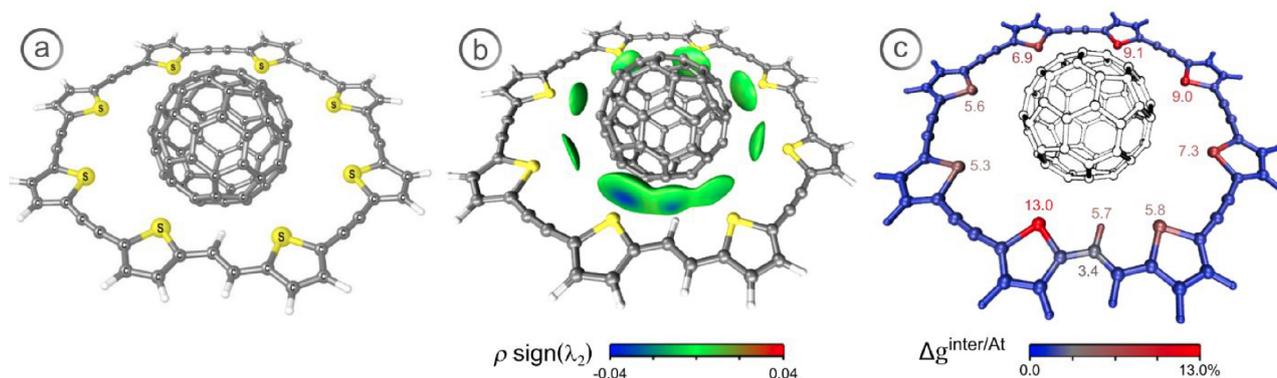
The atomic scheme decomposition  $\Delta g^{inter/At}$ , is an extension of the original IGM- $\delta g$  approach and is able to emphasize the most relevant contributions to the non-covalent interactions occurring between two fragments, after the integration of the local  $\delta g^{inter/At}$  index, complementary to  $\delta g^{inter}$ . After running the IGM calculation, a file entitled AtContribInter.dat is generated, which contains the contribution of each atom in %. Moreover, an associated file AtContribInter.vmd is provided to display in color these contributions within a vmd session. We highlight in Fig.32 this possibility on a trinuclear copper complex serving as a buckycatcher.

```
2
frag1.xyz
frag2.xyz
INCREMENTS 0.1 0.1 0.1
```



**Figure 32.** a) Optimized structure of copper complex; b) IGM analysis with an isosurface  $\delta g^{inter}$  of 0.01 a.u., BGR color code in the range  $-0.06 < \text{sign}(\lambda_2)\rho < 0.06$  a.u., c) Host colored according to the  $\Delta g^{inter/At}$  score using a BGryR color scale

A  $\Delta g^{inter/At}$  value of 23.1% has been obtained for the three copper atoms confirming their significant participation in the C<sub>60</sub> capture. However, the high overall value  $\Delta g^{inter/At}$  value associated with the three pyrazolate rings (48.0%) and the lateral CF<sub>3</sub> groups (28.9%) show that these fragments are also relevant in the C<sub>60</sub> inclusion. An additional example is given below in the field of supramolecular systems based on host-guest chemistry. On Fig.33 a macrocyclic oligothiophene incorporates C<sub>60</sub> in its inner cavity to form a unique Saturn-like complexes. The IGM- $\delta g$  reveals van der Waals interactions between the fullerene and the host sulfur atoms, as well as atomic contributions to this interaction.



**Figure 33.** Host-guest example (oligothiophene-C<sub>60</sub> complex);  $\delta g^{inter}$  isosurfaces and  $\Delta g^{inter/At}$  scores

This tool can be useful for teams devoted to the synthesis of functional supramolecular materials,

where a judicious structural modification may result in an advantageous tuning of the host-guest capabilities.

**Note that this atomic decomposition scheme is now available within the Quantum-Mechanical mode (using ED derived from QM calculations).**

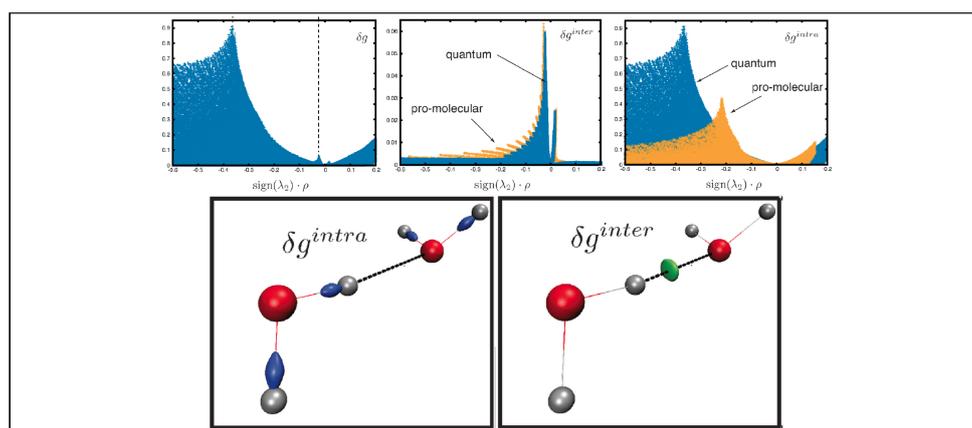
### 8.9 Example 9 (QM): $H_2O \dots H_2O$ dimer, FRAG1 keyword (test18)

In the QM mode (WFn/WFX) the electron density is obtained from a wave function, prior to the IGM calculation. The name of a single .wfn or .wfx file has to be supplied in that case. Without any specific keyword, the whole system will be considered as a single molecule, and the resulting  $\delta g^{inter}$  cube will be 0.0 (all interactions will be present in the  $\delta g^{intra}$  cube, like in the 1-fragment approach using the promolecular ED).

Instead, in order to separate the two domains (intra/inter), you can use the FRAG1 keyword in the param.igm input file and supply an atom selection pattern (see example below). For instance, in the provided .wfn file, if the first three atoms of the water dimer belong to the first water molecule and we want to study the interaction between the two water molecules, we just define the fragment 1 using the FRAG1 pattern: 'FRAG1 1-3'.

Be aware that if only FRAG1 keyword is specified in the param.igm input file, fragment 2 will be made of the remainder atoms, such that frag1+frag2 represents the whole system (see FRAG1/FRAG2 keywords details on all the available possibilities).

```
1
dimer.wfn
INCREMENTS 0.07 0.07 0.07
ONAME dimer
FRAG1 1-3
```



**Figure 34.** Water dimer studied by means of electron density coming from QM calculations (WFn/WFX mode)

Among other things, it can be noted (orange plot in Fig.34) that the promolecular ED (also reported here for comparison) clearly underestimates ED gradient drops for covalent bonding

( $\delta g_{max}^{intra}$ -pro O-H = 0.4 vs ( $\delta g_{max}^{intra}$ -QM = 0.9 a.u.). In contrast, in the domain of weak interactions, SCF quantum-mechanical ED and frozen promolecular ED have very similar features, which have served to validate the promolecular approach for weak interactions. IGMPlot allows you a bit of flexibility to handle cases where atoms in fragment 1 are not written in a contiguous way in the .wfn file. We give below some general guidelines for specifying atom numbers through FRAG1 keyword:

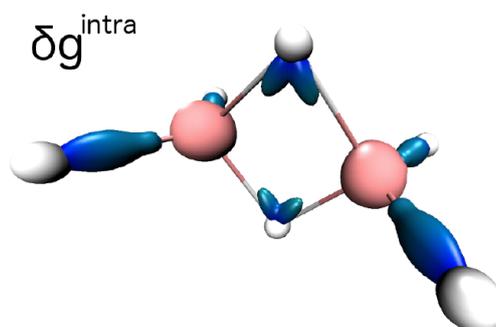
- You can specify a single atom number: FRAG1 5
- You can specify a group of non-sequential atoms by separating them by semicolons: FRAG1 4;8;19
- You can specify a sequential range of atoms by putting a dash between them: FRAG1 4-7
- You can mix the possibilities: FRAG1 6;2-4;22
- You can specify the range: FRAG1 4- (from atom 4 to the last atom in the wfn file).
- Any whitespace character will be ignored.

Note that AtContribInter.dat is generated, which contains the contribution of each atom in %. Moreover, the associated file AtContribInter.vmd is provided to display in color these contributions within a vmd session.

### 8.10 Example 10 (QM): $B_2H_6$ taken as one piece (test19)

The typical case of a three-center two-electron bond cannot be partitioned in two separated fragments. In that case, the appropriate solution is to perform a single fragment calculation. Remind that not using FRAG1 nor FRAG2 in the following param.igm input file triggers a single fragment calculations by default, but all the interactions will be revealed on the same picture (see Fig. 35).

```
1
mol.wfn
INCREMENTS 0.09 0.09 0.09
ONAME b2h6
```

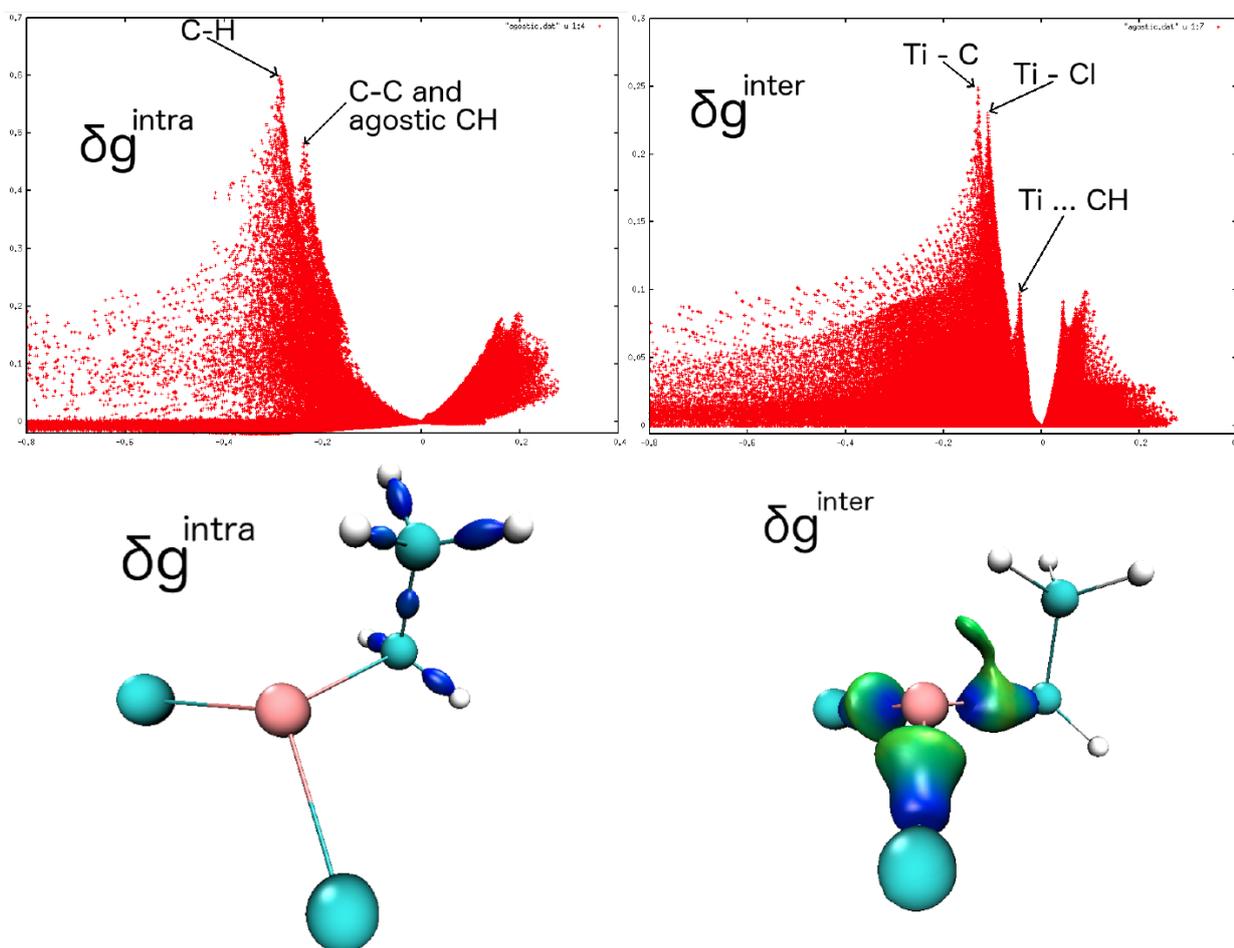


**Figure 35.**  $B_2H_6$   $\delta g^{intra} = 0.20$  a.u. isosurfaces

### 8.11 Example 11 (QM): Agostic interaction in the metallic complex $[TiCl_2CH_2CH_3]^+$ (test20)

Here is a smart use of the inter-uncoupling scheme: fragment 1 = Titanium atom (number 1), fragment 2 will be the rest of the system. This allows you to probe every interaction involved by the metallic centre. In addition to the 3 coordination bonds we can observe on Fig.36 that the  $\delta g^{inter}$  index is able to detect the agostic interaction between Ti and C-H.

```
1
agostic.wfn
INCREMENTS 0.09 0.09 0.09
ONAME agostic
FRAG1 1
```



