



**HAL**  
open science

# A Step toward the Quantification of Noncovalent Interactions in Large Biological Systems: The Independent Gradient Model-Extremely Localized Molecular Orbital Approach

Erna Wieduwilt, Jean-Charles Boisson, Giancarlo Terraneo, Eric Hénon,  
Alessandro Genoni

## ► To cite this version:

Erna Wieduwilt, Jean-Charles Boisson, Giancarlo Terraneo, Eric Hénon, Alessandro Genoni. A Step toward the Quantification of Noncovalent Interactions in Large Biological Systems: The Independent Gradient Model-Extremely Localized Molecular Orbital Approach. *Journal of Chemical Information and Modeling*, 2021, 61 (2), pp.795-809. <10.1021/acs.jcim.0c01188>. <hal-03114658>

**HAL Id: hal-03114658**

**<https://hal.univ-lorraine.fr/hal-03114658v1>**

Submitted on 19 Jan 2021

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

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



HAL Authorization

This document is the Accepted Manuscript version of a Published Work that appeared in final form in the *Journal of Chemical Information and Modeling*, copyright © American Chemical Society after peer review and technical editing by the publisher. To access the final edited and published work see <https://pubs.acs.org/doi/10.1021/acs.jcim.0c01188>.

# A Step Towards the Quantification of Non-Covalent Interactions in Large Biological Systems: the Independent Gradient Model-Extremely Localized Molecular Orbital Approach

Erna K. Wieduwilt,<sup>†</sup> Jean-Charles Boisson,<sup>‡</sup> Giancarlo Terraneo,<sup>¶</sup> Eric Hénon,<sup>\*,§</sup>  
and Alessandro Genoni<sup>\*,†</sup>

<sup>†</sup>*Université de Lorraine & CNRS, Laboratoire de Physique et Chimie Théoriques, UMR  
CNRS 7019, 1 Boulevard Arago, F-57078 Metz, France*

<sup>‡</sup>*CReSTIC EA 3804, Université de Reims Champagne-Ardenne, Moulin de la Housse,  
F-51687 Reims Cedex 02 BP39, France*

<sup>¶</sup>*Laboratory of Supramolecular and Bio-Nanomaterials (SupraBioNanoLab), Department of  
Chemistry, Materials and Chemical Engineering “Giulio Natta”, Politecnico di Milano, Via  
L. Mancinelli 7, I-20131 Milan, Italy*

<sup>§</sup>*Institut de Chimie Moléculaire de Reims UMR CNRS 7312, Université de Reims  
Champagne-Ardenne, Moulin de la Housse, F-51687 Reims Cedex 02 BP39, France*

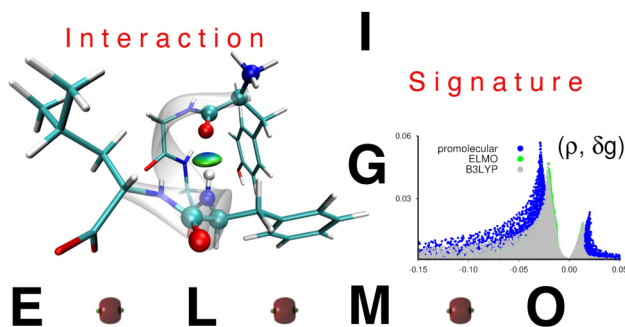
E-mail: [eric.henon@univ-reims.fr](mailto:eric.henon@univ-reims.fr); [Alessandro.Genoni@univ-lorraine.fr](mailto:Alessandro.Genoni@univ-lorraine.fr)

Phone: +33(0)3 26 91 84 97; +33(0)3 72 74 91 70

## Abstract

The Independent Gradient Model (IGM) is a recent electron density-based computational method that enables to detect and quantify covalent and non-covalent interactions. When applied to large systems, the original version of the technique still relies on promolecular electron densities given by the sum of spherically-averaged atomic electron distributions, which leads to approximate evaluations of the inter- and intramolecular interactions occurring in systems of biological interest. To overcome this drawback and perform IGM analyses based on quantum mechanically rigorous electron densities also for macromolecular systems, we coupled the IGM approach with the recently constructed libraries of extremely localized molecular orbitals (ELMOs) that allow fast and reliable reconstructions of polypeptide and protein electron densities. The validation tests performed on small polypeptides and peptide dimers have shown that the novel IGM-ELMO strategy provides results that are systematically closer to the fully quantum mechanical ones and outperforms the IGM method based on the crude promolecular approximation, but still keeping a quite low computational cost. The results of the test calculations carried out on proteins have also confirmed the trends observed for the IGM analyses conducted on small systems. This makes us envisage the future application of the novel IGM-ELMO approach to unravel complicated non-covalent interaction networks (e.g., in protein-protein contacts) or to rationally design new drugs through molecular docking calculations and virtual high-throughput screenings.

## TOC Graphic



# Introduction

Non-covalent interactions (NCIs) play a key role in biochemistry. They comprise a wide variety of attractive and repulsive forces, including van der Waals (vdW), hydrogen-bonding and electrostatic interactions. Besides being involved in ligand–protein biomolecular recognition,<sup>[1][3]</sup> which is an essential concept in the field of drug-design, NCIs are also crucial in protein-protein contacts,<sup>[4]</sup> which have been shown to be extremely important to understand biological processes in living cells. NCIs are also primarily responsible for the three-dimensional arrangement that biological polymers adopt. Such weak intra-molecular interactions occur between two chemical patterns located within the same chemical structure. They are ubiquitous in biomolecular systems and are responsible for structure, stability and dynamics of proteins and DNA.

Obtaining accurate energy values for non-covalent inter- and intra-molecular interactions would be desirable to get better and deeper insights into the nature of stabilizations in protein systems. Several theoretical methods have been implemented to quantify NCIs at various degrees of accuracy. At the lowest level, NCIs can be described as a function of the distances between interacting atoms. Some docking sampling algorithms use this fast computational approach employing a knowledge-based potential<sup>[5]</sup> or an empirical scoring function as, for example, the one implemented in Autodock Vina.<sup>[6]</sup> Furthermore, semi-empirically fitted force-fields exploit the Coulomb law and the Lennard-Jones potential to assess electrostatic and vdW interactions, respectively, in a still computationally attractive approach allowing for protein simulations.<sup>[7]</sup> To a higher degree there are also the methods based upon high-level quantum mechanical (QM) calculations, which are clearly the most sophisticated choices for the accurate description of NCIs. Among them we can mention the supermolecular approach, symmetry adapted perturbation theory<sup>[8][9]</sup> (SAPT) or the energy decomposition schemes,<sup>[10][11]</sup> which were developed to more easily investigate the nature of stabilizations via non-covalent interactions. In other respect it is essential to recall the enhanced possibilities offered by the Interacting Quantum Atom (IQA) method<sup>[12][13]</sup> in the context of molecular interactions. IQA

is a fully Quantum Chemical Topology (QCT) technique that uses a partition of the total electron density (ED)  $\rho$  into atomic contributions and writes the total energy of the system under investigation as sum of intra- and inter-atomic energies. The IQA approach is able to deliver a global interaction energy within a very useful decomposition scheme involving physically meaningful terms. IQA can be currently considered as the ultimate benchmark to perform quantitative studies of all kind of interactions (covalent, non-covalent and even steric<sup>[14]</sup>). However, its non-negligible computational cost currently limits its application to large systems and, for this reason, at the moment we have to resort to more computationally advantageous (but also significantly more approximate) techniques, such as those discussed below and the one proposed in this work.

In fact, in recent years, as complementary tools of the Quantum Theory of Atoms in Molecules (QTAIM),<sup>[15]</sup> electron density-based local approaches have been developed to detect regions of space where NCIs manifest. In particular, they exploit the gradient of the electron density ( $\nabla\rho$ ) in order to provide relevant insights and serve as interpretative techniques of interactions. Examples of such methods are the Density Overlap Region Indicator (DORI),<sup>[16][17]</sup> the so-called NCI tool,<sup>[18-23]</sup> and, more recently, the Independent Gradient Model (IGM) strategy,<sup>[24-27]</sup> which is used in this work. All of these techniques stand apart from the prominent QTAIM method. They do not aim either at characterizing the topological properties of the ED (critical points, bond paths, ...) or at dividing the space into non-overlapping atomic basins as in the QTAIM approach. On the contrary, they carry information on regions of space where the total ED results from an overlap of individual atomic or molecular electron distributions. These overlap-based approaches are generally accompanied by the construction of 3D plots consisting of isosurfaces of the used local descriptor ( $f(\nabla\rho)$ ). Such isosurfaces are practically always found as soon as electron densities of different entities clash, and interestingly correspond to troughs/peaks of the  $(\rho, f(\nabla\rho))$  plots, which occur in the presence of non-covalent interactions (see below and also in the Supporting Information where the IGM approach will be discussed in detail).

For the sake of completeness, based on the sign of the second eigenvalue of the ED hessian ( $\lambda_2$ ; see also Supporting Information), the NCI approach initially introduced a criterion to distinguish between the different types of interactions.<sup>[18]</sup> Bonded interactions are detected between two atoms when  $\lambda_2 < 0$ . This is in line with Bader’s QTAIM model where the presence of a (3,-1) critical point (CP) and its associated bond path of locally maximum density linking the nuclei in an equilibrium geometry is both necessary and sufficient for the two atoms to be bonded.<sup>[28]</sup> Remarkably, this bond path is mirrored by a line in the virial field (which is homeomorphic to the electron density), along which the potential energy density is maximally negative (stabilizing). Conversely, within the NCI framework, atoms are classified in a non-bonded contact when  $\lambda_2$  is locally found positive (ring or cage situations). Therefore, the criterion based on the sign of  $\lambda_2$  has been proposed to classify NCIs as strongly attractive (e.g., hydrogen bonds), weak (e.g., vdW interactions) and steric when the troughs/peaks of  $f(\nabla\rho)$  occur at negative, small negative/positive and positive values of  $\text{sign}(\lambda_2)\rho$ , respectively. The DORI approach adopted the same strategy to classify the interactions, revealing for instance steric crowding with the help of the sign of  $\lambda_2$ .<sup>[16]</sup> However, it is also fair to mention that the meaning of the sign of  $\lambda_2$  is sometimes over-interpreted and may erroneously lead to assert the existence of repulsive interactions in the middle of stable aromatic rings. Actually, a positive value of  $\lambda_2$  does not indicate a repulsive potential energy of interaction, but instead only a situation in which an excess of kinetic energy locally dominates the stabilizing potential energy. Within the IGM framework, the local descriptor  $\delta g$  accounts for the tendency of electrons to be shared between atoms (see below and the Supporting Information for more details). It measures to some extent (not in a direct manner) the kinetic energy lowering resulting from shared electrons being able to roam over a larger area in the real system than in non-interacting fragments. This difference of kinetic energy, which is detected by IGM between the non-interacting reference state and the real system, can take place in both bonded ( $\lambda_2 < 0$ ) and non-bonded ( $\lambda_2 > 0$ ) regions of the real molecular system, offering two complementary interpretations.

Bearing in mind the previous caveats, nowadays the ED gradient-based methods are anyway successful and pragmatic strategies to detect non-covalent interactions in molecular systems. As already mentioned above, they are less computationally demanding compared to fully QCT techniques and, for this reason, considering the power of the currently available computational resources, they are also the most adapted strategies to be extended to large macromolecules. In this work we will particularly see how this is possible for the Independent Gradient Model.

As briefly mentioned above, the IGM technique is based on a local descriptor,  $\delta g$ , which measures for two given fragments the difference between a virtual upper limit of the ED gradient norm ( $|\nabla\rho^{IGM}|$ ), which corresponds to a non-interacting system (Independent Gradient Model), and the true gradient norm ( $|\nabla\rho|$ ), which is associated with the real interacting system. Therefore,  $\delta g$  can be seen as a local measure of the electron sharing caused by the ED clash between subunits, thereby representing a step forward in the development of novel computational tools for the analysis of chemical interactions. Of course, it is worth bearing in mind that this can be seen as a convenient zero-order quantification/characterization of the interactions, while finer details (for instance, the purely quantum effects due to exchange) can be captured by only applying more sophisticated (and more expensive) QCT methods as IQA. A summary of theoretical details for the IGM approach is given in the Supporting Information.

