Extracting Quantitative Information at Quantum Mechanical Level from Noncovalent Interaction Index Analyses

Erna Wieduwilt, Roberto Boto, Giovanni Macetti, Rubén Laplaza, Julia Contreras-García, Alessandro Genoni

Extracting Quantitative Information of Quantum Mechanical Level from Noncovalent Interaction Index Analyses

Erna K. Wieduwilt[(1,3)*], Roberto A. Boto[(2,4)], Giovanni Macetti[(1,5)], Rubén Laplaza[(2,6)], Julia Contreras-García[(2)], Alessandro Genoni[(1)*]

(1) Université de Lorraine & CNRS, Laboratoire de Physique et Chimie Théoriques (LPCT), UMR CNRS 7019, 1 Boulevard Arago, F-57078 Metz, France.

(2) Sorbonne Université & CNRS, Laboratoire de Chimie Théorique (LCT), UMR CNRS 7616, 4 Place Jussieu, F-750005 Paris, France.

Present Addresses:

(3) Department of Physics, Chemistry and Pharmacy, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark.

(4) Donostia International Physics Center (DIPC), Paseo Manuel de Lardizabal 4, 20018, Donostia-San Sebastián, Spain.

(5) Dipartimento di Chimica, Università degli Studi di Milano, via Golgi 19, I-20133 Milano, Italy.

(6) Laboratory for Computational Molecular Design (LCMD), Institute of Chemical Sciences and Engineering (ISIC), École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland.

---

* Correspondence to:

- Alessandro Genoni, Université de Lorraine & CNRS, Laboratoire de Physique et Chimie Théoriques (LPCT), UMR CNRS 7019, 1 Boulevard Arago, F-57078 Metz, France. E-mail: alessandro.genoni@univ-lorraine.fr; Phone: +33 (0)3 72 74 91 70; Fax: +33 (0)3 72 74 91 87.

- Erna K. Wieduwilt, Department of Physics, Chemistry and Pharmacy, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark. E-mail: erna@sdu.dk.

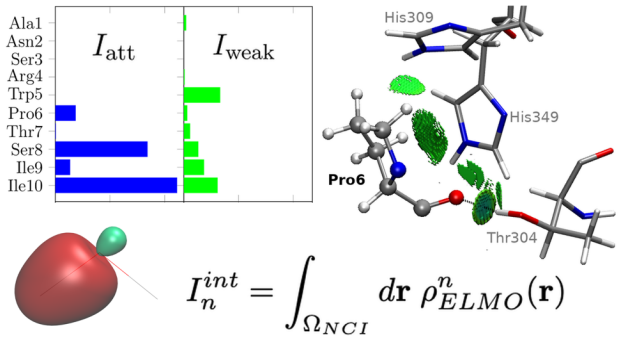
ABSTRACT

The noncovalent interaction (NCI) index is nowadays a well-known strategy to detect noncovalent interactions in molecular systems. Even though it initially provided only qualitative descriptions, the technique has been recently extended to also extract quantitative information. To accomplish this task, integrals of powers of the electron distribution were considered, with the requirement that the overall electron density can be clearly decomposed as sum of distinct fragment contributions to enable the definition of the (noncovalent) integration region. So far, this was done by only exploiting approximate promolecular electron densities, which are given by the sum of spherically averaged atomic electron distributions and, therefore, represent too crude approximations. Therefore, to obtain more quantum mechanically rigorous results from NCI-index analyses, in this work we propose to use electron densities obtained through the transfer of extremely localized molecular orbitals (ELMOs) or through the recently developed QM/ELMO (quantum mechanics / extremely localized molecular orbital) embedding technique. Although still approximate, the electron distributions resulting from the above-mentioned methods are fully quantum mechanical and, above all, are again partitionable into subunit contributions, which makes them completely suitable for the NCI integral approach. Therefore, we benchmarked the integrals resulting from NCI-index analyses (both those based on the promolecular densities and those based on ELMO electron distributions) against interaction energies computed at high quantum chemical level (in particular, Coupled Cluster level). The performed test calculations have indicated that the NCI integrals based on ELMO electron densities outperform the promolecular ones. Furthermore, it was observed that the novel quantitative NCI-(QM/)ELMO approach can be also profitably exploited both to characterize and evaluate the strength of specific interactions between ligand subunits and protein

residues in protein-ligand complexes, and to follow the evolution of noncovalent interactions along trajectories of molecular dynamics simulations. Although further methodological improvements are still possible, the new quantitative ELMO-based technique could be already exploited in situations in which fast and reliable assessments of noncovalent interactions are crucial, such as in computational high-throughput screenings for drug discovery.

KEYWORDS: noncovalent interactions, NCI-index, electron density, extremely localized molecular orbitals, QM/ELMO embedding method.

TOC GRAPHICS



1. INTRODUCTION

Many processes and phenomena in chemistry, biology and materials science are governed by subtle balances of many noncovalent interactions (NCIs). Well known examples are the structure of biological macromolecules, protein-ligand interactions, the packing of molecular crystals and the self-assembly of nanomaterials.[1-6]

Considering this aspect, over the years theoretical chemists have constantly proposed computational methods to reliably identify and quantify NCIs in systems of chemical and biochemical interest. The simplest approaches are those based on pairwise atomic distances, which are conveniently exploited in knowledge-based potentials or empirical scoring functions of docking algorithms.[7,8] Other computationally cheap strategies are the ones that use semi-empirically fitted force fields relying on the Lennard-Jones potential and the Coulomb law to describe van der Waals and electrostatic interactions, respectively.[9] At the other end of the spectrum, we have more expensive methods that are fully based on quantum mechanics. Examples of this kind are the sophisticated and powerful symmetry adapted perturbation theory (SAPT),[10,11] and the more intuitive supermolecular approach coupled with the counterpoise correction[12] for the elimination of the basis set superposition error (BSSE) or combined with energy decomposition schemes[13,14] to quantify the different contributions to the global interaction energy. Another important class of computational techniques that are exploited to study noncovalent interactions are the so-called real-space approaches. They make use of physically sound scalar or vector fields and, among them, we can certainly mention the popular Quantum Theory of Atoms in Molecules[15] (QTAIM) and the related Interacting Quantum Atom (IQA) technique. In particular, in addition to partitioning the total electron distribution into atomic components, the latter also gives the total energy of the investigated system as sum of intra- and inter-atomic contributions and was

successfully applied to unravel the nature of different types of noncovalent interactions, such as anion-π, hydrogen and halogen bonding.[16-18] Remaining in the context of real-space methods, other notable examples are the strategies that explicitly use the (reduced) gradient of the electron density for the detection and characterization of NCIs. In this regard, prominent examples are the NCI-index technique,[19-24] the density overall region indicator[25,26] (DORI) and the more recent independent gradient model[27-30] (IGM).

Of the last three methods, the first one has increasingly become a widely used strategy to detect and analyze inter- and intra-molecular noncovalent interactions in molecular systems. As we will briefly see in the Theory section, the NCI-index approach basically consists in looking for regions of real-space characterized by low values of the electron density ρ and, at the same time, by low values of the reduced density gradient s. These regions have been shown to correspond to intra- or inter-molecular noncovalent interactions, whose nature (i.e., strong interactions, weak contacts, and steric clash) is also defined according to the sign of the second eigenvalue λ_2 of the electron density Hessian. Nevertheless, despite its undeniable success in many and different applications, the original version of the NCI-index strategy suffered from two intrinsic limitations: i) the impossibility to have quantum mechanics-based NCI analyses for macromolecules; ii) the lack of quantitative results.

The former problem was related to the large computational cost associated with the quantum chemical calculations that are necessary to obtain macromolecular electron densities of quantum mechanical (QM) level. In fact, when applied to large systems, the NCI-index analyses used to rely only on the rough promolecular approximation (also known as independent atom model, IAM), which consists in constructing the global electron distribution as sum of tabulated spherically averaged atomic electron

densities. To overcome this limitation, the NCI-index approach has been recently interfaced with libraries of extremely localized molecular orbitals (ELMOs).[31-33] In fact, as we will see more in detail in Subsection 2.2, ELMOs are molecular orbitals strictly localized on small molecular subunits[34-36] (such as, atoms, bonds and functional groups) and, thanks to their absolute localization, they can be easily and reliably transferred from one molecule to another.[31,32,37-41] By taking advantage of this property, databanks of extremely localized molecular orbitals covering all the fragments of the twenty natural amino acids have been assembled. These libraries allow the almost instantaneous reconstruction of approximate (but quantum mechanically rigorous) wave functions, electron densities and electrostatic potentials of large polypeptides and proteins.[33] The coupling of the NCI-index strategy with the ELMO databases led to the development of the new NCI-ELMO approach,[42] through which it was possible to obtain improved qualitative descriptions of noncovalent interaction networks in systems of biological interest compared to corresponding descriptions resulting from original and standard promolecular-NCI analyses. Along the same line, the ELMO libraries have also been recently interfaced with the Independent Gradient Model, for which the ELMO-based analyses were shown to be better than the promolecular ones even from the quantitative point of view.[43]

As mentioned above, the second important limitation of the standard NCI-index descriptions used to be the complete absence of quantitative information. In response to this shortcoming, some of us have recently proposed an improved version of the NCI-index technique,[24] where integrals of powers of the electron density over regions of non-covalent interactions are exploited to extract quantitative values that can be directly associated with interaction energies (see Subsection 2.1 for more details). Nevertheless, the drawback of this approach is the unavoidable requirement of expressing the global

electron density of the investigated system as sum of two (or more) contributions corresponding to the electron densities of the interacting fragments. For example, in case of a dimer A:B, the method works only if it is possible to clearly distinguish the electron density of monomer A from the one of monomer B. Therefore, so far, the new integral-based NCI approach has been used only in conjunction with the atom-based promolecular approximation, which indeed allows the expression of each fragment contribution as sum of tabulated spherical electron densities centered on the atoms that belong to the fragment under exam.

Nevertheless, as also seen in the validation studies of the NCI-ELMO[42] and IGM-ELMO[43] techniques, the independent atom model is only a qualitatively acceptable approximation but does not provide results of QM level. To improve the quantitative description of the NCI strategy, the only possibility is to resort to localized quantum mechanical methods that still enable a clear partitioning of the global electron density into distinct fragment contributions. With this purpose in mind, in this work we propose the coupling of the NCI integral approach i) with the transfer of the above-mentioned extremely localized molecular orbitals[33] and ii) with the QM/ELMO (quantum mechanics / extremely localized molecular orbital) embedding method.[44-50] In the former case, the overall electron distribution can always be seen as given by the sum of terms (i.e., the square moduli of the molecular orbitals) that can be directly associated with well-defined subunits. In the latter situation, we will consider a recently developed embedding technique in which a region of the system under examination is treated at fully quantum mechanical level, while the rest is described by means of frozen extremely localized molecular orbitals previously transferred from the available ELMO databanks or from tailor-made model molecules. Also in this case, the electron density of the QM region can be easily distinguished from the one of the ELMO subunit. As

we will see, both proposed extensions of the integral-based NCI strategy will be particularly useful in the investigation of macromolecular systems of biological interest.

The article is organized as follows. In the next section we will briefly review the theoretical foundations of the computational strategies used and combined in this work (i.e., the NCI-index method, ELMOs and ELMO libraries, QM/ELMO technique). Afterwards, in Section 3, we will show and discuss the results of all the test calculations that we performed to assess the capabilities of the novel approach. Finally, conclusions and future perspectives will be presented in the last part of the paper (Section 4).

2. THEORY

In this section, we will focus on the essential theoretical aspects of the computational strategies considered in this investigation. In Subsection 2.1, we will present the main features of the NCI-index technique, with also a particular attention to its more recent integral-based version that allows the extraction of quantitative information from the analyses. In Subsection 2.2, we will discuss the concept of extremely localized molecular orbitals and of ELMOs transferability, we will briefly describe the recently constructed ELMO libraries and, finally, we will also introduce the philosophy of the QM/ELMO multiscale embedding method.

