JCTC Journal of Chemical Theory and Computation

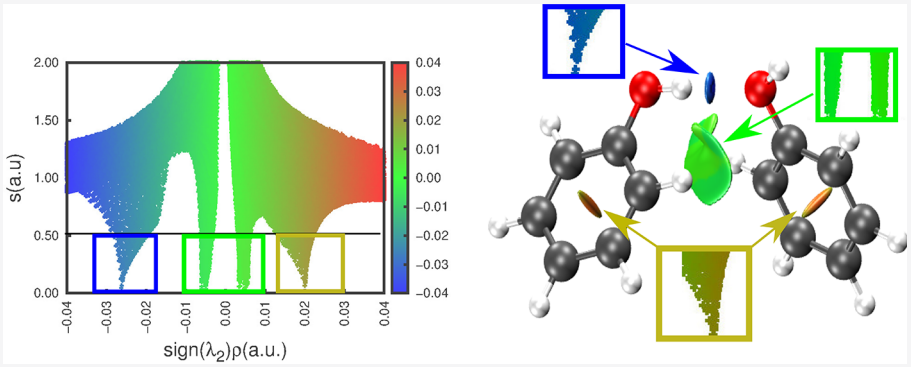
pubs.acs.org/JCTC Article

NCIPLOT4: Fast, Robust, and Quantitative Analysis of Noncovalent Interactions

Roberto A. Boto, Francesca Peccati, Rubén Laplaza, Chaoyu Quan, Alessandra Carbone, Jean-Philip Piquemal, Yvon Maday, and Julia Contreras-García*

Cite This: J. Chem. Theory Comput. 2020, 16, 4150–4158 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: The NonCovalent Interaction index (NCI) enables identification of attractive and repulsive noncovalent interactions from promolecular densities in a fast manner. However, the approach remained up to now qualitative, only providing visual information. We present a new version of NCIPLOT, NCIPLOT4, which allows quantifying the properties of the NCI regions (volume, charge) in small and big systems in a fast manner. Examples are provided of how this new twist enables characterization and retrieval of local information in supramolecular chemistry and biosystems at the static and dynamic levels.

INTRODUCTION

The proper evaluation of noncovalent interactions is of paramount importance to understanding and designing new compounds, from biological macromolecules to materials.[1,2] In this design quest, local analyses can provide valuable information. Knowing which parts of a molecular system are interacting (or not) and through which type of interaction is fundamental for an efficient inverse design. This kind of information, which is implicitly reflected in the binding energy, is important for both experimental and computational chemistry. In this context, real space approaches, which are based on a chemically grounded scalar or vector field, have enabled the revision of many classical chemical concepts. Among the most popular real space approaches, one can find the Quantum Theory of Atoms In Molecules (QTAIM),[3] the Electron Localization Function (ELF),[4] restricted space partitioning,[5] and the NonCovalent Interaction (NCI) index.[6]

The NCI index has become a widely used tool to analyze chemical interactions in the past decade. It was originally proposed as a visualization tool to identify and classify noncovalent interactions. However, the most important point that remained elusive was the relationship between NCI regions and the interaction energy. Several approaches have been put forward.[7,8] However, these studies were not able to provide a general approach to the energetics in terms of the NCI region and, just like other types of density analysis,[3] remained limited to intermolecular interactions within the same family of compounds.[9]

In this paper, we present a new version of the NCIPLOT code[10] which includes the up-to-now missing quantitative features. We define integrals of the density involved in the noncovalent interactions, which given the complexity of the calculations and the simplicity of the approach, show good correlation with the energy. Indeed, the density used for this linear regression is promolecular: $\rho(\mathbf{r}) = \sum_A \rho_A(\mathbf{r})$, where A runs over all atoms in the system. That is, these densities do not require an SCF (self-consistent field) calculation of the system under study. This enables application to very large systems where the wave function is not available, while keeping a reasonable accuracy.[11] Parallelization and multilevel grids have also been implemented, which enable a considerable

Received: January 19, 2020
Published: May 29, 2020

https://dx.doi.org/10.1021/acs.jctc.0c00063
J. Chem. Theory Comput. 2020, 16, 4150–4158

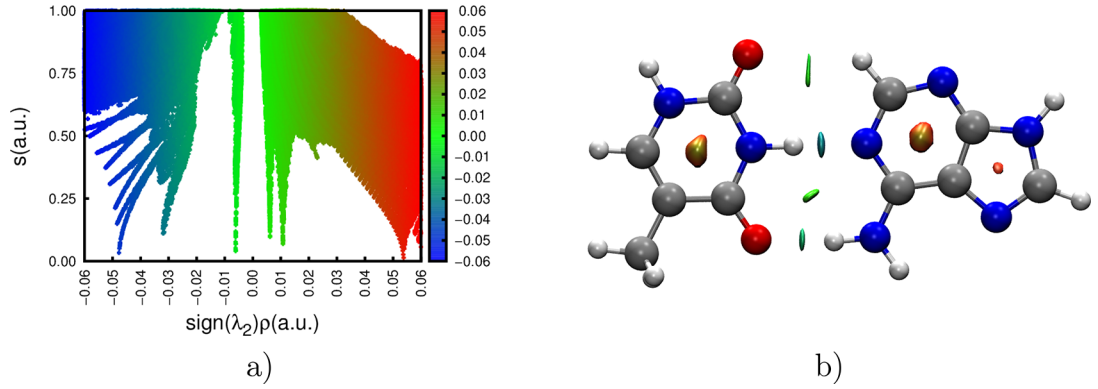


Figure 1. Standard NCI index representations for an adenine−thymine dimer using a promolecular density. (a) $s(\mathbf{r})$ plotted against $\text{sign}(\lambda_2)\rho(\mathbf{r})$. (b) Isosurfaces of $s(\mathbf{r}) = 0.3$ colored by $\text{sign}(\lambda_2)\rho(\mathbf{r})$.