By plotting the pair  $(\rho, \delta g)$  collected on the nodes of a sampling-grid, one can easily obtain specific two-dimensional “fingerprints” with peaks corresponding to chemical interactions that cover both covalent and non-covalent domains. Compared to other local descriptors, IGM- $\delta g$  has the advantage of intrinsically and directly carrying information on the strength of the interaction. In fact, a strong link between the integrated  $\delta g^{atom\ pair}$  bond signature and the physically grounded concept of bond force-constant has been recently demonstrated, leading to the new intrinsic bond strength index (IBSI).<sup>27</sup> Moreover, in the range of weak inter-molecular interactions, good correlations have been observed between  $\delta g$

peak heights and accurately calculated stabilization energies for different hydrogen bonds, not only at the potential energy minimum, but also at different non-equilibrium distances between monomers.<sup>24</sup> This indicates the capability of this method to quantitatively characterize NCIs, even in those cases in which they are very weak. Another appealing feature of the IGM approach is to provide an intra/inter uncoupling-scheme that automatically extracts the signature of intra- ( $\delta g^{intra}$ ) and inter- ( $\delta g^{inter}$ ) molecular interactions within and between two user-selected fragments, respectively. This possibility is particularly attractive to assess the role of non-covalent interactions that occur in an intra-molecular fashion between atoms or sub-fragments in one single molecule (for instance to characterize intra-molecular  $\pi - \pi$  stacking or hydrogen-bonding).

As anticipated above, the corner stone of the IGM approach is the electron density  $\rho$ . Nevertheless, although the IGM strategy relies on the knowledge of the ED, no topological analysis is required and the IGM investigation can be achieved with little preparation. The method was initially designed to work only with promolecular electron densities (PEDs),<sup>24</sup> which are non-relaxed electron distributions given by the sum of spherically-averaged neutral atomic densities. In the current *IGMPlot* program,<sup>29</sup> they are straightforwardly obtained as sum of simple exponential functions properly tabulated and stored for the atoms of the first four periods, thus making the IGM calculations very fast. In 2018 the Gradient-Based Partitioning (GBP) was also proposed to extend the IGM concept to electron densities derived from wave functions.<sup>25</sup> Therefore, up to now, two kinds of ED source (promolecular or relaxed QM-based ED) can be used to probe NCIs within the IGM approach, whilst the description of strong interactions (like covalent bonding) is restricted to the use of relaxed ED obtained from quantum mechanical computations.

Although approximate, the PED has been shown to provide qualitative results similar to the relaxed fully QM one, provided that the calculations remain in the non-covalent domain.<sup>25,26</sup> Thus, promolecular densities are useful to study large systems (typically ligand-protein complexes), requiring only the geometry of the macromolecular system as input.

However, in view of quantifying weak interactions through IGM integration schemes, the use of such a PED approximation may be unacceptable. Actually, beyond lacking relaxation, the promolecular electron density (which is also known as independent atom model), imposes a spherical form and electroneutrality of atoms. Therefore, the PED rather poorly describes amino acids having either a negative (Asp, Glu) or a positive charge (Lys, Arg) in their conventional protonation state at physiological pH. Furthermore, to a lower extent, it cannot characterize in a rigorous way the polar character of residues such as Gln, Asn, Tyr, Thr and Ser that are able to form hydrogen bonds as proton donors or acceptors. In all the situations mentioned above, the use of highly accurate QM methods would be necessary, but their large computational cost would unfortunately prevent the application of the IGM approach to extended biological systems.

Nowadays theoretical chemistry offers different possibilities to overcome this drawback. Some of them are represented by reliable machine-learning models of the electron density that have been also used to investigate biological systems.<sup>30-33</sup> Another option, which will be exploited in the present work, is given by the recently constructed libraries of extremely localized molecular orbitals (ELMOs).<sup>34-36</sup> ELMOs are molecular orbitals (MOs) strictly localized on small molecular subunits<sup>37-39</sup> (e.g., atoms, bonds or functional groups) and, due to their absolute localization, they can be easily and reliably transferred from one molecule to another.<sup>34-35-39-44</sup> On the basis of this property, libraries of ELMOs have been constructed with the goal of almost instantaneously obtaining approximate wave function and electron densities of molecules ranging from small polypeptides to large proteins.<sup>36</sup> They have been also recently exploited to perform the first successful quantum mechanics-based refinements of crystal structures of relatively large biological molecules<sup>45</sup> in the context of quantum crystallography<sup>46-51</sup> and to develop the novel QM/ELMO (quantum mechanics / extremely localized molecular orbital) technique,<sup>52-54</sup> namely an embedding strategy in which the chemically active region of the system is treated at fully quantum chemical level, while the environment is described through transferred and frozen ELMOs.

In the context of this paper, it is also worth noting that the ELMO databanks have been also lately coupled with the above-mentioned NCI tool, giving rise to the NCI-ELMO method and revealing significant improvements in the qualitative description of non-covalent interaction fingerprints compared to using approximate PEDs.<sup>55</sup> Therefore, coupling IGM with the ELMO libraries could be considered as a good way towards the quantitative assessment of non-covalent interactions present in large molecules. Hence, the aim of the present study is to perform for the first time IGM-ELMO calculations for systems of biological interest, observing the influence of using such improved EDs on the description of non-covalent interactions, at both qualitative and quantitative levels.

## Results and discussion

### Validation tests

In order to start testing capabilities and performances of the new IGM-ELMO approach, we initially performed a series of validation tests on relatively small polypeptide monomers and dimers. For each of the considered systems, we compared the results of IGM analyses based on corresponding DFT (B3LYP functional), promolecular and transferred-ELMO (from now on also indicated as ELMOdb) electron densities. The reason of initially focusing on polypeptide monomers and dimers was dictated by the need of considering relatively small systems that are still treatable at fully quantum mechanical level (for instance, through DFT computations) and, at the same time, characterized by non-covalent interactions similar to those usually present in large biological molecules, such as proteins.

At first, we focused on the following two peptides: i) Leu-enkephalin (86 atoms), for which a high-resolution X-ray crystal structure is available,<sup>56</sup> ii) a fifteen-residue synthetic peptide (275 atoms) corresponding to an intracellular sequence of the  $\beta$ -adrenergic receptor,<sup>57</sup> for which an NMR structure is deposited in the Protein Data Bank (PDB code: 1DEP). In the former we considered two interactions between the Tyr1 and Phe4 residues: the backbone

hydrogen-bond and the T-shaped  $\pi - \pi$  stacking occurring between the side-chain aromatic rings. In the latter we took into account the multiple hydrogen-bond interaction between the negatively charged residue Asp4 and the positively charged residues Arg1 and Arg11. The three analyzed cases allowed us to evaluate the performances of the novel IGM-ELMO technique in describing the most common types of NCIs occurring in biological systems, such as hydrogen-bonding and van der Waals interactions.

In Figures 1A, 1C and 1E, we reported the  $\delta g$ -isosurfaces corresponding to the intramolecular interactions mentioned above, as obtained from IGM calculations relying on ELMOdb/cc-pVDZ electron densities. These isosurfaces correctly identify the different NCIs that we decided to analyze and, moreover, they are hardly distinguishable from those obtained through IGM computations based on B3LYP/cc-pVDZ electron distributions (this is the reason why in Figure 1 we decided to plot only the IGM-ELMO results). On the contrary, using the same isovalue, the isosurfaces resulting from the IGM-promolecular analyses are more diffuse than the IGM-ELMO and IGM-DFT ones (see Figures 1B, 1D and 1F), although they still correctly detect the NCIs in the real space. Furthermore, for the hydrogen bond interactions, we observe that the IGM-promolecular isosurfaces are slightly shifted towards the hydrogen-bond acceptor compared to the QM-based ones (see Figures 1B and 1F). Completely analogous results were obtained when the other four basis-sets currently available for the ELMO libraries (i.e., 6-31G, 6-311G, 6-31G(d,p) and 6-311G(d,p)) were adopted for the IGM-DFT and IGM-ELMO calculations.

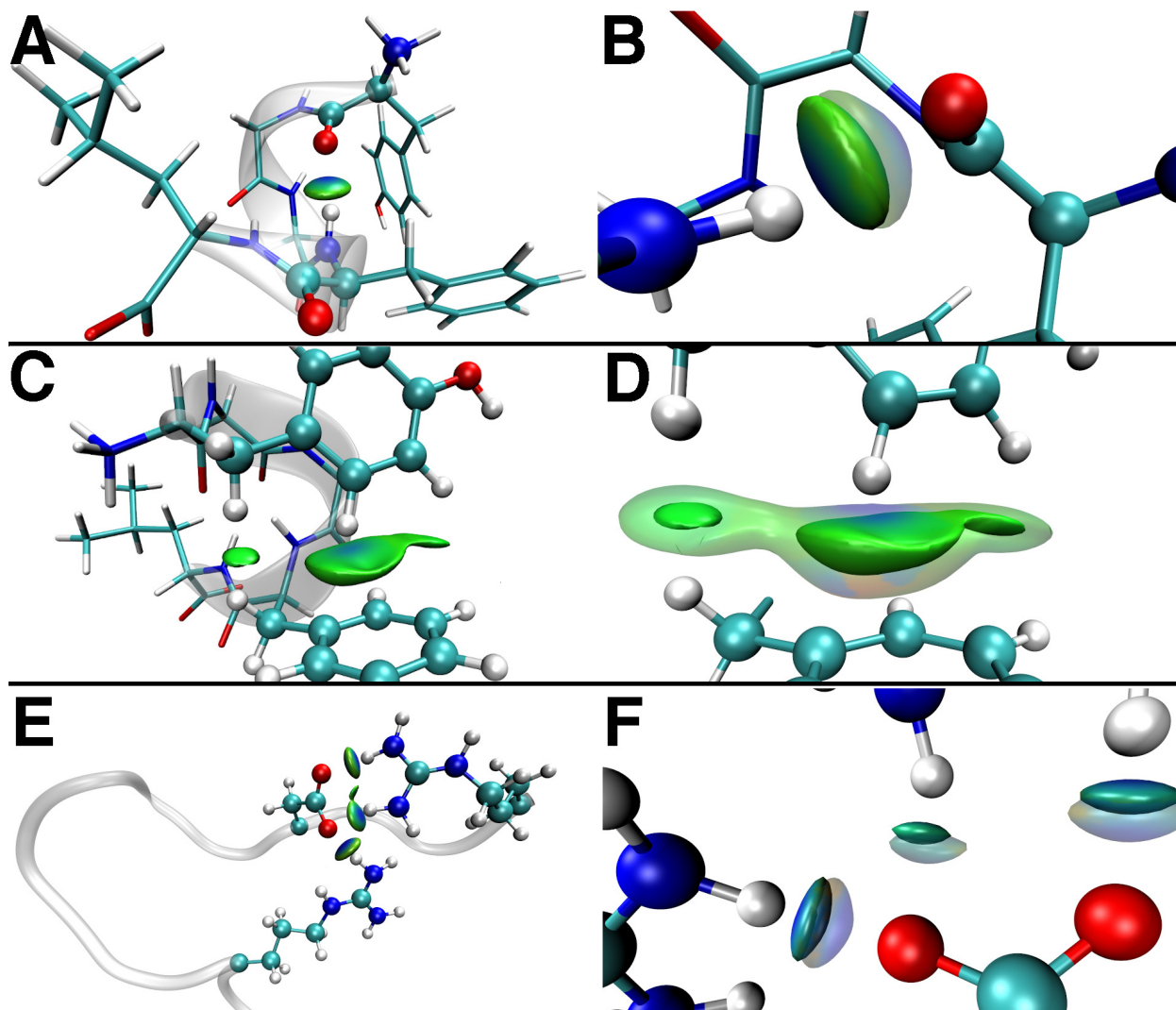


Figure 1:  $\delta g$ -isosurfaces associated with (A) the backbone hydrogen-bond interaction between residues Tyr1 and Phe4 in Leu-enkephalin at ELMOdb/cc-pVDZ level (zoomed view in (B), with superimposition of the IGM-promolecular isosurface in translucent representation; isovalues set equal to 0.01 a.u.), (C) the T-shaped  $\pi - \pi$  stacking between residues Tyr1 and Phe4 in Leu-enkephalin at ELMOdb/cc-pVDZ level (zoomed view in (D), with superimposition of the IGM-promolecular isosurface in translucent representation; isovalues set equal to 0.004 a.u.), and (E) the multiple hydrogen-bond between the charged residues Asp4, Arg1 and Arg11 in the fifteen-residue polypeptide 1DEP (zoomed view in (F), with superimposition of the IGM-promolecular isosurface in translucent representation; isovalues set equal to 0.015 a.u.). All the isosurfaces are colored according to the BGR scheme over the range  $-0.05 \text{ a.u.} < \text{sign}(\lambda_2)\rho < 0.05 \text{ a.u.}$

In order to better highlight and discuss the possible differences between the different IGM analyses, we afterwards decided to focus on the typical two-dimensional  $(\rho, \delta g)$  fingerprint-

plots. In Figure 2, for each investigated interaction, we have superimposed the 2D plots obtained at promolecular, DFT-B3LYP and ELMOdb levels, always considering the cc-pVDZ basis-set for the quantum mechanical calculations (separated plots have been reported in Figures S3-S5 of the Supporting Information). From the analyses of these graphs it clearly emerges that, for all the interactions, the description based on transferred and frozen extremely localized molecular orbitals is always closer to the DFT one. For the hydrogen-bonding in Leu-enkephalin (see Figures 2A and 2B), we can notice that in the promolecular description we have peaks that occur at greater (absolute) values of the signed electron density and that are higher in terms of  $\delta g$  value compared to the DFT-B3LYP case. On the contrary, the IGM-ELMO analysis gives peaks that are slightly shifted to lower (absolute) ED values and of lower intensity; however, despite these differences, the higher similarity between the IGM-DFT and IGM-ELMO ( $\rho$ ,  $\delta g$ ) fingerprint-plots is clear. The similarity is even higher in the case of the T-shaped  $\pi - \pi$  stacking interaction (see Figures 2C and 2D). In fact, while the promolecular description always provides higher peaks that are observed at greater (absolute) values of the signed electron distribution, the ELMO- and DFT-based plots are almost identical, with the associated peaks that practically overlap almost perfectly. Finally, concerning the multiple hydrogen-bond interaction between charged residues in the synthetic peptide 1DEP (see Figures 2E and 2F), we have a situation comparable to the one observed in Figures 2A and 2B, with the IGM-ELMO description that is clearly more similar to the DFT one compared to the promolecular case.

