



**HAL**  
open science

# NCI-ELMO: A New Method To Quickly and Accurately Detect Noncovalent Interactions in Biosystems

David Arias-Olivares, Erna K. Wieduwilt, Julia Contreras-García, Alessandro Genoni

► **To cite this version:**

David Arias-Olivares, Erna K. Wieduwilt, Julia Contreras-García, Alessandro Genoni. NCI-ELMO: A New Method To Quickly and Accurately Detect Noncovalent Interactions in Biosystems. *Journal of Chemical Theory and Computation*, 2019, 10.1021/acs.jctc.9b00658 . hal-02324015

**HAL Id: hal-02324015**

**<https://hal.univ-lorraine.fr/hal-02324015>**

Submitted on 10 May 2020

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

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

This document is the Accepted Manuscript version of a Published Work that appeared in final form in the *Journal of Chemical Theory and Computation*, copyright © American Chemical Society after peer review and technical editing by the publisher. To access the final edited and published work see <https://pubs.acs.org/doi/10.1021/acs.jctc.9b00658>.

# NCI-ELMO: a New Method to Quickly and Accurately Detect Non-Covalent Interactions in Biosystems

David Arias-Olivares,<sup>†,‡</sup> Erna K. Wieduwilt,<sup>¶</sup> Julia Contreras-García,<sup>‡</sup> and  
Alessandro Genoni<sup>\*,¶</sup>

<sup>†</sup>*Doctorado en Fisicoquímica Molecular, Facultad de Ciencias Exactas, Universidad Andrés Bello, Ave. Republica 275, Santiago, Chile*

<sup>‡</sup>*Sorbonne Université, Laboratoire de Chimie Thorique, UMR CNRS 7616, 4 Place Jussieu, F-75005 Paris, France*

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

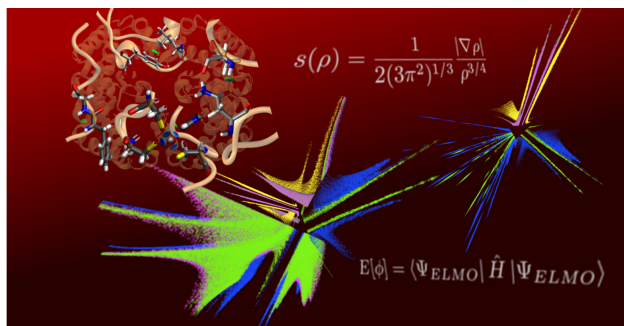
E-mail: Alessandro.Genoni@univ-lorraine.fr

Phone: +33 (0)3 72 74 91 70. Fax: +33 (0)3 72 74 91 87

## Abstract

Accurate determinations of non-covalent interactions in biological systems are fundamental to rationalize the structure and to get insights into the functions and the dynamics of macromolecules. Here we propose a new tool for the efficient identification of non-covalent interactions in proteins. The non-covalent interaction (NCI) method, a well-established strategy to detect non-covalent interactions in chemical systems, is coupled with the libraries of extremely localized molecular orbitals (ELMOs), which allow instantaneous reconstruction of quantum mechanically rigorous electron distributions of polypeptides and proteins. Test calculations performed on different interactions show that the new NCI-ELMO strategy always outperforms the original NCI method based on the promolecular approximation, while it is in close agreement with original NCI analyses based on fully quantum mechanical calculations. The new technique allows to unravel the network of interactions in biological systems and to rapidly monitor their evolutions with time, with possible consequences on the design of new drugs.

## TOC Graphic



# 1 Introduction

Non-covalent interactions play a fundamental role in the definition of the three-dimensional (3D) structure of macromolecules. Furthermore, they are also crucial in many biological processes in which biomolecules bind to small ligands<sup>1,2</sup> (e.g., inhibitors or cofactors) or to other large biological systems (e.g., protein-protein interactions<sup>3</sup>). Hence, a complete and accurate description of the non-covalent network is essential to shed light on the functions of biological molecules and on the dynamics of biological processes, with also a possible impact in the design of new and more efficient drugs. As an example, the discovery of new directional interactions mediated by halogen  $\sigma$ -holes has been crucial in the design and engineering of new drugs and in their crystallization.<sup>4</sup>

Due to their chemical and biological relevance, a detailed understanding of the physical origin and scope of non-covalent interactions has become a major goal of physical organic chemistry. To this purpose different strategies have been developed.

In this context, one of the most common methods<sup>5</sup> to quickly map and classify non-covalent interactions as van der Waals interactions, hydrogen bonds or steric clashes is based on the evaluation of pairwise distances between atoms computed through van der Waals radii.

Other more theoretically sound strategies are the Quantum Chemical Topology (QCT) methods, namely strategies based on electron and kinetic energy densities of the investigated systems. Among them, a prominent role is obviously occupied by the well known Quantum Theory of Atoms in Molecules (QTAIM).<sup>6</sup> In fact, this technique can also be used to find and characterize weak interactions, exploiting the fact that a critical point in the electron density naturally occurs in presence of two interacting atoms. More interestingly, strong correlations have been found between interaction energies and topological properties (e.g., electron density, Laplacian of the electron density and kinetic energy density) at the (non-covalent) bond critical points, especially for hydrogen bonds and van der Waals interactions.<sup>7-9</sup> Within the same framework, we can also mention other investigations that used the Electron Localization Function (ELF)<sup>10,11</sup> to approximately evaluate the strength

of hydrogen bonds.<sup>12,13</sup>

However, nowadays one of the most efficient strategies in theoretical chemistry and quantum crystallography<sup>14-17</sup> to detect all the (inter- and intra-molecular) non-covalent interaction existing in a chemical system is undoubtedly the Non-Covalent Interactions (NCI) method.<sup>18,19</sup> Fully belonging to the family of the QCT techniques, this approach relies on the analysis of electron density and of its reduced density gradient (RDG).

In the current version of the NCI technique, the underlying electron density can be obtained in two different ways: (i) through a fully quantum mechanical calculation on the molecule under exam; (ii) by means of the promolecular density approximation (also known as “independent atom model” (IAM) approximation). In other words, in the latter case, instead of using an electron density obtained through a traditional quantum chemistry computation, the electron density is given as the sum of independent and spherical atomic electron distributions, which were previously tabulated as a database of atomic shell exponentials. Calculations showed that the promolecular approximation provides results that are qualitatively very similar to those obtained through the application of fully quantum chemical methods. In this regard, in a recent work,<sup>20</sup> Contreras-García *et al.* pointed out that, even in presence of dispersion, NCI pictures resulting from IAM or Hartree-Fock calculations are sufficient to qualitatively detect the presence of the interactions when the geometry is fixed. However, the authors also stress that, depending on the adopted method, quantitative differences in the NCI results can be observed for particular interactions (e.g., steric repulsions) and especially when one evaluates the relative stability of different conformers with geometries resulting from optimizations that include correlation and dispersion effects.

While for small and medium-size systems NCI analyses are usually based on electron densities from fully quantum mechanical calculations, for macromolecules the promolecular approximation must be used in order to reduce the computational cost. Although the obtained results are qualitatively similar from the visual point of view, the promolecular approach is known to provide incorrect densities which lack relaxation and, on the basis of

what mentioned above, this might have repercussions also on the NCI analysis of the relative stabilities of representative polypeptides/proteins conformers obtained from Molecular Dynamics simulations exploiting force fields accounting for dispersion.

In this contribution, to overcome the limitations of the NCI method when applied to large molecules, we propose a new version of the technique that, by exploiting the computational advantages offered by the Extremely Localized Molecular Orbitals (ELMOs),<sup>21-27</sup> can be applied to unravel the non-covalent interactions networks of very large biological systems.

ELMOs are orbitals strictly localized on small molecular units (e.g., atoms, bonds and functional groups) and, for this reason, easily transferable from one molecule to another, provided that the subunits on which they are localized have the same chemical environment. Due to their intrinsic transferability,<sup>22,28-34</sup> databanks of ELMOs covering all the fragments of the twenty natural amino-acids have been recently constructed.<sup>33-35</sup> It has been shown that they allow fast reconstructions of quantum mechanically rigorous electron densities and electrostatic potentials of systems ranging from small polypeptides to very large proteins.<sup>35</sup> Therefore, we have decided to couple the ELMO databases with the NCI method, giving rise to the novel NCI-ELMO approach. As we will describe in this paper, the new technique enables to obtain accurate NCI features in biological molecules for a large variety of interactions, but at a very low computational cost.

The paper is organized as follows. We will initially analyse and discuss the outcomes of some benchmark tests on polypeptides aiming at comparing the results of NCI calculations exploiting promolecular, ELMO and fully quantum mechanical electron densities (Section 2). We will afterwards consider different non-covalent interactions in proteins which are interesting from the chemical and/or biological points of view and which, due to the large computational cost, cannot be characterised through traditional NCI analyses based on usual quantum chemical calculations (Section 3). In particular, in all the examined cases, we will carefully compare the NCI-ELMO and NCI-promolecular descriptions by showing how the former outperforms the latter in all the considered situations. Finally, we will draw general

conclusions and we will outline future perspectives to follow up this strategy (Section 4). For the sake of completeness, in the Appendix we will also briefly introduce the theoretical foundations of the NCI and ELMO techniques and we will provide a concise description of the ELMO libraries.

## 2 Benchmark calculations

To evaluate the capabilities of the new NCI-ELMO approach, we compared the results of NCI computations based on promolecular, DFT and *ELMOdb* electron densities. In particular, the comparisons were conducted on three polypeptides characterized by different types of non-covalent interactions. We have chosen to use polypeptides in our tests because they represent a good compromise: they are medium-size systems which show non-covalent interactions similar to those normally found in large proteins, while their fully quantum mechanical calculations (to be used as benchmarks) are still feasible.

As first test case we considered lactoferrampin, an antimicrobial peptide characterized by antibacterial and candidacidal activities, for which we have selected the first NMR molecular geometry in PDB file 2MD3.<sup>36</sup> The considered geometry was afterwards used to perform calculations at DFT level (B3LYP functional) and to transfer the ELMOs from the considered libraries. All the basis-sets currently available for the ELMO databanks were taken into account to also evaluate the possible dependence of the NCI results on the size of the adopted set of basis functions. Furthermore, the total promolecular density was calculated for the chosen geometry as sum of tabulated atomic densities within the *NCIplot* program.

We particularly focused on a classical weak hydrogen bond interaction occurring between residues Phe2 and Gly3, which is depicted through a typical green RDG isosurface of the NCI method shown in Figure 1A. The actual comparison was performed by analyzing the traditional 2D NCI plots of the reduced density gradient  $s$  in function of the electron density multiplied by the sign of the second Hessian eigenvalue ( $sign(\lambda_2) \rho$ ; hereinafter also indicated



as “signed” electron density). In Figure 1B we have simultaneously reported the results obtained exploiting the promolecular, B3LYP/6-31G and *ELMOdb*/6-31G electron densities, while, in analogous way, in Figure 1C we have overlapped the 2D graphs resulting from the use of promolecular, B3LYP/cc-pVDZ and *ELMOdb*/cc-pVDZ charge distributions. In Figures 1B and 1C it is easy to observe that all the performed NCI calculations provided two symmetric peaks, which generally are the “NCI fingerprints” for weak intermolecular interactions. Furthermore, while the promolecular density provides reduced density gradient peaks that are clearly shifted from the reference B3LYP ones, the electron distributions obtained through the simple transfer of ELMOs give NCI spikes practically indistinguishable from those associated with the DFT calculations.