2.1 The NCI-index approach. As briefly anticipated in the Introduction, the NCI-index is a computational tool that enables the identification of intra- and inter-molecular noncovalent interactions in the real-space.[19,20] To accomplish this task, this strategy exploits the reduced density gradient (RDG) $s(\boldsymbol{r})$ given by the following expression:

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

RDG is a well-known density functional theory (DFT) quantity that takes into account the inhomogeneity of the electron distribution and, for this reason, allows the detection of regions corresponding to electron density variations due to interatomic interactions. In fact, when $s(\boldsymbol{r})$ is plotted against the electron density, some characteristic troughs appear (see Figure 1A). They correspond to noncovalent interactions and can be characterized by multiplying the electron density values by the corresponding sign of the second eigenvalue of the electron density Hessian (λ_2): troughs occurring at negative values of the signed electron density indicate the presence of strongly attractive interactions (e.g., strong/conventional hydrogen bonds, such as O-H···O or N-H···O contacts), troughs at slightly negative/positive values generally correspond to weaker interactions (e.g., weak/non-conventional hydrogen bonds, such as C-H···O contacts, or van der Waals interactions), while troughs at positive values reveal steric clashes. A color-code is usually applied to distinguish the different cases, where the blue, green and red colors indicate strong interactions, weak contacts, and steric clashes, respectively (see again Figure 1B). Moreover, by plotting isosurfaces of the reduced density gradient and coloring them according to the above-indicated scheme, the different interactions (and their nature) can be visualized three-dimensionally (see Figure 1C).

The reduced density gradient features remain quite stable with respect to the computational method chosen to determine the underlying electron density. This is also true for promolecular electron distributions, which, for this reason, are usually exploited to carry out NCI-index analyses of biomolecular systems. However, as already mentioned in the Introduction, we have also shown that, by coupling the NCI-index strategy with libraries of extremely localized molecular orbitals (NCI-ELMO method), it is possible to obtain qualitative descriptions of noncovalent interaction networks in

biosystems that outperform the corresponding NCI-promolecular ones, but always keeping a reasonably low computational cost.[42]

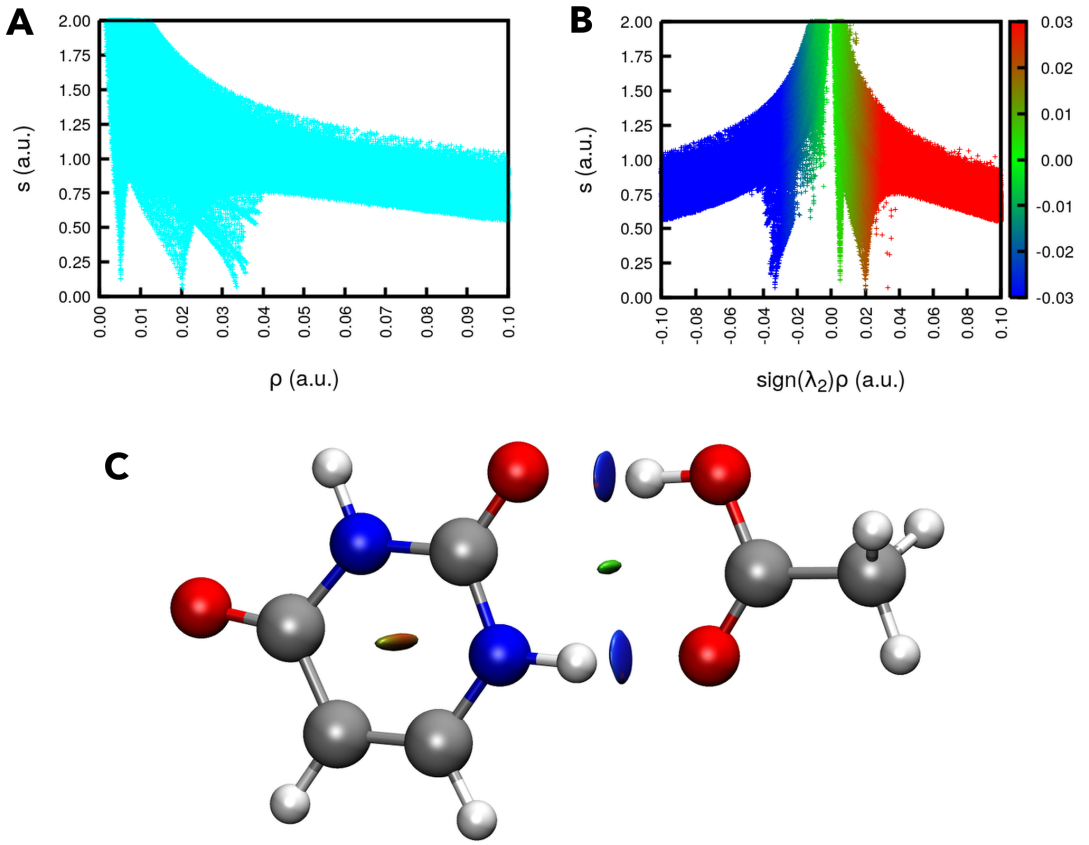


Figure 1. Example of NCI analysis for the uracil-acetic acid dimer: (A) reduced density gradient s plotted as a function of the electron density ρ; (B) reduced density gradient s plotted as a function of the signed electron density $\mathrm{sign}(\lambda_2)\rho$; (C) reduced density gradient isosurfaces in the real space. As mentioned in the text, the blue, green and red colors indicate strong interactions, weak contacts, and steric clashes, respectively.

Furthermore, given the absence of quantitative information in the original version of the NCI-index method, some of us have recently proposed to overcome this drawback by analyzing the correlation between interaction energies resulting from high-level quantum mechanical calculations and integrals of powers of the electron density over well-defined interaction regions.[24] In particular, the following integrals have been considered:

$$I_n = \int_{\Omega_{NCI}} d\boldsymbol{r}\ \rho^n(\boldsymbol{r}) \qquad (2),$$

where the interaction region Ω_{NCI} (hereinafter also indicated as "integration domain") is defined as that part of the real-space where the reduced density gradient $s(\boldsymbol{r})$ is low and where the global electron density is the result of significant contributions of both the two generic fragments A and B forming the complex A:B. In other words, the region Ω_{NCI} is constituted by the set of points $\boldsymbol{r}_i$ satisfying the following three conditions

$$\begin{cases} s(\boldsymbol{r}_i) < s_c \\ \rho_A(\boldsymbol{r}_i) < \gamma_{ref}\ \rho(\boldsymbol{r}_i) \\ \rho_B(\boldsymbol{r}_i) < \gamma_{ref}\ \rho(\boldsymbol{r}_i) \end{cases} \qquad (3)$$

Therefore, Ω_{NCI} depends on the proper choice of values for parameters s_c and γ_{ref}, and on the possibility of clearly defining distinct electron densities $\rho_A(\boldsymbol{r})$ and $\rho_B(\boldsymbol{r})$ associated with subunits A and B, respectively. It is also worth noting that s_c is a cutoff of the reduced density gradient, while γ_{ref} can be seen as the fraction of the total electron density coming from a single subunit/fragment.

In the previous study,[24] the definition of the interaction region was possible thanks to the adoption of the promolecular approximation. Moreover, integrals I_n (see equation (2)) were evaluated by choosing different values for n, γ_{ref} and s_c. Optimal values for these parameters were then selected by finding the best obtained correlations between the integrals I_n and the interaction energies computed at CCSD(T)/CBS level for the small noncovalent dimers constituting the well-known S66 data set.[51] As we will see in the next sections, by always using the energetics of the S66 database as references, in this paper we will re-determine the best values for parameters n, γ_{ref} and s_c more systematically, not only when the promolecular approximation is used, but also when extremely localized molecular orbitals are exploited.

2.2 Extremely localized molecular orbital-based techniques. ELMOs are molecular orbitals strictly localized on very small molecular subunits, such as atoms, bonds or functional groups. Over the years, several strategies have been proposed to determine this kind of orbitals, all of them practically based on the definition of a chemically plausible localization scheme before starting the calculations.[34,52-60]

The extremely localized molecular orbitals that were used in this work were obtained through a technique proposed by Stoll and coworkers.[34] In this method, the system under exam is preliminarily subdivided into fragments that may overlap, in most cases according to the Lewis diagram of the molecule. Owing to this subdivision, local basis sets $\beta_i = \{|\chi_{i\mu}\rangle\}_{\mu=1}^{M_i}$ (where i stands for the label of the generic i-th fragment) are automatically assigned to each subunit. This leads to molecular orbitals of each fragment being expressed as follows:

$$|\varphi_{i\omega}\rangle = \sum_{\mu\in\beta_i} C_{i\mu,i\omega}\,|\chi_{i\mu}\rangle \qquad (4).$$

These orbitals are afterwards used to construct the single Slater determinant that describes the investigated system (ELMO wave function), and the coefficients $\{C_{i\mu,i\omega}\}$ are simply obtained by variationally minimizing the energy associated with that wave function. As shown by Stoll *et al.*,[34] this is equivalent to solving coupled modified Hartree-Fock equations for each subunit:

$$\hat{F}_i\,|\varphi_{i\omega}\rangle = \varepsilon_{i\omega}\,|\varphi_{i\omega}\rangle \qquad (5),$$

with $\hat{F}_i$ as the modified Fock operator corresponding to the i-th fragment:

$$\hat{F}_i = \left(1 - \hat{\rho} + \hat{\rho}_i^{\dagger}\right)\hat{F}\,(1 - \hat{\rho} + \hat{\rho}_i) \qquad (6).$$

In the previous equation, $\hat{F}$ is the traditional Fock operator, $\hat{\rho}$ is the global density operator depending on all the occupied ELMOs of the system, and $\hat{\rho}_i$ is the density operator for the i-th subunit depending only on the occupied ELMOs of the fragment.

For the sake of completeness, we also mention that, since the predefined (possibly overlapping) subunits may share a part of their local basis sets, obtained ELMOs belonging to different fragments are non-orthogonal.

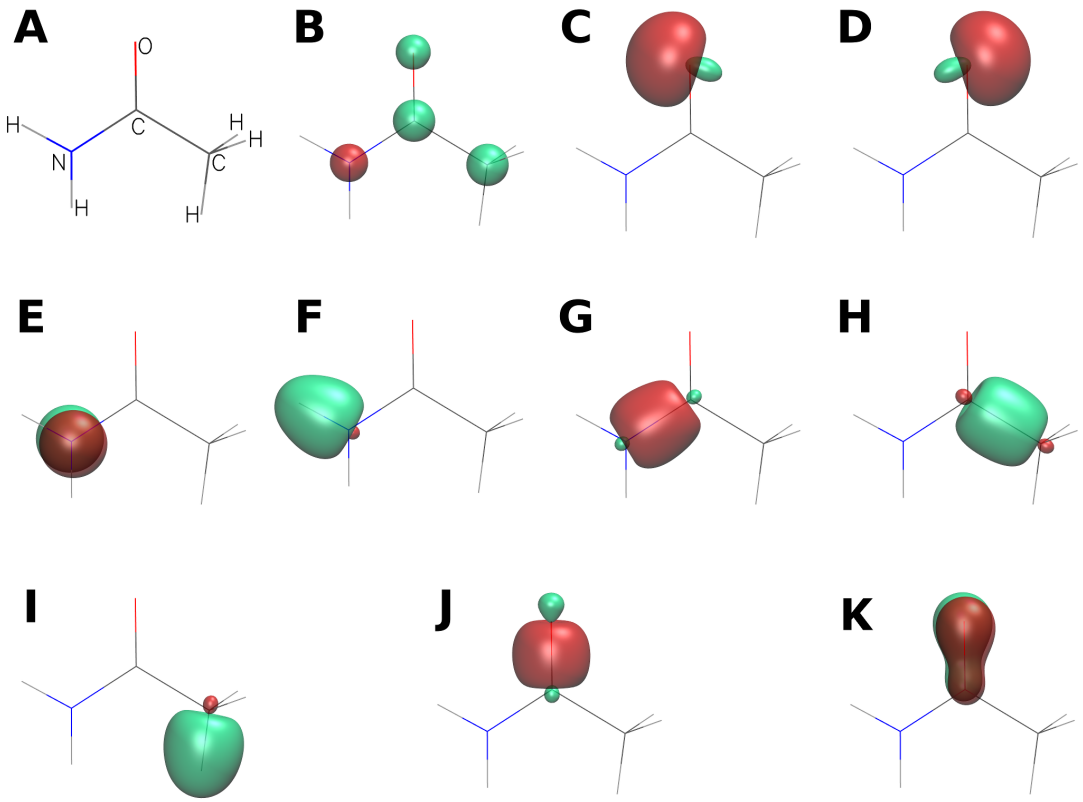


Figure 2. Examples of extremely localized molecular orbitals for acetamide using a localization scheme corresponding to the Lewis structure of the molecule: (A) acetamide structure, (B) ELMOs for the core electrons of the nitrogen, carbon and oxygen atoms of the molecule, (C, D) ELMOs describing the lone pairs of the oxygen atom, (E) ELMO associated with the lone pair of the nitrogen atom, (F) ELMO corresponding to one of the two N-H bonds, (G) ELMO for the N-C bond, (H) ELMO associated with the C-C bond, (I) ELMO corresponding to one of the three C-H bonds, (J, K) ELMOs describing the σ and π electrons of the C=O double bond. All the orbitals were computed with the cc-pVDZ basis set and plotted using 0.15 a.u. isosurfaces. Adapted from Figure 1 in ref. 61.