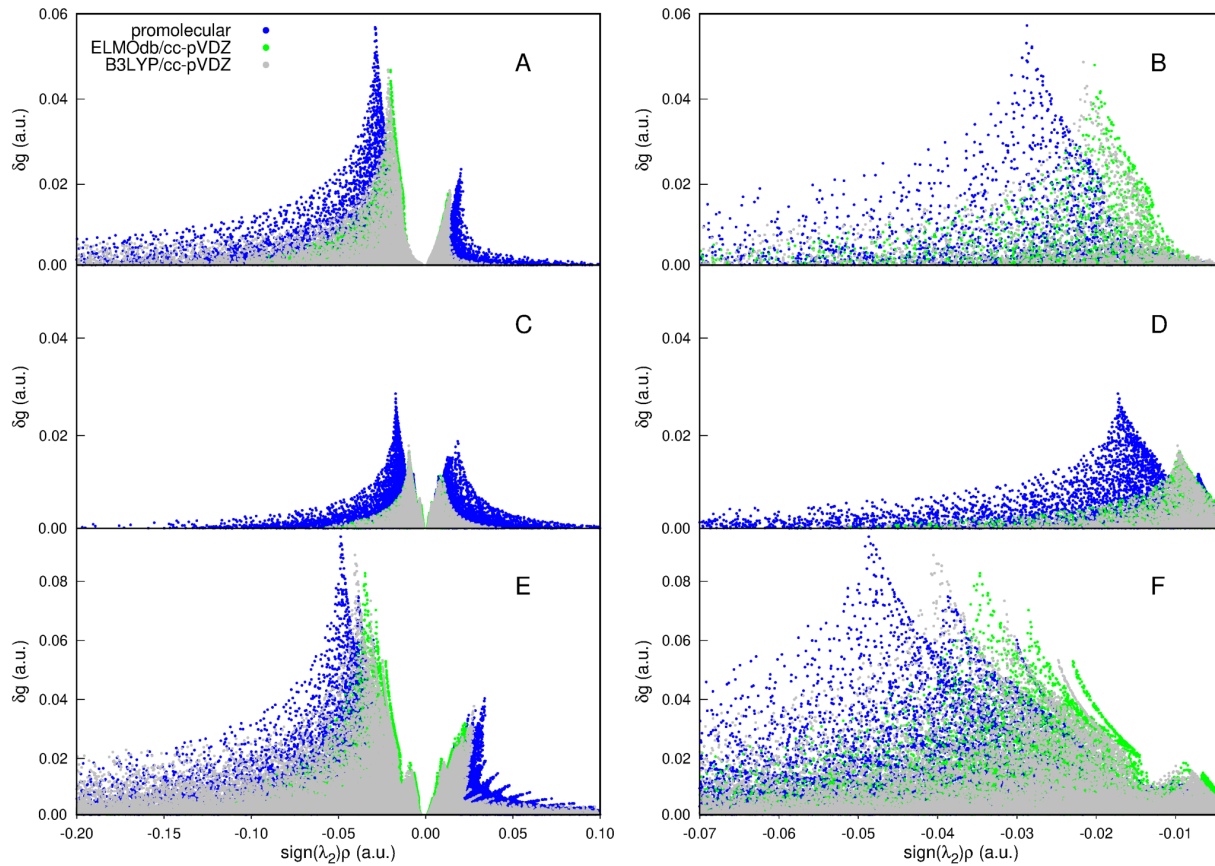


Figure 2: Comparison between the  $(\rho, \delta g)$  2D fingerprint-plots obtained at promolecular, ELMOdb/cc-pVDZ and B3LYP/cc-pVDZ levels, with zooms on the peaks (in the negative region) in the right panels: (A) and (B) local hydrogen-bond between residues Tyr1 and Phe4 in Leu-enkephalin, (C) and (D) T-shaped  $\pi - \pi$  stacking between Tyr1 and Phe4 in Leu-enkephalin, (E) and (F) multiple hydrogen-bond between charged residues Asp4, Arg1 and Arg11 in polypeptide 1DEP.

Results analogous to those shown above were also obtained with the other four basis-sets considered in our calculations. In this regard, in Figure 3 we showed how the two-dimensional fingerprint-plots change when different basis-sets are used. We can observe that, irrespective of the analyzed interaction, the basis-set has only a marginal influence on the IGM analysis, at least from a qualitative point of view. In fact, the  $(\rho, \delta g)$  fingerprint-plots obtained with the different basis-sets practically overlap and their differences are really negligible, at both DFT-B3LYP and ELMOdb levels.

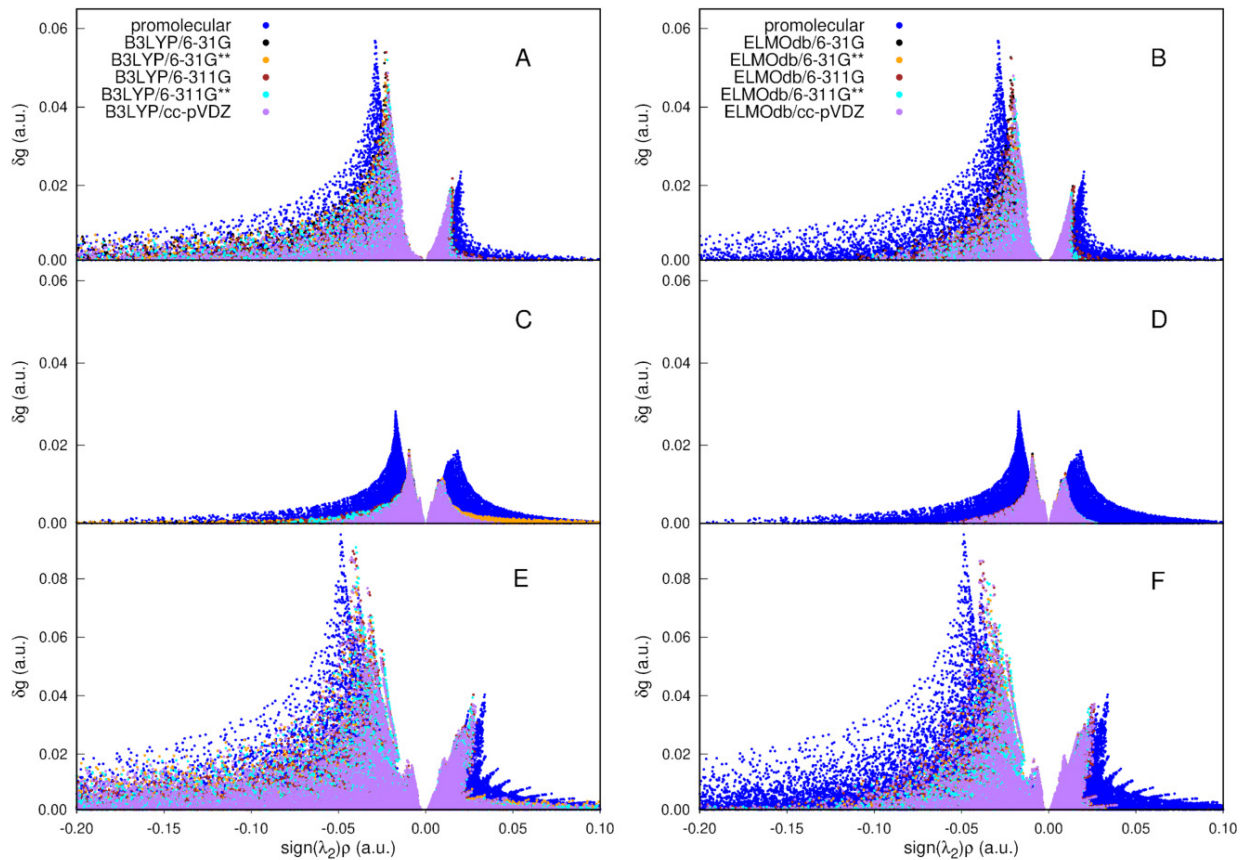


Figure 3: Basis set dependence of the  $(\rho, \delta g)$  2D-fingerprint plots associated with (A and B) the local hydrogen-bond between residues Tyr1 and Phe4 in Leu-enkephalin, (C and D) the T-shaped  $\pi - \pi$  stacking between Tyr1 and Phe4 in Leu-enkephalin, and (E and F) the multiple hydrogen-bond between charged residues Asp4, Arg1 and Arg11 in polypeptide 1DEP.

So far, we based our discussion only on qualitative aspects and the results confirmed the trends already observed for the recent NCI-ELMO approach, which was shown to provide reduced density gradient (RDG) isosurfaces and  $(\rho, \text{RDG})$  plots in close agreement with those resulting from NCI analyses based on fully quantum mechanical electron densities.<sup>55</sup> However, since the Independent Gradient Model also allows the evaluation of the strengths associated with the interactions, in this study we decided to extend our comparisons also to a quantitative level. To accomplish this task, we integrated the peaks appearing in the negative region of the obtained  $(\rho, \delta g)$  plots (for more details about this strategy, see eq. S4 in the Supporting Information). The obtained values (from now on indicated as  $\Delta g$ ) are reported in

Table 1 and practically confirm the trends already observed from the qualitative comparisons of the two-dimensional fingerprints-plots. In fact, for all the considered interactions, the reference DFT-B3LYP  $\Delta g$ 's are always closer to the ELMOdb values rather than to the promolecular ones. It is also confirmed that the adopted basis-sets have little influence on the IGM description. Furthermore, analyzing Table 1 more in detail, we can actually observe that, in absolute values, the promolecular results overestimate the DFT ones, whereas the ELMO  $\Delta g$ 's are always slightly lower than the DFT benchmarks, which is consistent with the obtained  $(\rho, \delta g)$  plots shown above.