**Figure 36.**  $[TiCl_2CH_2CH_3]^+$  complex optimized at the B3LYP/6-311++G(2d,2p) level of theory;  $\delta g^{intra} = 0.36$  a.u. isosurfaces;  $\delta g^{inter} = 0.06$  a.u. isosurfaces; color coding in the range  $-0.3 < \text{sign}(\lambda_2)\rho < 0.3$  a.u

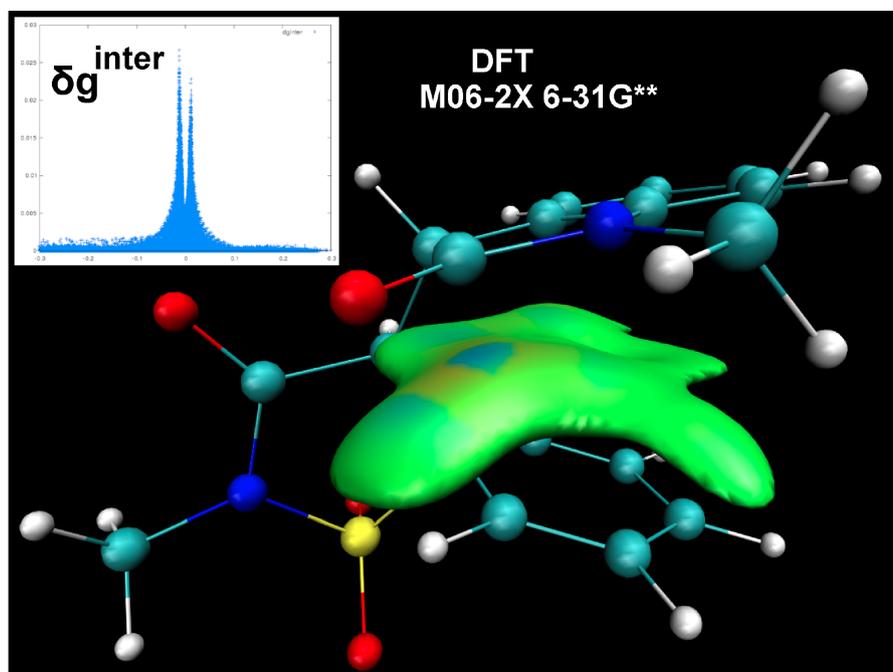
## 8.12 Example 12 (QM): FRAG1 and FRAG2 keywords: quantifying intramolecular $\pi$ -stacking (test21)

The combination of FRAG1 and FRAG2 keywords allows to probe interactions between two sub-parts of a given molecule (see for instance reference<sup>[16]</sup>). This possibility is very convenient to study intramolecular  $\pi$ -stacking. In order to do so, both FRAG1 and FRAG2 atom selections must be supplied in the param.igm input file. Be aware that, when FRAG1 and FRAG2 are specified, the sum (FRAG1 + FRAG2) can be smaller than the whole system. In the example below (Fig.37), FRAG1 and FRAG2 have been set to the phenyl ring and indole-based ring, respectively :

```
1
mol.wfn
FRAG1 11-21;35-42
FRAG2 1-6;26-30
```

Then, only the atomic orbitals of the FRAG1/FRAG2 selections are considered to achieve the IGM- $\delta g^{inter}$  calculation. Accordingly, the ED is limited to fragments 1 and 2 atomic sources.

**In very rare circumstances, using diffuse functions to perform an IGM- $\delta g^{inter}$  analysis might lead to spurious isosurfaces. In such a case, it is then advised to change the basis set by removing diffuse functions. Otherwise, the use of the HBP (Hirshfeld-based partition, keyword HIRSH, so-called IGMH framework) is an alternative to the GBP (default Gradient-based partition within the IGM frame-work) that might help in that case, but we generally do not recommend this IGM/Hirshfeld approach for non-covalent interactions.**



**Figure 37.** Probing intramolecular  $\pi$ -stacking at the DFT M062X/6-31G\*\* level of theory,  $\delta g^{inter} = 0.004$  a.u. isosurface; color coding in the range  $-0.08 < \text{sign}(\lambda_2)\rho < 0.08$  a.u.



As can be seen, the flat region of interaction between the two aromatic rings can be extracted from the wave function information. IGMPlot also quantifies this interaction through the integration of the  $\delta g^{inter}$  descriptor. The corresponding output emphasizes how many orbitals have participated to probe this internal interaction (Fig.38). This option should be useful in organic chemistry for mechanistic studies often revealing such non-covalent but intramolecular interactions during reactions.

```
* ----- *
* Density type          : QUANTUM *
* * * * *
* ED Grad Atom partition : Gradient-Based (GBP) *
* * * * *
* IGM Model            : dg, dgInter, dgIntra descriptors *
*                      + Atomic Degree of Interaction (DOI) *
*                      Cubic grid *
* * * * *
* NbFiles              : 1, fragmentation scheme *
* * * * *
* Chemical formula     : C18SN204H17 *
* Fragment 1 *
* ==> File name        : mol.wfn *
* ==> Number of atoms  : 19 / 42 *
* ==> Atom indexes     : 11-21 35-42 *
* ==> Primitives       : 364 / 843 *
* * * * *
* Fragment 2 *
* ==> File name        : mol.wfn *
* ==> Number of atoms  : 11 / 42 *
* ==> Atom indexes     : 1-6 26-30 *
* ==> Primitives       : 203 / 843 *
* * * * *
* Total number of atoms considered :      30 / 42 *
* Total number of primitives considered : 567 / 843 *
* * * * *
```

**Figure 38.** Example of text output for the combination of keywords FRAG1/FRAG2

The corresponding  $\pi$ -stacking interaction is quantified by IGMPlot thanks to the integral:  $\int \delta g^{inter} dV$  (given in the output). Hence, it is now for instance possible to monitor the effect of internal  $\pi$ -stacking during a chemical reaction across the stationary points of a given reaction path.

### 8.13 Example 13 (QM): CUBEFRAG keyword (test22)

Performing an IGM analysis on a large system from a WFN (or WFX) input file, may be CPU expensive. That is the reason why the option CUBEFRAG is proposed. It is intended to limit the atomic orbitals considered in the calculation to a reduced portion of the system. CUBEFRAG means that only the atoms within the working grid defined by the keyword CUBE (or RADIUS) will be considered. Hence, the fragment 1 will be a part of the whole system, and the fragment 2 will be empty. Accordingly, the  $\delta g^{inter}$  descriptor will be 0, and  $\delta g^{intra}$  will

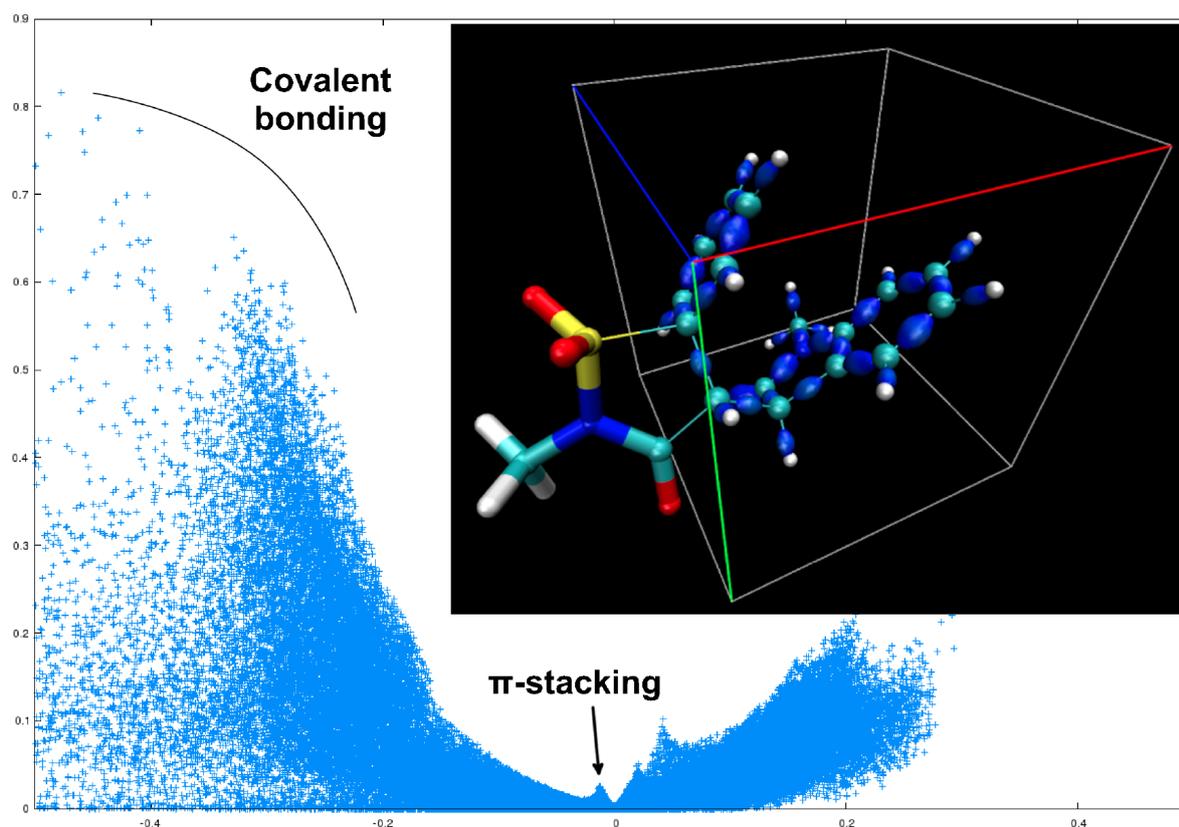
describe all the interactions present in fragment 1. Note that the ED is then limited to fragment 1.

```

1
mol.wfn
CUBE -1.4407 -4.2105 3.5878 5.6603 3.4265 -3.2688
CUBEFRAG
  
```

As above-mentioned, the complementary keyword CUBE (or RADIUS) must be supplied in order to position the grid in the system.

In this example, the CUBE keyword sets the grid around the two aromatic rings of the system. Then, specifying CUBEFRAG will limit the IGM analysis to those atoms inside the rectangular box. The covalent bonding isosurfaces are obtained by choosing a relative large isovalue (0.3 a.u. on Fig.39); furthermore, to get rid of the strong interaction domain, one could use the keyword CUTPLOT\_IGM or PEAKFOCUSINTRA, and then obtain the non-covalent interactions like  $\pi$ -stacking (not represented here). Of course, when one looks at the interaction between two subfragments, the best thing to do is to use the FRAG1/FRAG2 combination (rather than CUBEFRAG).