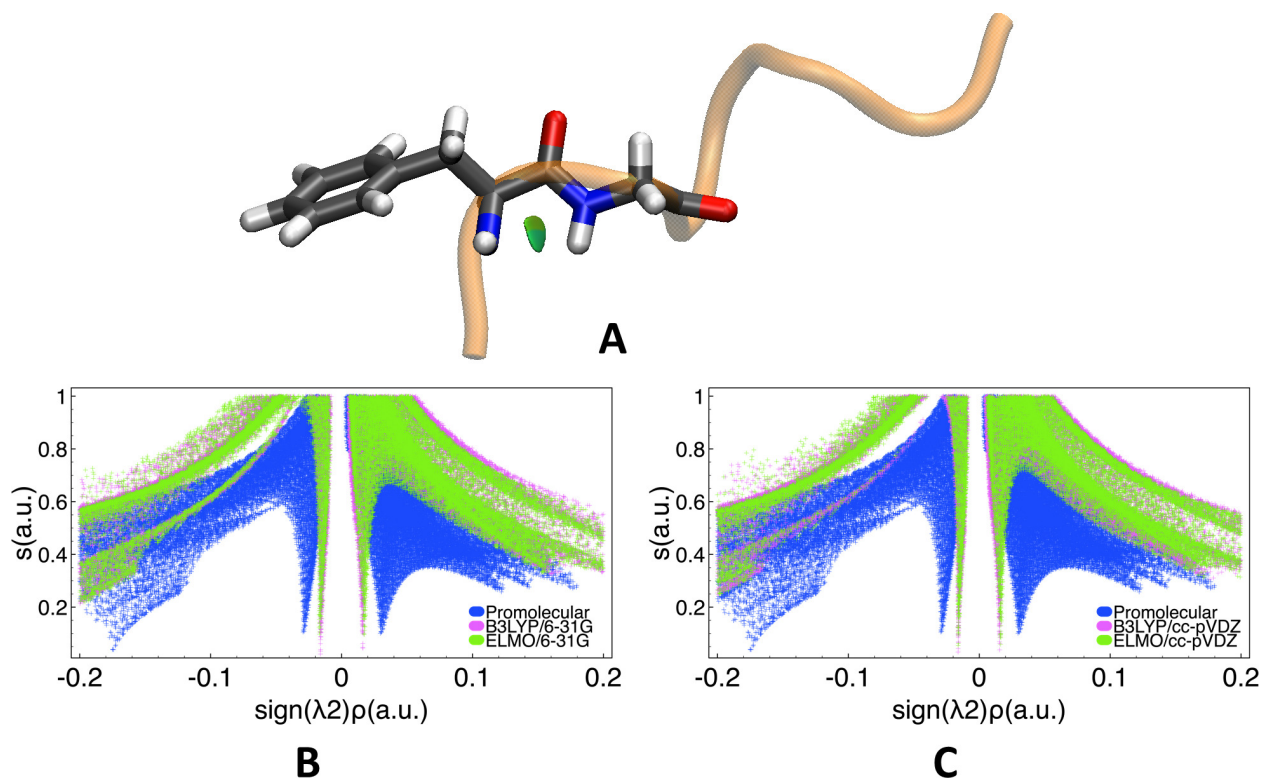


Figure 1: Weak hydrogen bond between Phe2 and Gly3 in lactoferrampin (PDB code: 2MD3): (A) RDG isosurface ( $s=0.6$  a.u., color scale:  $-0.03$  a.u.  $< \text{sign}(\lambda_2)\rho < 0.03$  a.u.) obtained at NCI-ELMO level and showing the interaction under examination; 2D RDG plots obtained at promolecular-NCI, NCI-B3LYP and NCI-ELMO levels for basis-sets (B) 6-31G and (C) cc-pVDZ

To better highlight the dependence of the NCI results on the adopted basis-set, in Figures

2A and 2B, we have simultaneously reported the 2D RDG plots obtained from the NCI calculations based on the DFT and *ELMOdb* electron distributions, respectively. It is easy to observe that the peaks completely overlap and all the basis-sets provide similar results.

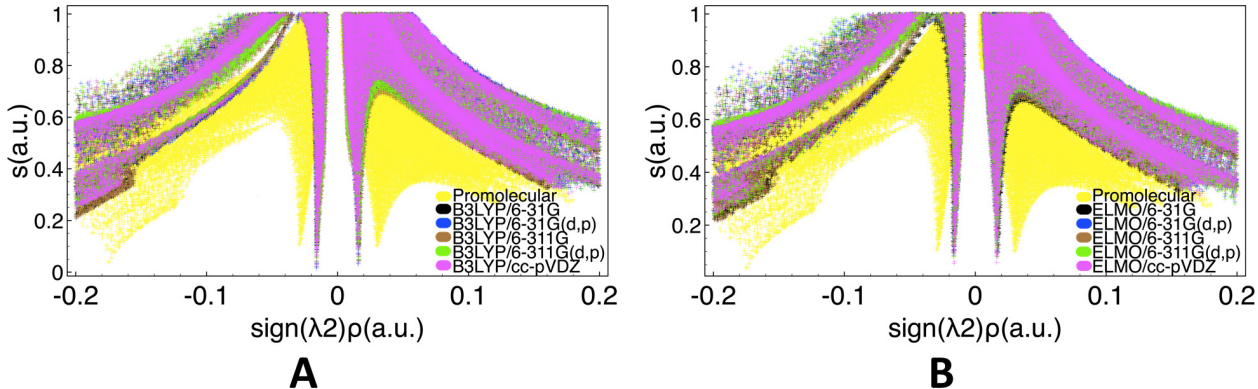


Figure 2: Weak hydrogen bond between Phe2 and Gly3 in lactoferrampin (PDB code: 2MD3): 2D RDG plots obtained at (A) NCI-B3LYP and (B) NCI-ELMO levels with all the considered basis-sets.

For our second benchmark-test, we considered a non-covalent interaction present in the high-resolution X-ray crystallographic structure of Leu-enkephalin.<sup>37</sup> In particular we considered the strong hydrogen bond between Tyr2 and Phe5, which is clearly highlighted by the RDG isosurface in Figure 3A. Also in this case we compared the 2D reduced density gradient plots obtained at promolecular, B3LYP and ELMO levels. As done for the interaction analyzed above, in Figure 3B and 3C we have reported the results for basis-sets 6-31G and cc-pVDZ, respectively. In this case, all the NCI calculations provided a single peak in the negative region of the “signed” electron density, as it is usually observed for stabilizing hydrogen bond interactions. However, both in Figure 3B and in Figure 3C, we can see that the promolecular peak associated with the hydrogen bond interaction is clearly shifted with respect to the B3LYP one, which, on the contrary, clearly matches more closely the ELMO spike. In particular, we can easily see that in the negative domain, which is fundamental to identify stabilizing non-covalent interactions, the NCI analysis based on the promolecular density gives a peak that is “more negative”, indicating a more stabilizing interaction than the one detected at DFT level. Conversely, the ELMO approximation slightly underesti-

mates the stability of the interaction. Identical results were observed adopting the other three basis-sets (see Figures S5-S7 in the Supporting Information).

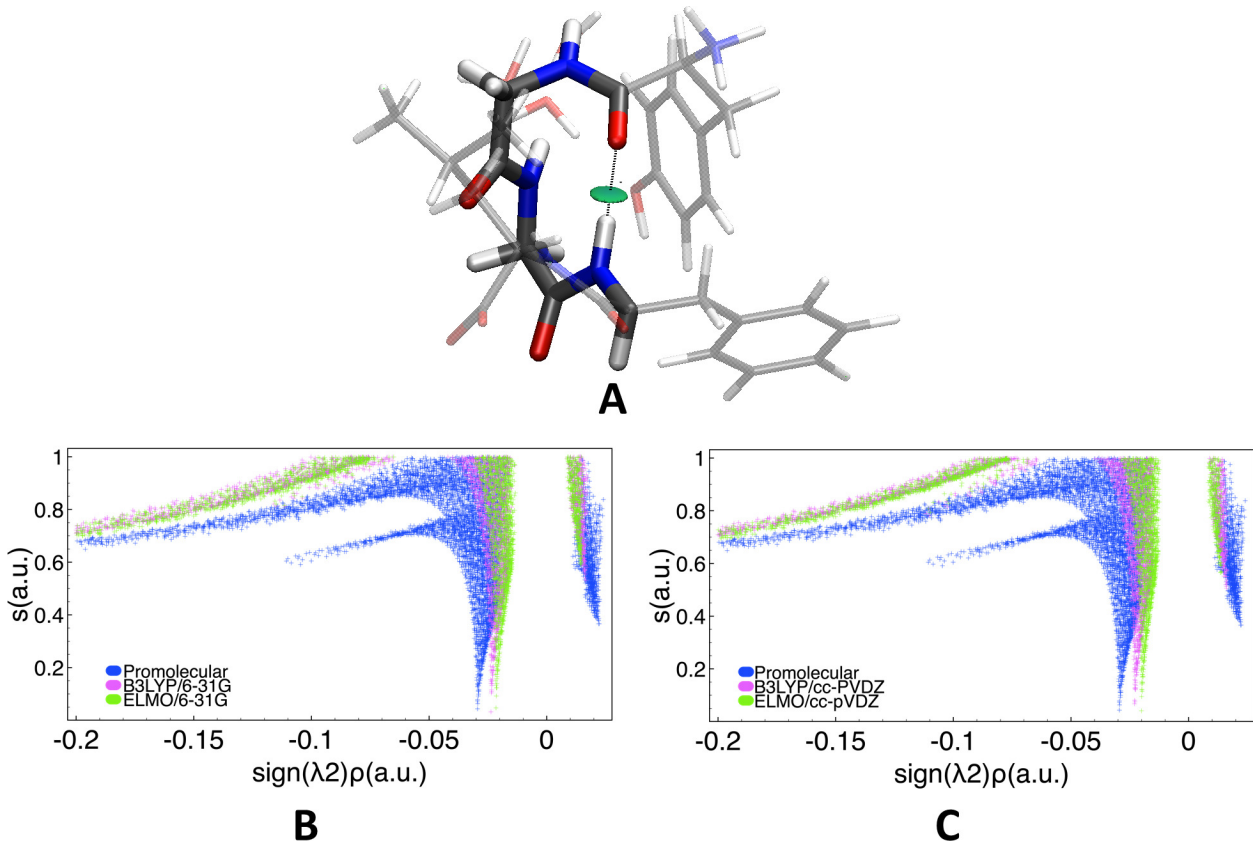


Figure 3: Strong hydrogen bond between Tyr2 and Phe5 in Leu-enkephalin<sup>37</sup>: (A) RDG isosurface ( $s=0.6$  a.u., color scale:  $-0.03$  a.u.  $< \text{sign}(\lambda_2)\rho < 0.03$  a.u.) obtained at NCI-ELMO level and showing the interaction under examination; 2D RDG plots obtained at promolecular-NCI, NCI-B3LYP and NCI-ELMO levels for basis-sets (B) 6-31G and (C) cc-pVDZ.

In Figure 4 we also report the dependence of the NCI computations on the used basis-set. In this case, the NCI-B3LYP and NCI-ELMO trends correspond (see Figure 4A and Figure 4B, respectively), with the NCI peak for the strong hydrogen bond that moves towards lower absolute values of the electron density as we consider larger basis-sets.

Finally, as third and final test calculation we considered an HIV Zinc fingerlike domain characterized by the presence of a metal center (Zinc) tetra-coordinated by four residues (particularly, three cysteine and one histidine residues). This example was chosen as a challenging case due to the presence of stronger interactions and of a charged metal center.

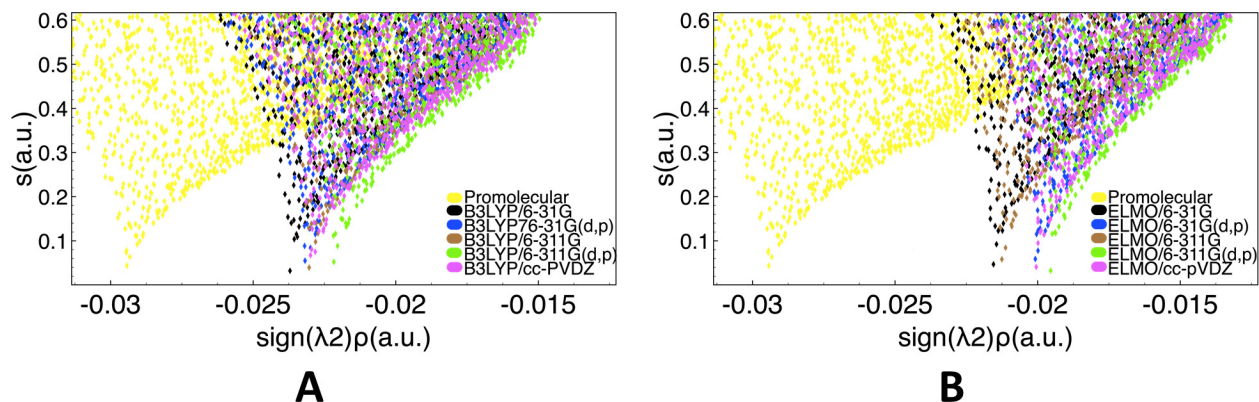


Figure 4: Strong hydrogen bond between Tyr2 and Phe5 in Leu-enkephalin<sup>37</sup>: 2D RDG plots obtained at (A) NCI-B3LYP and (B) NCI-ELMO levels with all the considered basis-sets.

The non-covalent interactions of Zinc with the surrounding residues (see Figure 5A) were indeed the focus of our test calculations, for which we considered the first NMR molecular geometry in PDB file 2ZNF.<sup>38</sup> As for the other polypeptides, the selected molecular geometry was used to perform calculations at DFT level, to transfer the ELMOs from the databanks and to construct the promolecular density. Unlike the tests conducted on lactoferrampin and Leu-enkephalin, the DFT calculations were performed using functional B3PW91, which is more suitable than B3LYP to study systems with metal centers. Furthermore, due to the presence of a metal, tailor-made ELMOs for Zinc were preliminary computed before the transfers. This was done by simply carrying out ELMO calculations on a model molecule mimicking the chemical environment of the metal in the polypeptide. Finally, it is also worth noting that, within the NCIPLOT program, the promolecular densities are generally constructed exploiting spherical densities of neutral atoms, but, in this specific case, for which the charge of the metal center is crucial, the promolecular density for the NCI analysis was obtained by considering the spherical atomic density associated with the  $\text{Zn}^{2+}$  cation.