As an example, in Figure 2 we have shown the extremely localized molecular orbitals obtained for acetamide by adopting a localization scheme corresponding to the Lewis structure of the molecule. It is easy to observe that the orbitals are absolutely localized on atoms and bonds. Due to this strict localization, ELMOs are easily transferable from

molecule to molecule, provided that the environments of the considered fragments in the two molecules are chemically similar. The transfers of extremely localized molecular orbitals are carried out by exploiting a strategy originally proposed by Philipp and Friesner.[31,62] The reliability of the ELMOs transferability have been extensively studied through several focused investigations.[31,32,36-41] These studies were preliminary to the construction of the recently proposed ELMO databanks[33] that currently cover all the elementary units of the twenty natural amino acids in all their possible protonation states and forms (namely, N-terminal, non-terminal and C-terminal) for five usual quantum chemistry basis sets (6-31G, 6-311G, 6-31G(d,p), 6-311G(d,p) and cc-pVDZ). These libraries comprise: i) ELMOs localized on single atoms to describe core and lone-pair electrons, ii) ELMOs localized on two-atom subunits for the treatment of ordinary two-electron/two-center bonds, and iii) ELMOs localized on three atoms to properly describe bonding situations in which it is necessary to take into account the delocalization of the electronic structure (e.g., π electrons in aromatic rings, carboxylate groups and peptide bonds). Moreover, the databases of extremely localized molecular orbitals and the associated *ELMOdb* program are also structured in such a way that tailor-made ELMOs computed on proper model molecules (using any desired basis set) can be read and transferred to the target system under examination. The ELMO databanks have been used not only to reconstruct approximate wave functions and electron densities of large biosystems, but they have also been the starting point for other methodological developments, such as the HAR-ELMO[63] (Hirshfeld atom refinement-ELMO) method in quantum crystallography[64-70] and, more importantly for the present work, the multiscale embedding QM/ELMO technique[44-50]. The QM/ELMO strategy is a novel approach that allows the treatment of the chemically crucial region of a large system through traditional methods of quantum chemistry for

both ground and excited states,[44-46,50] while the environment is described by properly transferred and frozen extremely localized molecular orbitals. After the subdivision in QM and ELMO subsystems and the transfer of the ELMOs to the ELMO subunit, the QM/ELMO procedure can be seen as consisting in two main parts: 1) preliminary orthogonalization of molecular orbitals and basis functions; 2) QM/ELMO self-consistent field (SCF) algorithm, which can also be optionally followed by post-HF/ELMO computations. Very recently, a third layer has been added to the QM/ELMO scheme, thus leading to the QM/ELMO/MM technique,[71] where the outermost subunit of the macrosystem under investigation is treated by means of a molecular mechanics force field.

Interested readers can find more details about the QM/ELMO and QM/ELMO/MM approaches in the Supporting Information or in the seminal papers of the two techniques.[44,45,71] Here we only restrict ourselves to stressing again that these novel multiscale embedding strategies provide well-defined QM and ELMO electron densities that are mainly localized on the corresponding regions, with only small tails that negligibly extend to the neighbouring subsystems. For this reason, also the QM/ELMO and QM/ELMO/MM methods can be exploited to obtain distinct electron densities to be used in quantitative NCI analyses. For instance, in the case of a protein-ligand complex, by treating the ligand quantum mechanically and the surrounding protein at ELMO level we would obtain two separable electron distributions that could be employed to extract quantitative information of quantum mechanical level through the integral-based NCI strategy discussed in Subsection 2.1.

3. TEST CALCULATIONS

We will now describe and discuss the test calculations that were carried out to evaluate the capabilities of the integral NCI-index approach based on the use of extremely localized molecular orbitals and on the application of the QM/ELMO technique. In Subsection 3.1 we will describe the strategy that we followed to systematically re-determine the values for the n, γ_{ref} and s_c parameters for both promolecular and ELMO electron densities. Afterwards, in Subsection 3.2, we will show how the ELMO-based NCI integrals can be exploited to evaluate the interactions between each residue of a ligand-polypeptide and a protein. In Subsection 3.3, we will evaluate the capabilities of the NCI-ELMO integrals in monitoring the evolution of noncovalent interactions along trajectories of a molecular dynamics-like simulations. Finally, in Subsection 3.4, we will focus on how, in some situations, the combination of the integral NCI-index strategy with the QM/ELMO method can be employed to characterize and quantify noncovalent interactions occurring between specific residues of a protein and a ligand.

3.1 Reparameterization. The goal of the reparameterization was to determine the best values for the n, γ_{ref} and s_c parameters (see equations (2) and (3)) for promolecular and ELMO electron densities. To this purpose, we considered the S66 dataset[51] of interaction energies computed at CCSD(T)/CBS level for small molecule dimers. For each type of underlying electron density used in the NCI-index analyses, we determined the set $\{n, \gamma_{ref}, s_c\}$ that provided the best correlation coefficient between the NCI integrals and the S66 interaction energies.

To accomplish this task, we exploited promolecular densities and electron distributions resulting from the transfer of extremely localized molecular orbitals. The ELMOs were previously determined on the single molecules (i.e., the monomers) constituting the

dimers of the S66 database. In particular, for each monomer we considered a localization scheme corresponding to its Lewis structure and, exploiting the molecular geometry optimized at B3LYP/6-311++G(d,p) level, we computed ELMOs using each of the five basis sets currently available in the ELMO libraries (6-31G, 6-311G, 6-31G(d,p), 6-311G(d,p) and cc-pVDZ). Afterwards, the obtained extremely localized molecular orbitals were transferred to the equilibrium geometries of the dimers in the S66 dataset and were orthogonalized following two different procedures. In one case (hereinafter indicated as monomer approximation), only the ELMOs transferred to the same monomer were Löwdin orthogonalized among themselves, thus leading to electron distributions strictly localized on each monomer. In the other case (hereinafter indicated as dimer approximation), all the ELMOs transferred to the dimer were Löwdin orthogonalized, thus entailing monomer electron distributions that are not anymore strictly localized. However, we must bear in mind that the Löwdin orthogonalization mainly preserves the absolutely localized nature of the ELMOs, with only negligible tails that extend beyond the original localization region. Therefore, also in the second case it was possible to separate the monomer electron densities to a very good approximation and, consequently, the integral NCI-index approach remained applicable. All the ELMO calculations were performed by exploiting a modified version of the *GAMESS-UK*[72] quantum chemistry package where the Stoll method[34] for extremely localized molecular orbitals is implemented,[35] while all the transfers of ELMOs were carried out through the *ELMOdb* program associated with the recently constructed ELMO libraries.[33] The previous details will remain valid also for all the other computations and transfers of ELMOs that will be discussed in the next subsections.

For each type of underlying electron distribution mentioned above, we afterwards determined the best correlation coefficient between the obtained NCI integrals and the S66 interaction energies. This was done by systematically varying n from 1.0 to 3.0 (0.1 step) and γ_{ref} from 0.51 to 0.99 (0.01 step). s_c was kept fixed to 1.0 because in the original study on the NCI integral approach this value was shown to provide quite stable integration regions comprising all relevant interactions. All NCI analyses and evaluations of NCI integrals mentioned in this and in the following subsections were performed by exploiting the new *NCIPLOT4.2* software.[23,24] Moreover, the correlation coefficient between two sets of N data $\mathrm{X} = \{x_i\}_{i=1}^{N}$ and $\mathrm{Y} = \{y\}_{i=1}^{N}$ was computed by using the following definition:

$$\text{Corr. Coeff. (X, Y)} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})\,(y_i - \bar{y})}{[\sum_{i=1}^{N}(x_i - \bar{x})\ \sum_{i=1}^{N}(y_i - \bar{y})]^{1/2}} \qquad (7),$$

where $\bar{x}$ and $\bar{y}$ are the average values for the two datasets X and Y, respectively.

The heatmaps depicted in Figure 3 report the trends in the correlation coefficient obtained by varying n and γ_{ref} when the promolecular and ELMO/6-31G(d,p) (monomer and dimer approximations) electron densities were used in the NCI integral analyses. We can easily notice that, when the ELMO electron densities are used, the correlation coefficient remains quite stable and high for most of the n and γ_{ref} values. On the contrary, when the promolecular electron distribution approximation is exploited, we observe a larger variability, with a reduction of the parametric space where the correlation coefficient is high and with even a region characterized by small negative values. Completely analogous results were also observed for the other basis sets (namely, 6-31G, 6-311G, 6-311G(d,p) and cc-pVDZ; see Figures S1-S4 in the Supporting Information).

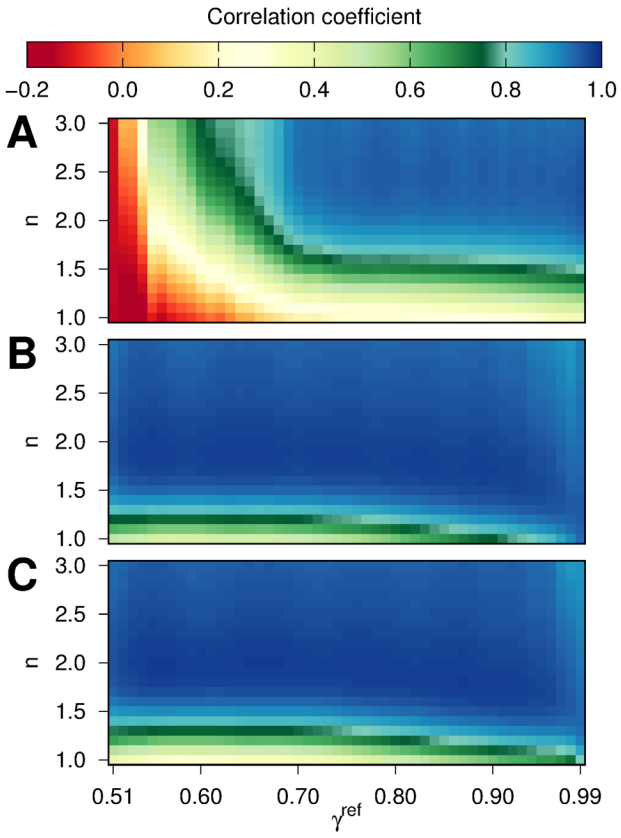


Figure 3. Heatmaps showing the variation of the correlation coefficient resulting from the fitting of the NCI integrals against the S66 interaction energies as a function of n and γ_{ref} (s_c kept fixed to 1.0) when the A) promolecular, B) ELMO/6-31G(d,p) (monomer approximation) and C) ELMO/6-31G(d,p) (dimer approximation) electron densities are used in the NCI analysis.

The larger stability of the correlation coefficient in the case of the ELMO electron densities can be explained with the corresponding invariance of the integration domain Ω_{NCI}. In Figure 4 we can indeed observe that the noncovalent interaction regions associated with the ELMO/6-31G(d,p) electron distributions are always chemically meaningful and their size (see also the volumes reported in Figure 4) remains more stable as γ_{ref} increases (s_c is always set equal to 1.0). Quite the opposite, for promolecular electron densities, the noncovalent interaction regions are almost or completely missing when γ_{ref} is low (which explains the obtained negative values for the correlation coefficient) and excessively increases when γ_{ref} is large (see again the

example reported in Figure 4). Similar results for the other basis sets are shown in the Supporting Information (see Figures S5-S8).