**Figure 39.** Example using the 1-fragment scheme and the tandem keywords CUBE and CUBEFRAG; 2D-plot  $\delta g^{intra}$  signature of atoms in the box;  $\delta g^{intra} = 0.3$  a.u. isosurface; color coding in the range  $-0.3 < \text{sign}(\lambda_2)\rho < 0.3$  a.u.



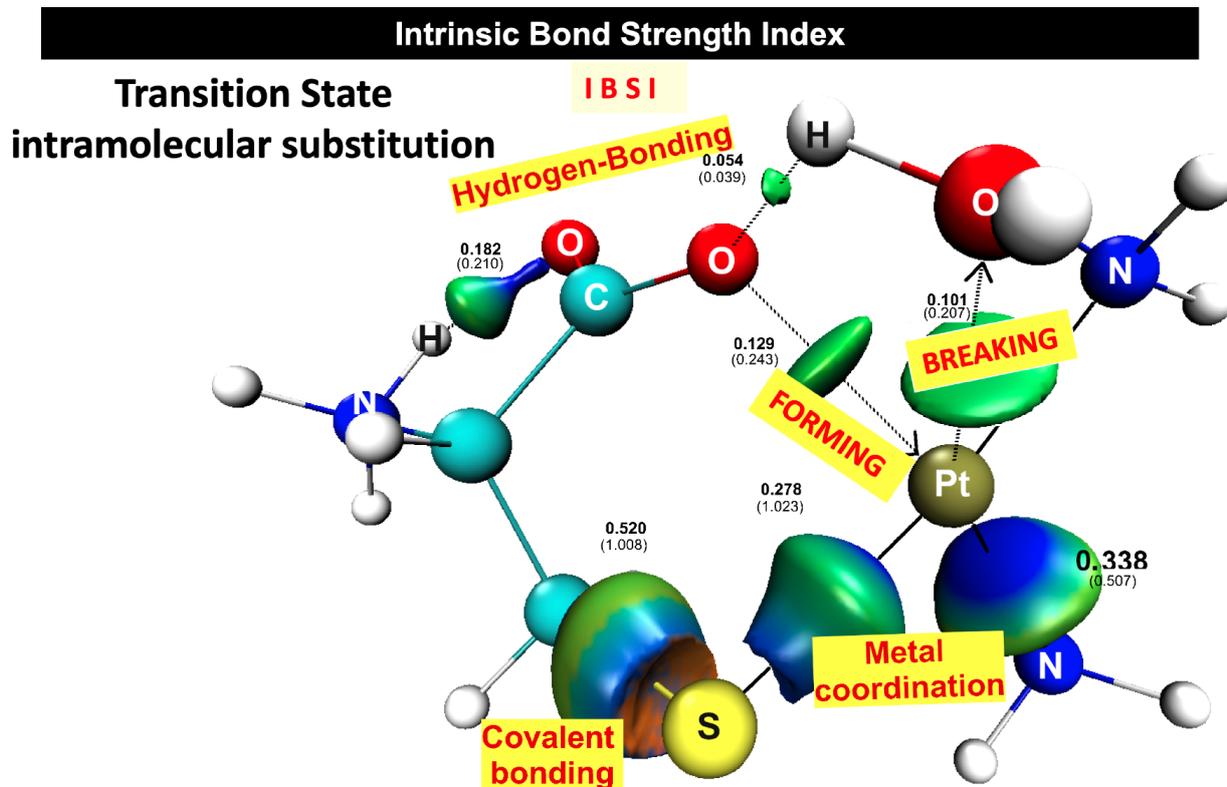
## 8.14 Example 14 (QM): IBSI keyword: a new way for probing bond strength (test23)

As above-mentioned, a new  $\delta g^{pair}$  descriptor has been built to focus the IGM analysis solely on the atomic ED gradients  $\nabla\rho_i$  brought by a given atom pair.<sup>[5]</sup> A bond by bond picture appears (see Fig.9), and it could be tempting to associate the strength of the bond with peak height of the 2D-plot  $\delta g^{pair}$  signature. However, each bond has its own individual profile suggesting that correlating the  $\delta g^{pair}$  descriptor should not be limited to the use of a local information (bond critical point for instance). The IBSI integrated score stems from this research and provides a quantitative bridge between a local electron density-based descriptor and the physically grounded bond strength concept. The results obtained on a large set of bonds underpin the IGM formalism and its two key-components: the non-interacting reference definition and the gradient based partition (GBP).

We refer the reader to the corresponding study,<sup>[5]</sup> where numerous examples are given. Here, only one example is given, covering all the interaction domains: non-covalent, metal coordination and covalent bonding. Fig.40 shows a transition state structure (TS) taken from a study on the reactivity of thiolates with cisplatin.<sup>[17]</sup> The mechanism proceeds with the intramolecular substitution of the coordinated water molecule by the  $-CO_2-$  group of the cysteine, which ultimately will act as a (S,O)-bidentate chelating ligand in the final product. The IBSI index and Mayer bond order are reported in bold for some atom pairs.

```
1
mol.wfn
IBSI
9 10
2 5
9 5
12 6
ENDIBSI
```

Theoretically, you can specify more than one bond in the IBSI/ENDIBSI section. However, you may prefer to distribute the IBSI calculations of several bonds (one by one) in parallel on several computation nodes.

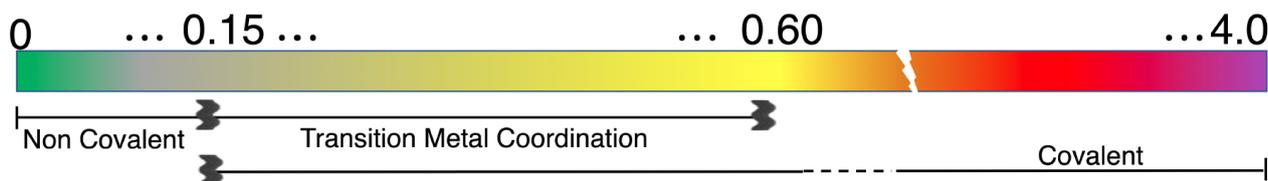


**Figure 40.** Bond strength Index (IBSI, bold) and  $\delta g^{pair} = 0.045$  a.u. isosurfaces for selected bonds in a TS involving a platinum complex (M06-2X LANL2DZ); color coding in the range  $-0.3 < \text{sign}(\lambda_2)\rho < 0.3$  a.u.; Mayer bond order in parentheses

From this figure we can see that IBSI and the Mayer bond order are complementary. For instance, according to the Mayer analysis (MBO), the C-S and S-Pt bonds exhibit an identical bond multiplicity of about 1.0. The IBSI provides an additional information: the covalent bond C-S is stronger (IBSI=0.520) than the S-Pt coordinate bond (IBSI=0.278). This is consistent with the strength of metal ligand bonds known to be intermediate between that of covalent bonds and non-bonded interactions.

In other respects, the MBO predicts a formal half-bond between the nitrogen and platinum atoms (0.507 pair) and a single bond between the sulfur and platinum atoms (1.023 pair). It may be then naively expected that the corresponding metal - ligand bond strengths should correlate with these bond orders at a ratio of 1(N-Pt):2(S-Pt). However, the IBSI predicts an inverse strength ratio with the N-Pt bond being significantly stronger (IBSI=0.338) than the S-Pt one (0.278). The IBSI stays attached to the restoring force concept, allowing for internally probing bond strength in TS structures and along reaction pathways for breaking and forming bonds. The IBSI covers a large range of interactions, from hydrogen-bonds to covalent bonding, as long as contragradience domains are well separated. However, there are limitations in interpreting IBSI for interactions with complex bonding scenarios where more than two atoms are involved in a bonding situation. Metallocenes for instance and even non-covalent stacking interaction, like in the benzene dimer, belong to this category. Actually, the IGM tool is fully able to reveal

any kind of interaction resulting from ED clash, however, whether the atompair IBSI score can be used is usually obvious by inspection of the molecular geometry and the shape of the contragradience isosurfaces. An indicative scale is provided (obtained for a set of 677 bonds of 235 molecules in their ground state):



From a practical point of view, the IBSI keyword does not generate cube files to visualize bond isosurfaces. Actually, an ultrafine cylindrical grid is employed by IGMPlot to calculate the IBSI score, which is not compatible with visualization program commonly used in theoretical chemistry. Instead of IBSI, one can use the FRAG1/FRAG2 tandem keywords to define two atomic fragments corresponding to the desired bond to be visualized (FRAG1 = one single atom, FRAG2 = one other single atom, generating cube files to be used with VMD).

**Note that since version 3.04, the cylindrical axis of the grid exactly matches the distance between the two atoms, while previously, a small empty gap was left between each atom of the atom pair and the beginning of the cylinder. This can lead to small differences with the numerical IBSI values produced by the previous version of the software.**

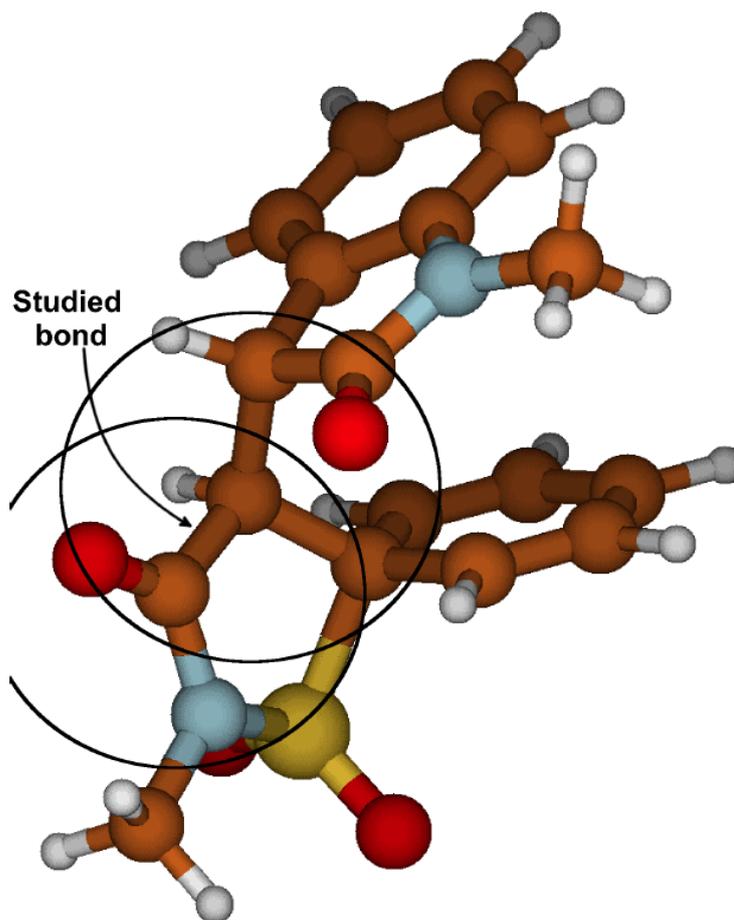
IBSI values come with PDA values, which measure the electron density asymmetry for the considered atom pair and its direction. This index (PDA) can be very useful to assess inductive effects in molecules.<sup>[1]</sup>

### 8.15 Example 15 (QM): IBSI keyword with cutoff option (test24)

Although the IBSI analysis is reduced to those atomic ED gradients  $\nabla\rho_i$  brought by the considered atom pair, the calculation of each atomic term  $\nabla\rho_i$  takes into account all the atomic orbitals of the entire system. It can be time-consuming. To speed-up the calculation, a cutoff can be applied around the considered bond (see Fig. 40). That way, only the atoms (and their primitives) within the cutoff radius are considered in the calculation. To do that, simply provide the cutoff value (in Å) after the keywords IBSI in the param.igm input file (3.0 Å in the example below).

```
1
mol.wfn
IBSI 3.0
9 10
ENDIBSI
```

The Table Tab.2 shows the IBSI scores obtained for the single C-C bond in the molecule of Fig.41 for several cutoff radii.