Also in this case we plotted and compared simultaneously the 2D reduced density gradient plots associated with the electron densities computed at different levels of theory. In Figures 5B and 5C we have reported the results obtained for basis-sets 6-31G and cc-pVDZ, respectively. In both situations it is evident that the promolecular charge distribution does not allow to reproduce the reference spikes obtained at B3PW91 level. In contrast to the

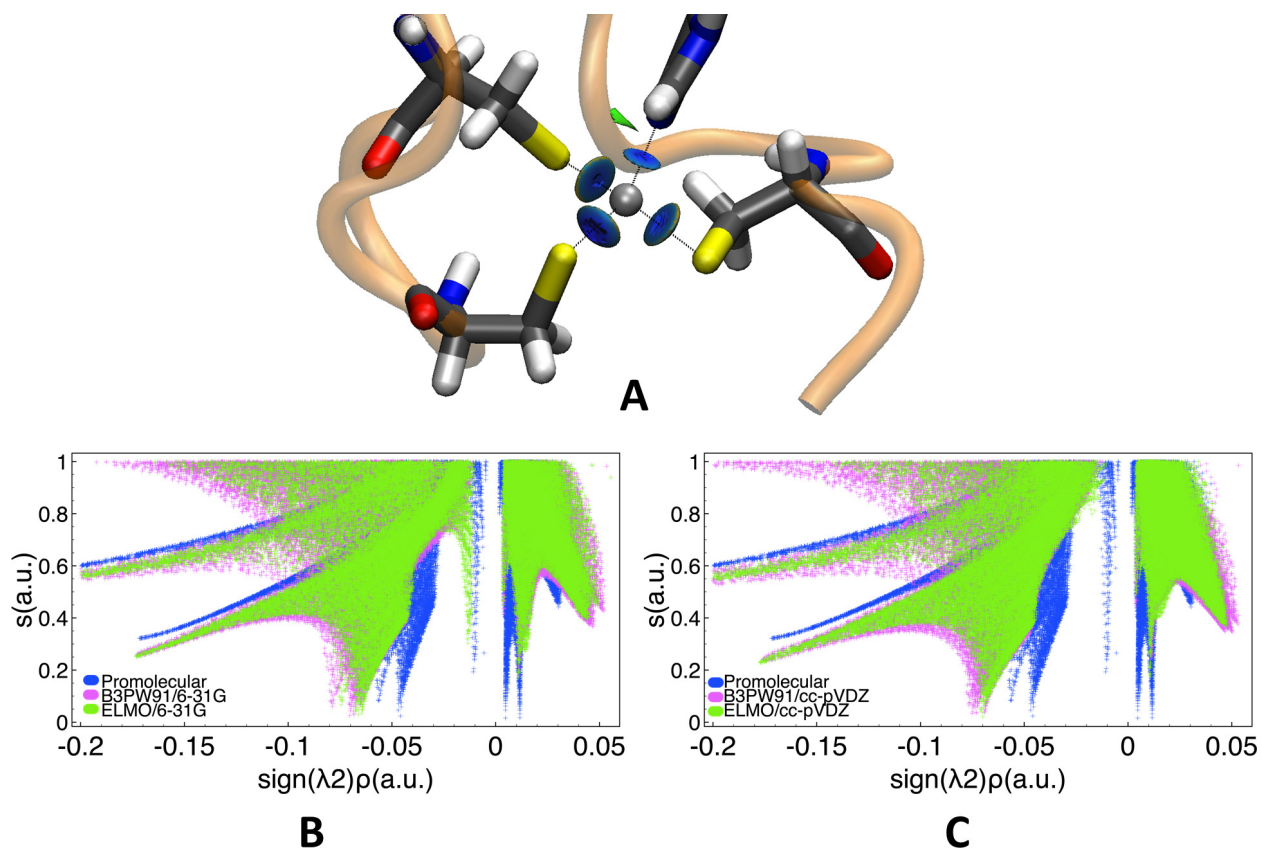


Figure 5: Interactions between the  $\text{Zn}^{2+}$  ion and the coordinating residues (Cys3, Cys6, Cys16 and His11) in an HIV Zinc fingerlike domain (PDB code: 2ZNF): (A) RDG isosurfaces ( $s=0.6$  a.u., color scale:  $-0.03$  a.u.  $< \text{sign}(\lambda_2)\rho < 0.03$  a.u.) obtained at NCI-ELMO level and showing the interaction under examination; 2D RDG plots obtained at promolecular-NCI, NCI-B3PW91 and NCI-ELMO levels for basis-sets (B) 6-31G and (C) cc-pVDZ.

results for the strong and weak hydrogen bonds, the NCI analysis based on the promolecular electron density gives peaks that are “more positive” compared to the DFT ones, thus revealing non-covalent interactions that are less stabilizing. Conversely, we can also clearly observe that there is a much better agreement between the B3PW91 and ELMO spikes, even if we do not observe the almost perfect overlap seen for the weak hydrogen bond interaction previously analyzed. In particular, we can observe that, when we use *ELMOdb* electron densities, the NCI analyses reveal non-covalent interactions that are slightly less stabilizing compared to the DFT situation. Analogous results have been observed for the other basis-sets (6-311G, 6-31G(d,p) and 6-311G(d,p); see Figures S8-S10 in the Supporting Information).

The preliminary test calculations on the three polypeptides have clearly shown that the new NCI-ELMO method is able to provide results very close to those obtained through NCI analyses that exploit traditional quantum chemical computations, independently of the adopted basis-set. Furthermore, and more importantly, it is also evident that the novel strategy outperforms the one based on the promolecular densities only at a slightly larger (and still completely acceptable) computational cost (See Table 1). Indeed, the improved performance comes at a little computational cost even for very large systems. For instance, for the investigated HIV Zinc fingerlike domain, the total computing time associated with the ELMO approach is of the order of minutes, whereas the SCF one would take hours or even days. On the basis of these results, we have thus decided to exploit the new method to detect and characterize different types of non-covalent interactions in quite large proteins, namely in systems for which fully quantum mechanical calculations are generally impractical or impossible. Also in those situations we have tried to highlight again the non-negligible differences existing between the promolecular-NCI and NCI-ELMO descriptions.

Table 1: Global CPU times (in the format hh:mm:ss.000) associated with all the benchmark NCI calculations. For the NCI-ELMO and NCI-DFT cases, the global CPU time is the sum of the time taken by the NCIPILOT program plus the time taken by the *ELMObd* or Gaussian09 calculations.

	Lactoferrampin	Leu-enkephalin	HIV Zinc fingerlike domain
Number of atoms	145	86	275
Number of electrons	522	326	1098
Promolecular-NCI	00:00:00.977	00:00:00.214	00:00:01.595
NCI-ELMO/6-31G	00:01:09.809	00:00:04.316	00:08:16.113
NCI-ELMO/6-311G	00:01:16.750	00:00:05.758	00:10:00.541
NCI-ELMO/6-31G(d,p)	00:01:43.015	00:00:05.980	00:10:41.436
NCI-ELMO/6-311G(d,p)	00:01:49.644	00:00:07.321	00:11:58.663
NCI-ELMO/cc-pVDZ	00:01:52.870	00:00:06.964	00:12:40.584
NCI-DFT/6-31G	02:29:19.071	00:31:31.273	04:46:49.380
NCI-DFT/6-311G	04:03:45.798	00:37:43.427	17:18:36.688
NCI-DFT/6-31G(d,p)	08:20:50.301	00:58:55.350	38:56:23.140
NCI-DFT/6-311G(d,p)	17:39:25.391	02:45:32.024	47:31:56.169
NCI-DFT/cc-pVDZ	20:53:03.176	01:55:34.030	51:36:05.193

## 3 Non-covalent interactions in proteins

### 3.1 Hydrogen bonds

Hydrogen bonds are fundamental interactions that significantly contribute to the determination of structure and function of biomolecules. They do not only play a crucial role in defining secondary, tertiary and quaternary structures of proteins and nucleic-acids, but they are also fundamental to establish the catalytic functions of enzymes. However, despite hydrogen bonding is a basic concept of general chemistry, its definition has greatly evolved over the years and it is nowadays fixed by the International Union of Pure and Applied Chemistry (IUPAC),<sup>39</sup> with the inclusion of different types of hydrogen bond donors and acceptors. In the present investigation, two types of hydrogen bond have been considered and characterized by applying the promolecular and ELMO versions of the NCI technique.

We initially focused on the strong O-H $\cdots$ O interaction, which is obviously one of the most conventional hydrogen bonds and which can be easily detected by means of IR and NMR experiments. In particular, we considered the D192N mutant of Rhamnogalacturonan acetyltransferase<sup>40</sup> (PDB code 3C1U), whose (1)H-NMR spectrum indicates the presence of a short and strong hydrogen bond between residues Asp75 and Asp87. To characterize this interaction, we performed NCI analyses exploiting promolecular and *ELMOdb* electron densities. Furthermore, for the NCI-ELMO computations, all the basis-sets currently available for the ELMO libraries were taken into account.

As expected, all the NCI calculations were able to identify a strong and attractive hydrogen bonding interaction, as shown by the blue RDG isosurface depicted in Figure 6A. To better compare the results obtained through the promolecular-NCI and NCI-ELMO calculations, we also considered the two-dimensional plots of the reduced density gradients (see Figure 6B). Each NCI computation provides a RDG peak occurring at a negative value of the “signed” electron density, which clearly corresponds to the hydrogen bond under examination. Furthermore, in analogy with what was previously observed in the benchmark



calculations for the strong hydrogen bond in Leu-Enkephalin (see Figures 3B and 3C), we can notice that the spike obtained at promolecular level indicates a more stabilizing/attractive interaction compared to the NCI-ELMO results.

The use of different sets of basis functions in the NCI-ELMO computations also showed that the RDG peak practically shifts towards less negative values as the size and the quality of the adopted basis-set increase (see Figure 6C), particularly with this order: 6-31G, 6-311G, 6-31G(d,p), 6-311G(d,p) and cc-pVDZ. Except for cc-pVDZ, the same basis-set trend was observed for the strong hydrogen bond in the Leu-Enkephalin polypeptide (see Figure 4), practically confirming the equivalence of the interactions taken into account in the two different cases.

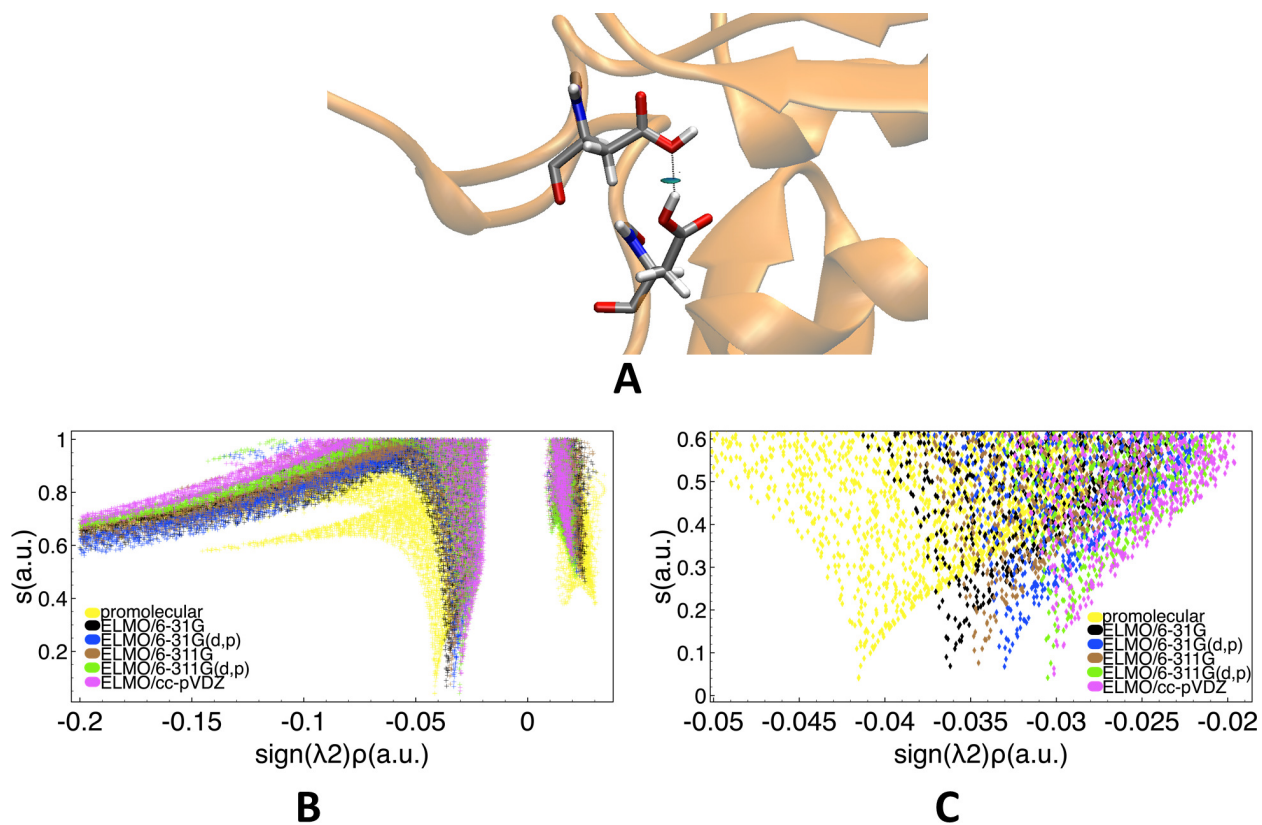


Figure 6: Strong hydrogen bond in the D192N mutant of Rhamnogalacturonan acetyltransferase between residues Asp75 and Asp87 (PDB code: 3C1IU): (A) RDG isosurface ( $s=0.6$  a.u., color scale:  $-0.03$  a.u.  $< \text{sign}(\lambda_2) \rho < 0.03$  a.u.) obtained at NCI-ELMO level and showing the interaction under examination; (B) 2D RDG plots obtained at promolecular-NCI and NCI-ELMO levels for all the considered basis-sets; (C) Zoomed 2D RDG plots obtained at promolecular-NCI and NCI-ELMO levels for all the considered basis-sets.