**Table 1:  $\Delta g$  values (in a.u.) resulting from the IGM analyses performed for the preliminary validation tests on the Leu-enkephalin and 1DEP polypeptides. The promolecular values are reported on the 6-31G row.**

| basis       | Leu-enk (H-bond) |        |         | Leu-enk ( $\pi - \pi$ ) |        |         | 1DEP   |        |         |
|-------------|------------------|--------|---------|-------------------------|--------|---------|--------|--------|---------|
|             | DFT              | ELMO   | Promol. | DFT                     | ELMO   | Promol. | DFT    | ELMO   | Promol. |
| 6-31G       | -0.055           | -0.049 | -0.076  | -0.082                  | -0.071 | -0.180  | -0.238 | -0.203 | -0.305  |
| 6-311G      | -0.051           | -0.047 | -       | -0.084                  | -0.075 | -       | -0.219 | -0.194 | -       |
| 6-31G(d,p)  | -0.057           | -0.050 | -       | -0.084                  | -0.073 | -       | -0.249 | -0.202 | -       |
| 6-311G(d,p) | -0.053           | -0.046 | -       | -0.087                  | -0.078 | -       | -0.229 | -0.188 | -       |
| cc-pVDZ     | -0.056           | -0.052 | -       | -0.082                  | -0.078 | -       | -0.247 | -0.205 | -       |

Therefore, on the basis of all the above-discussed validation tests, it is clear that the IGM analyses of non-covalent interactions based on electron densities resulting from the transfer of ELMOs are very close to those resulting from the use of fully quantum mechanical electron distributions (notably, DFT calculations with B3LYP functional). Moreover, and more importantly, we have also observed that the new IGM-ELMO method practically outperforms the IGM approach based on promolecular electron densities for all the examined systems. This is even more noteworthy considering the fact that the IGM-ELMO results are obtained only at a slightly larger computational cost compared to the promolecular ones (see Table S1 in the Supporting Information). In fact, it emerges that the promolecular calculations are practically instantaneous, the IGM-ELMO computations take only few seconds, while

the timings associated with the corresponding IGM-DFT analyses are already of the order of minutes for the relatively small polypeptides taken into account for the preliminary test calculations. In this regard, it is also worth stressing that, for all the performed IGM analyses, the real IGM calculations are always extremely fast (see again Table S1 in the Supporting Information).

Before concluding the discussion on these first validation tests, we also want to stress that the differences in terms of computational cost between IGM-ELMO and fully quantum mechanical IGM analyses are certainly bound to increase if macromolecules are taken into account. Therefore, this aspect makes us envisage future applications of the proposed IGM-ELMO approach to large biological systems, for which a good trade-off between accuracy and computational cost is desirable.

In the second part of our validation tests, we decided to assess the capabilities of the IGM-ELMO approach in evaluating the strength of NCIs occurring in similar systems in order to determine possible rankings among the analyzed interactions. This could become extremely advantageous in view of a possible future use of the IGM-ELMO method in the context of virtual high throughput screenings (vHTS) or docking calculations for the rational design of new drugs. To accomplish this task, we considered the dimers of two halogenated polypeptides deriving from the bromination and chlorination of the tyrosine residue in position Y10 of the sequence 7-12 (DSGYEV) of the flexible amyloid- $\beta$  ( $A\beta$ ) peptide.<sup>58</sup> For the sake of precision, we point out that, in those cases, only two hydrogen atoms of tyrosine are substituted with halogens, namely, those at the ortho-positions with respect to the hydroxyl group in the aromatic ring of tyrosine.

From the X-ray crystal structures of the brominated and chlorinated peptides (from now on also indicated as DSGY(Br)EV and DSGY(Cl)EV, respectively),<sup>58</sup> we extracted dimers (170 atoms) characterized by a  $\pi - \pi$  interaction between the aromatic rings of the halogenated tyrosine residues (for the structures of the dimers, see Figure 4 and Figure S6 in the Supporting Information). For each of them, we also substituted the halogen with hydrogen

atoms using the software MolProbity (see the section dedicated to the Computational Details in Appendix for more details). This allowed us to have completely de-halogenated dimers and compare the strengths of their  $\pi - \pi$  interactions to those of the corresponding interactions in the halogenated cases. In analogy with the validation tests shown above, for each couple of halogenated/de-halogenated dimers we have carried out Independent Gradient Model computations based on promolecular, DFT-B3LYP and ELMOdb electron densities. Moreover, also in this case, for the QM-based analyses we considered all the basis-sets currently available in the ELMO databanks. For the sake of clarity, we want to stress that the structures (and consequently also the IGM results) of the two de-halogenated peptides of identical size and identical chemical composition considered in this investigation are different. In fact, each of them derives from the halogen-to-hydrogen substitution in one of the two crystal structures of the halogenated peptides taken into account. Therefore, in the discussion below, the obtained results will be compared within each pair of halogenated/de-halogenated dimers.

From each of the considered dimers, we also extracted further reduced structures consisting only of the tyrosine side-chains cut at the carbon  $C\beta$  and afterwards saturated with hydrogen atoms, whose positions were properly optimized keeping fixed the rest of the system. The obtained geometries were then exploited to perform counterpoise calculations of the interaction energies for the reduced dimers (see Table S2 in the Supporting Information for the actual values), which were afterwards taken as references for the IGM analyses and especially for the obtained  $\Delta g$  values. The computations were carried out at B3LYP-D3(BJ) level (B3LYP functional with Grimme’s D3 empirical correction for dispersion and Becke-Johnson damping) with all the basis-sets considered for the IGM and IGM-ELMO analyses.

In Figure 4 we depicted the obtained  $\delta g$ -isosurfaces associated with the  $\pi - \pi$  interaction in the dimer of the brominated peptide DSGY(Br)EV (see Figures 4A-4C) and in the dimer of the corresponding de-brominated peptide DSGYEV (see Figures 4D-4F), as ob-

tained through IGM calculations based on the DFT-B3LYP/cc-pVDZ, ELMOdb/cc-pVDZ and promolecular electron distributions. We can observe that, for both the brominated and de-brominated case, the IGM approach is able to correctly detect the presence of the  $\pi - \pi$  interaction, regardless of the nature of the underlying electron density. Furthermore, as one should expect, for all the levels of IGM analysis, the isosurfaces are more extended for the brominated dimer, which correlates well with the greater strength of its  $\pi - \pi$  interaction (see Table S2 in the Supporting Information). Finally, it is also interesting to note that, for both the DSGY(Br)EV and DSGYEV dimers, the IGM-DFT and IGM-ELMO  $\delta g$ -isosurfaces are generally more similar to each other rather than to the more diffuse promolecular one. Analogous results were also obtained in the case of the chlorinated/de-chlorinated dimers (see the corresponding  $\delta g$ -isosurfaces in Figure S6 of the Supporting Information).

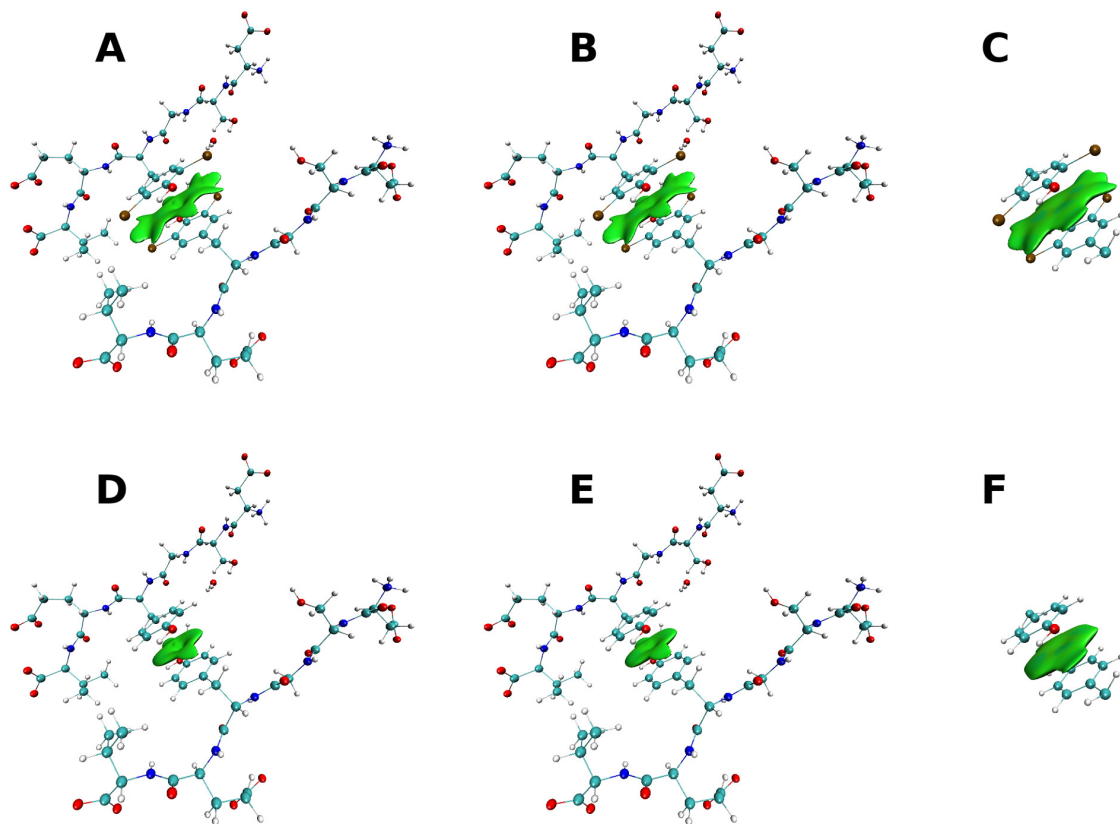


Figure 4:  $\delta g$ -isosurfaces associated with the  $\pi - \pi$  interactions in the dimers of the (A-C) brominated and (D-F) de-brominated polypeptides based on B3LYP/cc-pVDZ (A and D), ELMOdb/cc-pVDZ (B and E) and promolecular (C and F) electron densities. All the isosurfaces correspond to the 0.004 a.u. isovalue and are colored according to the BGR scheme over the range  $-0.08 \text{ a.u.} < \text{sign}(\lambda_2)\rho < 0.08 \text{ a.u.}$

We have afterwards considered the  $(\rho, \delta g)$  fingerprint-plots that are depicted in Figure 5 (B3LYP/cc-pVDZ, ELMOdb/cc-pVDZ and promolecular levels; separated plots are also reported in Figures S7 and S8 of the Supporting Information). For the brominated/de-brominated dimers (see Figures 5A and 5B), we easily notice a larger similarity between the 2D plots at DFT and ELMO levels. In particular, as for the first validation tests, we observe that the promolecular description provides higher peaks that occur at larger absolute values of the signed electron density, while the IGM-DFT and IGM-ELMO peaks are approximately equivalent, with the latter slightly higher and, to a smaller extent, shifted towards larger absolute values of the electron density. For the chlorinated and de-chlorinated dimers, we

obtained completely similar results (see Figures 5C and 5D), with the only difference that the IGM-ELMO and IGM-DFT peaks overlap almost completely. Also in these cases, the basis-set has a marginal influence on the two-dimensional IGM plots, as it can be evinced from the inspection of Figures S9 and S10 in the Supporting Information.

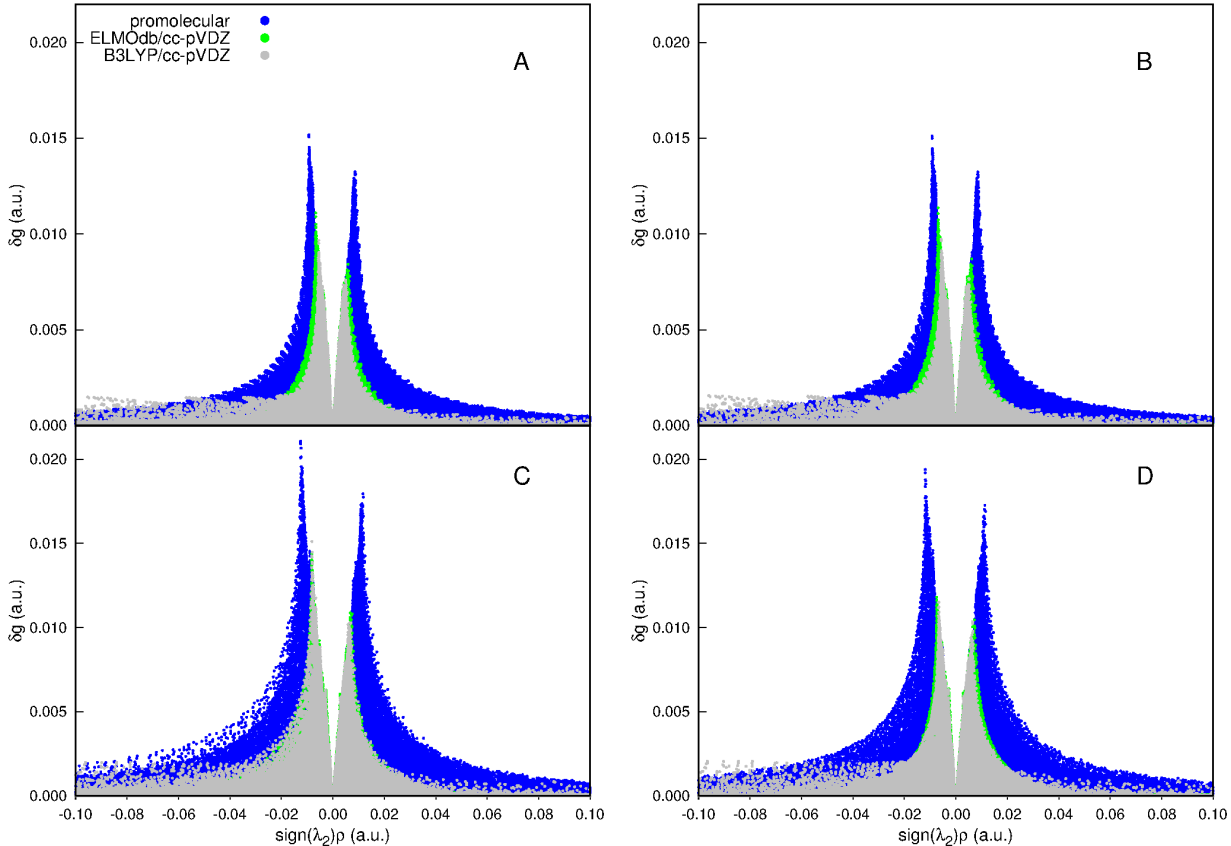


Figure 5: Comparison between the  $(\rho, \delta g)$  2D-fingerprint plots obtained at B3LYP/cc-pVDZ, ELMOdb/cc-pVDZ and promolecular levels for the  $\pi - \pi$  interactions in the dimers of the (A) brominated, (B) de-brominated, (C) chlorinated, and (D) de-chlorinated polypeptides.