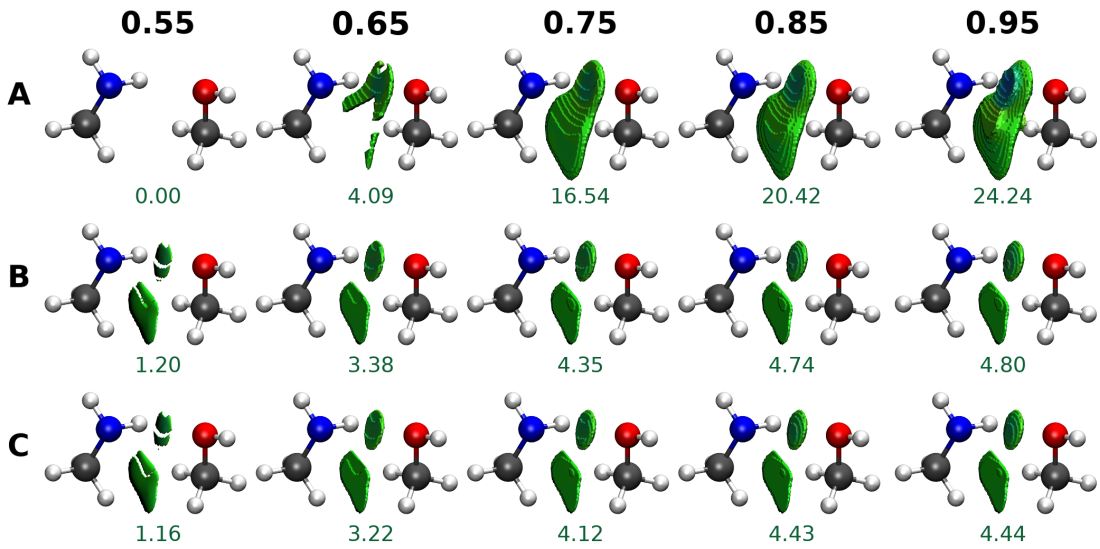


Figure 4. Integration domains Ω_{NCI} (for the $MeNH_2$-MeOH dimer) as a function of γ_{ref} (s_c kept fixed to 1.0) when the A) promolecular, B) ELMO/6-31G(d,p) (monomer approximation) and C) ELMO/6-31G(d,p) (dimer approximation) electron densities are used in the NCI analysis. For each case, the volume (in $bohr^3$) of the integration region is also reported.

Analogous trends can also be observed considering the corresponding 2D NCI plots (i.e., s vs. $\mathrm{sign}(\lambda_2)\rho$ plots; see Figure 5), where each depicted point satisfies the three conditions reported in equation (3). In fact, also in those cases we can see that the overall shape of the ELMO/6-31G(d,p) graphs remains rather invariant as γ_{ref} varies, while the one associated with the promolecular plots significantly change when γ_{ref} increases. Similar plots for the other basis sets are also reported in the Supporting Information (see Figures S9-S12).

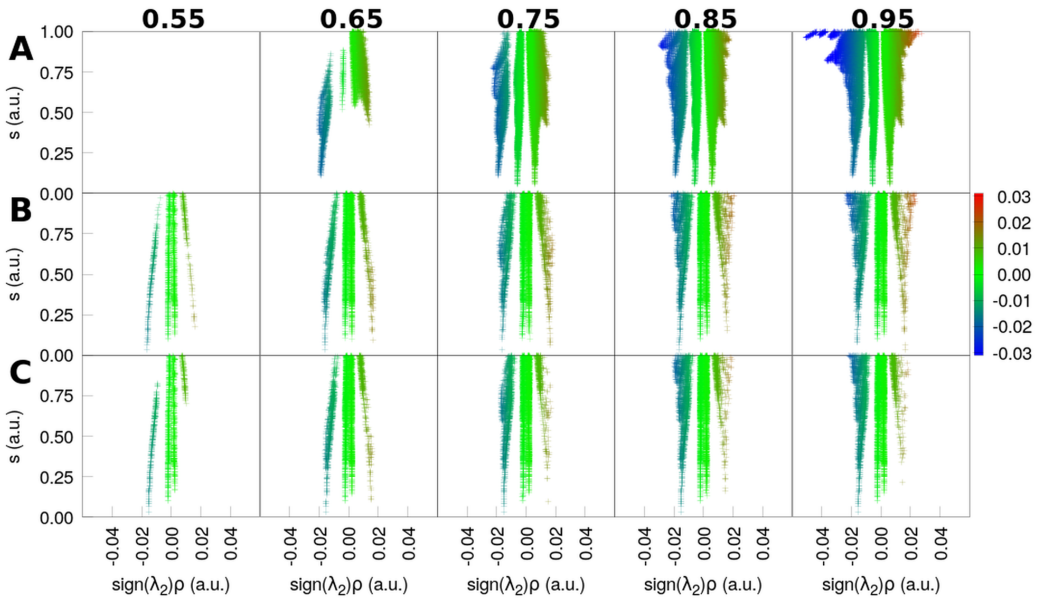


Figure 5. 2D NCI plots (for the $MeNH_2$-MeOH dimer) as a function of γ_{ref} (see the values in bold above the graphs) when the A) promolecular, B) ELMO/6-31G(d,p) (monomer approximation) and C) ELMO/6-31G(d,p) (dimer approximation) electron densities are used in the NCI analysis.

From a more quantitative point of view, we can see that the ELMO electron densities always provide better correlations with the reference S66 interaction energies compared to the promolecular case, as it can also be observed in the heatmaps of Figure 3 zoomed on the regions where the values of the correlation coefficient are maximal (see Figure 6 for basis set 6-31G(d,p) and Figures S13-S16 in the Supporting Information for the other sets of basis functions). This is confirmed in Table 1, where, for each underlying electron density, we reported the best correlation coefficient along with the corresponding optimal values of the n and γ_{ref} parameters.

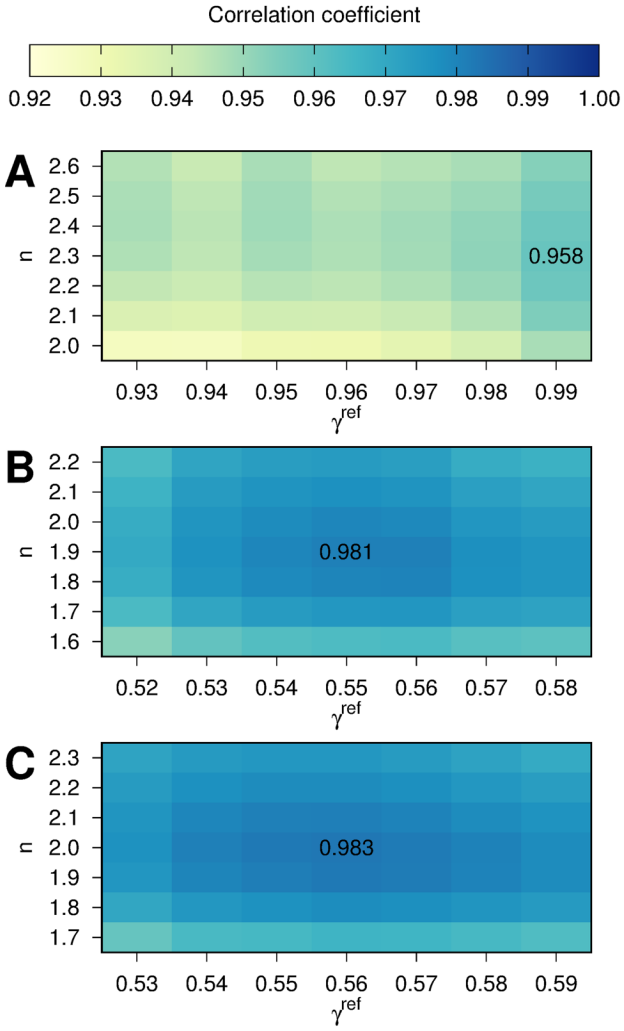


Figure 6. Heatmaps showing the variation of the correlation coefficient resulting from the fitting of the NCI integrals against the S66 interaction energies as a function of n and γ_{ref} (s_c kept fixed to 1.0) and zoomed over the maximum regions when the A) promolecular, B) ELMO/6-31G(d,p) (monomer approximation) and C) ELMO/6-31G(d,p) (dimer approximation) electron densities are used in the NCI analysis.

Table 1. Best correlation coefficient (Corr. Coeff.) and optimal values of the n and γ_{ref} parameters for each underling electron density. The promolecular values are reported in the columns of the monomer approximation for the ELMO electron densities.

Electron Density	Monomer Approximation			Dimer Approximation		
	Corr. Coeff.	n	γ_{ref}	Corr. Coeff.	n	γ_{ref}
Promolecular	0.958	2.3	0.99			
ELMO/6-31G	0.974	1.8	0.60	0.975	1.9	0.56
ELMO/6-311G	0.973	1.9	0.63	0.975	2.0	0.65
ELMO/6-31G(d,p)	0.981	1.9	0.55	0.983	2.0	0.56
ELMO/6-311G(d,p)	0.978	2.0	0.58	0.981	2.1	0.59
ELMO/cc-pVDZ	0.974	2.0	0.57	0.976	2.1	0.58

Finally, we compared the correlation coefficients, mean absolute errors (MAEs) and mean relative errors (MREs) resulting from the integral NCI-ELMO analyses to those of standard quantum chemistry calculations performed at different levels of theory (DFT-PBE, DFT-PBE with D3 dispersion correction,[73] and DFT-M06-2X) with various sets of basis functions (namely, the five basis sets of the ELMO libraries plus 6-31+G(d,p), 6-311+G(d,p) and aug-cc-pVDZ). In all cases, the CCSD(T)/CBS results in the S66 dataset were used as references. All DFT computations were performed using the *Gaussian09* quantum chemistry package.[74]

From the results collected in Table 2 we can observe that the basic PBE calculations provided correlations and errors that are systematically worse than the NCI-ELMO ones, regardless of the chosen basis set. Interestingly, we can also see that the MAEs associated with the integral NCI-ELMO analyses are always lower than the chemical accuracy limit (1.0 kcal/mol). The situation changes with the introduction of Grimme's D3 dispersion correction[73] in the PBE calculations and with the DFT computations

performed by exploiting the meta-hybrid GGA functional M06-2X. For those cases we can notice that, when Pople basis sets without polarization and diffuse functions were used (6-31G and 6-311G), the traditional quantum chemical calculations gave again results slightly worse than those obtained through the integral NCI-ELMO strategy. On the contrary, when polarization and diffuse basis functions were introduced, the quantum chemical computations provided better results. The same can be observed for the PBE-D3 and M06-2X calculations carried out with correlation consistent basis sets (cc-pVDZ and aug-cc-pVDZ).

The reason why the integral NCI-ELMO approach generally performs worse than PBE-D3 and M06-2X computations with polarization and diffuse functions is ascribable to the fact that ELMOs are transferred from model molecules to the target system and are afterwards kept frozen without adapting to the new chemical environment. To improve the results, we are currently planning to develop an *a posteriori* technique that allows the polarization/relaxation of the ELMOs after the transfer, as a response to the surrounding chemical groups in the target molecule. To accomplish this task, we could imagine the exploitation of virtual extremely localized molecular orbitals, which are already available in the ELMO libraries, and which can be already transferred/rotated along with the occupied ones by exploiting the *ELMOdb* program.

Table 2. Correlation coefficients (Corr. Coeffs.), mean absolute errors (MAEs) and percentage mean relative errors (MREs) with respect to the S66 reference energy values as obtained from integral NCI-ELMO analyses and from traditional quantum chemistry DFT calculations.

Method /basis set	Corr. Coeff.	MAE (kcal/mol)	MRE (%)
NCI-ELMO/6-31G (monomer)	0.974	0.84	18.3
NCI-ELMO/6-311G (monomer)	0.973	0.83	18.2
NCI-ELMO/6-31G(d,p) (monomer)	0.981	0.83	18.3
NCI-ELMO/6-311G(d,p) (monomer)	0.978	0.83	18.2
NCI-ELMO/cc-pVDZ (monomer)	0.974	0.84	18.3
NCI-ELMO/6-31G (dimer)	0.975	0.83	18.2
NCI-ELMO/6-311G (dimer)	0.975	0.83	18.2
NCI-ELMO/6-31G(d,p) (dimer)	0.983	0.83	18.2
NCI-ELMO/6-311G(d,p) (dimer)	0.981	0.83	18.2
NCI-ELMO/cc-pVDZ (dimer)	0.976	0.83	18.3
PBE/6-31G	0.889	2.55	65.2
PBE/6-311G	0.897	2.40	61.9
PBE/6-31G(d,p)	0.910	2.22	59.8
PBE/6-311G(d,p)	0.920	2.14	56.7
PBE/6-31+G(d,p)	0.918	2.11	56.3
PBE/6-311+G(d,p)	0.921	2.10	55.5
PBE/cc-pVDZ	0.920	2.35	61.1
PBE/aug-cc-pVDZ	0.929	2.16	57.1
PBE-D3/6-31G	0.958	1.35	24.1
PBE-D3/6-311G	0.965	1.18	20.1
PBE-D3/6-31G(d,p)	0.976	0.79	15.8
PBE-D3/6-311G(d,p)	0.983	0.58	11.7
PBE-D3/6-31+G(d,p)	0.983	0.62	11.8
PBE-D3/6-311+G(d,p)	0.985	0.57	10.7
PBE-D3/cc-pVDZ	0.983	0.45	9.0
PBE-D3/aug-cc-pVDZ	0.987	0.44	8.0
M06-2X/6-31G	0.961	1.14	23.6
M06-2X /6-311G	0.971	0.88	16.2
M06-2X /6-31G(d,p)	0.980	0.62	15.0
M06-2X /6-311G(d,p)	0.988	0.32	6.4
M06-2X /6-31+G(d,p)	0.988	0.37	7.5
M06-2X /6-311+G(d,p)	0.990	0.34	7.1
M06-2X /cc-pVDZ	0.986	0.71	16.0
M06-2X /aug-cc-pVDZ	0.991	0.29	6.3