As briefly mentioned above, in the last decades, new functional groups have been found able to act as donors or acceptors of hydrogen bonds, giving rise to less conventional definitions of hydrogen bonding. Examples are C-H $\cdots$ O, C-H $\cdots\pi$ ,<sup>41</sup> and N-H $\cdots\pi$  interactions, which are ubiquitous in macromolecular structures and have consequently attracted enormous interest in chemistry and structural biology. However, notwithstanding their indisputable importance, we have to still clarify until to which extent these interactions are analogous to more traditional types of hydrogen bond.

To partially address the previous question and to further assess the capabilities of the new NCI-ELMO method, we have thus decided to analyze the CH- $\pi$  interaction between the methyl group of Leu50 and the aromatic ring of Tyr59 in the crystallographic structure of human erythrocytic ubiquitin<sup>42</sup> (PDB code: 1UBQ), a structure already analyzed in details by other authors.<sup>43,44</sup> The presence of the above mentioned CH- $\pi$  interaction has been also confirmed through Nuclear Magnetic Resonance spectroscopy,<sup>45</sup> which showed the qualitative structure of this kind of hydrogen bond in macromolecules.

From all our NCI calculations performed on the ubiquitin crystallographic structure, a green RDG isosurface was detected, both exploiting the traditional promolecular charge density and using the novel *ELMOdb* electron distributions (see Figure 7A). The 3D plot shows a more delocalized (not a circular shape) hydrogen bond due to the interaction with the whole  $\pi$  cloud<sup>46</sup>. The corresponding 2D plots of the reduced density gradient are reported in Figure 7B. They show a very high similarity with the RDG plots obtained for the traditional hydrogen-bond in lactoferrampin (see Figures 1B and 1C), with the promolecular-NCI and NCI-ELMO spikes having similar shapes but with the former observed at higher absolute density values compared to the latter.

As can be seen in Figure 7B, the different basis-sets associated with the ELMO libraries have been taken into account also for the NCI study of this interaction. In analogy with what was already observed for the weak hydrogen bond in lactoferampin (see Figure 2B), the different sets of basis functions basically give identical results and it is difficult to detect

a trend in function of the adopted basis-set, thus leading to the conclusion that, for the NCI computations, the choice of the correct strategy to obtain the density (ELMO vs. promolecular) is generally more important than the choice of the basis-set.

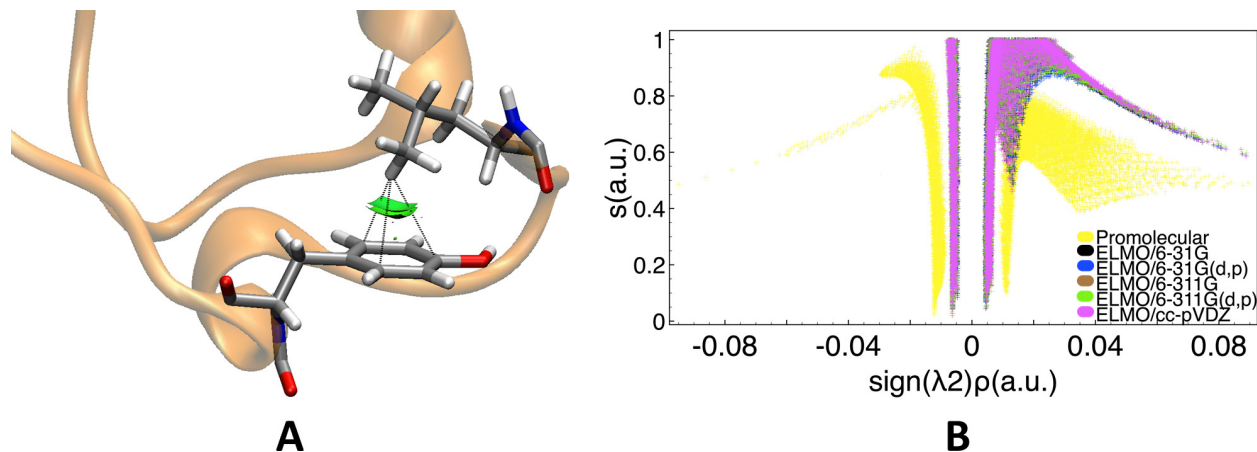


Figure 7: CH- $\pi$  interaction in the human erythrocytic ubiquitin between Leu50 and the aromatic ring in Tyr59 (PDB code: 1UBQ): (A) RDG isosurface ( $s=0.6$  a.u., color scale:  $-0.03$  a.u.  $< \text{sign}(\lambda_2)\rho < 0.03$  a.u.) obtained at NCI-ELMO level and showing the interaction under examination; (B) 2D RDG plots obtained at promolecular-NCI and NCI-ELMO levels for the considered basis-sets.

## 3.2 Cation- $\pi$ and anion- $\pi$ interactions

Another fundamental class of interactions that define structures and functions of biomolecules is represented by the cation- $\pi$  and anion- $\pi$  interactions. In particular, across a wide range of structural types, the former have been shown to have a crucial role not only in determining protein structures, but also in characterizing protein-ligand interactions. An important implication of the cation- $\pi$  interaction and of the related polar- $\pi$  interactions is that phenylalanine, tyrosine, and tryptophan should not be considered simply as hydrophobic conventional residues such as valine, leucine and isoleucine. In fact, they are actually expected to play a unique role in protein structure and folding.

First of all, we focused on cation- $\pi$  interactions, which generally occur between aromatic and ammonium groups<sup>47</sup> and particularly between sidechains of positively charged amino acids (e.g., lysine and arginine) and sidechains of amino acids bearing aromatic groups (e.g., phenylalanine, tyrosine and tryptophan).<sup>48</sup> In our investigation, this kind of interactions was analyzed by considering the X-ray crystallographic structure of glucoamylase-471 complexed with d-gluco-dihydroacarbose<sup>49</sup> (PDB code: 1GAI), which is characterized by four non-covalent interactions between the positively charged ammonium group of Lys108 with the aromatic sidechains of Trp52, Trp120, Tyr50 and Tyr116.

To study these interactions we performed promolecular-NCI and NCI-ELMO (cc-pVDZ basis-set) calculations. Both of them provided green delocalized low-density RDG isosurfaces (see Figure 8A) clearly corresponding to the above mentioned cation- $\pi$  interactions. We then compared the obtained 2D reduced density gradient plots, which are shown in Figures 8B (general plot) and 8C (zoomed plot). Compared to the hydrogen bond cases discussed in the previous subsection, the plots are more convoluted due to the simultaneous presence of four weak non-covalent interactions. However, also in this situation, the description given by the promolecular approximation is quite different from the one based on the *ELMOdb* electron density. As already observed for other interactions, the promolecular peaks are generally observed at larger absolute values of the “signed” density and, therefore, they lead to an

overestimation of the strength of the associated interactions (especially if we consider the negative region of the “signed” electron density). Finally, we can also observe that ELMO-NCI spikes are generally sharper than the promolecular ones and hence the corresponding surfaces are more localized in the 3D picture.

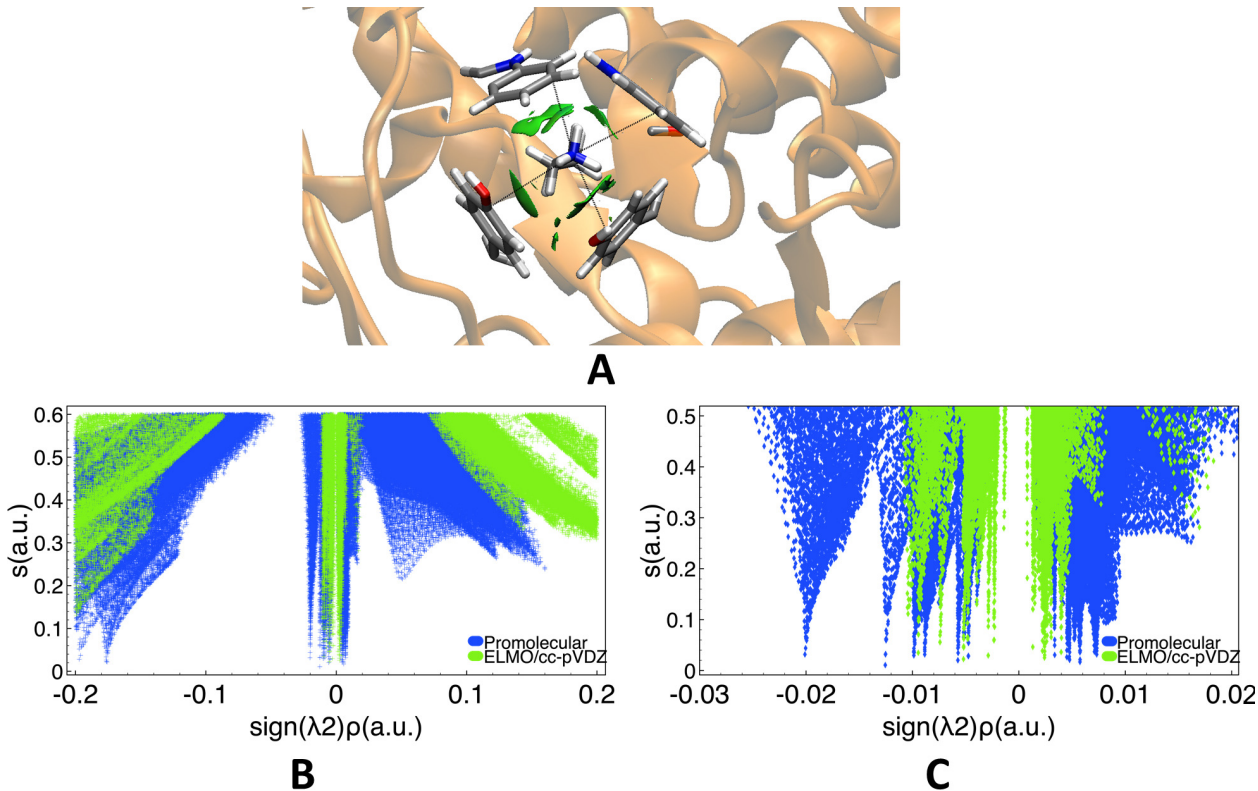


Figure 8: Cation- $\pi$  interaction in the glucoamylase-471 complexed with d-glucodihydrocarbose between the positively charged ammonium group of Lys108 with the aromatic sidechains of Trp52, Trp120, Tyr50 and Tyr116 (PDB code: 1GAI): (A) RDG isosurfaces ( $s=0.6$  a.u., color scale:  $-0.03$  a.u.  $< \text{sign}(\lambda_2)\rho < 0.03$  a.u.) obtained at NCI-ELMO level and showing the interaction under examination; (B) 2D RDG plots obtained at promolecular-NCI and NCI-ELMO/cc-pVDZ levels; (C) Zoomed 2D RDG plots obtained at promolecular-NCI and NCI-ELMO/cc-pVDZ levels.

At a second stage we took into account anion- $\pi$  interactions, which are among the most recent interactions classified as non-covalent bonding contacts.<sup>50,51</sup> For this purpose we considered the crystallographic structure of the antigene-binding fragment of the catalytic antibody 15A9 in complex with phosphopyridoxyl-L-alanine<sup>52</sup> (PDB code: 1WC7), for which we analyzed the non-covalent interaction occurring between Glu58 and Tyr94, already indicated

as an anion- $\pi$  contact by Lucas *et al.*<sup>53</sup> by using geometrical criteria.

Also for this system we performed NCI analyses based on the promolecular density and on the charge distribution obtained by transferring ELMOs from the constructed libraries (cc-pVDZ basis-set). As expected, the two NCI computations consistently detected the presence of a weak non-covalent interaction between the carboxylic group of residues Glu58 and the aromatic ring of Tyr94, which is depicted as a green reduced density gradient isosurface in Figure 9A. At the same time, in Figure 9B, the 2D RDG plots obtained at promolecular and ELMO/cc-pVDZ levels are also shown, with the promolecular peaks that always occur at higher absolute values of the “singled” electron density compared to the ELMO ones. Interestingly, we can notice again that the NCI-ELMO spikes are sharper and better-defined than those resulting from the use of the promolecular electron density.