To provide a more quantitative analysis, in Table 2 we also reported the  $\Delta g$  values resulting from the integration of the peaks appearing in the  $(\rho, \delta g)$  plots. It is worth reminding that the value of the  $\Delta g$  descriptor can be strictly associated with the strength of the interaction: the larger its absolute value, the stronger the associated interaction. In our cases, independently of the chosen level for the electron density (B3LYP, ELMOdb or promolecular), the  $\Delta g$  values show the same trends observed for the BSSE-corrected interaction

energies computed for the reduced dimers (see Table S2 of the Supporting Information), thus confirming that the  $\pi - \pi$  interaction between halogenated peptides is stronger than the one between the corresponding de-halogenated peptides, and also showing that the experimental geometry of the brominated dimer is more stable than the experimental geometry of the analogous chlorinated dimer. Finally, for all the adopted basis-sets, the IGM-ELMO and IGM-DFT  $\Delta g$  values are quite close to each other, which is consistent with the similarities that we have previously observed in the two-dimensional plots. Moreover, we see again that the use of different types of basis-sets has little influence on the IGM descriptions at quantum mechanical levels, with the largest discrepancies observed for the cc-pVDZ basis-set when transferred-ELMO electron densities are exploited.

**Table 2:**  $\Delta g$  values (in a.u.) for the  $\pi - \pi$  interactions between tyrosine residues as resulting from the IGM analyses performed on the halogenated/de-halogenated peptide dimers.

|                    | Brominated<br>dimer | De-brominated<br>dimer | Chlorinated<br>dimer | De-chlorinated<br>dimer |
|--------------------|---------------------|------------------------|----------------------|-------------------------|
| Promolecular       | -0.340              | -0.255                 | -0.331               | -0.281                  |
| ELMOdb/6-31G       | -0.194              | -0.126                 | -0.179               | -0.140                  |
| ELMOdb/6-311G      | -0.194              | -0.125                 | -0.181               | -0.142                  |
| ELMOdb/6-31G(d,p)  | -0.194              | -0.126                 | -0.174               | -0.139                  |
| ELMOdb/6-311G(d,p) | -0.194              | -0.127                 | -0.178               | -0.144                  |
| ELMOdb/cc-pVDZ     | -0.206              | -0.132                 | -0.190               | -0.150                  |
| B3LYP/6-31G        | -0.205              | -0.127                 | -0.180               | -0.135                  |
| B3LYP/6-311G       | -0.204              | -0.126                 | -0.180               | -0.134                  |
| B3LYP/6-31G(d,p)   | -0.207              | -0.130                 | -0.176               | -0.139                  |
| B3LYP/6-311G(d,p)  | -0.205              | -0.128                 | -0.175               | -0.137                  |
| B3LYP/cc-pVDZ      | -0.212              | -0.125                 | -0.178               | -0.133                  |

In conclusion, the second set of our validation tests have confirmed the high similarity between the IGM-ELMO and IGM-DFT analyses and, more importantly, they have also shown the capability of the novel IGM-ELMO approach in correctly evaluating (and ranking) the strength of inter-molecular interactions in systems very similar from the chemical point of view.

## Application to proteins

After the preliminary calculations on peptides and peptide dimers, we afterwards tested the new IGM-ELMO method on macromolecules in order to confirm or not the trends observed on the previously considered smaller systems. In particular, to accomplish this task, we took into account two proteins that were already used also to assess the capabilities of the related NCI-ELMO method:<sup>[55]</sup> i) human erythrocytic ubiquitin<sup>[59]</sup> (1231 atoms) and ii) human carbonic anhydrase II<sup>[60]</sup> (1477 atoms). In the former we considered the crystal structure corresponding to the Protein Data Bank (PDB) file 1UBQ and we only focused on the  $\text{CH}\cdots\pi$  interaction between the methyl group of Leu50 and the aromatic ring of Tyr59. In the latter we took into account the crystal structure deposited in the PDB with the code 3KS3 and we analyzed three different types of non-covalent interactions: i) a hydrogen bond between residues Asn61 and Gly63; ii) an  $n - \pi^*$  interaction between the lone-pair on the oxygen atom of the Asn61 sidechain and the  $\pi^*$  molecular orbital associated with the carbonyl group of the Asn61 backbone; iii) a multiple hydrogen-bond interaction between the charged residues Arg58 and Glu69. In all these cases, we carried out IGM analyses based only on promolecular and ELMOdb electron densities (for all the basis-sets currently available in the ELMO libraries). DFT calculations were impractical for the large systems taken into account.

Let us initially focus on the  $\text{CH}\cdots\pi$  interaction between residues Leu50 and Tyr59 in the ubiquitin protein. In Figure [6A](#) we depicted the  $\delta g$ -isosurface obtained at ELMOdb/cc-pVDZ level, which correctly indicates the presence of the interaction under examination. As one should expect, we observe that also the IGM analysis based on the associated promolecular density correctly detects the considered  $\text{CH}\cdots\pi$  interaction. However, the corresponding  $\delta g$ -isosurface is significantly more diffuse than the ELMOdb one (see Figure [6B](#)), as it was already observed also in many of the preliminary test calculations discussed above. In order to better highlight the differences between the IGM-ELMO and IGM-promolecular analyses, we have then considered the associated  $(\rho, \delta g)$  fingerprint-plots, which are plotted

on the same graphs in Figures 6C and 6D, with the latter showing the details of the IGM-ELMO and IGM-promolecular peaks in the negative region of the signed electron density. Again in analogy with the validation tests, also in this case the 2D fingerprint-plot at IGM-promolecular level presents signature peaks that are higher and that occur at greater absolute values of the signed electron density compared to the IGM-ELMO case. The observed differences between the promolecular and ELMOdb ( $\rho$ ,  $\delta g$ ) plots reflect on the calculated  $\Delta g$  values for the interaction. In fact, in Table 3, we can observe that, in absolute value, the promolecular  $\Delta g$  is much larger than all the  $\Delta g$ 's resulting from the use of transferred-ELMO electron densities, regardless of the adopted basis-set. This is again consistent with the results of the benchmark computations carried out on smaller systems.

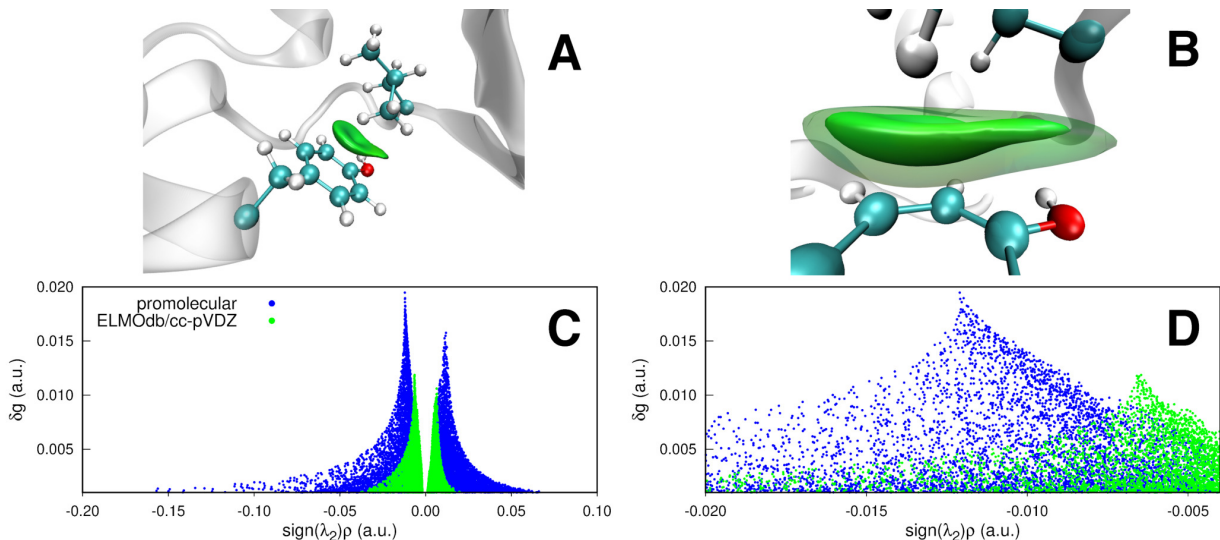


Figure 6: CH  $\cdots$   $\pi$  interaction between residues Leu50 and Tyr59 in ubiquitin (PDB ID: 1UBQ): (A)  $\delta g = 0.005$  a.u. isosurface obtained with the ELMOdb/cc-pVDZ electron density and colored according to the BGR scheme over the range  $-0.05$  a.u.  $< \text{sign}(\lambda_2)\rho < 0.05$  a.u.; (B) ELMOdb versus promolecular (translucent)  $\delta g = 0.005$  a.u. isosurfaces; (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).

We will now discuss the results of the IGM analyses for the considered interactions in the protein human carbonic anhydrase II. First of all, let us take into account the local hydrogen

**Table 3:**  $\Delta g$  values (in a.u.) associated with the non-covalent interactions taken into account in the IGM analyses carried out on proteins.

|                    | 1UBQ      | 3KS3         |                         |                 |
|--------------------|-----------|--------------|-------------------------|-----------------|
|                    | CH- $\pi$ | Local H-bond | n $\rightarrow$ $\pi^*$ | Multiple H-bond |
| Promolecular       | -0.137    | -0.080       | -0.103                  | -0.194          |
| ELMObd/6-31G       | -0.063    | -0.053       | -0.042                  | -0.125          |
| ELMObd/6-311G      | -0.064    | -0.050       | -0.042                  | -0.117          |
| ELMObd/6-31G(d,p)  | -0.064    | -0.054       | -0.041                  | -0.127          |
| ELMObd/6-311G(d,p) | -0.067    | -0.051       | -0.042                  | -0.115          |
| ELMObd/cc-pVDZ     | -0.068    | -0.056       | -0.040                  | -0.127          |

bond between residues Gly63 and Asn61. As in the previous case, also in this situation the IGM-ELMO method correctly detects the interaction under exam (Figure 7A), as indicated by the  $\delta g$ -isosurface between the N-H bond of the glycine backbone and the carbonyl group of the asparagine sidechain. Moreover, in analogy with what was observed for the CH $\cdots\pi$  interaction in ubiquitin, the corresponding promolecular isosurface is more spread out (see Figure 7B), and, in this case, slightly shifted towards the carbonyl group of the residue Asn63. These results can be explained on the basis of the  $(\rho, \delta g)$  signature plots reported in Figures 7C and 7D, where we can see that the peaks resulting from the IGM-promolecular analysis are again higher, wider (especially the one in the negative region) and always shifted towards greater absolute values of the signed electron distribution. Furthermore, the integration of the peaks provided results consistent with the previous observations. In fact, in Table 3 we can notice again that the promolecular  $\Delta g$  is greater than the corresponding IGM-ELMO values, always independently of the basis-set chosen to reconstruct the electron distribution through the transfer of extremely localized molecular orbitals from the ELMO libraries. All of these observations fully confirm the trends already seen for the hydrogen-bond interaction considered in the validation test on the Leu-enkephalin polypeptide, which can be considered analogous to the Gly63-Asn61 interaction taken into account in the human carbonic anhydrase II protein. This leads us to hypothesize that, if quantum mechanical calculations (particularly DFT calculations) were practically feasible on the large protein under exam, we would obtain IGM results close to those based on the transferred-ELMOs approximation.