To conclude the reparameterization, we also analyzed the 2D NCI plots obtained for all the dimers of the S66 dataset by performing quantitative NCI analyses at promolecular and ELMO (monomer and dimer approximations) levels using the optimal values for γ_{ref} reported in Table 1. All these graphs are reported in the Supporting Information (see Figures S17-S27). We can clearly observe that, in the promolecular case, the troughs associated with weak interactions (i.e., weak/non-conventional hydrogen bonds or van der Waals contacts) fall in the $[-0.02\ \text{a.u.}, 0.02\ \text{a.u.}]$ range of $\text{sign}(\lambda_2)\rho(\boldsymbol{r})$, while strongly attractive interactions (i.e., strong/conventional hydrogen bonds) and steric clashes occur when $\text{sign}(\lambda_2)\rho(\boldsymbol{r}) < -0.02$ a.u. and $\text{sign}(\lambda_2)\rho(\boldsymbol{r}) > 0.02$ a.u., respectively (see Figure S17). Differently, when ELMOs are used, the peaks corresponding to the weak non-covalent contacts can be found in the $[-0.01\ \text{a.u.}, 0.01\ \text{a.u.}]$ interval, while troughs associable with strongly attractive interactions and steric clashes are observed for $\text{sign}(\lambda_2)\rho(\boldsymbol{r}) < -0.01$ a.u. and $\text{sign}(\lambda_2)\rho(\boldsymbol{r}) > 0.01$ a.u., respectively (see Figures S18-S27). These repartitions of the NCI peaks will be extremely important in the next two subsections, where we will exploit the integral NCI-(QM/)ELMO approach to determine the different types of intermolecular interactions in selected protein-ligand complexes.

3.2 Quantitative NCI-ELMO analysis. As a second step of our investigation, we tested the capabilities of the integral NCI-index approach purely based on ELMO electron densities in evaluating the interactions of each residue of a ligand-polypeptide with a target protein.

To accomplish this task, we considered a complex between the iCAL36 polypeptide (namely, the ligand) and the CFTR associated ligand (CAF) PDZ domain (namely, the target protein). From an experimental crystal structure (PDB code: 4E34)[75] we extracted a 550-atom model. In particular, we considered the iCAL36 polypeptide plus

the PDZ domain residues and the water molecules having at least one atom within a radius of 5 Å around the ligand, leading to a final system consisting of iCAL36, 47 residues and two water molecules (see Figure 7A). The terminal residues were saturated by means of N-methyl amino (CH_3-NH-) and acetyl (CH_3-CO-) groups using the *ProScrs.py* program provided with the *AMBER* package[76]. Afterwards, exploiting the *ELMOdb* program, we reconstructed the global electron density of the examined system by simply transferring the required ELMOs from the ELMO libraries to the model target structure of the protein-ligand complex (cc-pVDZ basis-set). The resulting electron distribution was then used to perform the NCI-ELMO integral analysis exploiting the optimal parameters previously determined for the cc-pVDZ set of basis functions in the case of the dimer approximation (see Table 1). Furthermore, based on the analysis reported in the concluding paragraph of Subsection 3.1, the NCI integrals were evaluated by considering the global range $[-0.1\ \text{a.u.}, 0.01\ \text{a.u.}]$ of the signed electron density ($\text{sign}(\lambda_2)\rho(\boldsymbol{r})$) comprising both strongly attractive interactions ($-0.1\ \text{a.u.} \leq \text{sign}(\lambda_2)\rho(\boldsymbol{r}) < -0.01\ \text{a.u.}$) and weak noncovalent contacts ($-0.01\ \text{a.u.} \leq \text{sign}(\lambda_2)\rho(\boldsymbol{r}) \leq 0.01\ \text{a.u.}$), whose contributions were also quantified separately.

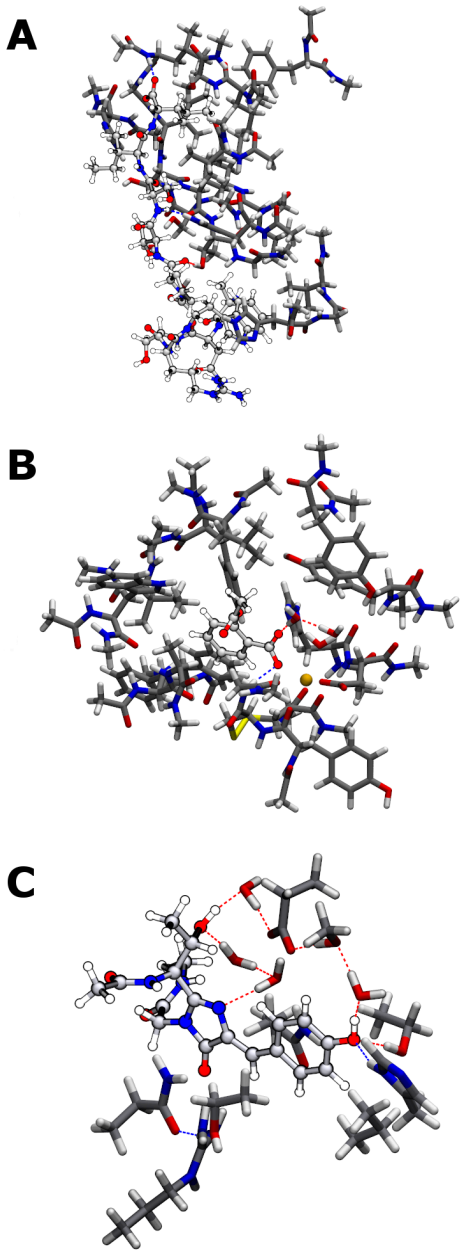


Figure 7. Graphical representation of the models for the protein-ligand complexes investigated in Section 3: (A) 550-atom model of the PDZ-iCAL36 domain complex, (B) 413-atom model of the complex between aspirin and the phospholipase A2 protein (frame 651 of the considered binding trajectory), (C) 161-atom model of the complex between the Green Fluorescent Protein and the chromophore p-hydroxybenzylidene-imidazolinone. For all complexes, the protein and the ligand are shown in licorice and ball-and-stick representations, respectively. Hydrogen, carbon, nitrogen, oxygen, sulphur, and calcium atoms are depicted in light grey/white (protein/ligand), dark grey/grey (protein/ligand), blue, red, yellow, and orange, respectively.

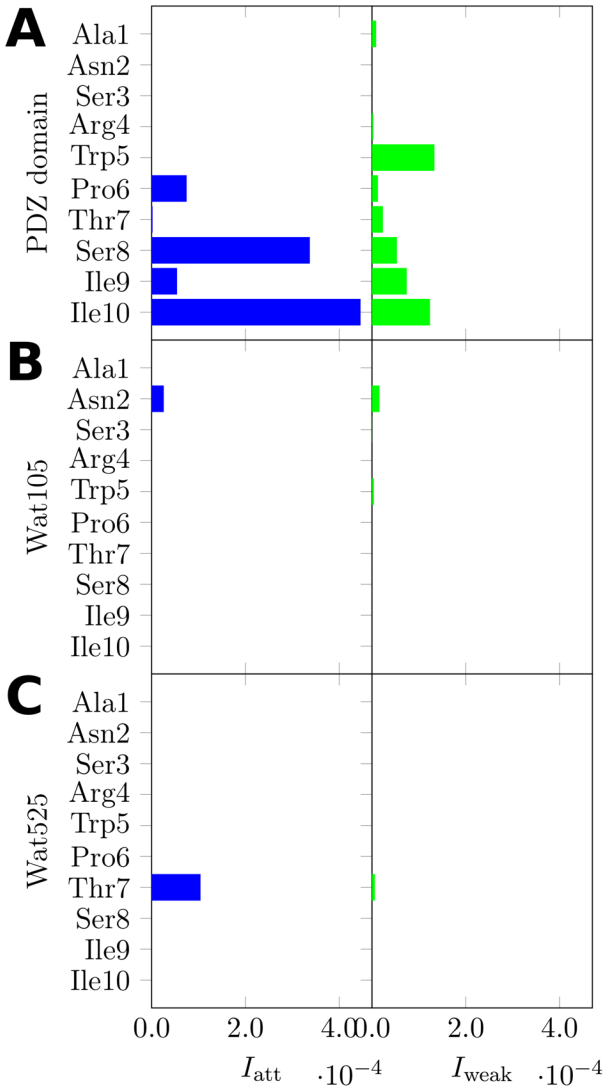


Figure 8. Histograms graphically depicting the strengths of the strongly attractive (I_{att}) and weak (I_{weak}) interactions established by each residue of the iCAL36 polypeptide with (A) the PDZ domain, (B) Wat105 and (C) Wat525, as obtained from the NCI-ELMO integral analysis with basis set cc-pVDZ (dimer approximation).

The analysis allowed us to characterize and evaluate the strength of the interactions established by each residue of the iCAL36 polypeptide with the PDZ domain and with two selected water molecules. The results are given in Figure 8, where the strongly attractive and weak contributions for all residues of iCAL36 are graphically depicted. The corresponding numerical values are reported in Table S1 of the Supporting Information, along with hydrogen-acceptor distances when plausible strong/conventional hydrogen bonds are present. Moreover, in Figure S28 of the

Supporting Information, we also included the 3D integration domains associated with the occurring intermolecular interactions.

By inspecting Figure 8A, we can immediately observe that Ile10 is the ligand residue that establishes the strongest interaction with the PDZ domain, with the largest contribution among the detected strong hydrogen bond interactions and the second largest contribution among the observed weak noncovalent contacts. The large strongly attractive contribution for Ile10 is certainly ascribable to the hydrogen bonding with residues Leu299, Gly300 and Ile301 of the PDZ domain (Figure S28-F), while the non-negligible contribution corresponding to weak interactions is due to the hydrophobic contacts with the same three residues mentioned above and Ile303, Val353, Leu356, Ser357 and Phe365 (Figure S28-F). Therefore, the NCI-ELMO analysis revealed an extended network of several noncovalent interactions between Ile10 and the PDZ domain. This result is remarkably in line with previous observations by Amacher *et al.*,[75] who identified the terminal isoleucine residue of the ligand polypeptide as critical for the binding selectivity. In Figure 8A, we can also observe that Ser8 is the other subunit of iCAL36 that significantly interacts with the target protein (second largest contribution for the strongly attractive contacts). This is mainly due to the hydrogen bond interaction with His349 and Ile303 (Figure S28-D) of the PDZ domain, which is again in agreement with the findings of the previous study conducted by Amacher and coworkers.[75] Figure 8A shows that the interaction of Ser8 with the target protein is also characterized by a non-negligible contribution due to weak noncovalent interactions, mainly ascribable to the hydrophobic contact with residue Val353 and to a weak/non-conventional hydrogen bond (i.e., C-H···O contact) with Ser302 (Figure S28-D). Pro6 is another residue characterized by a strong hydrogen bonding with the PDZ domain, which is due to its interaction with Thr304 of the target protein. Moreover, out of the

six residues that mainly interact with the PDZ domain (Trp5, Pro6, Thr7, Ser8, Ile9 and Ile10), Pro6 has the smallest integral value in the range of weak interactions, which is due to a CH-π interaction between Pro6 and the imidazole ring of His349 and other hydrophobic interactions with His309 and His349 (Figure S28-B). Ile9 is a residue that shows a non-negligible contribution both in terms of strong and weak contacts. The former originates from an interaction with residue His319; this is a C-H···H contact that is analogous to a contact already observed through an NCI analysis in a previous work,[77] but, in this case, it is surprisingly strong, probably due to the acidic nature of the involved histidine hydrogen atom. The latter weak contacts correspond to hydrophobic interactions between Ile9 and residues Ile301, Ser302 and His319 (Figure S28-E). It is also worth noting that the residue with the largest contribution due to weak interactions is Trp5, which is characterized by a weak/non-conventional hydrogen bond with residue Thr304 and by hydrophobic contacts with residues Gly305, His309, Val311 and Leu314 (Figure S28-A). This aspect also perfectly agrees with the work by Amacher and collaborators, where it is explicitly mentioned that Trp5 "interacts with a mostly hydrophobic ledge on the surface of the CALP (*editor's note:* PDZ) domain".[75] For the sake of completeness, we also mention that residue Thr7 interacts with residues Ser302, Ile303, Thr304, Ser316 and His349 of the target protein only through non-conventional hydrogen bonds and van der Waals contacts (see also Figure S28-C), while strongly attractive interactions are negligible. Finally, it is interesting to observe that the first four residues (Ala1, Asn2, Ser3 and Arg4) do not show significant interactions with the target protein (only a negligible weak interaction is observed for Ala1), in agreement with previous findings indicating that the PDZ domain can interact with up to seven ligand residues.[78]

To complete the analysis, we also considered the interactions of the polypeptide residues with the two water molecules in our 550-atom model of the protein-ligand complex. The obtained results (see Figures 8B and 8C) show that Wat105 and Wat525 practically interact only with residues Asn2 and Thr7, respectively. Both interactions are essentially due to strong hydrogen bonds (see Figures S28-G and S28-H). In the case of the Thr7-Wat525 interaction (Figure S28-H), a conventional hydrogen bond is formed between OG1 in Thr7 and H2 in Wat525, leading to the significant attractive interaction detected through the NCI-ELMO integral strategy (Figure 8C). Moreover, an additional weak non-conventional hydrogen bond is also detected between HG22 of Thr7 and the oxygen atom of Wat525, which explains the very small contribution in the weak interaction range (Figure 8C). Finally, for the Asn2-Wat105 interaction (Figure S28-G), a conventional hydrogen bond is formed between the backbone oxygen atom of Asn2 and H1 of Wat105. With a donor acceptor distance of 2.309 Å, this interaction lies on the borderline between attractive and weak interactions, which explains why it contributes to both ranges in Figure 8B. This could be improved in the future through the definition of smaller interaction ranges able to better distinguish between strong hydrogen bonds, weak hydrogen bonds and van der Waals contacts (see the end of Section 3.1 for the current definition).