**Figure 41.** Cutoff radius taken around each atom of the considered bond for the IBSI calculation when the value of the radius is specified after the IBSI keyword

Default is to use no cutoff ("No limit", full calculation taking into account all the atomic orbitals), leading here to the IBSI (C-C) = 0.829. However, providing a cutoff radius of 3.5 Å seems quite reasonable (IBSI = 0.829), saving much time.

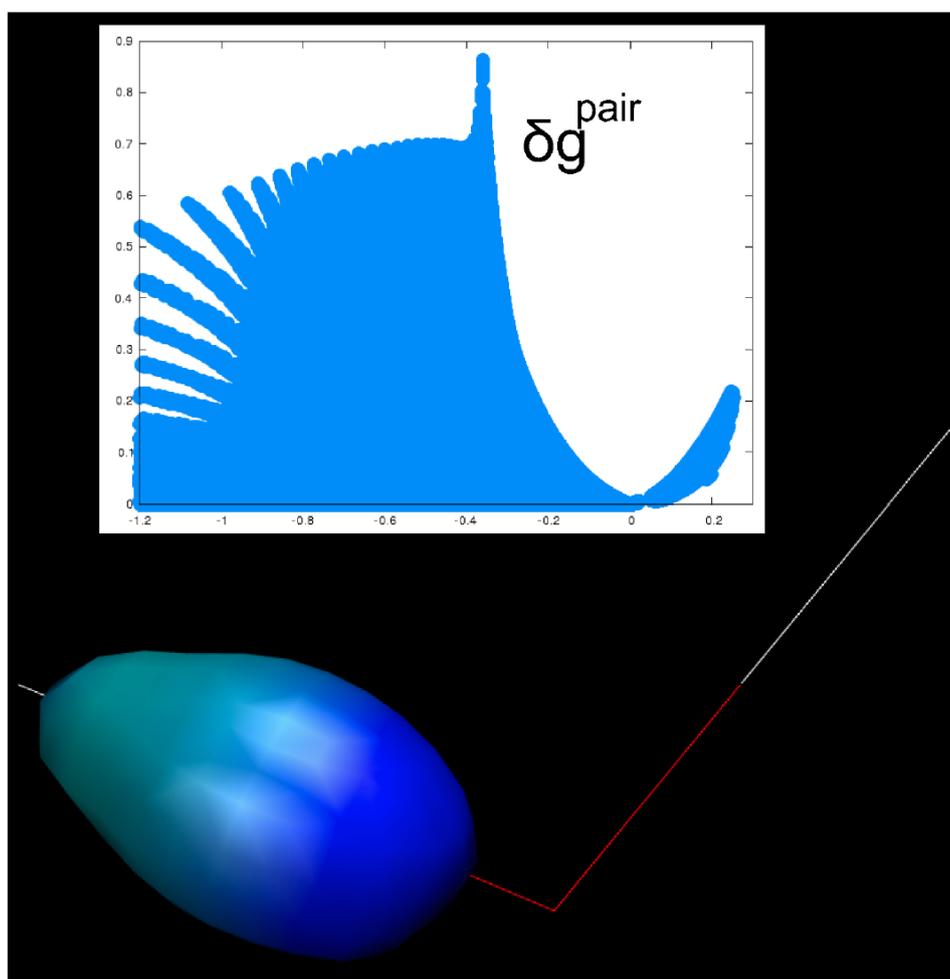
Cutoff radius (in Å)	2.0	2.5	3.0	3.5	4.0	4.5	5.0	No limit
IBSI (C-C)	0.837	0.835	0.829	0.829	0.829	0.829	0.829	0.829
primitives	175	210	402	528	640	668	773	843
Execution time (s)	148	193	694	1197	1710	1856	2432	2815

**Table 2.** Effect of several cutoff radii applied to the IBSI calculation for a single C-C bond in a molecule

This way, IBSI calculations can be performed on very large systems by focusing on the considered atom pair, taking only a few minutes while maintaining accuracy.

### 8.16 Example 16 (QM): Bond $\delta g^{pair}$ signature and atom pair isosurface (test25)

If you are interested in visualizing the  $\delta g^{pair}$  2D-plot and associated bond isosurfaces for a given atom pair, you must use the FRAG1/FRAG2 combination in order to generate a  $\delta g^{inter}$  cube file, and then you must remove the IBSI keyword from the param.igm input file. Actually, the IBSI score is obtained using an ultrafine cylindrical grid. Ideally, this grid should be used to generate bond isosurfaces. However, software commonly used in Chemistry like VMD are only able to build three-dimensional isosurfaces from scalar data store in cube file (gaussian cube file), not a cylindrical one. You can run a FRAG1/FRAG2 IGM calculation (without the IBSI keyword) where FRAG1 represents the first atom and FRAG2 the second atom of the bond. In that case, it is strongly advised to set the CUTOFFS and CUTPLOT\_IGM keywords to : "CUTOFFS 10.0 100.0" and "CUTPLOT\_IGM 0.3 10.0" in order to obtain a complete isosurface (even close to the atoms where the electron density is large, but  $\delta g^{pair}$  is small). This is illustrated on Fig.42.



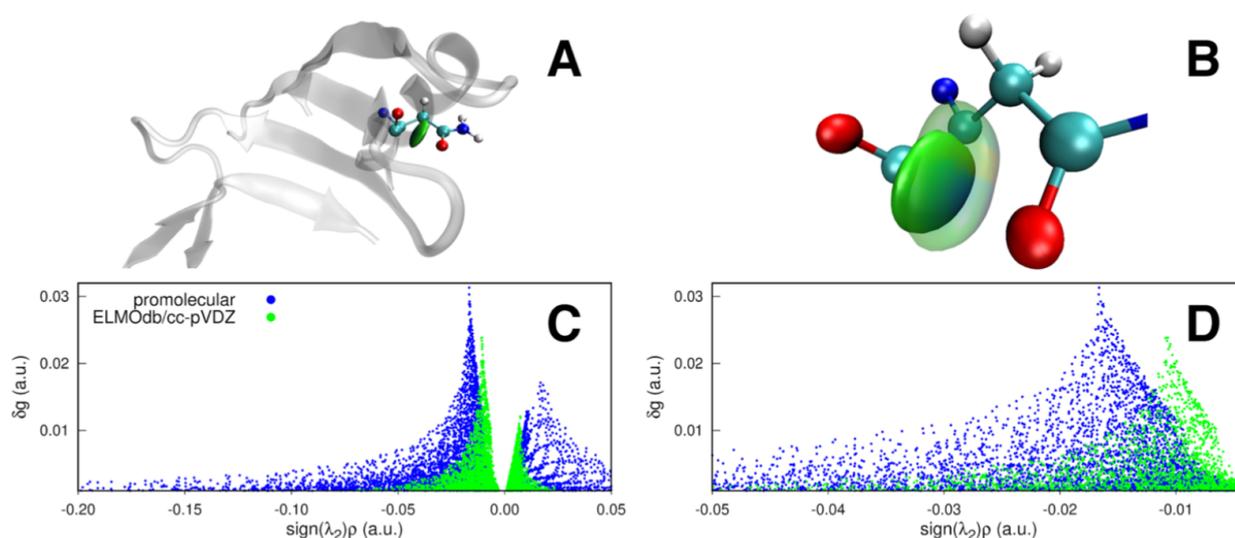
**Figure 42.**  $\delta g^{pair}$  2D-plot and  $\delta g^{pair} = 0.4$  a.u. isosurface for the selected OH bond of water; color coding in the range  $-0.6 < \text{sign}(\lambda_2)\rho < 0.6$  a.u.

One has to keep in mind that using the tandem FRAG1/FRAG2 keywords means that only the

primitives coming from the two selected atoms are considered in the calculation of  $\delta g^{pair}$  and of the ED used to color the  $\delta g^{pair}$  isosurface. The visual rendering is however consistent with what is obtained from a full calculation. On the contrary, using the IBSI keyword to calculate accurately the bond strength will make every primitive be considered in the calculation (unless an IBSI cutoff radius is employed).

### 8.17 ELMO possibility

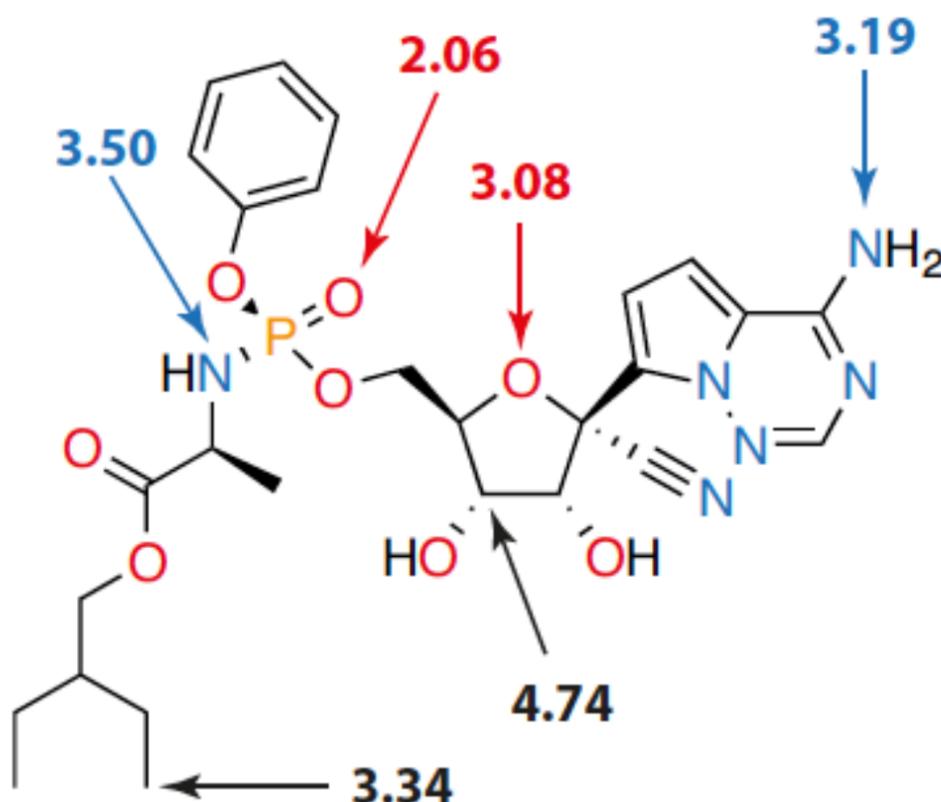
A step has been done towards the quantification of noncovalent interactions in large biological systems thanks to the coupling between the Independent Gradient Model and the Extremely Localized Molecular Orbital approach.<sup>[6]</sup> ELMO databanks of quantum mechanics/extremely localized molecular orbitals allows going beyond the promolecular approximation at lower cost than full QM calculations for large systems. Results are always very close to those obtained through corresponding IGM analyses that exploit fully quantum mechanical electron distributions (particularly DFT electron distributions). This has been observed both qualitatively and quantitatively. Therefore, rigidly assembling the global electron distribution as collection of ELMOs localized on chemically meaningful fragments rather than as combination of spherically averaged and neutral atomic electron densities (namely, exploiting the promolecular approximation) is a critical step toward the accurate description of weak inter- and intramolecular interactions in large systems. The example below on the human carbonic anhydrase (1477 atoms) unveils the difference between  $\delta g$  isosurfaces obtained either with promolecular ED (translucent on Fig.43) or ELMO data base of QM molecular orbitals (solid). This difference is also conspicuous from the superimposition of the two 2D-signatures.