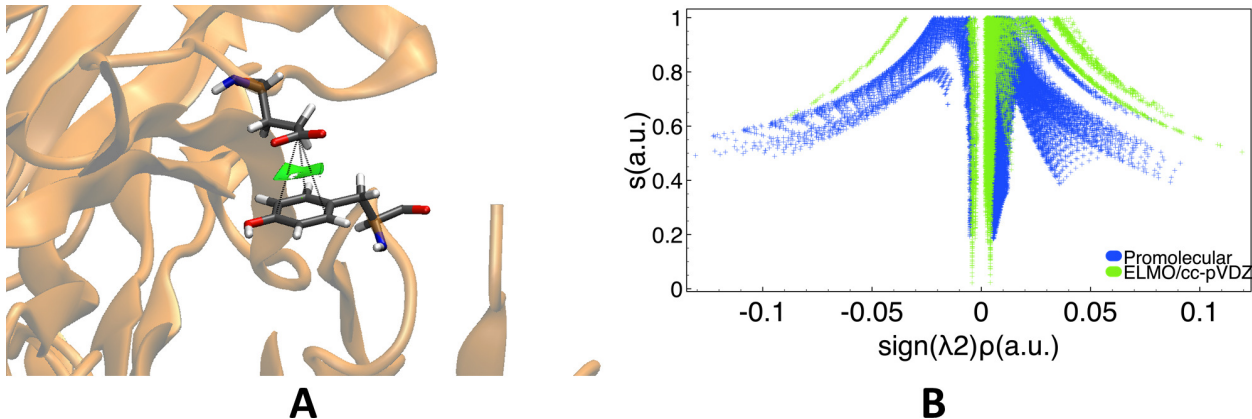


Figure 9: Anion- $\pi$  interactions in the antigen-binding fragment of the catalytic antibody 15A9 in complex with phosphopyridoxyl-L-alanine between Glu58 and Tyr94 (PDB code: 1WC7): (A) RDG isosurface ( $s=0.6$  a.u., color scale:  $-0.03$  a.u.  $< \text{sign}(\lambda_2)\rho < 0.03$  a.u.) obtained at NCI-ELMO level and showing the interaction under examination; (B) 2D RDG plots obtained at promolecular-NCI and NCI-ELMO/cc-pVDZ levels.

### 3.3 Lone pair ( $n-\pi^*$ ) interactions

Lone-pair ( $n-\pi^*$ ) interactions are another class of non-covalent interactions that significantly contribute to the stability of proteins and DNA conformations and that are extremely important for the binding of inhibitors in the pockets of biochemical receptors. As observed

for a large variety of proteins,<sup>54</sup> the  $n-\pi^*$  interactions often come along with hydrogen bonds and, together, they contribute to stabilize local backbone-sidechain contacts. In particular, according to Bartlett *et al.*,<sup>54</sup> there are two kinds of hydrogen bonds involved in the interplay with the  $n-\pi^*$  interactions: i) the local ones, which are usually found at turns of protein secondary structures and “bind” almost consecutive residues (namely,  $i$  and  $i+2$  residues); and ii) the non-local ones, which generally involve residues that are separated by at least five residues in the aminoacidic sequence of the protein.

In this investigation, the promolecular-NCI and NCI-ELMO (cc-pVDZ basis-set) techniques were exploited to analyze  $n-\pi^*$  interactions in presence of local and non-local hydrogen bonds. To address the first case, we considered the crystallographic structure of the protein human carbonic anhydrase II<sup>55</sup> (PDB code: 3KS3), which is characterized by a local hydrogen bond between residues Asn61 and Gly63 and by an  $n-\pi^*$  interaction between residues Asn61 and Asn59. At the same time, to study the second situation, we resorted to the crystallographic structure of leucyl/phenylalanyl-tRNA-protein transferase<sup>56</sup> (PDB code: 2CXA), which presents a non-local hydrogen bond between residues Asn24 and Asp62 and an  $n-\pi^*$  interaction between the oxygen atom (OD1) in the side chain of Asn24 and the carbon atom (C) of the carboxylic group in the backbone of Asn24. It is worthwhile to note that, both of these networks of interactions were also investigated by Bartlett and coworkers by performing NBO (Natural Bond Orbital) analyses.<sup>54</sup>

Both the promolecular and the *ELMOdb* electron densities allowed to correctly detect the networks of non-covalent interactions in the two systems under examination. In fact, both in Figure 10A and in Figure 10B, it is easy to observe a blue reduced density gradient isosurface associated with the local/non-local hydrogen bond and, at the same time, a green reduced density gradient isosurface indicating the simultaneous presence of the  $n-\pi^*$  interaction.

We have also compared the 2D RDG plots resulting from the different NCI analyses by separating the RDG peaks corresponding to the hydrogen bond from those corresponding to the  $n-\pi^*$  interaction. In Figures 11A and 11C we report the 2D RDG plots related to the

hydrogen bonds detected in protein structures 3KS3 and 2CXA, respectively. We can always notice the same trend observed for other conventional strong hydrogen bonds analyzed in the present study (for example, see the benchmark calculations on the strong hydrogen bond in the Leu-Enkephalin polypeptide). In other words, the single hydrogen bond spike obtained at promolecular level is clearly shifted towards higher absolute values of the “signed” electron density compared to the corresponding NCI-ELMO ones. Alike the previous cases, this means that the promolecular approximation tends to overestimate the strength of the hydrogen bond interactions compared to the ELMO approximation and all the quantum mechanical calculations. Finally, in Figures 11B and 11D, we show the reduced density gradient plots associated with the  $n-\pi^*$  interactions. For both systems under investigation we observe again the same trend seen for other weak non-covalent interactions considered in this study. In fact, these interactions are characterized by the presence of two symmetric peaks that, in the NCI-ELMO cases, are sharper and better-defined and that, above all, always occur at lower absolute values of the “signed” electron density.

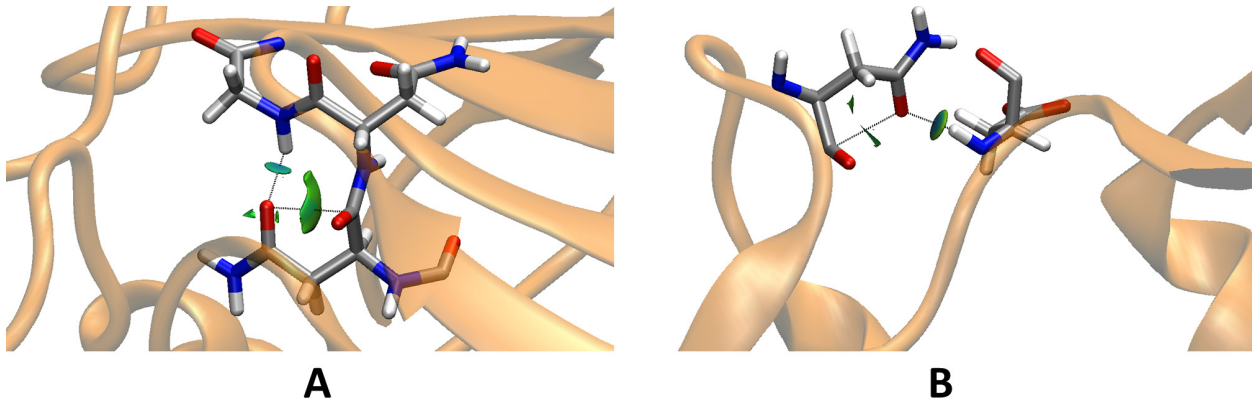


Figure 10: RDG isosurfaces ( $s=0.6$  a.u., color scale:  $-0.03$  a.u.  $< \text{sign}(\lambda_2)\rho < 0.03$  a.u.) obtained at NCI-ELMO level and showing the interaction under examination: (A) local hydrogen bond between residues Asn61 and Gly63 and  $n-\pi^*$  interaction between residues Asn61 and Asn59 in the human carbonic anhydrase II protein (PDB code: 3KS3); (B) non-local hydrogen bond between Asn24 and Asp62 and  $n-\pi^*$  interaction between the oxygen atom (OD1) in the side chain of Asn24 and the carbon atom (C) of the carboxylic group in the backbone of Asn24 in the leucyl/phenylalanyl-tRNA-protein transferase (PDB code: 2CXA).



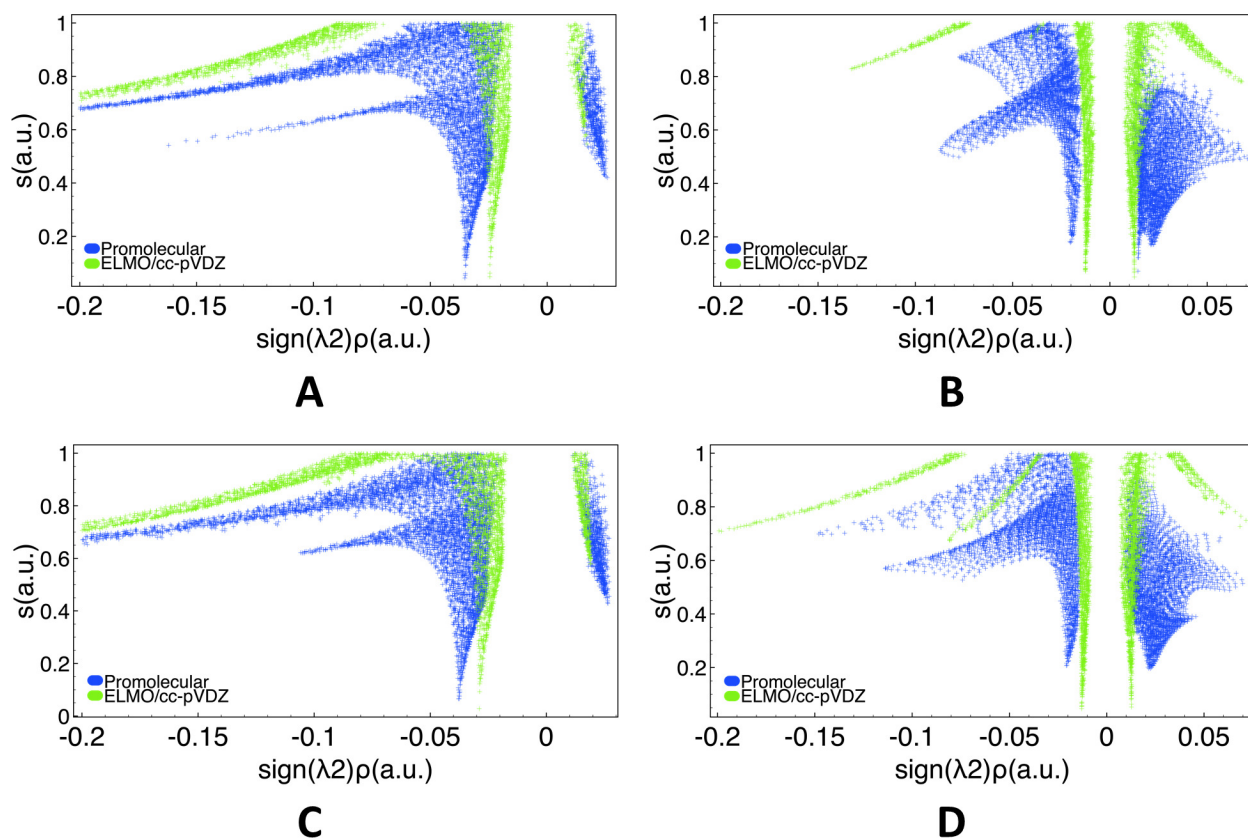


Figure 11: 2D RDG plots obtained at promolecular-NCI and NCI-ELMO/cc-pVDZ levels: (A) local hydrogen bond between residues Asn61 and Gly63 and (B)  $n-\pi^*$  interaction between residues Asn61 and Asn59 in the human carbonic anhydrase II protein (PDB code: 3KS3); (C) non-local hydrogen bond between Asn24 and Asp62 and (D)  $n-\pi^*$  interaction between the oxygen atom (OD1) in the side chain of Asn24 and the carbon atom (C) of the carboxylic group in the backbone of Asn24 in the leucyl/phenylalanyl-tRNA-protein transferase (PDB code: 2CXA).

### 3.4 Interactions with metal centers

To conclude our study, we have finally decided to consider the interactions with metal centers, which, after hydrogen bonds, are probably the non-covalent interactions that have the most important role in determining protein structure and enzymatic activity. Unfortunately, the diffuse nature of the electron density and the multiple oxidation states of metal centers generally hamper the determination of non-covalent interactions in apo-enzymes and in ligand-protein complexes.

The promolecular-NCI and NCI-ELMO (cc-pVDZ basis-set) strategies were used to identify and characterize the interactions established by Zn(II) with the E2 domain of the amyloid precursor protein (APP), whose experimental evidence has been reported in a recent study.<sup>57</sup> To accomplish this task, we particularly considered the X-ray structure of the APP E2 domain in complex with zinc<sup>58</sup> (PDB code: 3UMI). Furthermore, as done for the benchmark computations on the HIV Zinc fingerlike domain, the ELMOs for the Zn<sup>2+</sup> center were preliminarily calculated on a model molecule mimicking the chemical environment of the metal in the target system for this interaction. This enabled us to have all the necessary Extremely Localized Molecular Orbitals before carrying out the NCI-ELMO analysis. Finally, also in this case, for the computation of the promolecular density, we have considered a spherical atomic electron density for the cation Zn<sup>2+</sup>.

All the NCI calculations (promolecular and NCI-ELMO) were able to successfully detect the four interactions in which the central Zn<sup>2+</sup> ion is involved, namely the interactions with three histidine residues His382, His432 and His 436 and one water molecule. They are shown in Figure 12A, where the blue color indicates their stabilizing nature. Furthermore, each interaction corresponds to a specific spike in the negative domain of the 2D reduced density gradient plots (see Figure 12B). In this case it is interesting to observe that the NCI-ELMO peaks are observed at higher absolute values compared to the promolecular ones, which means that, according to the NCI-ELMO approach, the considered interactions are more attractive and stabilizing. This is in full agreement with the trend already seen for the test

calculations performed on the HIV Zinc fingerlike domain (see Figures 5B and 5C), another system characterized by interactions with a metal center.