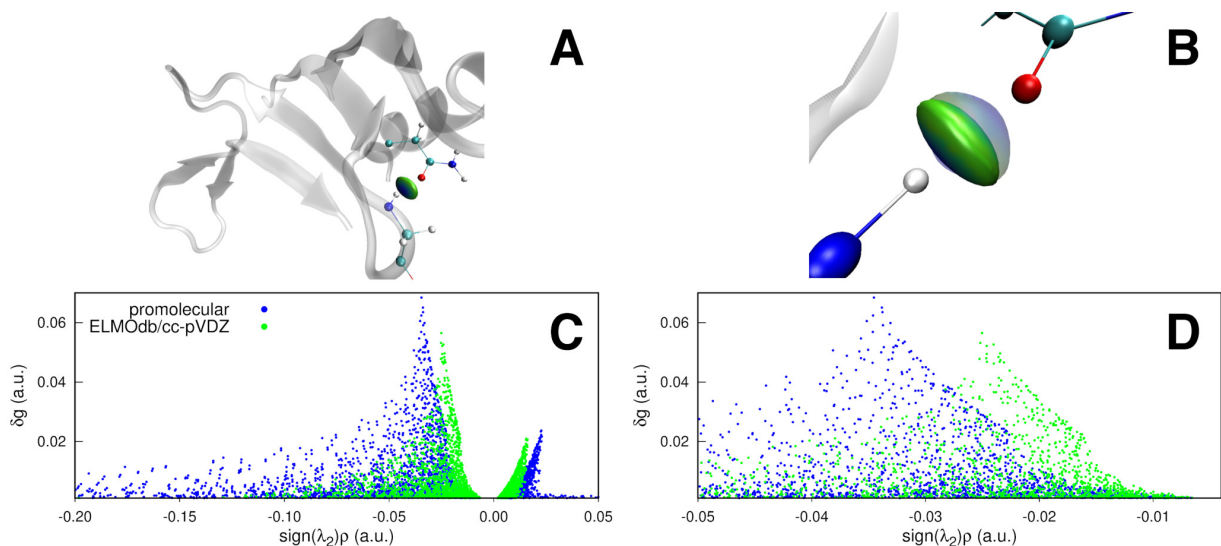


Figure 7: Local hydrogen-bond between residues Asn61 and Gly63 in Human Carbonic Anhydrase II (PDB ID: 3KS3): (A)  $\delta g = 0.01$  a.u. isosurface obtained with the ELMOdb/cc-pVDZ electron density and colored according to the BGR scheme over the range  $-0.05$  a.u.  $< \text{sign}(\lambda_2)\rho < 0.05$  a.u.; (B) ELMOdb versus promolecular (translucent)  $\delta g = 0.01$  a.u. isosurfaces; (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).

At a second stage, we considered the  $n - \pi^*$  interaction involving two fragments belonging to the backbone and the sidechain of the residue Asn61.  $n - \pi^*$  interactions play a crucial role in the stabilization of local backbone-sidechain contacts<sup>61,62</sup> and are often associated with hydrogen bonds. In our case, the examined  $n - \pi^*$  interaction is to be considered coupled with the local hydrogen-bond discussed in the previous paragraph. The results of our IGM analyses for this particular non-covalent interaction are collected in Figure 8 and the trends are completely analogous to those obtained in the other situations. In fact, both the IGM-promolecular and IGM-ELMO analyses successfully detect the  $n - \pi^*$  interaction (see Figures 8A and 8B), with the former giving  $\delta g$ -isosurfaces more diffuse and here slightly differently oriented in the real-space. This finds correspondence in the corresponding 2D signature plots, where we can see that the peaks for the IGM analysis based on the ELMOdb/cc-pVDZ electron density are always lower, narrower and occurring at lower absolute values of the electron density (see Figures 8C and 8D). This consequently reflects on the quantitative level, with the IGM-ELMO  $\Delta g$ 's that are always lower (by more than a factor of two) than

the corresponding IGM-promolecular value (see again Table 3).

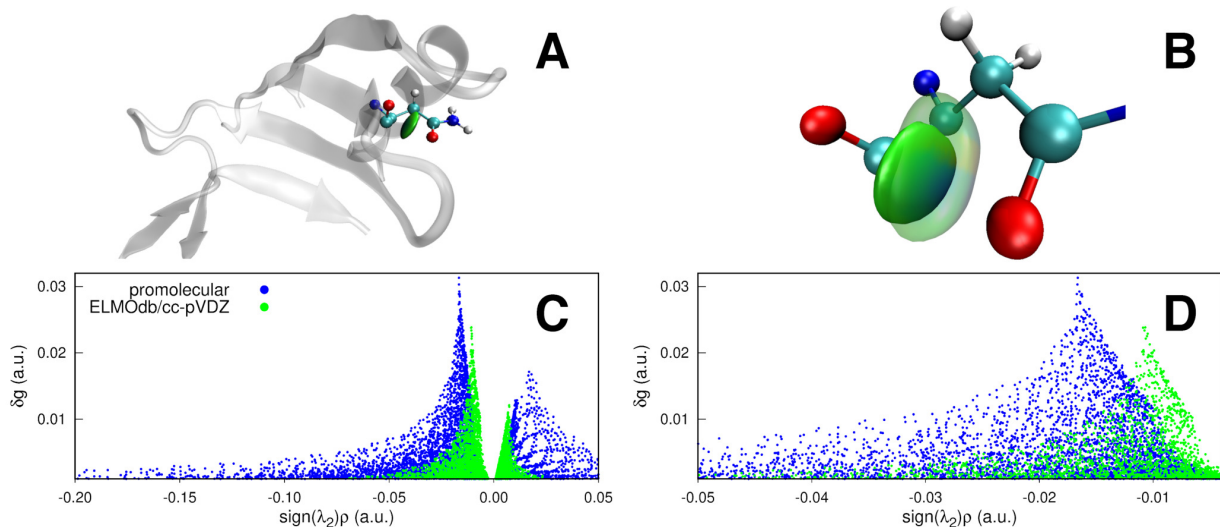


Figure 8:  $n - \pi^*$  interaction between the lone pair on the oxygen atom of the Asn61 sidechain and the  $\pi^*$  molecular orbital associated with the carbonyl group of the Asn61 backbone in Human Carbonic Anhydrase II (PDB ID: 3KS3): (A)  $\delta g = 0.005$  a.u. isosurface obtained with the ELMOdb/cc-pVDZ electron density and colored according to the BGR scheme over the range  $-0.05$  a.u.  $< \text{sign}(\lambda_2)\rho < 0.05$  a.u.; (B) ELMOdb versus promolecular (translucent)  $\delta g = 0.005$  a.u. isosurfaces; (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).

As a final application of the IGM-ELMO approach to interactions occurring in proteins, we considered the multiple hydrogen-bond between residues Arg58 and Glu69, which can be considered analogous to the interaction between charged residues analyzed in the preliminary validation test on the fifteen-residue synthetic peptide 1DEP. As usual, both the new IGM-ELMO method and the IGM-promolecular approach identify the interaction under exam, with the IGM-ELMO  $\delta g$ -isosurfaces that are always more localized (see Figures 9A and 9B). The more diffuse nature of the IGM-promolecular isosurfaces can be always rationalized through the analysis of the associated  $(\rho, \delta g)$  plots (see Figures 9C and 9D), where we notice that the peaks corresponding to the IGM-promolecular analysis are higher, wider and occurring at greater absolute values of the signed electron density compared to the IGM-ELMO case. This is in line with what we previously observed for polypeptide 1DEP and also

with the quantitative analysis based on the integration of the  $\delta g$  peaks, according to which the IGM-promolecular  $\Delta g$  value is always greater than the IGM-ELMO ones (see Table 3).

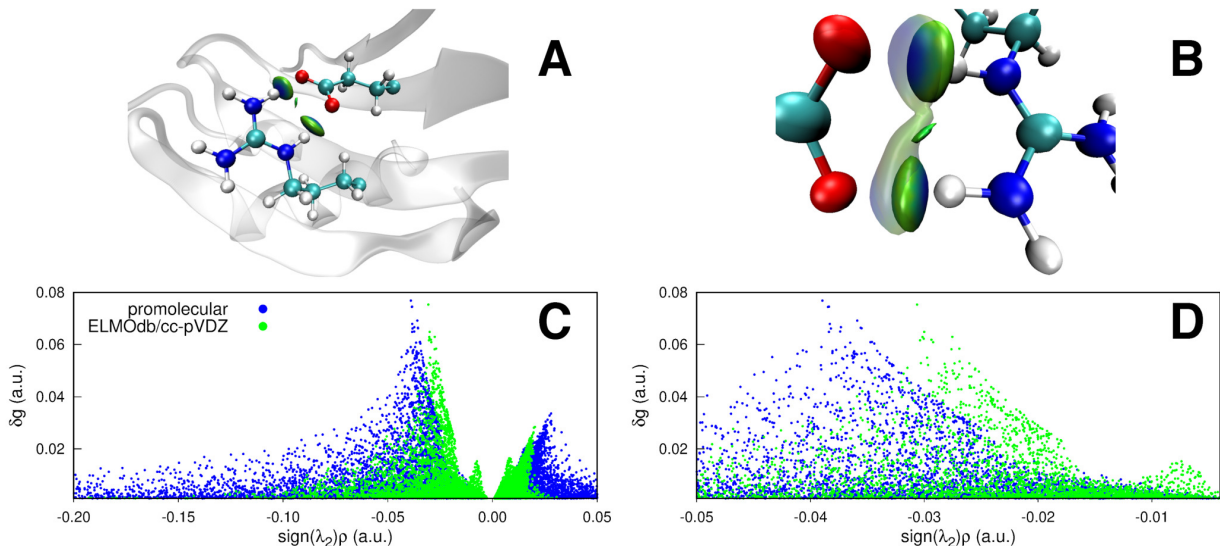


Figure 9: Multiple hydrogen-bonding interaction between residues Arg58 and Glu69 in Human Carbonic Anhydrase II (PDB ID: 3KS3): (A)  $\delta g = 0.014$  a.u. isosurface obtained with the ELMOdb/cc-pVDZ electron density and colored according to the BGR scheme over the range  $-0.05 \text{ a.u.} < \text{sign}(\lambda_2)\rho < 0.05 \text{ a.u.}$ ; (B) ELMOdb versus promolecular (translucent)  $\delta g = 0.014$  a.u. isosurfaces; (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).

## Conclusions

In this work, we have proposed the coupling of the Independent Gradient Model technique with the recently constructed libraries of extremely localized molecular orbitals. The final goal was to extend the applicability of IGM analyses based on quantum mechanically rigorous electron distributions to large macromolecules, for which only IGM computations relying on approximate promolecular densities were previously possible.

The validation tests carried out on small polypeptides and peptide dimers have consistently proved that the novel IGM-ELMO method provides results that are always very close to those obtained through corresponding IGM analyses that exploit fully quantum me-

chanical electron distributions (particularly DFT electron distributions). On the contrary, although revealing the region of interaction, the results provided by IGM computations based on promolecular densities are always markedly different, exaggerating the interaction signature. This has been observed both qualitatively and quantitatively. Furthermore, the first tests have also indicated the capability of the new IGM-ELMO technique in correctly assessing the relative strengths of non-covalent interactions when they occur in chemically similar systems. The better results obtained through the new strategy stem from the electron density relaxation that is intrinsically present in the ELMO building-blocks. 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 intra-molecular interactions in large systems and substantially improves the performance of the IGM approach.

As originally expected, the preliminary test calculations have also shown that the computational cost associated with the IGM-ELMO analyses does not significantly increase compared to the one of the calculations exploiting promolecular electron densities. This clearly indicated the possibility of applying the novel quantum mechanical version of the IGM method to large systems of biological interest. The novel IGM-ELMO technique has been thus exploited to investigate non-covalent interactions in macromolecules. The performed analyses gave results fully consistent with those previously obtained from the first validation tests on small peptides, with analogous trends observed in both qualitative and quantitative IGM descriptors for similar types of interactions.