3.3 NCI-ELMO analysis along an MD-like trajectory. As further step of our test calculations, we decided to assess the capabilities of the integral NCI-index approach based on ELMO electron densities in monitoring the evolution of protein-ligand interactions along trajectories of molecular dynamics-like simulations.

To this purpose, we considered the binding trajectory of the phospholipase A2 protein with aspirin. This trajectory was downloaded from the PELE (protein energy landscape exploration) web server and was calculated by applying the PELE method,[79,80] a

strategy that combines a Monte Carlo stochastic approach with protein structure prediction algorithms to accurately reproduce long time-scale processes in only few hours of CPU time. Of the 651 available frames we considered the last 252 ones, which were those characterized by the largest variations in terms of binding poses. We have afterwards compared the evolution of the protein-ligand interaction energy provided at the OPLS2005 force field level by the PELE technique with the evolution of the sum of the strongly attractive and weak contact contributions to the NCI integrals (hereinafter also simply indicated as $I_{att}+I_{weak}$) resulting from the transfer of ELMOs.

The integral NCI-index approach is based on the integration of noncovalent interaction regions in the real space that are well-localized at the interface of the interacting subsystems. Therefore, without losing generality, from each considered frame of the PELE binding trajectory we extracted substructures consisting of aspirin plus the rest of the system (i.e., phospholipase A2 protein, solvent water molecules and Ca^{2+} cation) that is within a radius of 4.0 Å from the barycenter of the ligand (see Figure 7B).

Concerning the transfer of ELMOs, the extremely localized molecular orbitals describing the fragments of the investigated protein and of the surrounding water molecules were taken from the currently available ELMO libraries (6-31G(d,p) basis set), while those for the Ca^{2+} cation and the aspirin fragments were preliminarily determined on the isolated ion and on the isolated aspirin molecule, respectively, always using the 6-31G(d,p) set of basis functions. Consequently, all the NCI integrals considered in this subsection were evaluated using the optimal parameters for the 6-31G(d,p) basis set (dimer approximation) reported in Table 1.

Finally, as already done above for the other case-study, also in this situation we evaluated the global NCI-integrals along the MD-like trajectory by considering the range $[-0.1\ \text{a.u.}, 0.01\ \text{a.u.}]$ of the signed electron density, distinguishing between

strongly attractive (-0.1 a. u. $\leq \mathrm{sign}(\lambda_2)\rho(\boldsymbol{r}) < -0.01$ a. u.) and weak (-0.01 a. u. $\leq \mathrm{sign}(\lambda_2)\rho(\boldsymbol{r}) \leq 0.01$ a. u.) interactions.

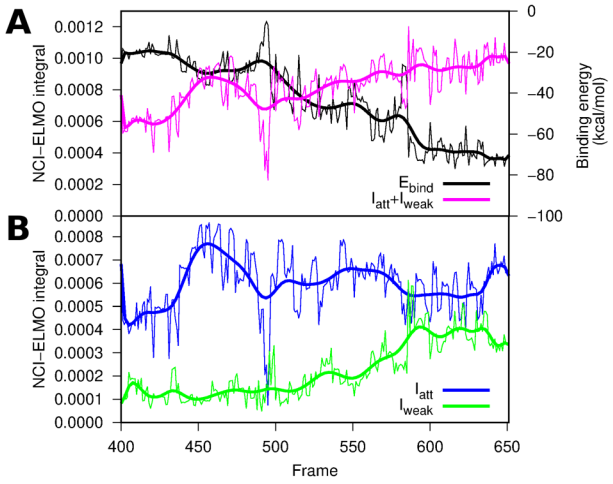


Figure 9. (A) Evolutions of the OPLS2005 binding energy and of the sum of the strongly attractive and weak contributions to the NCI-ELMO integrals (I_{att}+I_{weak}), and (B) separate evolutions of the strongly attractive (I_{att}) and weak (I_{weak}) contributions along the PELE binding trajectory of aspirin with the phospholipase A2 protein. For each plot, other than reporting the actual values for each frame, we also show the result of the Bezier smoothing applied to each set of data.

The overall results are shown in Figure 9. In Figure 9A, on the same graph, we reported both the evolution of the OPLS2005 binding energy and the evolution of the sum of the strongly attractive and weak noncovalent contributions to the NCI-ELMO integrals along the PELE binding trajectory. As already noticed in a previous study on the NCI integral approach based on promolecular electron densities,[23] it is worth observing that the trend of I_{att}+I_{weak} mirrors the one of the OPLS2005 binding energy. In fact, for each decrease/increase of the binding energy we have a corresponding increase/decrease of I_{att}+I_{weak}, with the highest value of the sum of the strongly attractive and weak components corresponding to the most negative (i.e., the most favorable) interaction energy. Moreover, if we also consider the curves corresponding to the Bezier smoothing applied to each set of data, we can also see that, in the very first steps that we considered, higher values of the force field binding energy generally correspond to

lower values of $I_{att}+I_{weak}$. Afterwards, the depicted curves indicate that the OPLS2005 binding energy overall decreases, while, on average, the NCI-descriptor increases. All the previous observations thus confirm again that the NCI integrals can be reliably exploited to rank protein-ligand binding poses.

In Figure 9B, the strongly attractive and weak contributions are separately shown along the trajectory. We can observe that the strongly attractive / hydrogen bond interactions are overall dominant compared to the weak ones. However, it is also worth noting that, although for the first frames (especially around frame 450) the strongly attractive contributions are larger, the two contributions afterwards become more similar. In particular, towards the final frames of the trajectory, the I_{att} and I_{weak} values are more or less equivalent, even if the former still predominate over the latter.

3.4 QM/ELMO-based analysis of noncovalent interactions. In the last part of our work, we also studied the possibility of interfacing the integral NCI-index approach with the QM/ELMO method,[44-50,71] which could be seen as a way of obtaining more accurate quantitative information of QM level from the NCI analyses. In some situations, this could be the option of choice to better analyze selected binding poses extracted from a molecular dynamics simulation or a docking computation for a protein-ligan complex. In fact, ligands may require a more rigorous quantum mechanical description than the one based on the simple ELMO technique. For example, if the ligand was a chromophore, ELMOs would not be enough to describe the possible conjugated (and consequently highly delocalized) system. In this regard, the QM/ELMO technique would be particularly useful as the fully quantum chemistry method used for the QM region would provide a more suitable electron density for the ligand. Therefore, in this subsection, a protein-ligand complex will be studied through the combined integral NCI-QM/ELMO approach. It is worth stressing again that the

use of the QM/ELMO technique still guarantees the separability of the electron density distributions associated with the interacting units (in this case, a protein and a ligand), which is the fundamental working hypothesis at the basis of the quantitative NCI strategy.

To investigate this possibility, we took into account the complex formed by the Green Fluorescent Protein (GFP) with its chromophore (p-hydroxybenzylidene-imidazolinone, pHBDI). To this purpose, we used a 161-atom model (see Figure 7C) previously extracted by Kaila and collaborators[81] from a crystal structure (PDB code: 1EMB) and already used to test the capabilities of multiscale embedding methods for excited states.[46,82] In addition to the chromophore, the extracted substructure includes the side chains (cut at the C_β atom) of nine residues of the Green Fluorescent Protein (Thr62, Gln69, Gln94, Arg96, His148, Val150, Thr203, Ser205 and Glu222) along with four crystallization water molecules that play an active role in the network of noncovalent interactions. Since the chromophore is directly linked to the protein, backbone atoms of Phe64 (C, O and C_α) and Val68 (N and C_α) are also part of the model system.

We performed a QM/ELMO calculation on the above-mentioned reduced model, with the chromophore corresponding to the quantum mechanical subunit treated at M06-2X/cc-pVDZ level of theory, and with the remaining part of the system coinciding with the ELMO region described through extremely localized molecular orbitals transferred from the currently available ELMO libraries (cc-pVDZ basis set). For the sake of completeness, the QM/ELMO computation was carried out by exploiting a modified version of the *Gaussian09*[74] package where the QM/ELMO approach was implemented.

The QM/ELMO calculation provided the subsystem electron densities that were afterwards used to carry out the integral NCI analysis, which allowed the extraction of information on the strength and the nature of the interactions between the chromophore and single residues in the binding pocket of the GFP. Since the cc-pVDZ basis set was used, in the analysis described in this subsection all the NCI integrals were evaluated using the optimal parameters for the cc-pVDZ set of basis functions (dimer approximation) shown in Table 1. Also in this case, to assess the contributions due to the different types of interactions, the NCI integrals were evaluated considering two subdomains of integration according to the value of the "signed" electron density $\mathrm{sign}(\lambda_2)\rho(\boldsymbol{r})$ (see again the final part of Subsection 3.1), with the $[-0.1\ \mathrm{a.u.}, -0.01\ \mathrm{a.u.}[$ range corresponding to strongly attractive interactions (mainly strong/conventional hydrogen bonds) and with the $[-0.01\ \mathrm{a.u.}, 0.01\ \mathrm{a.u.}]$ interval for weak noncovalent contacts (i.e., weak/non-conventional hydrogen bonds and van der Waals interactions).

Concerning the results, we decided to investigate and to better unravel the nature of plausible noncovalent interactions occurring between the chromophore and the surrounding residues of the above-described model for the GFP. These interactions are schematically depicted in Figure 10A in a LIGPLOT[83,84] diagram obtained through the LigPlot+ program[84], which allows to sketch 2D interaction plots for protein-ligand complexes and which exploits the HBPLUS software[85] to detect protein-ligand interactions on the basis of purely geometrical criteria. In those diagrams, hydrogen bonds are drawn as green dashed lines, while non-bonded (e.g., van der Waals) interactions are represented as arcs with spokes. These interactions were afterwards analyzed by means of integral NCI-QM/ELMO computations. The main results of these calculations are shown in Figure 10B, where we have graphically depicted the strongly

attractive and weak noncovalent contributions for each considered interaction. The corresponding exact numerical values are given in Table S2, where we have also reported the hydrogen-acceptor distances for plausible hydrogen bonds. For the sake of completeness, in Figure S29 we have shown the integration domains associated with the analyzed intermolecular interactions.

By inspecting Figure 10B, we can see that strong hydrogen bond interactions largely predominate over the weak noncovalent contacts. The largest contributions due to strong attractive interactions are observed for water molecules Wat14, Wat11 and Wat12, which establish direct hydrogen bonds with the chromophore, as also indicated by the disk-shaped integration domains with blue centres depicted in Figures S29-A and S29-B. Among the three above-mentioned water molecules, the largest I_{att} values are obtained for Wat14, which is involved in two hydrogen bonds with pHBDI (see again Figure S29-A). Furthermore, again in Figure S29-A, a green integration domain corresponding to a non-conventional hydrogen bond contact is located below Wat14, which is the reason for the non-zero contribution to the weak noncovalent interactions reported in Figure 10B and Table S2. Concerning Wat11 and Wat12, the attractive contributions are slightly lower than for Wat14, which can be explained with the fact that both form only one hydrogen bond with the chromophore (Figures S29-A and S29-B). Finally, by comparing the attractive interactions associated with Wat11 and Wat12, a larger I_{att} value is observed for the former, which is indeed the water molecule involved in a shorter hydrogen bond (see Table S2).