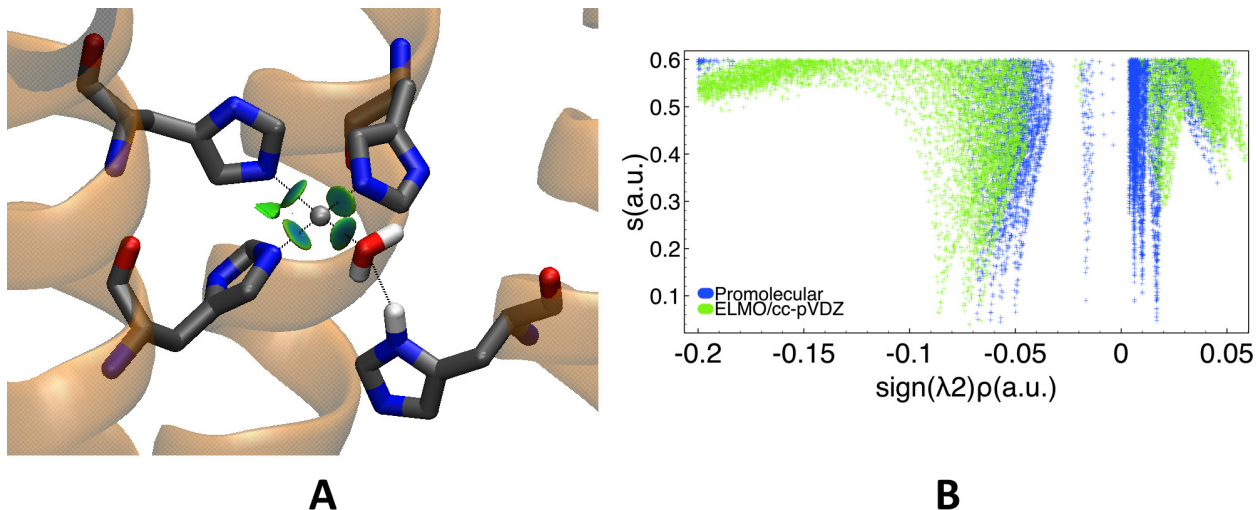


Figure 12: Interactions between the  $\text{Zn}^{2+}$  ion and coordinating residues (one water molecule and three histidine residues His382, His432 and His 436) in APP (PDB code: 3UMI): (A) RDG isosurfaces ( $s=0.6$  a.u., color scale:  $-0.03$  a.u.  $< \text{sign}(\lambda_2)\rho < 0.03$  a.u.) obtained at NCI-ELMO level and showing the interaction under examination; (B) 2D RDG plots obtained at promolecular-NCI and NCI-ELMO levels for basis-set cc-pVDZ.

## 4 Conclusions

In this paper, we have introduced an extension of the NCI method for the accurate detection of non-covalent interaction in systems of biological and biochemical interest. Since the new technique is based on quantum mechanically rigorous electron densities obtained through the transfer of Extremely Localized Molecular Orbitals from properly constructed libraries, the method clearly represents an advancement compared to the old version of the NCI strategy for macromolecules, which mainly exploited promolecular electron distributions. All the test calculations performed on small polypeptides and large proteins for a large variety of non-covalent interactions have confirmed this fact, showing that, in all the situations, even at fixed geometries, the novel NCI-ELMO method is definitely superior to the previous promolecular-NCI technique, without any significant increase of the associated computational cost. The

observed improvements in all the obtained 2D plots are ascribable to the fact that the ELMO electron densities are certainly more realistic charge distributions compared to the promolecular ones. In fact, the ELMO description is able to fully describe the anisotropic features of the electron density, not only for covalent bonding, but also for non-covalent interactions.

For this reason, we believe that the proposed strategy can be fruitfully used to rapidly and reliably define the networks of non-covalent interactions in biological molecules and, particularly, in very large proteins. Furthermore, in the future, the improvement introduced in the NCI method through the ELMO approximation might be exploited to better quantitatively assess the relative stabilities of proteins/polypeptides conformers or to rank poses of docking experiments, also in light of recent advancements in obtaining NCI-based interaction energies<sup>59</sup>.

Finally, since proteins are inherently flexible systems in constant motion among different conformational substates and their dynamic properties are frequently associated with molecular recognition, function and catalysis,<sup>60-64</sup> the new NCI-ELMO approach encloses great potentiality for monitoring the evolution of non-covalent interactions with time, for instance through a proper coupling with Molecular Dynamics simulations. In particular, analyzing representative samples of the structures for the investigated systems (properly extracted from trajectories of Molecular Dynamics simulations), it could be possible to provide time-averaged descriptors associated with the existing non-covalent interactions, with the goal of assigning the overall dynamic importance of each interaction and, possibly, with the goal of proposing a new semi-quantitative strategy to construct pharmacophores able to take into account the dynamic features of the macromolecules.<sup>60,61</sup> Work in this direction is currently under way.

## 5 Further Computational Details

The ELMOs transfers from the current ELMO databanks to the selected protein/polypeptide structures were carried out exploiting the recently implemented *ELMOdb* program,<sup>35</sup> which requires a PDB file as input. In particular, before starting the transfer of the orbitals, the original PDB file downloaded from the Protein Data Bank was processed i) by selecting only one of the possible conformers for the disordered regions of the polypeptide/protein, ii) by assigning, when necessary, the correct protonation states to the different residues according to the pH of crystallization/experiment, and, if necessary, iii) by adding the missing hydrogen atoms by exploiting suitable software (in our case, the *tleap* module of the AMBER Molecular Dynamics package<sup>65</sup>). The *ELMOdb* program provided a Gaussian09<sup>66</sup> formatted checkpoint file and wavefunction files that were used for the following NCI analyses (for more details see the ELMO subsection in the final Appendix).

The traditional ELMO calculations to obtain tailor-made Extremely Localized Molecular Orbitals for the Zinc metal centers ( $\text{Zn}^{2+}$  ion) were carried out using a program<sup>22</sup> that implements the Stoll technique within a development version of the GAMESS-UK package.<sup>67</sup> Finally, for the sake of completeness, all the DFT computations for our benchmark tests were performed by exploiting the quantum chemical software Gaussian09.<sup>66</sup>

The ELMO libraries and the associated *ELMOdb* program are available upon request to the corresponding author, while the NCI PLOT program is available for download (<https://github.com/aoterodelaroza/nciplot>) and online through the NCI webserver (<http://www.lct.jussieu.fr/pagesperso/contrera/nciweb/index.php>). The NCI-ELMO method will be also made available through the latter platform.

## 6 Appendix

In this Appendix, we will give a general and brief theoretical overview of the NCI and ELMO methods, which have been coupled to develop the novel NCI-ELMO strategy. We

will review the basics of the NCI method, paying special attention to its current application to large systems. In the ELMO subsection, the first part will explain the basic theory of the Extremely Localized Molecular Orbitals, while the second one will mainly focus on the main features of the recently constructed ELMO libraries.

## 6.1 NCI

The NCI index (for Non Covalent Interactions)<sup>18,19</sup> is a tool to reveal non-covalent interactions in three-dimensional (3D) space. It is based on a Density Functional Theory (DFT) quantity, the reduced density gradient (RDG, from now on also simply indicated as  $s$ ), which accounts for density inhomogeneity with respect to the atomic density and, therefore, identifying the regions of density change due to interatomic interactions. The reduced density gradient shows a very characteristic shape when it is plotted against the electron density, namely the appearance of peaks that highlight the presence of interactions in the system, as illustrated in Figure 13A for the Pyridoxine-Aminopyridine dimer. However, if only the electron density is taken into account, the peaks correspond both to favorable and unfavorable interactions. Therefore, the absolute value of the electron density is multiplied by the corresponding sign of the second eigenvalue of the electron density Hessian (see Figure 13B) and a color scheme is applied according to that sign. Additionally, the interactions can be visualized in 3D by plotting RDG isosurfaces exploiting the same color scheme, as shown in Figure 13C. In the example of the Pyridoxine-Aminopyridine dimer reported in Figure 13, we can see that the peaks (Figure 13B) and isosurfaces (Figure 13C) are related to intermolecular hydrogen bonds, van der Waals interactions and intramolecular steric clashes within the rings. Blue highlights strong interactions (i.e. hydrogen bonds). These are also characterized by being very localized, so that their 3D surface is compact, whereas the peak in the two-dimensional (2D) plot occupies a bigger surface. Repulsive interactions appear in red and are also localized. Instead, very weak interactions, in this case van der Waals, are displayed in green and are showcase delocalized interactions: their 2D peaks are very steep

and their 3D shape is more “amorphous”, even occupying big surfaces in the 3D plot when the interaction is very delocalized.

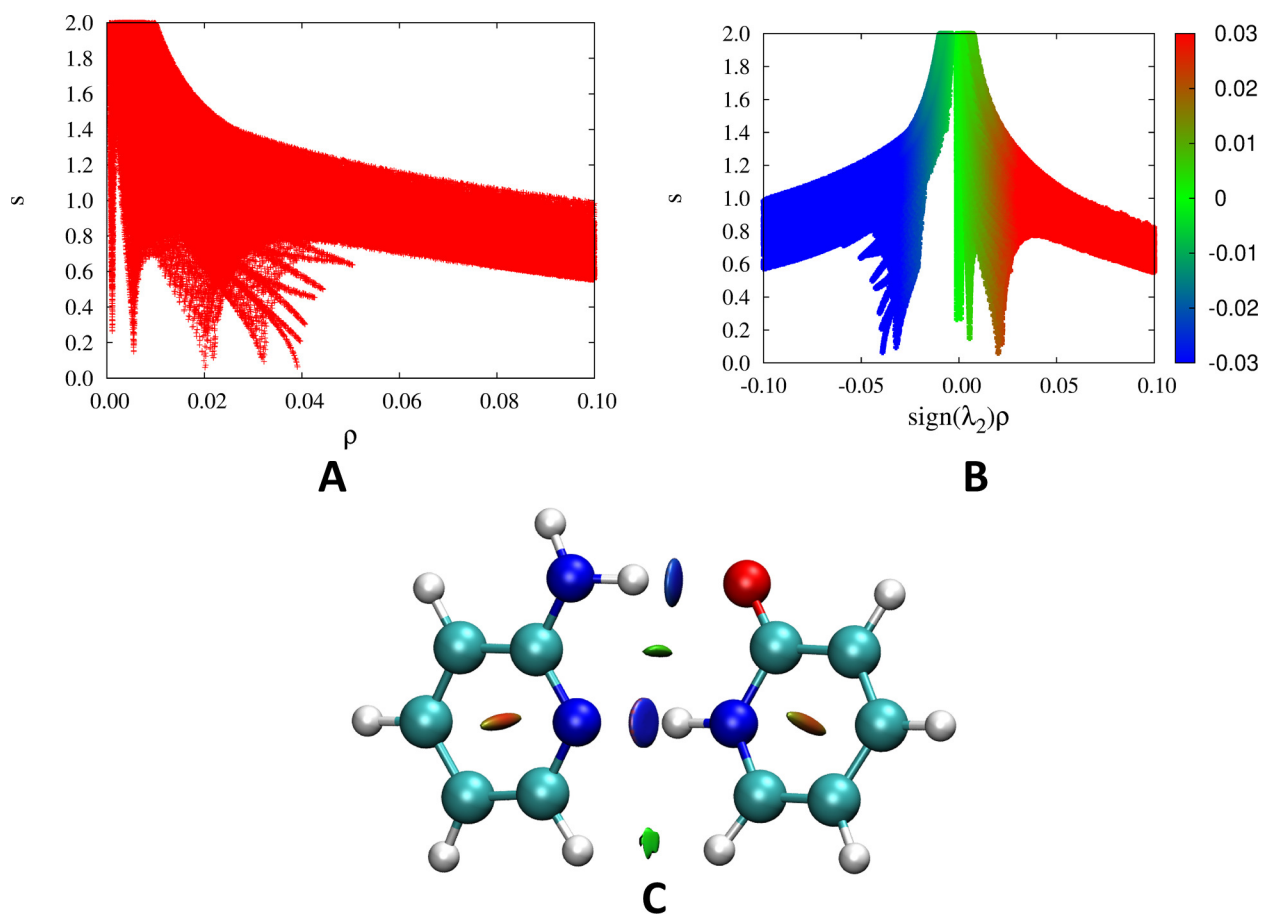


Figure 13: Example of NCI analysis for the Pyridoxine-Aminopyridine dimer: (A) 2D plot of the reduced density gradient against the electron density ( $\rho$ ); (B) 2D plot of the reduced density gradient against the “signed” electron density ( $\text{sign}(\lambda_2)\rho$ ) (color scale:  $-0.03$  a.u.  $< \text{sign}(\lambda_2)\rho < 0.03$  a.u.); (C) reduced density gradient isosurfaces ( $s=0.6$  a.u., color scale:  $-0.03$  a.u.  $< \text{sign}(\lambda_2)\rho < 0.03$  a.u.).