Therefore, on the basis of the encouraging results obtained from our preliminary test calculations, we believe that the new IGM-ELMO method proposed in this paper could be successfully applied to unravel networks of non-covalent interactions in very large biosystems, such as, for example, in the case of complicated protein-protein contacts. Finally, since the new strategy has been shown i) to cheaply provide quantitative results very close to those

obtained through fully quantum mechanical IGM analyses and ii) to correctly assess and rank the relative strengths of non-covalent interactions in chemical systems, we also envisage the future use of the IGM-ELMO technique in connection with molecular docking computations or virtual high-throughput screenings in the context of rational drug-design investigations.

## Appendix: computational details

### Preparation of the target structures

For the first set of test calculations, we considered a high-resolution crystal structure of Leu-enkephalin<sup>56</sup> and the NMR molecular geometry of the fifteen-residue polypeptide corresponding to an intracellular sequence of the  $\beta$ -adrenergic receptor<sup>57</sup> (PDB code: 1DEP). In those cases, the positions of the hydrogen atoms were already reliably determined experimentally and, therefore, the available structures were directly used to perform all the calculations and IGM analyses that will be described below.

Concerning the second part of the validation tests, we initially extracted the geometries of the dimers from X-ray crystal structures deposited with the Cambridge Structural Database under the numbers CCDC-1884415 and CCDC-1884416 for the chlorinated and brominated peptides, respectively. The hydrogen atoms were removed from the original geometries and afterwards re-added by employing the software *MolProbability*,<sup>63,65</sup> which fixes the positions of the hydrogen atoms on the basis E-H (element-hydrogen) distances resulting from neutron diffraction measurements. The geometries of the de-halogenated dimers were simply obtained by removing the halogen atoms from the structures of the halogenated dimers and by adding the hydrogen atoms again exploiting *MolProbability*.

For the proteins considered in the last part of our work (human erythrocytic ubiquitin<sup>59</sup> and human carbonic anhydrase II<sup>60</sup>), the structures were extracted from the corresponding PDB files (1UBQ and 3KS3, respectively), which underwent the following preparation-steps: i) selection of only one possible conformation for the disordered regions; ii) assignment of

the correct protonation state to the different residues on the basis of the pH of crystallization/experiment; iii) addition of the missing hydrogen atoms through the *tleap* module of the AMBER Molecular Dynamics package.<sup>[66]</sup>

## IGM analyses

The program *IGMPlot*<sup>[29]</sup> was exploited to perform all the IGM analyses presented in this study. Within the IGM approach, probing and quantifying interactions between two sub-fragments FRAG1 and FRAG2 in a molecular system requires to supply the definition of these fragments in terms of atom indices and to specify the grid parameters (position, size, increments). A grid-step of 0.1 Å was taken for all the IGM analyses carried out in this study. For the sake of completeness, all the other parameters and fragment definitions used in this investigation have been reported in the Supporting Information.

## Quantum mechanical calculations

The ELMO electron density distributions used for all the performed IGM-ELMO analyses on peptides, peptide dimers and proteins considered in this study were obtained by transferring extremely localized molecular orbitals from the current version of the recently constructed ELMO-libraries<sup>[36]</sup> to the examined target structures.

For the sake of precision, for the polypeptides considered in the first validation tests (Leu-enkephalin and 1DEP), the de-halogenated peptides of the second set of test calculations, and the proteins taken into account in the third part of our computational assessment (human erythrocytic ubiquitin and human carbonic anhydrase II), we transferred ELMOs corresponding to standard amino acids and, therefore, already stored in the standard folders of the current ELMO libraries. For the halogenated peptides (DSGY(Br)EV and DSGY(Cl)EV), we also transferred tailor-made ELMOs corresponding to the fragments of the brominated and chlorinated tyrosine residues. These ELMOs were computed on proper model molecules (see Figure S11 in the Supporting Information) for all the five standard basis-sets currently

available for the ELMO databanks (6-31G, 6-311G, 6-31G(d,p), 6-311G(d,p) and cc-pVDZ).

The transfers of the ELMOs were performed exploiting the software *ELMOdb*,<sup>36</sup> which implements the strategy originally proposed by Philipp and Friesner for the rotation of strictly localized bond orbitals that are used to describe frontier regions in quantum mechanics / molecular mechanics techniques.<sup>67</sup> The wave function files (.wfx format) provided in output by the *ELMOdb* program were then passed to the *IGMPlot* software<sup>29</sup> for the subsequent IGM analyses. All the basis-sets currently available in the ELMO libraries were considered for the transfer of the ELMOs. The *ELMOdb* program and the associated ELMO databanks are available upon motivated requests to A.G. ([Alessandro.Genoni@univ-lorraine.fr](mailto:Alessandro.Genoni@univ-lorraine.fr)).

Finally, to obtain the electron densities necessary for the DFT-based IGM analyses of the preliminary validation tests, we carried out B3LYP calculations on the examined structures with the 6-31G, 6-311G, 6-31G(d,p), 6-311G(d,p) and cc-pVDZ basis-sets. For these computations, the quantum chemistry package *Gaussian16* was used.<sup>68</sup>

## Acknowledgement

A.G. and E.K.W gratefully acknowledge the French Research Agency (ANR) for funding through the Young Researcher Project QuMacroRef (Grant No. ANR-17-CE29-0005-01). E.H. and J.-C.B. thank the MaSCA (Maison de la Simulation de Champagne-Ardenne, France) for various supports including computing facilities (<http://romeo.univ-reims.fr>) and the CRIANN computational center (<http://www.criann.fr>) for additional support.

## Supporting Information Available

Details about the IGM approach and about theory, transfer and libraries of extremely localized molecular orbitals (with Figure S1 reporting some examples of plotted ELMOs and Figure S2 showing the schematic representation of reference frames and atomic triads required for the ELMOs rotation). Figures S3-S5 reporting the individual ( $\rho$ ,  $\delta g$ ) fingerprint-plots

obtained at different levels of theory for the analyzed hydrogen-bond and T-shaped  $\pi - \pi$  interactions in Leu-enkephalin and for the multiple hydrogen-bond between charged residues in the 1DEP polypeptide. Figure S6 depicting the  $\delta g$  isosurfaces associated with the  $\pi - \pi$  interactions in the dimers of the chlorinated and de-chlorinated polypeptides. Figures S7 and S8 reporting the individual  $(\rho, \delta g)$  fingerprint-plots obtained at different levels of theory for the  $\pi - \pi$  interactions in the dimers of the halogenated and de-halogenated polypeptides. Figures S9 and S10 showing the basis-set dependence in the  $(\rho, \delta g)$  fingerprint-plots associated with the  $\pi - \pi$  interactions in the dimers of the halogenated/de-halogenated polypeptides. Figure S11 depicting the model molecules used for the computations of the tailor-made ELMOs describing the fragments belonging to the aromatic rings of the brominated and chlorinated tyrosine residues. Table S1 reporting the computational cost (i.e., the CPU times) associated with each step of the IGM calculations carried out on the Leu-enkephalin and 1DEP polypeptides. Table S2 providing the interaction energies computed at B3LYP-D3(BJ) level for the  $\pi - \pi$  interactions between the side-chains of the tyrosine residues in the dimers of the halogenated and de-halogenated polypeptides.

## References

- (1) Panigrahi, S. K.; Desiraju, G. R. Strong and Weak Hydrogen Bonds in the Protein-Ligand Interface. *Proteins* **2007**, *67*, 128–141.
- (2) Kollman, P. Non-Covalent Forces of Importance in Biochemistry. *New Compr. Biochem.* **1984**, *6*, 55–71.
- (3) Leckband, D.; Israelachvili, J. Intermolecular Forces in Biology. *Q. Rev. Biophys.* **2001**, *34*, 105–267.
- (4) Keskin, O.; Tuncbag, N.; GURSOY, A. Predicting Protein-Protein Interactions from the Molecular to the Proteome Level. *Chem. Rev.* **2016**, *116*, 4884–4909.
- (5) Muegge, I.; Martin, Y. C. A General and Fast Scoring Function for Protein-Ligand Interactions: A Simplified Potential Approach. *J. Med. Chem.* **1999**, *42*, 791–804.
- (6) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461.
- (7) Ponder, J. W.; Case, D. A. Force Fields for Protein Simulations. *Adv. Protein Chem.* **2003**, *66*, 27–85.
- (8) Jeziorski, B.; Moszynski, R.; Szalewicz, K. Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van der Waals Complexes. *Chem. Rev.* **1994**, *94*, 1887–1930.
- (9) Szalewicz, K. Symmetry-Adapted Perturbation Theory of Intermolecular Forces. *WIREs Comput. Mol. Sci.* **2012**, *2*, 254–272.
- (10) Phipps, M. J. S.; Fox, T.; Tautermann, C. S.; Skylaris, C.-K. Energy Decomposition Analysis Approaches and Their Evaluation on Prototypical Protein–Drug Interaction Patterns. *Chem. Soc. Rev.* **2015**, *44*, 3177–3211.

- (11) Zhao, L.; von Hopffgarten, M.; Andrada, D. M.; Frenking, G. Energy Decomposition Analysis. *WIREs Comput. Mol. Sci.* **2018**, *8*, e1345.
- (12) Blanco, M. A.; Martín Pendás, A.; Francisco, E. Interacting Quantum Atoms: A Correlated Energy Decomposition Scheme Based on the Quantum Theory of Atoms in Molecules. *J. Chem. Theory Comput.* **2005**, *1*, 1096–1109.
- (13) Maxwell, P.; Pendás, M.; Popelier, P. L. A. Extension of the Interacting Quantum Atoms (IQA) Approach to B3LYP Level Density Functional Theory (DFT). *Phys. Chem. Chem. Phys.* **2016**, *18*, 20986–21000.
- (14) Wilson, A. L.; Popelier, P. L. A. Exponential Relationships Capturing Atomistic Short-Range Repulsion from the Interacting Quantum Atoms (IQA) Method. *J. Phys. Chem. A* **2016**, *120*, 9647–9659.
- (15) Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; Oxford University Press: Oxford, U.K., 1990.
- (16) de Silva, P.; Corminboeuf, C. Simultaneous Visualization of Covalent and Noncovalent Interactions Using Regions of Density Overlap. *J. Chem. Theory Comput.* **2014**, *10*, 3745–3756.
- (17) Meyer, B.; Barthel, S.; Mace, A.; Vannay, L.; Guillot, B.; Smit, B.; Corminboeuf, C. DORI Reveals the Influence of Noncovalent Interactions on Covalent Bonding Patterns in Molecular Crystals Under Pressure. *J. Chem. Phys. Lett.* **2019**, *10*, 1482–1488.
- (18) Johnson, E. R.; Keinan, S.; Mori-Sanchez, P.; Contreras-García, J.; Cohen, A. J.; Yang, W. Revealing Noncovalent Interactions. *J. Am. Chem. Soc.* **2010**, *132*, 6498–6506.
- (19) Contreras-García, J.; Johnson, E. R.; Keinan, S.; Chaudret, R.; Piquemal, J.-P.; Be-

- ratan, D. N.; Yang, W. NCIPLOT: a Program for Plotting Noncovalent Interaction Regions. *J. Chem. Theory Comput.* **2011**, *7*, 625–632.
- (20) Peccati, F. NCIPLOT4 Guide for Biomolecules: An Analysis Tool for Noncovalent Interactions. *J. Chem. Inf. Mod.* **2020**, *60*, 6–10.
- (21) Boto, R. A.; Peccati, F.; Laplaza, R.; Quan, C.; Carbone, A.; Piquemal, J.-P.; Mayday, Y.; Contreras-García, J. NCIPLOT4: Fast, Robust, and Quantitative Analysis of Noncovalent Interactions. *J. Chem. Theory Comput.* **2020**, *16*, 4150–4158.
- (22) Saleh, G.; Gatti, C.; Lo Presti, L.; Contreras-García, J. Revealing Non-covalent Interactions in Molecular Crystals through Their Experimental Electron Densities. *Chem. Eur. J.* **2012**, *18*, 15523–15536.
- (23) Saleh, G.; Lo Presti, L.; Gatti, C.; Ceresoli, D. *NCImilano*: an Electron-Density-Based Code for the Study of Noncovalent Interactions. *J. Appl. Cryst.* **2013**, *46*, 1513–1517.
- (24) Lefebvre, C.; Rubez, G.; Khartabil, H.; Boisson, J.-C.; Contreras-García, J.; Hénon, E. Accurately Extracting the Signature of Intermolecular Interactions Present in the NCI Plot of the Reduced Density Gradient versus Electron Density. *Phys. Chem. Chem. Phys.* **2017**, *19*, 17928–17936.
- (25) Lefebvre, C.; Khartabil, H.; Boisson, J.-C.; Contreras-García, J.; Piquemal, J.-P.; Hénon, E. The Independent Gradient Model: A New Approach for Probing Strong and Weak Interactions in Molecules from Wave Function Calculations. *ChemPhysChem* **2018**, *19*, 724–735.
- (26) Ponce-Vargas, M.; Lefebvre, C.; Boisson, J.-C.; Hénon, E. Atomic Decomposition Scheme of Noncovalent Interactions Applied to Host–Guest Assemblies. *J. Chem. Inf. Mod.* **2020**, *60*, 268–278.