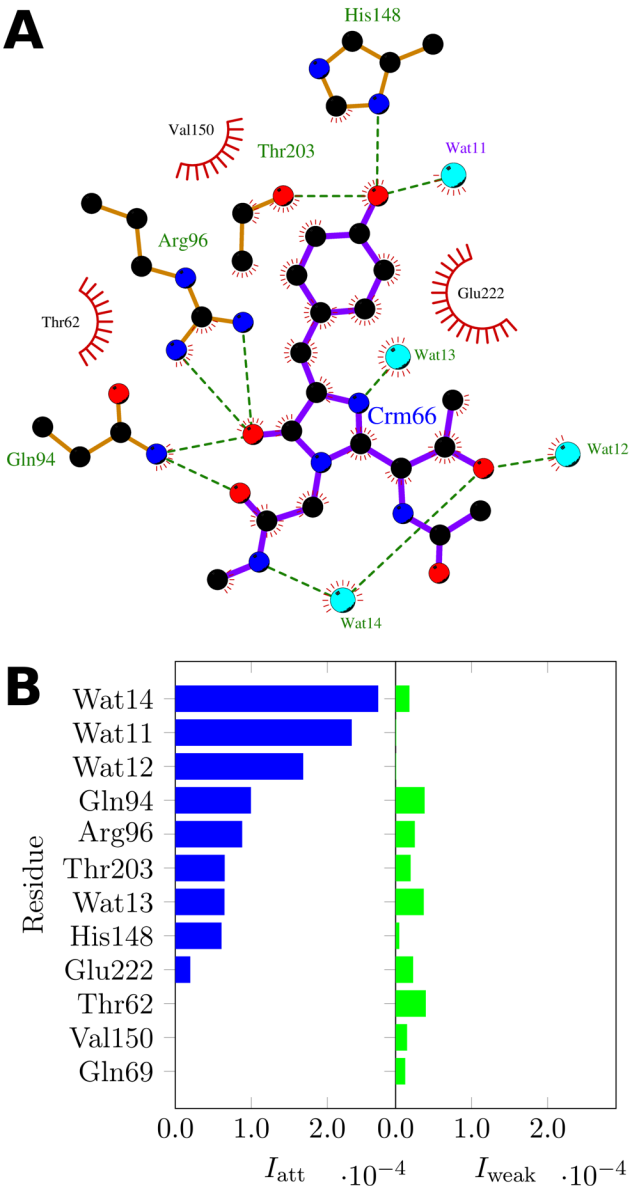


Figure 10. (A) LIGPLOT diagram graphically showing the interactions between the pHDBI chromophore and the surrounding residues of the Green Fluorescent Protein (hydrogen bonds are depicted through green dotted lines, while hydrophobic contacts are represented by means of red arcs with spokes); (B) histograms graphically depicting the strengths of the strongly attractive (I_{att}) and weak (I_{weak}) interactions formed by residues of the Green Fluorescent Protein with the pHDBI chromophore in the binding pocket, as resulting from the NCI-QM/ELMO integral analysis with basis set cc-pVDZ (dimer approximation).

In addition to the previous noncovalent contacts, pHBDI establishes other important strong hydrogen bond interactions with residues Gln94, Arg96, Thr203, His148 and water molecule Wat13. All of them are less strong than the interactions with Wat14,

Wat11 and Wat12. This can be rationalized by noticing that, in all these cases, the hydrogen acceptor distances are significantly longer (see again Table S2). Furthermore, it is interesting to observe that the I_{att} values for Gln94 and Arg96 are approximately equivalent and larger than those corresponding to the other above-mentioned residues, which can be explained by considering that both Gln94 and Arg96 are involved in a double hydrogen bond with the chromophore (see Figures S29-C and S29-D). Gln94 and Arg96 are also characterized by non-negligible I_{weak} values, mainly ascribable to the quite large 3D integration domains depicted in Figures S29-C and S29-D that extend beyond the typical "hydrogen bond ring-shaped domains" (in particular, see the central parts of the interaction regions). Weaker attractive interactions are established by residues Thr203 and His148, which indeed form only a single hydrogen bond with pHBDI (see Figures S29-E and S29-F). Thr203 is also characterized by a non-negligible contribution due to weak noncovalent contacts, which corresponds to the extended green integration domain depicted in Figure S29-E and indicating the presence of a hydrophobic interaction. Quite the opposite, I_{weak} is marginal for His148, which correlates well with the small integration domain of van der Waals type depicted in FigureS29-F. Finally, we observe that the attractive interaction established by Wat13 with the chromophore is as strong as those formed by Thr203 and His148 and is clearly ascribable to a hydrogen bond (see again Figure S29-A). This water subunit is also characterized by non-marginal weak interactions, probably related to a weak/non-conventional hydrogen bond with pHBDI (Figure S29-A). For the sake of completeness, it is worth mentioning that also Glu222 establishes a non-negligible (although weaker) attractive interaction with the chromophore, even if, for this residue, the contribution due to weak noncovalent contacts is slightly more significant (see Table S2 and Figure S29-G).

To conclude the discussion of the results reported in Figure 10B, it is also important to observe that pHBDI establishes purely weak noncovalent interactions (mainly van der Waals contacts) with four residues: Val150, Thr62 and Gln69. Among them, Thr62 is the most strongly interacting, as also indicated by the larger integration domain reported in Figure-S29-H (in particular, compare Figures S29-H, S29-I and S29-J).

Although the results shown above are quite promising, it is worth noting that there is room for improvement (especially from the quantitative point of view). In fact, also in the current version of the QM/ELMO approach, the extremely localized molecular orbitals that describe the ELMO subsystem remain frozen after the transfer and do not adapt to the presence of the fully quantum mechanical ground state electron density of the QM region. As already mentioned in Subsection 3.1, a possible strategy to improve the results could consist in allowing the transferred ELMOs to relax in response to the environment (in this case to the electron distribution of the QM subunit). In this way we will probably obtain better electron densities for the ELMO regions, thus improving the quality of the quantitative integral NCI analyses based on QM/ELMO computations.

4. CONCLUSIONS AND PERSPECTIVES

In this work, we have presented an extension of the integral NCI-index approach with the final goal of extracting quantitative information from NCI analyses of noncovalent interaction networks also when one uses electron densities of quantum mechanical nature. The only prerequisite is to have electron distributions that, in good approximation, can be partitioned into distinct contributions corresponding to the interacting fragments. For this reason, the new integral NCI-index strategy has been coupled with methods based on extremely localized molecular orbitals: i) the technique

based on the simple transfer of ELMOs, and ii) the more recent multiscale QM/ELMO embedding approach.

From the preliminary validation tests that used the high-level S66 dimer interaction energies as reference values, it clearly emerged that the novel ELMO-based NCI integral technique gives better correlations compared to the original NCI integral strategy relying on promolecular electron densities. It was also interestingly observed that the values of the NCI-ELMO integrals remain more stable with respect to the variation of the integral parameters. It was proved that this is related to the invariance of the noncovalent interaction region over which the NCI integration is performed. This is due to the greater chemical reliability of the ELMO description compared to the promolecular one.

The new NCI-ELMO integral approach was then exploited to characterize the interactions of each subunit of a ligand in a protein-ligand complex. The method gave very reasonable results in agreement with findings of a previous investigation, also allowing to easily distinguish between strong hydrogen bonds and weak noncovalent contacts (i.e., weak/non-conventional hydrogen bonds and van der Waals interactions). The technique was afterwards used to monitor the variation of the interaction energy between a protein and a ligand along a molecular dynamics-like trajectory (in our case, a binding trajectory). We have shown that the obtained energy profile is in optimal agreement with the corresponding one based on the use of a molecular mechanics force field, thus paving the way to the possibility of easily and quickly extracting energetic information of quantum mechanical level from molecular dynamics simulations. Finally, we have seen that the integral NCI-index strategy can be also profitably interfaced with the more recent QM/ELMO embedding technique to perform more reliable quantitative NCI analyses, especially when the examined ligand needs to be

treated through a traditional quantum chemistry technique. The approach has been applied to a protein-ligand complex and allowed us to easily quantify and reveal the nature of the interactions occurring between the whole ligand and the surrounding protein residues in the protein binding pocket. Therefore, when necessary, this version of the NCI integral method could become very useful to carry out more detailed NCI analyses of key configurations or docking poses extracted from molecular dynamics simulations or docking calculations, respectively.

All the obtained results are quite promising, which makes us envisage future applications of the proposed (QM/)ELMO-based NCI integral strategy. In addition to monitoring the evolution of noncovalent interactions along molecular dynamics trajectories, we believe that the novel techniques could be successfully used to investigate complicated NCI-networks (e.g., in protein-protein complexes) or in conjunction with docking computations and high-throughput virtual screenings for the rational design of new drugs.

Finally, as already mentioned in the discussion of the obtained results, it is worth pointing out that, notwithstanding the encouraging results obtained so far, the NCI-ELMO integral methods also have room for important algorithmic and methodological improvements. First, a more automatized and direct coupling of the ELMO libraries and of the QM/ELMO technique with the *NCIPLOT4* software would be beneficial to make the analyses described in this work more user friendly and available to a larger number of researchers. Moreover, in the near future the introduction of polarizable ELMO libraries will become crucial to make the transferred extremely localized molecular orbitals adapt to the environment in the target system and, consequently, to carry out more quantum mechanically rigorous (QM/)ELMO-based NCI integral analyses.

ASSOCIATED CONTENT

Supporting Information. Details about the QM/ELMO and QM/ELMO/MM methods. Figures S1-S4 depicting heatmaps that show the variation of the correlation coefficient as a function of the n and γ_{ref} parameters when promolecular and ELMO electron densities (basis sets: 6-31G, 6-311G, 6-311G(d,p) and cc-pVDZ) are used. Figures S5-S8 showing the variations of the integration region Ω_{NCI} as a function of the γ_{ref} parameter when promolecular and ELMO electron densities (basis sets: 6-31G, 6-311G, 6-311G(d,p) and cc-pVDZ) are used. Figures S9-S12 depicting the variations of the 2D NCI plots as a function of the γ_{ref} parameter when promolecular and ELMO electron densities (basis sets: 6-31G, 6-311G, 6-311G(d,p) and cc-pVDZ) are used. Figures S13-S16 showing heatmaps that report the variation of the correlation coefficient as a function of the n and γ_{ref} parameters and zoomed over the maximum regions when promolecular and ELMO electron densities (basis sets: 6-31G, 6-311G, 6-311G(d,p) and cc-pVDZ) are used. Figures S17-S27 reporting the 2D NCI plots obtained for the sixty-six dimers of the S66 databank through the integral NCI approach at promolecular and ELMO (all basis sets / monomer and dimer approximations) levels with the corresponding optimal value of the parameter γ_{ref}. Figures S28 and S29 showing the integration domains associated with the intermolecular interactions occurring in the protein-ligand complexes analyzed through the NCI-ELMO and NCI-QM/ELMO methods. Tables S1 and S2 reporting the values of the strongly attractive (I_{att}) and weak (I_{weak}) contributions to the interactions detected in the investigated protein-ligand complexes through the performed NCI-ELMO and NCI-QM/ELMO analyses (hydrogen-acceptor distances for plausible hydrogen bonds are also given).

AUTHOR INFORMATION

Notes

The authors declare no competing financial interests.

ACKNOWLEDGEMNTS

A.G., E.K.W. and G.M. gratefully acknowledge the French Research Agency (ANR) for the financial support of the Young Investigator Project *QuMacroRef* (Grant No. ANR-17-CE29-0005-01), the High-Performance Computing Center *EXPLOR* of the University of Lorraine for providing computing time through the projects 2019CPMXX0966, 2019CPMXX0886 and 2019CPMXX1332, and Fabien Pascale for the set-up and maintenance of our local cluster, which was used to perform most of the calculations reported in this paper. Finally, A.G. and J.C.-G. also thank the GDR Solvate and the French Network of Theoretical Chemistry (Réseau Français de Chimie Théorique, RCTF) for financial support of the collaboration.

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. Chapter 2 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] Raucci, R.; Laine, E.; Carbone, A. Local Interaction Signal Analysis Predicts Protein-Protein Binding Affinity. *Structure* **2018**, *26*, 905-915.