Being based on the electron density, this approach has the advantage that it can be applied to diverse chemical systems while taking the environment into account (alike parameterized van der Waals radii). The reduced density gradient features are quite stable with respect to the computation method, to such an extent that they are already contained in the sum of atomic densities (*aka* the promolecular density). This has the advantage of not needing a quantum chemical calculation, so that it has been the option of choice for biomolecular systems. Thus, qualitative NCI analysis is applicable to extremely large sys-

tems, such as proteins and DNA, where describing the interplay of attractive and repulsive interactions is crucial for understanding structure and functionality. However, promolecular densities lack relaxation. Hence, the semi-quantitative information of the relative strength of bonds is lost. Moreover, only main group elements have been fitted in NCIPLLOT-1.0, where the zero oxidation state is a reasonable assumption.

## 6.2 Extremely Localized Molecular Orbitals and their libraries

Although chemists often use Lewis structures (namely, they refer to a localized picture of electronic structure) to rationalize structure and reactivity of molecules, molecular orbitals usually obtained from Hartree-Fock calculations are generally delocalized over the whole system under consideration. Therefore, over the years, one of the many common ways of connecting traditional chemical concepts with quantum chemistry computations consisted in developing methods to localize the obtained canonical Hartree-Fock molecular orbitals through *a posteriori* unitary transformations<sup>68–72</sup> or in proposing *a priori* strategies to determine molecular orbitals strictly localized on small molecular subunits.<sup>21,73–82</sup>

For the construction of the libraries of Extremely Localized Molecular Orbitals used in the present study, we have particularly exploited the method developed by Stoll.<sup>21–23</sup> This technique preliminarily subdivides the system under examination into small fragments (i.e., atoms, bonds or larger functional groups) according to the traditional chemical intuition or computational needs. These fragmentations generally consist in choosing a localization scheme corresponding to the Lewis structure of the molecule and automatically lead to the definition of local basis-sets for each subunit. These local basis-sets are exclusively constituted by the atomic orbitals centred on the atoms belonging to the fragment and are used to expand only the molecular orbitals describing the electrons of the corresponding subsystems. Therefore, by imposing this preliminary constraint, ELMOs are simply obtained by variationally minimizing the energy of the single Slater determinant constructed with them (see Supporting Information for further theoretical details about the Stoll technique



to obtain Extremely Localized Molecular Orbitals).

Unlike the traditional localized molecular orbitals of quantum chemistry,<sup>68-72</sup> ELMOs are absolutely localized on small molecular fragments, without orthogonalization tails that extend beyond the main localization region (see Figure S1 in the Supporting Information). For this reason, ELMOs can be easily and univocally associated with elementary molecular units and straightforwardly transferred from one molecule to another. For the sake of completeness and clarity, we report here (see Figure 14) representative Extremely Localized Molecular Orbitals obtained through a traditional ELMO calculation performed on methanol with a localization pattern corresponding to the Lewis structure of the system. As you can see, traditional fragments, such as C-H, O-H, C-O bonds and oxygen lone pairs are defined *a priori*, leading to ELMOs corresponding to these “Lewis entities”.

Exploiting the Stoll method to obtain individual ELMOs (as briefly outlined above) and taking advantage of the easy transferability of the ELMOs, libraries of Extremely Localized Molecular Orbitals have been recently assembled.<sup>35</sup> This has been done by performing a large number of ELMO calculations on model molecules that reproduce the suitable molecular environments of the small fragments on which we strictly localized the molecular orbitals to be stored in our databanks. The current ELMO databases consider all the possible elementary units of the twenty natural amino acids in all their possible protonation states and forms (i.e., N-terminal, non-terminal and C-terminal forms). They were built for five standard quantum chemistry basis-sets (6-31G, 6-311G, 6-31G(d,p), 6-311G(d,p) and cc-pVDZ). In particular, the current version of the databanks comprise i) ELMOs localized on one-atom fragments to describe core and lone-pair electrons, ii) ELMOs localized on two-atom fragments to deal with ordinary bond electron pairs, and also iii) ELMOs localized on three-atom fragments for all those situations in which it is crucial to properly take into account the delocalized nature of the electronic structure (e.g., carboxylic groups, peptide bonds, aromatic rings).

The ELMO libraries are associated with the *ELMOdb* program that efficiently imple-

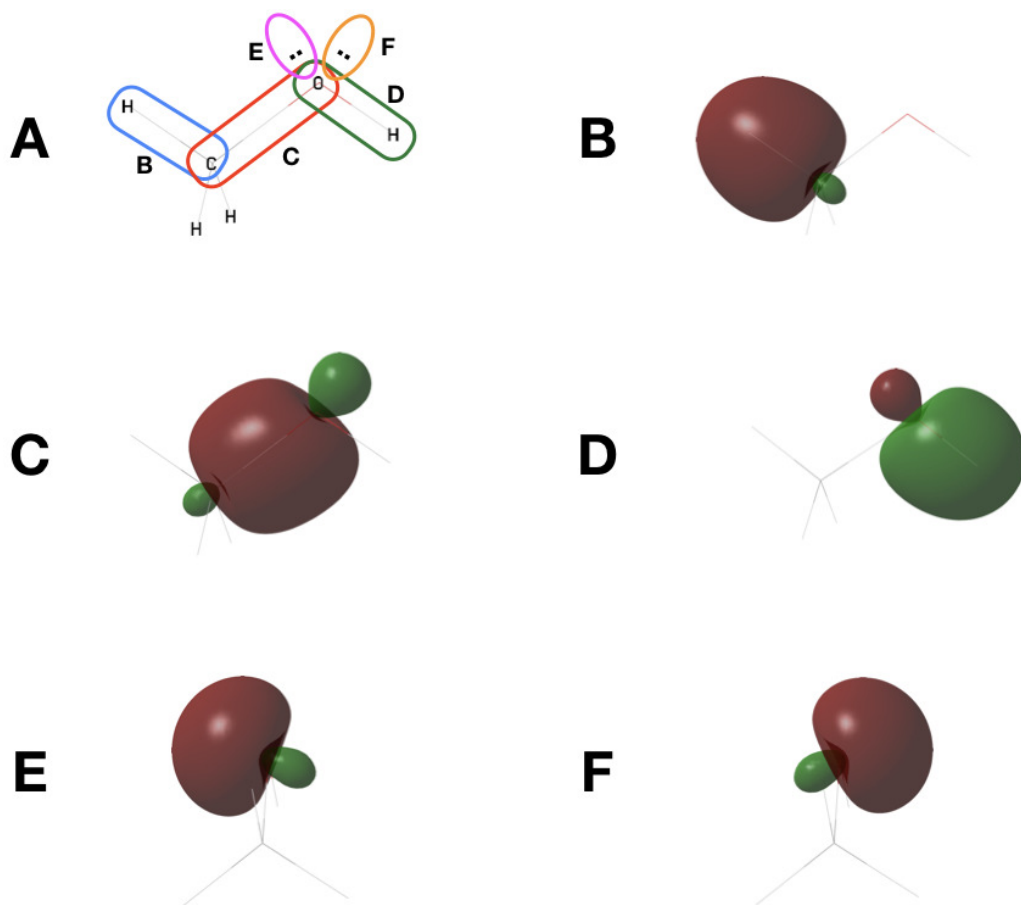


Figure 14: Example of Extremely Localized Molecular Orbitals for methanol that can be obtained when a localization scheme corresponding to the Lewis structure of the molecule is used: (A) 3D Lewis structure of methanol with explicit indication of the fragments for which ELMOs are depicted in Figures 2B-2F, (B) ELMO describing one of the C-H bonds of the methyl group, (C) ELMO describing the C-O bond, (D) ELMO describing the O-H bond, (E)/(F) ELMOs describing the lone-pairs of the oxygen atom (back view of the methanol molecule). Since ELMOs describing the core electrons of the carbon and oxygen atoms simply correspond to spherical distributions around the nuclei, they are omitted for the sake of simplicity. The orbitals have been computed using the cc-pVDZ basis-set and have been depicted using a 0.1 a.u. (C-H, C-O and O-H ELMOs) or a 0.2 a.u. (lone-pair ELMOs) isosurface.

ments the transfer of the extremely localized molecular orbitals from the databases to the protein target structure according to the rotation matrices originally introduced by Philipp and Friesner.<sup>33,83</sup> Test calculations have shown and confirmed that the ELMO databanks really enable to obtain approximate (but quantum mechanically rigorous) electron densities of large macromolecules at a reduced computational cost.<sup>35</sup> From now on, the electron distributions obtained from the ELMO databanks will be indicated as *ELMOdb* electron densities.

Furthermore, it is worth considering that, since the ELMO databases are currently limited to fragments of amino acids, the *ELMOdb* program has been organized in such a way that ELMOs not currently included in the libraries, but specifically computed for the systems under investigation (e.g., ELMOs associated with the subunits of a protein ligand or ELMOs associated with a metal center in a metallo-protein) can be easily read in input and used in the reconstruction of the global wave function and electron density.

Exploiting the ELMOs, which are either stored in the database or, if necessary, obtained from tailor-made ELMO calculations, the *ELMOdb* program computes the final one-particle density matrix of the system under investigation and, at the end of the calculation, it provides corresponding Gaussian09<sup>66</sup> formatted checkpoint (FChk) file and wavefunction files (in the .wfn and .wfx formats), which can be directly read by the NCIPLLOT program to detect and analyze non-covalent interactions (see previous subsection).

## Acknowledgement

A. G. gratefully acknowledges the French Research Agency (Agence Nationale de la Recherche, ANR) for financial support of the Young Researcher Project *QuMacroRef* through Grant No. ANR-17-CE29-0005-01. J. C.-G. and A. 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. D.A.-O. acknowledges the “Programa de Cooperacin Cientfica

ECOS-CONICYT” (N° ECOS170045) for financial support.

## Supporting Information Available

Details about the theory of Extremely Localized Molecular Orbitals; Figure S1 showing the differences between a canonical Hartree-Fock Molecular Orbital, a Pipek-Mezey Molecular Orbital and an Extremely Localized Molecular Orbital; Figures S2-S4 showing 2D RDG plots obtained at promolecular-NCI, NCI-B3LYP and NCI-ELMO levels with basis-sets 6-311G, 6-31G(d,p) and 6-311G(d,p) for lactoferrampin; Figures S5-S7 showing 2D RDG plots obtained at promolecular-NCI, NCI-B3LYP and NCI-ELMO levels with basis-sets 6-311G, 6-31G(d,p) and 6-311G(d,p) for Leu-enkephalin; Figures S8-S10 showing 2D RDG plots obtained at promolecular-NCI, NCI-B3LYP and NCI-ELMO levels with basis-sets 6-311G, 6-31G(d,p) and 6-311G(d,p) for the HIV Zinc fingerlike domain. This material is available free of charge via the Internet at <http://pubs.acs.org/>.

## References

- (1) Chandler, D. Interfaces and the driving force of hydrophobic assembly. *Nature* **2005**, *437*, 640–647.
- (2) Panigrahi, S. K.; Desiraju, G. R. Strong and weak hydrogen bonds in the protein-ligand interface. *Proteins* **2007**, *67*, 128–141.
- (3) Keskin, O.; Tuncbag, N.; Gursoy, A. Predicting Protein-Protein Interactions from the Molecular to the Proteome Level. *Chem. Rev.* **2016**, *116*, 4884–4909.
- (4) Cavallo, G.; Metangolo, P.; Milani, R.; Pilati, T.; Priimagi, A.; Resnati, G.; Terraneo, G. The Halogen Bond. *Chem. Rev.* **2016**, *116*, 2478–2601.
- (5) Klein, R. A. Modified van der Waals atomic radii for hydrogen bonding based on electron density topology. *Chem. Phys. Lett.* **2006**, *425*, 128 – 133.
- (6) Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; Oxford University Press: Oxford, U.K., 1990.
- (7) Espinosa, E.; Molins, E.; Lecomte, C. Hydrogen Bond Strengths Revealed by Topological analyses of Experimentally Observed Electron Densities. *Chem. Phys. Lett.* **1998**, *285*, 170–173.
- (8) Espinosa, E.; Souhassou, M.; Lachekar, H.; Lecomte, C. Topological analysis of the electron density in hydrogen bonds. *Acta Cryst. B* **1999**, *55*, 563–572.
- (9) Grabowski, S. J. Ab Initio Calculations on Conventional and Unconventional Hydrogen Bonds Study of the Hydrogen Bond Strength. *J. Phys. Chem. A* **2001**, *105*, 10739–10746.
- (10) Becke, A. D.; Edgecombe, K. E. A simple measure of electron localization in atomic and molecular systems. *J. Chem. Phys.* **1990**, *92*, 5397–5403.

- (11) Silvi, B.; Savin, A. Classification of chemical bonds based on topological analysis of electron localization functions. *Nature* **1994**, *371*, 683–686.
- (12) Alikhani, M. E.; Fuster, F.; Silvi, B. What Can Tell the Topological Analysis of ELF on Hydrogen Bonding? *Struct Chem* **2005**, *16*, 203–210.
- (13) Contreras-García, J.; Recio, J. M. Electron delocalization and bond formation under the ELF framework. *Theor. Chem. Acc.* **2011**, *128*, 411–418.
- (14) Genoni, A.; Bucinsky, 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. .; Massa, L.; Matta, C. F.; Merz Jr., K. M.; Nakashima, P. N. H.; Ott, H.; Ryde, U.; Schwarz, K.; Sierka, M.; Grabowsky, S. Quantum Crystallography: Current Developments and Future Perspectives. *Chem. Eur. J.* **2018**, *24*, 10881–10905.
- (15) Grabowsky, S.; Genoni, A.; Bürgi, H.-B. Quantum crystallography. *Chem. Sci.* **2017**, *8*, 4159–4176.
- (16) Massa, L.; Matta, C. F. Quantum Crystallography: A perspective. *J. Comput. Chem.* **2017**, *39*, 1021–1028.
- (17) Tsirelson, V. Early days of quantum crystallography: A personal account. *J. Comput. Chem.* **2017**, *39*, 1029–1037.
- (18) Johnson, E. R.; Keinan, S.; Mori-Sanchez, P.; Contreras-Garcia, J.; Cohen, A. J.; Yang, W. Revealing noncovalent interactions. *J. Am. Chem. Soc.* **2010**, *132*, 6498–6506.
- (19) Contreras-García, J.; Johnson, E. R.; Keinan, S.; Chaudret, R.; Piquemal, J.-P.; Beratan, D. N.; Yang, W. NCIPLOT: a program for plotting noncovalent interaction regions. *J. Chem. Theory Comput.* **2011**, *7*, 625–632.