**Figure 43.** n-p\* interaction in the Asn61 residue of the human anhydrase II (PDB ID : 3KS3). IGM- $\delta g = 0.005$  a.u. isosurface obtained with the ELMOdb/cc-pVDZ electron density and colored according to the BGR scheme over the range  $-0.05 < \text{sign}(\lambda_2)\rho < 0.05$  a.u., (B) ELMOdb versus promolecular (translucent)  $\delta g = 0.005$  a.u. isosurfaces, and (C) comparison between the  $(\rho, \delta g)$  2D fingerprint plots obtained at promolecular and ELMOdb/cc-pVDZ levels, with a zoom on the peaks (negative region) in (D)

### 8.18 Example 17 (QM): DOI : the atomic Degree Of Interaction (test26, $CH_3NH_2$ )

This new index<sup>[7]</sup> is shown to characterize the strength of attachment of an atom to its molecular neighbourhood (see Fig.44). A molecular map of atomic DOIs (last column of output file mol-AtomDOI.dat) is generated within an automatic workflow, provided a wave function input file (covering single- and multi-determinant wave functions). The atomic DOI is sensitive to the chemical environment of the considered atom and to the nature of the bonds around the atom. No correlation has been found between the atomic DOI and 40 other atomic properties. In particular, we have shown that this atomic score does not reduce to a property like the sum of bond orders. This makes it a specific source of chemical information, encoding the attachment strength of an atom in a molecule. The effect of perturbations (light, adding or removing electrons, ...) on atoms could be investigated in terms of change of their strength of attachment in the molecule, in complement with more conventional indices like the changes in atomic partial charges or changes in atomic spin localisation. We hope this tool will find utility in applications in the chemistry community (energetics material, organic chemistry, ...).



**Figure 44.** atomic DOI values for selected atoms on the Remdesivir molecule

From a reaction mechanism perspective, a strong connection has been established between this electron density-based index and the scalar reaction path curvature, the cornerstone of the benchmark Unified Reaction Valley Approach (URVA). We observe that reaction path curvature peaks appear when atoms experience an acceleration phase of electron density sharing during



the reaction, detected by peaks of the DOI second derivative either in the forward or reverse direction.

The DOI calculation is automatically performed during any IGM calculation on a molecule, generating the text file AtomDOI.dat.

From a theoretical perspective, the atomic DOI is obtained as follows: we partition a molecular system in an unusual way: a given atom  $i$  on the one hand (FRAG1) and the rest of the atoms on the other hand (FRAG2). This way, the resulting integrated score  $IGM-\Delta g_i$  (=DOI) reflects the degree of ED sharing between atom  $i$  and its surroundings.

Finally, restricting the DOI analysis to the low electron density domain allows for automatically identifying those atoms contributing to weak interactions inside a single molecule. When no fragments are defined (single fragment mode), this leads to the file weakAtomContrib.vmd coloring the atoms according to their contribution to weak interactions (which can be useful to detect atoms involved in weak interactions in a single molecule). This possibility is currently limited to molecules with atoms of periods 1-3. The corresponding atomic values can be found in the output file mol-AtomDOI.dat (column 3 'Weak\_dgSum').

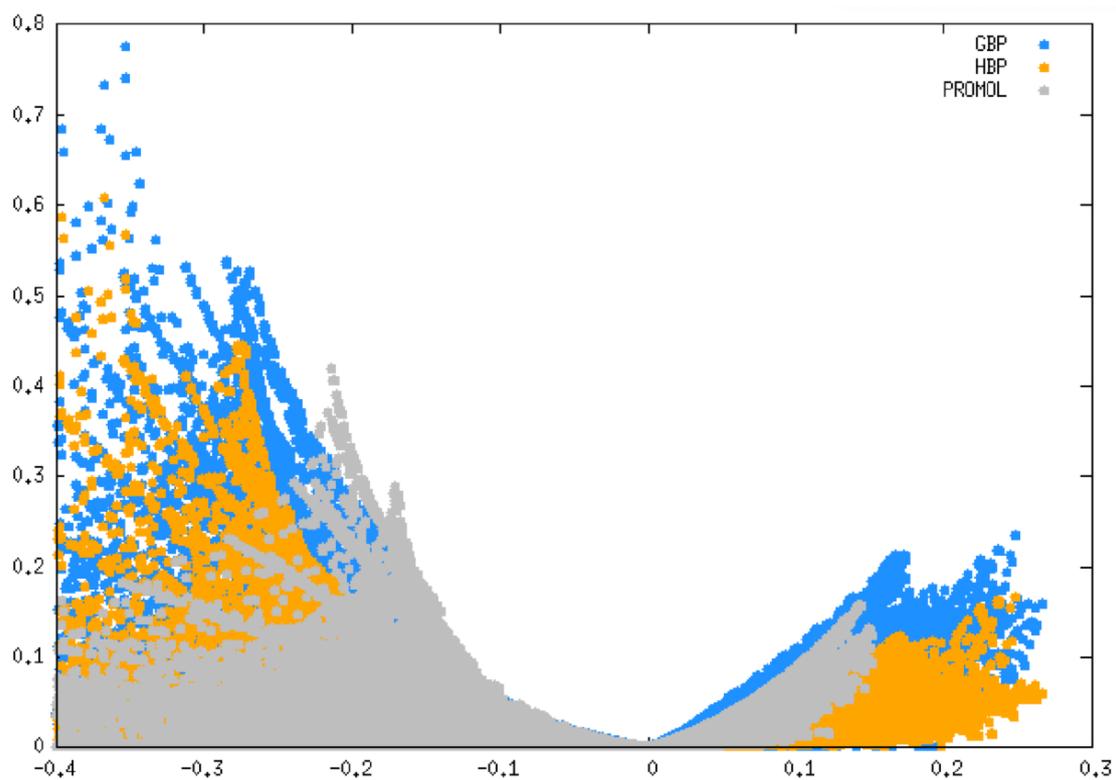
Finally, note that a file called 'AtomDOI.dat' is generated with values of atomic 'DOI' reported in the last column), but also with values 'Weak\_dgSum' (third column) quantifying the involvement of each atom to possible weak interactions (0 means no weak interactions for this atom). The latter information is obtained by partitioning the 2D-plot signature in two domains (weak and strong interaction peaks), and some times these two domains may overlap. This possibility is detected thanks to the column labelled 'Strong\_Weak\_Ovlap' (0 means that strong and weak domains are well separated).

Atom	Element	Weak_dgSum	Strong_Weak_Ovlap	DOI
1	C	0.000000	0.000000	3.217116
2	N	0.000000	0.000000	2.971655
3	H	0.000000	0.000000	0.901194
4	H	0.000000	0.000000	0.909946
5	H	0.000000	0.000000	0.910676
6	H	0.000000	0.000000	1.038879
7	H	0.000000	0.000000	1.040244

**Figure 45.** Atomic DOI values for the CH<sub>3</sub>NH<sub>2</sub> molecule, together with the 'Weak\_dgSum' field pointing atoms involved in weak interactions.

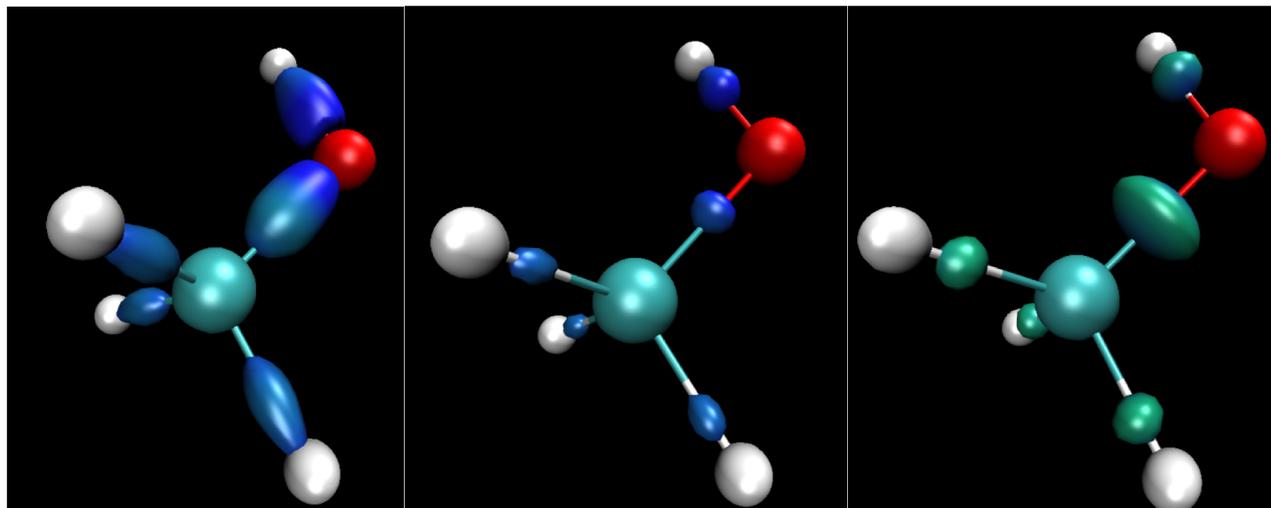
### 8.19 Example 18 (QM): IGMH (test27, CH<sub>3</sub>OH)

Using the Hirshfeld-Based Partition (HBP) of the electron density gradient, instead of the Gradient-Based Partition (GBP) initially developed within the IGM methodology, leads to the IGMH methodology, an extension of the IGM approach published by Tian Lu and Qinxue Chen.<sup>[8]</sup> This possibility has been coded in IGMPlot, and can be run using the HIRSH keyword.



**Figure 46.** Gradient-Based partition (blue), Hirshfeld-BAsed partition (orange) and promolecular (gray) IGM- $\delta g$  2D-plot signatures for the methanol molecule.

Although peaks are observed at the same ED position on the 2D-plot for the two QM approaches: GBP (conventional IGM approach) and HBP (so-called IGMH approach), using the Hirshfeld-based partition (orange) leads to less intense peaks compared with the GBP 2D-plot signature (blue), as can be seen on Fig.46.



**Figure 47.**  $IGM-\delta g = 0.35$  a.u. isosurface obtained for the methanol molecule with GBP (left), HBP (center) and promolecular (right,  $IGM-\delta g = 0.20$ ) approaches and colored according to the BGR scheme over the range  $-0.40 < \text{sign}(\lambda_2)\rho < 0.40$  a.u.