[6] Lu, W.; Cham, M. C. W.; Zhu, N.; Che, C.-M.; He, Z.; Wong, K.-Y. Structural Basis for Vapoluminescent Organoplatinum Materials Derived from Noncovalent Interactions as Recognition Components. *Chem. - Eur. J.* **2003**, *9*, 6155-6166.

[7] 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.

[8] 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.

[9] Ponder, J. W.; Case, D. A. Force Fields for Protein Simulations. *Adv. Protein Chem.* **2003**, *66*, 27-85.

[10] 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.

[11] Szalewicz, K. Symmetry-Adapted Perturbation Theory of Intermolecular Forces. *WIREs Comput. Mol. Sci.* **2012**, *2*, 254-272.

[12] Boys, S. F.; Bernardi, F. The Calculation of Small Molecular Interactions by the Differences of Separate Total Energies. Some Procedures with Reduced Errors. *Mol. Phys.* **1970**, *19*, 553-566.

[13] 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.

[14] Zhao, L.; von Hopffgarten, M.; Andrada, D. M.; Frenking, G. Energy Decomposition Analysis. *WIREs Comput. Mol. Sci.* **2018**, *8*, e1345.

[15] Bader, R. F. W. Atoms in Molecules: A Quantum Theory; Oxford University Press: Oxford, U.K., 1990.

[16] 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.

[17] Maxwell, P.; Pendás, Á. 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.

[18] 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.

[19] Johnson, E. R.; Keinan, S.; Mori-Sánchez, P.; Contreras-García, J.; Cohen, A. J.; Yang, W. Revealing Noncovalent Interactions. *J. Am. Chem. Soc.* **2010**, *132*, 6498-6506.

[20] Contreras-García, J.; Johnson, E. R.; Keinan, S.; Chaudret, R.; Piquemal, J.-P.; Beratan, D. N.; Yang, W. NCIPLOT: a Program for Plotting Noncovalent Interaction Regions. *J. Chem. Theory Comput.* **2011**, *7*, 625-632.

[21] 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.

[22] Saleh, G.; Gatti, C.; Lo Presti, L. Energetics and Non-Covalent Interactions from Electron and Energy Density Distributions. *Comput. Theor. Chem.* **2015**, *1053*, 53-59.

[23] Peccati, F. NCIPLOT4 Guide for Biomolecules: An Analysis Tool for Noncovalent Interactions. *J. Chem. Inf. Model.* **2020**, *60*, 6-10.

[24] Boto, R. A.; Peccati, F.; Laplaza, R.; Quan, C.; Carbone, A.; Piquemal, J.-P.; Maday, Y.; Contreras-García, J. NCIPLOT4: Fast, Robust, and Quantitative Analysis of Noncovalent Interactions. *J. Chem. Theory Comput.* **2020**, *16*, 4150-4158.

[25] 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.

[26] 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. Phys. Chem. Lett.* **2019**, *10*, 1482-1488.

[27] 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.

[28] 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.

[29] Ponce-Vargas, M.; Lefebvre, C.; Boisson, J.-C.; Hénon, E. Atomic Decomposition Scheme of Noncovalent Interactions Applied to Host−Guest Assemblies. *J. Chem. Inf. Model.* **2019**, *60*, 268-278.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] Stoll, H.; Wagenblast, G.; Preuss, H. On the Use of Local Basis Sets for Localized Molecular Orbitals. *Theor. Chim. Acta* **1980**, *57*, 169−178.

[35] 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.

[36] Sironi, M.; Genoni, A.; Civera, M.; Pieraccini, S.; Ghitti, M. Extremely Localized Molecular Orbitals: Theory and Applications. *Theor. Chem. Acc.* **2007**, *117*, 685-698.

[37] 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.

[38] 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.

[39] 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.

[40] Genoni, A.; Merz. K. M., Jr.; Sironi, M. A Hylleras functional based perturbative technique to relax extremely localized molecular orbitals. *J. Chem. Phys.* **2008**, *129*, 054101.

[41] 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.

[42] 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.

[43] Wieduwilt, E. K.; Boisson, J.-C.; Terraneo, G.; Hénon, E.; Genoni, A. A Step toward the Quantification of Noncovalent Interactions in Large Biological Systems: The Independent Gradient Model-Extremely Localized Molecular Orbital Approach. *J. Chem. Inf. Model.* **2021**, *61*, 795-809.

[44] 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.

[45] 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.

[46] 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.

[47] Wieduwilt, E. K.; Macetti, G.; Genoni, A. Climbing Jacob's Ladder of Structural Refinement: Introduction of a Localized Molecular Orbital-Based Embedding for Accurate X-ray Determinations of Hydrogen Atom Positions. *J. Phys. Chem. Lett.* **2021**, *12*, 463-471.

[48] Macetti, G.; Wieduwilt, E. K.; Genoni, A. QM/ELMO: a Multi-Purpose Fully Quantum Mechanical Embedding Scheme Based on Extremely Localized Molecular Orbitals. *J. Phys. Chem. A* **2021**, *125*, 2709-2726.

[49] Macetti, G.; Genoni, A. Quantum Mechanics / Extremely Localized Molecular Orbital Embedding Technique: Theoretical Foundations and Further Validation. *Adv. Quantum Chem.* **2021**, *83*, 269-285.

[50] Macetti, G.; Genoni, A. Initial Maximum Overlap Method for Large Systems by the Quantum Mechanics/Extremely Localized Molecular Orbital Embedding Technique. *J. Chem. Theory Comput.* **2021**, *17*, 4169-4182.

[51] Řezáč, J.; Riley, K. E.; Hobza, P. S66: A well-balanced database of benchmark interaction energies relevant to biomolecular structures. *J. Chem. Theory Comput.* **2011**, *7*, 2427-2438.

[52] McWeeny, R. The Density Matrix in Many-Electron Quantum Mechanics. 1. Generalized Product Functions. Factorization and Physical Interpretation of the Density Matrices. *Proc. R. Soc. London Ser. A* **1959**, *253*, 242-259.

[53] Adams, W. H. On the Solution of the Hartree-Fock Equations in Terms of Localized Orbitals. *J. Chem. Phys.* **1961**, *34*, 89-102.

[54] Huzinaga, S.; Cantu, A. A. Theory of Separability of Many-Electron Systems. *J. Chem. Phys.* **1971**, *55*, 5543-5549.

[55] Gilbert, T. L. Multiconfiguration self-consistent-field theory for localized orbitals. II. Overlap constraints, Lagrangian multipliers, and the screened interaction field. *J. Chem. Phys.* **1974** *60*, 3835-3844.

[56] Matsuoka, O. Expansion methods for Adams-Gilbert equations. I. Modified Adams-Gilbert equation and common and fluctuating basis sets. *J. Chem. Phys.* **1977**, *66*, 1245-1254.

[57] Smits, G. F.; Altona, C. Calculation and properties of non-orthogonal, strictly local molecular-orbitals. *Theor. Chim. Acta* **1985**, *67*, 461-475.

[58] Francisco, E.; Martín Pendás, A.; Adams, W. H. Generalized Huzinaga Building-Block Equations for Nonorthogonal Electronic Groups: Relation to the Adams-Gilbert Theory. *J. Chem. Phys.* **1992**, *97*, 6504-6508.

[59] Ordejón, P.; Drabold, D.; Grumbacj, M.; Martin, R. Unconstrained Minimization Approach for Electronic Computations that Scales Linearly with System Size. *Phys. Rev. B* **1993**, *48*, 14646-14649.

[60] Couty, M.; Bayse, C. A.; Hall, M. B. Extremely localized molecular orbitals (ELMO): a non-orthogonal Hartree-Fock method. *Theor. Chim. Acta* **1997**, *97*, 96-109.

[61] Wieduwilt, E.K.; Macetti, G.; Scatena, R.; Macchi, P.; Genoni, A. Extending Libraries of Extremely Localized Molecular Orbitals to Metal Organic Frameworks: A Preliminary Investigation, *Crystals* **2021**, *11*, 207.

[62] 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.

[63] Malaspina, L. A.; Wieduwilt, E. K.; Bergmann, J.; Kleemiss, F.; Meyer, B.; Ruiz-López, 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.

[64] Grabowsky, S.; Genoni, A.; Bürgi, H.-B. Quantum Crystallography. *Chem. Sci.* **2017**, *8*, 4159-4176.

[65] 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. J.; Matta, C. F.; Merz, K. M., Jr.; 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.

[66] Massa, L.; Matta, C. F. Quantum Crystallography: A perspective. *J. Comput. Chem.* **2018**, *39*, 1021-1028.

[67] Tsirelson, V. Early days of quantum crystallography: A personal account. *J. Comput. Chem.* **2018**, *39*, 1029-1037.

[68] Genoni, A.; Macchi, P. Quantum Crystallography in the Last Decade: Developments and Outlooks. *Crystals* **2020**, *10*, 473.

[69] 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 *21st Century Challenges in Chemical Crystallography II – Structural Correlations and Data Interpretation*. *Structure and Bonding*, vol. 186, pp 65-144; Mingos, D. M. P., Rathby, P., Eds.; Springer: Berlin, Heidelberg; DOI: 10.1007/430_2020_62.

[70] Macchi, P. The connubium between crystallography and quantum mechanics. *Crystallogr. Rev.* **2020**, *26*, 209-268.

[71] Macetti, G; Genoni, A. Three-Layer Multiscale Approach Based on Extremely Localized Molecular Orbitals to Investigate Enzyme Reactions. *J. Phys. Chem. A* **2021**, *125*, 6013-6027.

[72] Guest, M. F.; Bush, I. J.; van Dam, H. J. J.; Sherwood, P.; Thomas, J. M. H.; van Lenthe, J. H.; Havenith, R. W. A.; Kendrick, J. The GAMESS-UK Electronic Structure Package: Algorithms, Developments and Applications. *Mol. Phys.* **2005**, *103*, 719−747.

[73] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-F) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

[74] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revision D.01; Gaussian, Inc., Wallingford, CT, USA, 2009.

[75] Amacher, J. F.; Cushing, P. R.; Bahl, C. D.; Beck, T.; Madden, D. R. Stereochemical Determinants of C-terminal specificity in PDZ Peptide-binding Domains. *J. Biol. Chem.* **2013**, *288*, 5114−5126.

[76] Case, D. A.; Betz. R. M.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Luchko, T.; Luo, R.; Madej, B.;

Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. L.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao. L.; Kollman. P.A. *AMBER 2016*; University of California San Francisco: San Francisco, CA, USA, 2016.

[77] Boto, R. A.; Contreras-García, J.; Calatayud, M. The role of dispersion forces in metal-supported self-assembled monolayers. *Comput. Theor. Chem.* **2015**, *1053*, 322-327.

[78] Tonikian, R.; Zhang, Y.; Sazinsky, S. L.; Currell, B.; Yeh, J. H.; Reva, B.; Held, H. A.; Appleton, B. A.; Evangelista, M.; Wu, Y.; Xin, X.; Chan, A. C.; Seshagiri, S.; Lasky, L. A.; Sander, C.; Boone, C.; Bader, G. D.; Sidhu, S. S. A specificity map for the PDZ domain family. *PLoS Biol.* **2008**, *6*, e239.

[79] Borrelli, K. W.; Vitalis, A.; Alcantara, R.; Guallar, V. PELE: Protein Energy Landscape Exploration. A Novel Monte Carlo Based Technique. *J. Chem. Theory Comput.* **2005**, *1*, 1304−1311.

[80] Madadkar-Sobhani, A.; Guallar, V. PELE Web Server: Atomistic Atudy of Biomolecular Systems at your Fingertips. *Nucleic Acids Res.* **2013**, *41*, W322−W328.

[81] Kaila, V. R. I.; Send, R.; Sundholm, D. Electrostatic spectral tuning mechanism of the green fluorescent protein. *Phys. Chem. Chem. Phys.* **2013**, *15*, 4491−4495.

[82] Wen, X.; Graham, D. S.; Chulhai, D. V.; Goodpaster, J. D. Absolutely Localized Projection-based Embedding for Excited-States. *J. Chem. Theory Comput.* **2020**, *16*, 385−398.

[83] Wallace, A.C.; Laskowski, R. A.; Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng. Des. Sel.* **1995**, *8*, 127–134.

[84] Laskowski, R. A.; Swindells, M. B. LigPlot+: Multiple Ligand-Protein Interaction Diagrams for Drug Discovery, *J. Chem. Inf. Model.* **2011**, *51*, 2778–2786.

[85] McDonald, I. K.; Thornton, J. M. Satisfying hydrogen bonding potential in proteins, *J. Mol. Biol.* **1994**, *238*, 777–793.