- (20) Contreras-García, J.; Boto, R. A.; Izquierdo-Ruiz, F.; Reva, I.; Woller, T.; Alonso, M. A benchmark for the Non-Covalent Interaction index (NCI) or... is it really all in the geometry? *Theor. Chem. Acc.* **2016**, *135*, 242.
- (21) Stoll, H.; Wagenblast, G.; Preuss, H. On the use of local basis sets for localized molecular orbitals. *Theor. Chem. Acc.* **1980**, *57*, 169–178.
- (22) 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.
- (23) Sironi, M.; Genoni, A.; Civera, M.; Pieraccini, S.; Ghitti, M. Extremely Localized Molecular Orbitals: Theory and Applications. *Theoretica Chimica Acta* **2007**, *117*, 685–698.
- (24) Genoni, A. Molecular Orbitals Strictly Localized on Small Molecular Fragments from X-ray Diffraction Data. *J. Phys. Chem. Lett.* **2013**, *4*, 1093–1099.
- (25) Genoni, A. X-ray Constrained Extremely Localized Molecular Orbitals: Theory and Critical Assessment of the New Technique. *J. Chem. Theory Comput.* **2013**, *9*, 3004–3019.
- (26) Dos Santos, L. H. R.; Genoni, A.; Macchi, P. Unconstrained and X-ray constrained extremely localized molecular orbitals: analysis of the reconstructed electron density. *Acta Cryst. A* **2014**, *70*, 532–551.
- (27) Genoni, A.; Meyer, B. X-Ray Constrained Wave Functions: Fundamentals and Effects of the Molecular Orbitals Localization. *Adv. Quantum Chem.* **2016**, *73*, 333–362.
- (28) Sironi, M.; Famulari, A.; Raimondi, M.; Chiesa, S. The transferability of extremely localized molecular orbitals. *J. Mol. Struct. (THEOCHEM)* **2000**, *529*, 47–54.

- (29) Genoni, A.; Sironi, M. A Novel Approach to Relax Extremely Localized Molecular Orbitals: the Extremely Localized Molecular Orbital-Valence Bond Method. *Theoretica Chimica Acta* **2004**, *112*, 254–262.
- (30) Genoni, A.; Fornili, A.; Sironi, M. Optimal Virtual Orbitals to Relax Wave Functions Built Up with Transferred Extremely Localized Molecular Orbitals. *Journal of Computational Chemistry* **2005**, *26*, 827–835.
- (31) 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.
- (32) 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.
- (33) 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.
- (34) 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.
- (35) 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.
- (36) Ghosh, A.; Datta, A.; Jana, J.; Kar, R. K.; Chatterjee, C.; Chatterjee, S.; Bhunia, A. Sequence context induced antimicrobial activity: insight into lipopolysaccharide permeabilization. *Mol. BioSyst.* **2014**, *10*, 1596–1612.



- (37) Pichon-Pesme, V.; Lecomte, C.; Wiest, R.; Bernard, M. Modeling fragments for the ab initio determination of electron density in polypeptides. An experimental and theoretical approach to the electron distribution in Leu-enkephalin trihydrate. *J. Am. Chem. Soc.* **1992**, *114*, 2713–2715.
- (38) Summers, M. F.; South, T. L.; Kim, B.; Hare, D. R. *Biochemistry* **1990**, *29*, 329–340.
- (39) Arunan, E.; Desiraju, G. R.; Klein, R. A.; Sadlej, J.; Scheiner, S.; Alkorta, I.; Clary, D. C.; Crabtree, R. H.; Dannenberg, J. . J.; Hobza, P.; Kjaergaard, H. G.; Legon, A. C.; Mennucci, B.; Nesbitt, D. J. Definition of the hydrogen bond(IUPAC Recommendations 2011). *Pure Appl. Chem.* **2011**, *83*, 1637–1641.
- (40) Langkilde, A.; Kristensen, S. M.; Lo Leggio, L.; Mølgaard, A.; Jensen, J. H.; Houk, A. R.; Navarro Poulsen, J.-C.; Kauppinen, S.; Larsen, S. Short strong hydrogen bonds in proteins: a case study of rhamnogalacturonan acetylerase. *Acta Cryst. D* **2008**, *64*, 851–863.
- (41) Nishio, M.; Hirota, M.; Umezawa, Y. *The CH/π interaction: evidence, nature, and consequences*; John Wiley & Sons, 1998; Vol. 21.
- (42) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. Structure of ubiquitin refined at 1.8 Å resolution. *J. Mol. Biol.* **1987**, *194*, 531–544.
- (43) Nishio, M.; Umezawa, Y.; Fantini, J.; Weiss, M. S.; Chakrabarti, P. CH–π hydrogen bonds in biological macromolecules. *Phys. Chem. Chem. Phys.* **2014**, *16*, 12648–12683.
- (44) Huang, G.-B.; Liu, W.-E.; Valkonen, A.; Yao, H.; Rissanen, K.; Jiang, W. Selective recognition of aromatic hydrocarbons by endo-functionalized molecular tubes via C/N-H · · π interactions. *Chin. Chem. Lett.* **2018**, *29*, 91–94.
- (45) Plevin, M. J.; Bryce, D. L.; Boisbouvier, J. Direct detection of CH/π interactions in proteins. *Nat. Chem.* **2010**, *2*, 466.

- (46) 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.
- (47) Burley, S.; Petsko, G. Amino-aromatic interactions in proteins. *FEBS letters* **1986**, *203*, 139–143.
- (48) Gallivan, J. P.; Dougherty, D. A. Cation- $\pi$  interactions in structural biology. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9459–9464.
- (49) Aleshin, A. E.; Stoffer, B.; Firsov, L. M.; Svensson, B.; Honzatko, R. B. Crystallographic complexes of glucoamylase with maltooligosaccharide analogs: relationship of stereochemical distortions at the nonreducing end to the catalytic mechanism. *Biochemistry* **1996**, *35*, 8319–8328.
- (50) Mascal, M.; Armstrong, A.; Bartberger, M. D. Anion-aromatic bonding: A case for anion recognition by  $\pi$ -acidic rings. *J. Am. Chem. Soc.* **2002**, *124*, 6274–6276.
- (51) Estarellas, C.; Frontera, A.; Quiñero, D.; Deyà, P. M. Relevant anion- $\pi$  interactions in biological systems: The case of urate oxidase. *Angew. Chem. Int. Ed.* **2011**, *50*, 415–418.
- (52) Golinelli-Pimpaneau, B. Structure of a pseudomerohedrally twinned monoclinic crystal form of a pyridoxal phosphate-dependent catalytic antibody. *Acta Cryst. D* **2005**, *61*, 472–476.
- (53) Lucas, X.; Bauzá, A.; Frontera, A.; Quiñero, D. A thorough anion- $\pi$  interaction study in biomolecules: on the importance of cooperativity effects. *Chem. Sci.* **2016**, *7*, 1038–1050.
- (54) Bartlett, G. J.; Newberry, R. W.; VanVeller, B.; Raines, R. T.; Woolfson, D. N. Inter-

- play of hydrogen bonds and  $n \rightarrow \pi^*$  interactions in proteins. *J. Am. Chem. Soc.* **2013**, *135*, 18682–18688.
- (55) Avvaru, B. S.; Kim, C. U.; Sippel, K. H.; Gruner, S. M.; Agbandje-McKenna, M.; Silverman, D. N.; McKenna, R. A short, strong hydrogen bond in the active site of human carbonic anhydrase II. *Biochemistry* **2009**, *49*, 249–251.
- (56) Dong, X.; Kato-Murayama, M.; Muramatsu, T.; Mori, H.; Shirouzu, M.; Bessho, Y.; Yokoyama, S. The crystal structure of leucyl/phenylalanyl-tRNA-protein transferase from *Escherichia coli*. *Protein Sci.* **2007**, *16*, 528–534.
- (57) Young, T. R.; Wedd, A. G.; Xiao, Z. Evaluation of Cu (I) binding to the E2 domain of the amyloid precursor protein - a lesson in quantification of metal binding to proteins via ligand competition. *Metallomics* **2018**, *10*, 108–119.
- (58) Dahms, S. O.; Könnig, I.; Roeser, D.; Gührs, K.-H.; Mayer, M. C.; Kaden, D.; Multhaupt, G.; Than, M. E. Metal binding dictates conformation and function of the amyloid precursor protein (APP) E2 domain. *J. Mol. Biol.* **2012**, *416*, 438–452.
- (59) Peccati, F.; Desmedt, E.; Contreras-García, J. A regression approach to accurate interaction energies using topological descriptors. *Comp. Theo. Chem.* **2019**, *1159*, 23–26.
- (60) Morra, G.; Genoni, A.; Neves, M. A. C.; Merz Jr., K. M.; Colombo, G. Molecular Recognition and Drug-Lead Identification: What Can Molecular Simulations Tell Us? *Curr. Med. Chem.* **2010**, *17*, 25–41.
- (61) Genoni, A.; Penanti, M.; Morra, G.; Zaffaroni, N.; Colombo, G. Ligand selection from the analysis of protein conformational substates: new leads targeting the N-terminal domain of Hsp90. *RSC Adv.* **2012**, *2*, 42684282.
- (62) Morra, G.; Genoni, A.; Colombo, G. Mechanisms of Differential Allosteric Modula-

- tion in Homologous Proteins: Insights from the Analysis of Internal Dynamics and Energetics of PDZ Domains. *J. Chem. Theory Comput.* **2014**, *10*, 5677–5689.
- (63) Legnani, L.; Compostella, F.; Sansone, F.; Toma, L. ConeCalix[4]arenes with Orientable Glycosylthioureido Groups at the Upper Rim: An In-Depth Analysis of Their Symmetry Properties. *J. Org. Chem.* **2015**, *80*, 7412–7418.
- (64) Ferraro, M.; D’Annessa, I.; Moroni, E.; Morra, G.; Paladino, A.; Rinaldi, S.; Compostella, F.; Colombo, G. Allosteric Modulators of HSP90 and HSP70: Dynamics Meets Function through Structure-Based Drug Design. *J. Med. Chem.* **2019**, *62*, 60–87.
- (65) Case, D.; Ben-Shalom, I.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E.; Cruzeiro, V.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. Amber 2018. University of California, San Francisco.
- (66) 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.; Strat-

- mann, 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, .; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian09 Revision D.01. Gaussian Inc. Wallingford CT 2009.
- (67) 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.
- (68) Boys, S. F. Construction of Some Molecular Orbitals to be Approximately Invariant for Changes from One Molecule to Another. *Rev. Mod. Phys* **1960**, *32*, 296–299.
- (69) Foster, J. M.; Boys, S. F. Canonical Configurational Interaction Procedure. *Rev. Mod. Phys.* **1960**, *32*, 300–302.
- (70) Edmiston, C.; Ruedenberg, K. Localized Atomic and Molecular Orbitals. *Rev. Mod. Phys.* **1963**, *35*, 457–465.
- (71) Edmiston, C.; Ruedenberg, K. Localized Atomic and Molecular Orbitals. II. *J. Chem. Phys.* **1965**, *43*, S97–S116.
- (72) Pipek, J.; Mezey, P. G. A Fast Intrinsic Localization Procedure Applicable for Ab Initio and Semiempirical Linear Combination of Atomic Orbital Wave Functions. *J. Chem. Phys.* **1989**, *90*, 4916–4926.
- (73) 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.
- (74) McWeeny, R. Some Recent Advances in Density Matrix Theory. *Rev. Mod. Phys.* **1960**, *32*, 335–369.

- (75) Adams, W. H. On the Solution of the Hartree-Fock Equations in Terms of Localized Orbitals. *J. Chem. Phys.* **1961**, *34*, 89–102.
- (76) Huzinaga, S.; Cantu, A. A. Theory of Separability of Many-Electron Systems. *J. Chem. Phys.* **1971**, *55*, 5543–5549.
- (77) 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.
- (78) 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.
- (79) Smits, G. F.; Altona, C. Calculation and properties of non-orthogonal, strictly local molecular-orbitals. *Theoretica Chimica Acta* **1985**, *67*, 461–475.
- (80) 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.
- (81) 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.
- (82) Couty, M.; Bayse, C. A.; Hall, M. B. Extremely localized molecular orbitals (ELMO): a non-orthogonal Hartree-Fock method. *Theoretica Chimica Acta* **1997**, *97*, 96–109.
- (83) 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.