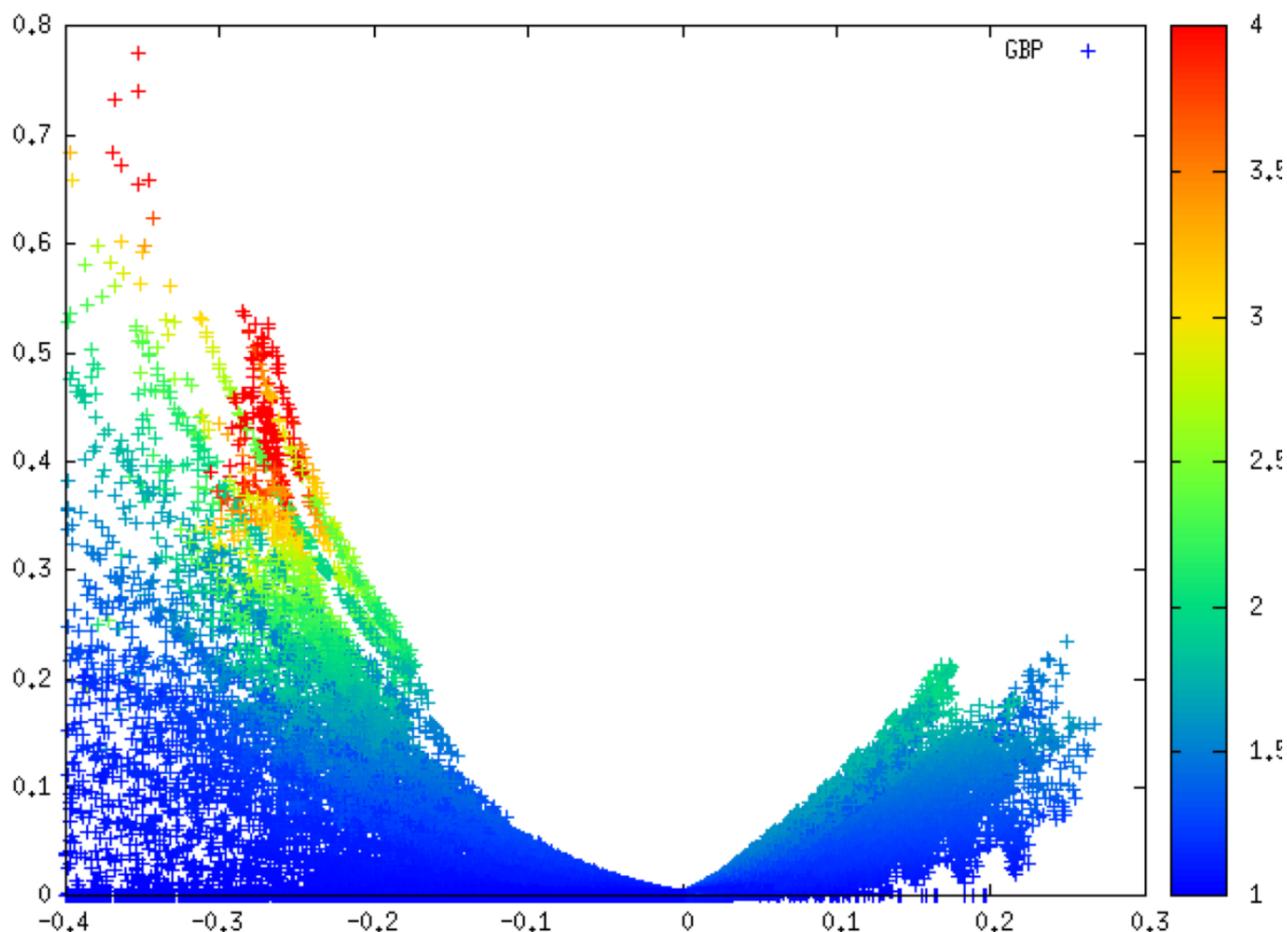
Subsequently, isosurfaces obtained within IGMH approaches turn out to be smaller, much less "stretched" along the bond axis as illustrated on Fig.47. It is just like IGMH isosurfaces display less "covalent" features, but rather exhibiting some promolecular features, which is not surprising since it uses an hybrid QM/promolecular partition of the ED gradient. We are still investigating the comparison between the original IGM approach using the GBP ED gradient partition and the IGMH variation. Meanwhile, we strongly advise using the GBP partition of the ED gradient (default in IGMPlot).

We would like to draw the user's attention that the version of the IGMH variant implemented in the Multiwfn program does not follow the equations presented in the IGMH original article,<sup>[8]</sup> and in fact, does not use the Hirshfeld partition. Unfortunately, the authors introduced an error in the implementation of this approach in the Multiwfn code (as stated in their own erratum<sup>[14]</sup>). From our perspective, the resulting atomic ED gradient partition used in the Multiwfn lacks physical meaning.

## 8.20 Example 19 (QM): qg descriptor (test28, $CH_3OH$ )

In order to better highlights peaks in 2D-plots  $\delta g$  signatures, the new qg descriptor was devised to color points of this plot (like on Fig.48). qg is obtained by performing the ratio  $\frac{\nabla_{\rho}^{IGM}}{\nabla_{\rho}}$ , instead of doing the difference  $\delta g = \nabla_{\rho}^{IGM} - \nabla_{\rho}$ . It is then much more sensitive to ED contragradiance

situation, tending to infinity as  $\nabla_{\rho}$  approaches 0 close to critical points, even for weak interactions like hydrogen-bonding.



**Figure 48.** IGM- $\delta g$  2D-plot for the methanol molecule with GBP and colored according to  $qg$  value in the scale [1:4].

**This provides a much higher resolution in identifying visually the peaks on the 2D-plot  $\delta g$  signature.** Two new columns have subsequently been automatically added (in the QM mode) to the `igm.dat` output file to generate more detailed 2D-plot signatures using a gnuplot command like this for instance:

```
set palette rgbformulae 22,13,-31 → choose a color palette
set cbrange[1:4] → color according to the palette on this range
plot "mol-igm.dat" u 1:3:5 palette
```

column 1 = signed electron density  
 column 3 =  $\delta g^{intra}$  descriptor  
 column 5 = qgintra descriptor

Similarly:

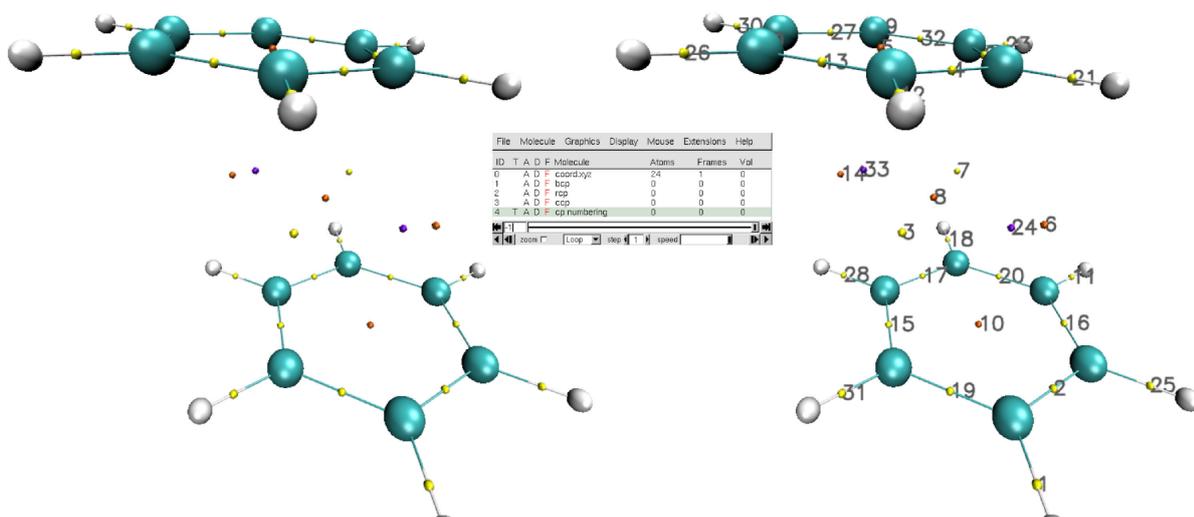
```
set palette rgbformulae 22,13,-31
set cbrange[1:4]
plot "mol-igm.dat" u 1:4:6 palette
```

column 1 = signed electron density  
 column 4 =  $\delta g^{inter}$  descriptor  
 column 6 = qginter descriptor

Note that IGMPlot automatically generates gnuplot scripts to draw  $\delta g^{intra}$  and  $\delta g^{inter}$  signatures.

## 8.21 Example 20 (QM): Critical point analysis (test29, benzene dimer)

A new feature of IGMPlot is to perform a critical point (cp) search.<sup>[1]</sup> Determining such critical points in the molecule is of high importance in much studies aiming at characterizing bonds by computing properties at cp like energy densities (kinetic, potential) or the ED Laplacian, or ellipticity, ... This analysis can be obtained very easily in IGMPlot by simply typing the keyword **CRITIC** in the param.igm input file. The file cp.vmd is generated to display cp with the VMD program (see Fig.49) and a summary of the analysis is provided in the text outfile igm.log (see Fig.50).



**Figure 49.** IGM- $\delta g$  Critical points in the benzene dimer (numbering given on the right); VMD GUI



```

* #####
* ##### CRITICAL POINT #####
* ##### ANALYSIS #####
* #####
* -----
* Seeding algorithm:
*           - IGM Promolecular Qgmax points
*             with medium grid
*           - midpoint between atom pairs
*
* BCP + RCP + CCP           :      33
*
* Number of critical points of each type           Signature
*
*   Bond critical points (BCP) :      26           (3,-1)
*   Ring critical points (RCP) :       5           (3,+1)
*   Cage critical points (CCP)  :       2           (3,+3)
*
* Poincare-Hopf relationship  NCP - BCP + RCP - CCP = 1
* Verification                24   26   5   2 = 1
* relationship satisfied
*
* -----

```

Figure 50. IGM- $\delta g$  Critical points analysis summary reported in the igm.log file

Properties calculated at critical points are reported in the file cp.txt (as illustrated on Fig.51):

CP	Signat.	X (Angs)	Y (Angs)	Z (Angs)	Rho (e-/bohr <sup>3</sup> )	GradRho ..(e-/bohr <sup>4</sup> )..	qg	L1 .....(e-/bohr <sup>5</sup> ).....	L2 (e-/bohr <sup>5</sup> )	L3 (e-/bohr <sup>5</sup> )	Laplacian (e-/bohr <sup>5</sup> )	G ....Hartree/bohr <sup>3</sup> ...	V	H	Ellipticity (3,-1) (L1/L2 - 1.0)	
1	(3,-1)	-0.929	-2.353	-1.803	2.777e-01	6.60e-09	0.607	1.66e+12	-0.739	-0.730	0.502	-0.967	0.037	-0.316	-0.279	0.012
2	(3,-1)	-0.934	-2.354	-0.605	3.098e-01	1.89e-13	0.631	3.80e+13	-0.634	-0.527	0.322	-0.839	0.099	-0.408	-0.309	0.202
3	(3,-1)	0.000	0.000	-0.969	6.304e-03	7.33e-15	0.013	4.38e+11	-0.002	-0.002	0.022	0.017	0.004	-0.003	0.001	0.107
4	(3,-1)	-0.934	1.410	-0.603	3.106e-01	6.47e-15	0.633	6.32e+13	-0.636	-0.530	0.322	-0.844	0.099	-0.409	-0.310	0.199
5	(3,+1)	0.001	1.883	-0.000	2.086e-02	3.35e-10	0.051	1.14e+12	-0.016	0.092	0.092	0.168	0.033	-0.024	0.009	
6	(3,+1)	-1.202	-0.578	-0.000	3.392e-03	5.41e-11	0.007	2.51e+10	-0.001	0.000	0.012	0.002	-0.002	0.001		
7	(3,-1)	-0.000	0.000	0.969	6.304e-03	5.49e-12	0.013	4.38e+11	-0.002	-0.002	0.022	0.017	0.004	-0.003	0.001	0.107
8	(3,+1)	0.000	-0.000	0.000	5.287e-03	7.96e-14	0.011	2.55e+11	-0.003	0.001	0.018	0.016	0.003	-0.003	0.001	
9	(3,-1)	0.934	2.354	-0.605	3.098e-01	1.26e-08	0.631	3.81e+13	-0.634	-0.527	0.322	-0.839	0.099	-0.408	-0.309	0.202
10	(3,+1)	-0.001	-1.883	-0.000	2.086e-02	3.82e-10	0.051	7.02e+10	-0.016	0.092	0.092	0.168	0.033	-0.024	0.009	
11	(3,-1)	-0.929	-2.353	1.803	2.777e-01	4.53e-14	0.607	1.66e+12	-0.739	-0.730	0.502	-0.967	0.037	-0.316	-0.279	0.012
12	(3,-1)	-0.928	1.410	-1.803	2.780e-01	5.86e-14	0.608	1.77e+12	-0.741	-0.733	0.502	-0.971	0.037	-0.316	-0.279	0.011

Figure 51. Properties calculated at critical points reported in cp.txt

The seeding strategy is twofold in IGMPlot: (1) a promolecular calculation of qg (the 'quotient' twin of  $\delta g$ ) is performed to identify those regions of space with electron contragradiance (large values of qg), serving as guess for the cp search using a Newton-Raphson optimization. To do this, a grid is built with medium grid steps, (2) additional points are added to the list of seeds taken as the midpoint of pairs of atoms less than 15 bohr apart.