- (27) Klein, J.; Khartabil, H.; Boisson, J.-C.; Contreras-García, J.; Piquemal, J.-P.; Hénon, E. New Way for Probing Bond Strength. *J. Phys. Chem. A* **2020**, *124*, 1850–1860.
- (28) Bader, R. F. W. A Bond Path: A Universal Indicator of Bonded Interactions. *J. Phys. Chem. A* **1998**, *102*, 7314–7323.
- (29) Klein, J.; Pluot, E.; Rubez, G.; Boisson, J. C.; Hénon, E. IGMPlot Revision 2.6.7. 2020.
- (30) Grisafi, A.; Fabrizio, A.; Meyer, B.; Wilkins, D. M.; Corminboeuf, C.; Ceriotti, M. Transferable Machine-Learning Model of the Electron Density. *ACS Cent. Sci.* **2019**, *5*, 57–64.
- (31) Fabrizio, A.; Grisafi, A.; Meyer, B.; Ceriotti, M.; Corminboeuf, C. Electron Density Learning of Non-Covalent Systems. *Chem. Sci.* **2019**, *10*, 9424–9432.
- (32) Cuevas-Zuviría, B.; Pacios, L. F. Analytical Model of Electron Density and Its Machine Learning Inference. *J. Chem. Inf. Mod.* **2020**, *60*, 3831–3842.
- (33) Kamal, D.; Chandrasekaran, A.; Batra, R.; Ramprasad, R. A Charge Density Prediction Model for Hydrocarbons Using Deep Neural Networks. *Mach. Learn.: Sci. Technol.* **2020**, *1*, 025003.
- (34) Meyer, B.; Guillot, B.; Ruiz-Lopez, M. F.; Genoni, A. Libraries of Extremely Localized Molecular Orbitals. 1. Model Molecules Approximation and Molecular Orbitals Transferability. *J. Chem. Theory Comput.* **2016**, *12*, 1052–1067.
- (35) Meyer, B.; Guillot, B.; Ruiz-Lopez, M. F.; Jelsch, C.; Genoni, A. Libraries of Extremely Localized Molecular Orbitals. 2. Comparison with the Pseudoatoms Transferability. *J. Chem. Theory Comput.* **2016**, *12*, 1068–1081.
- (36) Meyer, B.; Genoni, A. Libraries of Extremely Localized Molecular Orbitals. 3. Construction and Preliminary Assessment of the New Databanks. *J. Phys. Chem. A* **2018**, *122*, 8965–8981.

- (37) Stoll, H.; Wagenblast, G.; Preuß, H. On the Use of Local Basis Sets for Localized Molecular Orbitals. *Theor. Chim. Acta* **1980**, *57*, 169–178.
- (38) Fornili, A.; Sironi, M.; Raimondi, M. Determination of Extremely Localized Molecular Orbitals and their Application to Quantum Mechanics/Molecular Mechanics Methods and to the Study of Intramolecular Hydrogen Bonding. *J. Mol. Struct.: THEOCHEM* **2003**, *632*, 157–172.
- (39) Sironi, M.; Genoni, A.; Civera, M.; Pieraccini, S.; Ghitti, M. Extremely Localized Molecular Orbitals: Theory and Applications. *Theor. Chem. Acc.* **2007**, *117*, 685–698.
- (40) Genoni, A.; Sironi, M. A Novel Approach to Relax Extremely Localized Molecular Orbitals: the Extremely Localized Molecular Orbital-Valence Bond Method. *Theor. Chem. Acc.* **2004**, *112*, 254–262.
- (41) Genoni, A.; Fornili, A.; Sironi, M. Optimal Virtual Orbitals to Relax Wave Functions Built Up with Transferred Extremely Localized Molecular Orbitals. *J. Comput. Chem.* **2005**, *26*, 827–835.
- (42) Genoni, A.; Ghitti, M.; Pieraccini, S.; Sironi, M. A Novel Extremely Localized Molecular Orbitals Based Technique for the One-Electron Density Matrix Computation. *Chem. Phys. Lett.* **2005**, *415*, 256–260.
- (43) Genoni, A.; Merz Jr., K. M.; Sironi, M. A Hylleraas Functional Based Perturbative Technique to Relax the Extremely Localized Molecular Orbital Wavefunction. *J. Chem. Phys.* **2008**, *129*, 054101.
- (44) Sironi, M.; Ghitti, M.; Genoni, A.; Saladino, G.; Pieraccini, S. DENPOL: a New Program to Determine Electron Densities of Polypeptides Using Extremely Localized Molecular Orbitals. *J. Mol. Struct.: THEOCHEM* **2009**, *898*, 8–16.

- (45) Malaspina, L. A.; Wieduwilt, E. K.; Bergmann, J.; Kleemiss, F.; Meyer, B.; Ruiz-Lopez, M. F.; Pal, R.; Hupf, E.; Beckmann, J.; Piltz, R. O.; Edwards, A. J.; Grabowsky, S.; Genoni, A. Fast and Accurate Quantum Crystallography: From Small to Large, from Light to Heavy. *J. Phys. Chem. Lett.* **2019**, *10*, 6973–6982.
- (46) Grabowsky, S.; Genoni, A.; Bürgi, H.-B. Quantum Crystallography. *Chem. Sci.* **2017**, *8*, 4159–4176.
- (47) Genoni, A.; Bučinský, L.; Claiser, N.; Contreras-García, J.; Dittrich, B.; Dominiak, P. M.; Espinosa, E.; Gatti, C.; Giannozzi, P.; Gillet, J.-M.; Jayatilaka, D.; Macchi, P.; Madsen, A. Ø.; Massa, L.; Matta, C. F.; Merz Jr., K. M.; Nakashima, P. N. H.; Ott, H.; Ryde, U.; Schwarz, K.; Sierka, M.; Grabowsky, S. Quantum Crystallography: Current Developments and Future Perspectives. *Chem. - Eur. J.* **2018**, *24*, 10881–10905.
- (48) Massa, L.; Matta, C. F. Quantum crystallography: A perspective. *J. Comput. Chem.* **2018**, *39*, 1021–1028.
- (49) Tsirelson, V. Early Days of Quantum Crystallography: A Personal Account. *J. Comput. Chem.* **2018**, *39*, 1029–1037.
- (50) Genoni, A.; Macchi, P. Quantum Crystallography in the Last Decade: Developments and Outlooks. *Crystals* **2020**, *10*, 473.
- (51) Grabowsky, S.; Genoni, A.; Thomas, S. P.; Jayatilaka, D. The Advent of Quantum Crystallography: Form and Structure Factors from Quantum Mechanics for Advanced Structure Refinement and Wavefunction Fitting. In *Structure and Bonding. 21st Century Challenges in Chemical Crystallography 2 - Structural Correlations and Data Interpretation*; Mingos, D. M. P., Rathby, P., Eds.; Springer: Berlin, Heidelberg; DOI:10.1007/430\_2020\_62.

- (52) Macetti, G.; Genoni, A. Quantum Mechanics/Extremely Localized Molecular Orbital Method: A Fully Quantum Mechanical Embedding Approach for Macromolecules. *J. Phys. Chem. A* **2019**, *123*, 9420–9428.
- (53) Macetti, G.; Wieduwilt, E. K.; Assfeld, X.; Genoni, A. Localized Molecular Orbital-Based Embedding Scheme for Correlated Methods. *J. Chem. Theory Comput.* **2020**, *16*, 3578–3596.
- (54) Macetti, G.; Genoni, A. Quantum Mechanics/Extremely Localized Molecular Orbital Embedding Strategy for Excited States: Coupling to Time-Dependent Density Functional Theory and Equation-of-Motion Coupled Cluster. *J. Chem. Theory Comput.* **2020**, *16*, 7490–7506.
- (55) Arias-Olivares, D.; Wieduwilt, E. K.; Contreras-García, J.; Genoni, A. NCI-ELMO: A New Method To Quickly and Accurately Detect Noncovalent Interactions in Biosystems. *J. Chem. Theory Comput.* **2019**, *15*, 6456–6470.
- (56) Pichon-Pesme, V.; Lecomte, C.; Wiest, R.; Bernard, M. Modeling Fragments for the Ab Initio Determination of Electron Density in Polypeptides. An Experimental and Theoretical Approach to the Electron Distribution in Leu-Enkephalin Trihydrate. *J. Am. Chem. Soc.* **1992**, *114*, 2713–2715.
- (57) Münch, G.; Dees, C.; Hekman, M.; Palm, D. Multisite Contacts Involved in Coupling of the  $\alpha$ -Adrenergic Receptor with the Stimulatory Guanine-Nucleotide-Binding Regulatory Protein. *Eur. J. Biochem.* **1991**, *198*, 357–364.
- (58) Maiolo, D.; Pizzi, A.; Gori, A.; Gazzera, L.; Demitri, N.; Genoni, A.; Baggi, F.; Moda, F.; Terraneo, G.; Baldelli Bombelli, F.; Metrangolo, P.; Resnati, G. Halogenation of the N-Terminus Tyrosine 10 Promotes Supramolecular Stabilization of the Amyloid- $\beta$  Sequence 7–12. *ChemistryOpen* **2020**, *9*, 253–260.

- (59) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. Structure of Ubiquitin Refined at 1.8 Å Resolution. *J. Mol. Biol.* **1987**, *194*, 531–544.
- (60) Avvaru, B. S.; Kim, C. U.; Sippel, K. H.; Gruner, S. M.; Agbandje-McKenna, M.; Silverman, D. N.; McKenna, R. A Short, Strong Hydrogen Bond in the Active Site of Human Carbonic Anhydrase II. *Biochemistry* **2009**, *49*, 249–251.
- (61) Hinderaker, M. P.; Raines, R. T. An Electronic Effect on Protein Structure. *Protein Sci.* **2003**, *12*, 1188–1194.
- (62) Khatri, B.; Majumder, P.; Nagesh, J.; Penmatsa, A.; Chatterjee, J. Increasing Protein Stability by Engineering the  $n \rightarrow \pi^*$  Interaction at the  $\beta$ -Turn. *Chem. Sci.* **2020**, *11*, 9480–9487.
- (63) Williams, C. J.; Headd, J. J.; Moriarty, N. W.; Prisant, M. G.; Videau, L. L.; Deis, L. N.; Verma, V.; Keedy, D. A.; Hintze, B. J.; Chen, V. B.; Jain, S.; Lewis, S. M.; Arendall III, W. B.; Snoeyink, J.; Adams, P. D.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. MolProbity: More and Better Reference Data for Improved All-Atom Structure Validation. *Protein Sci.* **2018**, *27*, 293–315.
- (64) Chen, V. B.; Arendall, W. B., III; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. *MolProbity*: All-Atom Structure Validation for Macromolecular Crystallography. *Acta Cryst., Sect. D: Biol. Crystallogr.* **2010**, *66*, 12–21.
- (65) Davis, I. W.; Leaver-Fay, A.; Chen, V. B.; Block, J. N.; Kapral, G. J.; Wang, X.; Murray, L. W.; Arendall, I., W. Bryan; Snoeyink, J.; Richardson, J. S.; Richardson, D. C. MolProbity: All-Atom Contacts and Structure Validation for Proteins and Nucleic Acids. *Nucleic Acids Res.* **2007**, *35*, W375–W383.
- (66) Case, D.; Ben-Shalom, I.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E.; Cruzeiro, V.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.;

Greene, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. Amber 2018. University of California, San Francisco.

- (67) Philipp, D. M.; Friesner, R. A. Mixed Ab Initio QM/MM Modeling Using Frozen Orbitals and Tests with Alanine Dipeptide and Tetrapeptide. *J. Comput. Chem.* **1999**, *20*, 1468–1494.
- (68) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Menucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian16 Revision C.01. 2016; Gaussian Inc. Wallingford CT.