For those situations where this strategy fails, optional keywords (CRITIFINE, CRITICUL-TRAFINE) employing finer grids are proposed. In addition, the optional keyword (CRITI-CaddSEEDS) is proposed to add complementary seeds taken from atom triplets.



## 8.22 Example 21 (QM): Investigating inductive effect with the PDA index (test30, test31)

In order to provide users with a simple tool to assess inductive effects on specific bonds in molecules the Pair Density Asymmetry (PDA) index has been devised (see reference<sup>[1]</sup>). It replaces the old index called BAF, which turned out to be especially CPU expensive, limited to the atoms of the periods 1-3 and and too much dependent on the basis set employed.

In a homonuclear diatomic molecule, in the absence of external perturbation, electrons are equally distributed between the two atoms leading to a symmetrical picture of the electron density (ED) along the bond. In contrast, when different atoms bond together the ED is accumulated unequally between atoms. The PDA index tells you how much the bond is asymmetrical and the direction of the asymmetry.

The PDA index takes the asymmetry information from the electron density gradient. First, a cylinder is chosen, with its longitudinal  $z$  axis coinciding with the A-B bond axis, and a regular grid is built to encompass the region between the atoms (as for the calculation of the IBSI index in IGMPlot). Next, at each point of this grid the ED derivative with respect to  $z$  ( $\frac{\partial \rho}{\partial z}$ ) is calculated and its atomic partition is performed thanks to the Gradient-Based Partition (GBP), such that  $\frac{\partial \rho}{\partial z} = \sum_{i=1}^{N_{\text{atoms}}} \frac{\partial \rho_i}{\partial z}$ . Then, focusing on the two atoms forming the considered AB pair, the following quantity is computed:  $\frac{\partial \rho_A}{\partial z} + \frac{\partial \rho_B}{\partial z}$ . The two terms correspond to the atomic ED derivative in the  $z$  direction for atom A and for atom B, respectively. Going from A to B, the ED source A decreases ( $\frac{\partial \rho_A}{\partial z} < 0$ ) while the ED source B increases ( $\frac{\partial \rho_B}{\partial z} > 0$ ). In other words, the ED contragradience takes place in between the atoms (only along the  $z$  direction). Hence, for a totally symmetrical bond, on average, these two contributions cancel, leading to the measure:  $\int_V \left( \frac{\partial \rho_A}{\partial z} + \frac{\partial \rho_B}{\partial z} \right) dV = 0$ . In contrast, one obtains a negative value when, on average, the atom A derivative  $\frac{\partial \rho_A}{\partial z}$  dominates over  $\frac{\partial \rho_B}{\partial z}$ . Finally, one obtains a positive value when the atom B derivative dominates.

To compare atom pairs involving different bond lengths, one divides the integrated value by the A-B distance  $d$ , leading to the Pair Density Asymmetry index :  $[\int_V \left( \frac{\partial \rho_A}{\partial z} + \frac{\partial \rho_B}{\partial z} \right) dV] / d$ , which is measured in  $e^-/\text{bohr}^2$ .

The following PDA ( $10^{-1}$ ) values and directions are obtained for the following bonds (see Fig. 52).

Method	Basis set \ A-B	C←B	C-C	C→N	C→O	C→F
		CH <sub>3</sub> -BH <sub>2</sub>	CH <sub>3</sub> -CH <sub>3</sub>	CH <sub>3</sub> -NH <sub>2</sub>	CH <sub>3</sub> -OH	CH <sub>3</sub> -F
DFT M06-2X	6-31G	-39.2	0.0	35.1	69.5	105.9
"	6-31G**	-38.5	0.0	34.5	69.3	107.1
"	6-311G**	-38.0	0.0	34.5	69.4	108.4
"	6-311++G**	-38.0	0.0	33.4	67.4	108.4
"	Def2SVP	-35.2	0.0	34.4	71.1	111.2
"	Def2TZVP	-37.5	0.0	34.3	68.9	105.7
"	Def2QZVP	-38.3	0.0	33.1	65.4	102.6
"	Aug-cc-pVDZ	-26.0	0.0	36.7	74.7	116.0
"	Aug-cc-pVTZ	-36.0	0.0	30.2	65.9	106.5
"	Aug-cc-pVQZ	-40.5	0.0	25.8	58.3	95.5
	SD	4.1	0.0	3.1	4.3	5.3
	RSD (%)	11.0	-	9.3	6.3	5.0
HF	6-31G**	-35.7	0.0	35.9	73.4	113.4
MP2	6-31G**	-35.9	0.0	34.7	71.1	110.8

*SD* = Standard deviation, *RSD* = Relative SD

**Figure 52.** Pair density asymmetry for several bonds at different levels of theory

Regarding the basis set effect on the PDA value on 4 different bonds and for a total of ten different basis sets, the observed relative standard deviation (RSD) is reasonable, being in the range 5 to 11%. Regarding the quantum method employed, the PDA values reported in the above table are very stable regardless of the method employed here. However, we recommend of course to employ the same level of theory to compare PDA values.

The atom bringing the most electrons in the bond axis (*z*) direction will lead to the largest ED slope, so will lead to the largest  $\frac{\partial \rho_{A \text{ or } B}}{\partial z}$  contribution. Hence, the direction of asymmetry as measured by the PDA points towards the atom bringing the most electrons in the bond axis (*z*) direction. Three factors govern the magnitude and direction of the Pair Density Asymmetry (PDA), in descending order:

- The period of each atom decides of the resulting direction and magnitude for the PDA. For instance, the PDA value of *C* – *H* bond in CH<sub>4</sub> ( $228.5 \cdot 10^{-1}$ ) is much larger than the PDA of *C* – *F* ( $107.1 \cdot 10^{-1}$ ). This makes sense since in CH<sub>4</sub>, the 1s atomic core orbital of C brings large electron density gradient in the bond axis direction (among others) while H has no core counterpart. Similarly, *H* – *Cl* ( $631.4 \cdot 10^{-1}$ ) shows a larger asymmetry than *H* – *F* ( $440.5 \cdot 10^{-1}$ ).
- The electronegativity difference between the two atoms of the considered pair plays a major role. Across the above-reported *C* – *X* series involving the atoms X, all belonging to the second period, the *C* – *F* bond features the largest PDA value, directed toward F. This is mainly due to the larger electronegativity of F polarizing the electron density towards F rather than towards C and giving rise, on average, to larger ED slopes  $\frac{\partial \rho_F}{\partial z}$  compared to  $\frac{\partial \rho_C}{\partial z}$ .

- The surroundings of the atom pair. That is the whole point of the PDA tool: to examine and quantify the effect of external factors on a given bond. For instance, the electronic effect caused by one, two or three fluorine atoms on the adjacent  $C - C$  bond of ethane initially purely symmetrical is highlighted in Fig.53.

Method	Basis set	0F	1F	2F	3F	1F	2F	3F
		(ethane)	PDA( $10^{-1}$ ) gross value			PDA ratio		
DFT M06-2X	6-31G**	0.0	6.2	11.6	15.8	1.0	1.9	2.5
	6-31++G**	0.0	5.3	9.7	13.3	1.0	1.8	2.5
	6-311++G**	0.0	3.3	6.4	8.1	1.0	1.9	2.5
	Aug-cc-pvdz	0.0	3.0	5.0	5.7	1.0	1.7	1.9
	Aug-cc-pvtz	0.0	3.8	7.9	13.1	1.0	2.1	3.4
	Aug-cc-pvqz	0.0	9.4	17.6	25.3	1.0	1.9	2.7
	Def2SVP	0.0	4.1	7.1	8.6	1.0	1.7	2.1
	Def2TZVP	0.0	4.5	8.2	11.5	1.0	1.8	2.6
	Def2QZVP	0.0	7.1	12.7	16.9	1.0	1.8	2.4

**Figure 53.** Assessment of the inductive effect of 1, 2 and 3 fluorine atoms on the central bond  $C - C$  of ethane, through the Pair Density Asymmetry index

Since the electronic effect probed here is very small, larger differences between basis sets are observed. Fortunately, the ratio between the effect of 2 or 3 fluorine atoms and the effect of one single atom F is more stable and indicative of the cumulative electron withdrawing effect of several fluorine atoms on the adjacent  $C - C$  bond.

Furthermore, in the following Fig.54, the PDA assesses the electronic effect caused by different chemical groups on the adjacent  $C - C$  bond in ethane and is in line with the relative free atom electronegativities of F, O, N, Cl and S. Other perturbations like electric field or light-induced

PDA ( $10^{-1}$ )	$H_3C-CH_3$	$H_3C-CH_2F$	$H_3C-CH_2OH$	$H_3C-CH_2NH_2$	$H_3C-CH_2Cl$	$H_3C-CH_2SH$
M06-2X/ 6-31g**	0.0	6.2	4.2	1.9	2.4	1.2

**Figure 54.** Assessment of the inductive electronic effect caused by different chemical groups on the adjacent  $C - C$  bond in ethane through the Pair Density Asymmetry index

excited state may trigger asymmetry or emphasize existing asymmetry in certain bonds of a molecule, which can be probed and quantified through the PDA tool.

We hope this investigating tool will find utility in applications in the chemistry community, in complement with other tools like atomic partial charges generally used to estimate the unequal electron distribution between atoms.

Noteworthy, the PDA index captures the electron distribution asymmetry directly attached to

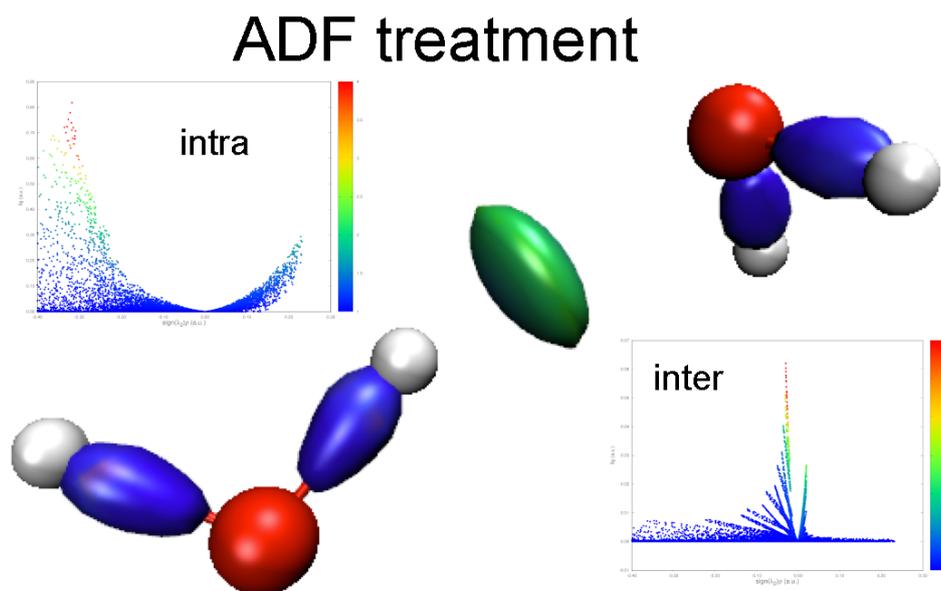


the atom pair. It is tackled directly at the gradient level. It has nothing to do with the bond polarity derived from a conventional point charge analysis. Actually, the partial atomic charge analysis is not based on the ED gradient (but rather on the ED and nuclear charges) and the resulting bond polarity incorporates, to some extents, the ED distribution asymmetry (whenever there is) of all neighbors beyond the examined atom pair. However, both analyses (PDA and partial atomic charges) can be complementary tools to probe the ED distribution and the effect on it caused by perturbations.

## 8.23 Example 22 (QM): ADF users (test32)

The main difference between ADF and most of the other QM packages lies in the use of Slater Types Orbitals (STO:  $x^{k_x} y^{k_y} z^{k_z} r^{k_r} e^{-\alpha r}$ ) instead of the more commonly used Gaussian Type Orbitals (GTO:  $x^{k_x} y^{k_y} z^{k_z} e^{-\alpha r^2}$ ). Then, compared to the use of GTOs, in addition to the  $\alpha$  exponents and type assignments (triplet  $k_x, k_y, k_z$ ), values of  $k_r$  have to be supplied as well to specify primitives used in the description of the wave function. Accordingly, WFN and WFX files cannot be used by IGMPlot to implement an STO description of the primitives. Rather, ADF provides users with the so-called `adf.rkf` binary file.

In order to provide ADF users with IGM calculations, a C++ reader has been implemented and incorporated to IGMPlot to parse this binary `adf.rkf` result file describing the electronic structure of a studied system obtained from the ADF execution. The KFReader library (under the terms of the GNU Lesser General Public License as published by the Free Software Foundation version 3 of the License) developed by Alexei Yakovlev is used. SCM owns the intellectual property right for the associated routines.



**Figure 55.** Use of `adf.rkf` binary file generated by ADF to study the water dimer through the use of STOs;  $\delta g = 0.4$  and  $0.03$  a.u. iso-surfaces for intra- and inter-molecular interactions, respectively; BGR colorscale in the range  $-0.30 < \text{sign}(\lambda_2)\rho < +0.30$  and  $-0.08 < \text{sign}(\lambda_2)\rho < +0.08$  for intra- and inter-molecular interactions, respectively; DFT(VWN)/DZ level of theory.

An example is provided in example 22 (test32) together with a binary `mol.rkf` file (coming from ADF execution), which is used by IGMPlot to investigate the non-covalent interaction within the water dimer. Note that further tests have been performed to cover: restricted, unrestricted cases, as well as the frozen core possibility offered by ADF (the innermost atomic shells are kept frozen). As a final test, IGM critical points have been successfully compared to the critical points found by ADF program for many cases.



All these tests reveal interaction signatures and 3D iso-surfaces of the  $\delta g$  descriptor that are very similar to those obtained using GTO primitives. So far, the aforementioned indicative  $\delta g$  peak scale and IBSI scale hold within the use of STO basis functions.

For bond strength calculations based upon STO family, the descriptor IBSI has been normalized to 1 for  $H_2$  at the DFT(M06-2X)/TZP level of theory in ADF.

Important points :

- ADF parser in IGMPlot does not handle point group symmetry (other than  $C_1$ , no symmetry).
- ADF has its own internal atom numbering, different from the input ordering provided by the user. The IGMPlot analysis uses the ADF internal numbering.
- To know the ADF inner atom ordering, you can first proceed for instance with a critical point analysis (fast) using the IGMPlot keyword CRITIC. This will generate the file 'mol-coord.xyz' indicating the proper atomic numbering to use with IBSI or FRAG1 (FRAG2) analysing schemes.
- A 'mol-rkf.txt' text file is automatically generated by IGMPlot after reading the binary mol.rkf file. It contains a summary of the electronic structure (primitives, molecular orbital informations, ...) together with a WFN section mimicking a WFN file. But this one is just for information purpose, it can not be used to describe STO with IGMPlot since it is not formatted to provide the set of  $k^r$  (the powers of r of the cartesian functions).

### 8.23.1 FULLAOACC

This keyword will enable the full accuracy level for the calculation of the atomic orbitals.

Indeed, particularly when evaluating local descriptors from electron density, the computational bottleneck arises from the evaluation of atomic orbitals (AOs, and their derivatives) at each point of a numerical grid that covers the entire molecule. To accelerate this process, we have implemented the computationally efficient atomic orbital pruning procedure<sup>[10,11]</sup> originally proposed by D. Kozłowski and J. Pilmé and implemented in TopChem2. It is based on a precomputed threshold for each atomic orbital. Specifically, we calculate a radial cutoff distance  $R_{threshold}$  for each AO, defined as the distance at which the integral of the AO's radial function:

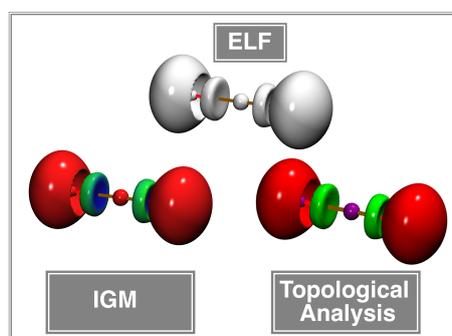
$$\int_0^{R_{threshold}} \phi(r)\phi(r)dr = 0.999999 \quad (8)$$

This threshold represents the radius within which 99.9% of the electron density of the AO is confined. During the grid evaluation, if the distance between the grid point and the nucleus of the atom hosting the AO exceeds this threshold, the AO is considered negligible and is skipped in the calculation, saving significant computational time. This approach, known as AO pruning, effectively reduces the number of AOs evaluated at each grid point, leading to a substantial performance improvement in large-scale computations.

S

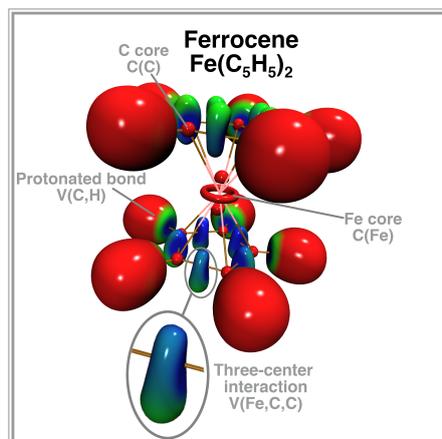
## 8.24 Example 23 (QM): ELF&IGM(test33)

We introduced a novel ELF&IGM methodology that combines the Electron Localization Function (ELF) and the Independent Gradient Model (IGM) to enable the rapid visual identification of ELF basin characteristics. Until now, computational approaches for analyzing ELF have fallen into two distinct categories: (i) comprehensive topological methods capable of rigorously defining ELF basin boundaries and integrating relevant quantities but at a high computational cost, and (ii) more accessible algorithms that merely compute ELF values and display isosurfaces without providing insight into basin nature. The ELF&IGM approach bridges this gap by offering an intermediate solution. It maintains the ability to determine basin characteristics while achieving computational efficiency gains of approximately two orders of magnitude compared to full topological analysis. This makes it a pragmatic and accessible alternative for high-throughput studies while preserving essential qualitative insights (see for example Fig. 56).



**Figure 56.** ELF localization domains (ELF=0.8) of carbon dioxide. ELF&IGM: ELF isosurfaces have been colored with the IGM  $\Gamma g(\mathbf{r})$  descriptor value using a BGR color scheme in the range [0.2:0.8]. Topological Analysis: ELF isosurfaces have been colored according to the synaptic order (purple = core, red = monosynaptic, green = disynaptic). ELF was computed over the same rectangular parallelepiped uniform grid for both calculations with a constant spacing of 0.1 Bohr between neighboring grid points. All calculations were performed at the B3LYP/6-31G\* level of theory.

The novel methodology offers a ‘low barrier’ entry point into the realm of ELF interpretation, serving as a foundation for more advanced topological analyses using sophisticated tools like TopChem2.<sup>[10,11]</sup> The ELF&IGM approach enables quick exploration of challenging systems ranging from transition metal complexes to large biological assemblies (beyond the reach of traditional methods due to prohibitive computational costs). Remarkably, in challenging systems like transition metal complexes (see Fig. 57), the ELF&IGM approach not only complements traditional topological analysis but also reveals nuanced electronic structural insights that conventional algorithms might struggle to capture.



**Figure 57.** Ferrocene investigation; 0.83 ELF isosurface colored with the IGM  $\Gamma g(\mathbf{r})$  descriptor value using a BGR color scheme in the range [0.2:0.8]. All calculations were performed at the M06-2X/aug-cc-pVDZ level of theory.

In addition, ELF<sup>&</sup>IGM provides a complementary perspective that challenges conventional understanding of chemical bonding (see reference<sup>[12]</sup> for more information).

## 9 Future prospects and outlook

Accelerating the code is one of our next concerns. This will be addressed in two ways. First, in addition to the OpenMP current possibility, a GPU version of IGMPlot is under progress, which will extend the parallel execution on several nodes on a computational center. This work is made in collaboration with the Irish Centre for High-End Computing (<https://www.ichec.ie>). In addition, the algorithm will be improved such to perform a pre-filtering of the grid points to reduce the number of calculations.

Deep-learning possibilities are currently investigated to obtain an interaction energy scoring function from fast promolecular calculations.

## References

- [1] C. Lefebvre, J. Klein, H. Khartabil, J.-C. Boisson, E. Hénon, <https://doi.org/10.1002/jcc.27123>, *J. Comp. Chem.*, **2023**, *44*, 1750–1766.
- [2] C. Lefebvre, G. Rubez, H. Khartabil, J.-C. Boisson, J. Contreras-García, E. Hénon, *Phys. Chem. Chem. Phys.* **2017**, *19*, 17928–17936.
- [3] C. Lefebvre, H. Khartabil, J.-C. Boisson, J. Contreras-García, J.-P. Piquemal, E. Hénon, *ChemPhysChem*, **2018**, *19*, 724–735.
- [4] M. Ponce-Vargas, C. Lefebvre, J.-C. Boisson, E. Hénon, *J. Chem. Inf. Model.* **2020**, *60*, 268–278.
- [5] J. Klein, H. Khartabil, J.-C. Boisson, J. Contreras-García, J.-P. Piquemal, E. Hénon, *J. Phys. Chem. A* **2020**, *124*, 1850–1860.



- [6] E. K. Wieduwilt, J.-C. Boisson, G. Terraneo, E. Hénon, A. Genoni, *J. Chem. Inf. Model.* **2021**, *61*, 795–809.
- [7] C. Lefebvre, H. Khartabil, E. Hénon, *Phys. Chem. Chem. Phys.*, just accepted, <https://doi.org/10.1039/D2CP02839E> **2023**.
- [8] T. Lu, Q. Chen, *J. Comp. Chem.* **2022**, *43*, 539–555.
- [9] E. R. Johnson, S. Keinan, P. Mori-Sanchez, J. Contreras-García, A. J. Cohen, W. Yang, *J. Am. Chem. Soc.* **2010**, *132*, 6498–6506.
- [10] D. Kozłowski, J. Pilmé, *J. Comp. Chem.* **2011**, *32*, 3207–3217.
- [11] H. Chevreau, J. Pilmé, *J. Comp. Chem.* **2023**, *44*, 1505–1516.
- [12] H. Khartabil, A. Rajamani, C. Lefebvre, J. Pilmé, E. Hénon, *J. Comp. Chem.* **2025**, Submitted for publication.
- [13] T. Lu, F. Chen, *J. Comput. Chem.* **2012**, *33*, 580–592.
- [14] T. Lu, Q. Chen, *ChemRxiv* **2022**, DOI: 10.26434/chemrxiv-2022-g1m34, this content is a preprint and has not been peer-reviewed.
- [15] A. T. Fenley, N. M. Henriksen, H. S. Muddana, M. K. Gilson, *J. Chem. Theor. Comput.* **2014**, *10*, 4069–4078.
- [16] H. Khartabil, L. Doudet, I. Allart-Simon, M. Ponce-Vargas, S. Gérard, E. Hénon, *Org. Biomol. Chem.* **2020**, *18*, 6840–6848.
- [17] T. Minervini, B. Cardey, S. Foley, C. Ramseyer, M. Enescu, *Metallomics*, **2019**, *11*, 833–844.