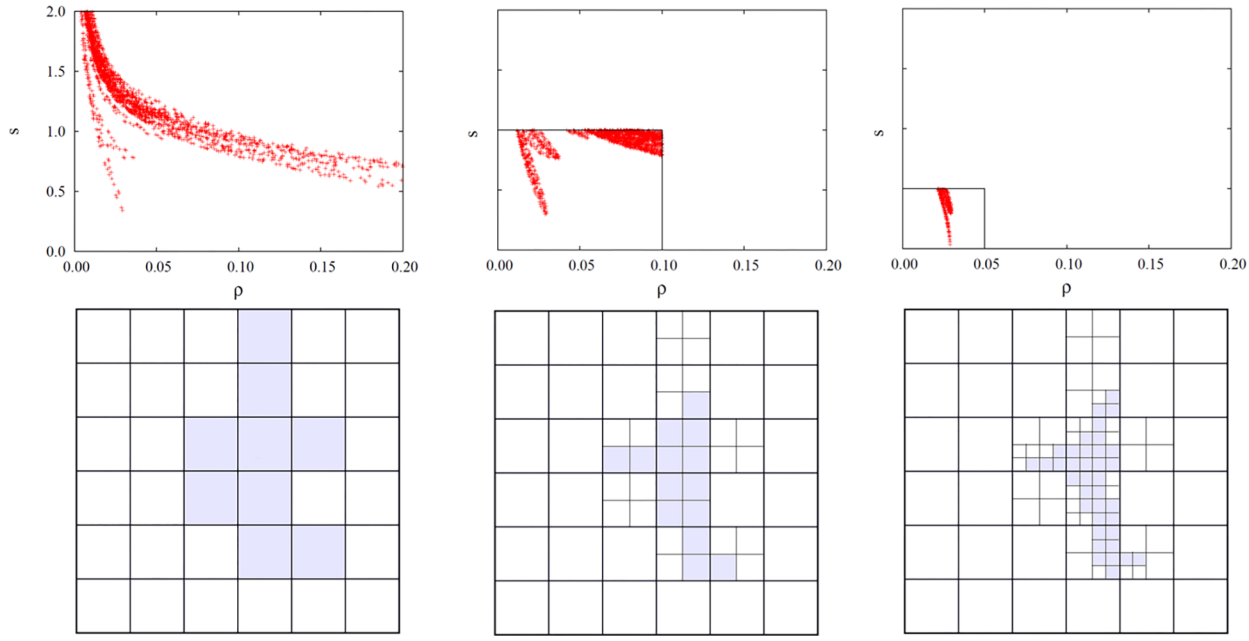


Figure 2. Two-dimensional schematic diagram of multilevel grids, from coarse (left) to fine grids (right) with 4:2:1 increments and $\rho_c = 0.05$ and $s_c = 1.0$. The density and the s values are computed at the active points of the coarse, medium, and fine grids (from left to right). Top, cutoff on (ρ, s); bottom, size of the boxes. The small blue boxes are the active boxes of the finest grid, which really contribute to forming the isosurface.

speedup of the calculations. Applications to macromolecular and biological systems are presented. NCI derived analysis contains structural and energetic information, including which atoms are contributing to the stabilization and how. We believe this new implementation should be of use for the analysis of MD trajectories, the construction of new scoring functions, and the understanding of large-scale systems where NCIs are key.

THE NCI METHOD

The NCI method is firmly rooted on the properties of the reduced density gradient.[12] It is a function of the electron density, $\rho(\mathbf{r})$, and its reduced density gradient

$$s(\mathbf{r}) = \frac{1}{C_s}\frac{|\vec{\nabla}\rho(\mathbf{r})|}{\rho(\mathbf{r})^{4/3}} \quad (1)$$

with $C_s = 2(3\pi^2)^{1/3}$. It highlights the interactions in a given system in the shape of peaks in an $s(\rho)$ diagram (Figure 1a). Hence, plotting s isosurfaces leads to very intuitive pictures where noncovalent interactions appear as surfaces in between the involved atoms (Figure 1b). Generally, steric clashes are shown as red isosurfaces, van der Waals interactions are shown as thin, delocalized green regions, and strong attractive interactions (i.e., hydrogen bonds) are shown as localized blue lentils (cf. Figure 1b). This provides very valuable local information for the comprehension and design of noncovalent interactions.

ADAPTING THE CODE FOR MOLECULAR DYNAMICS

Acceleration with Multilevel Grids. It should be noted that the total density in the noncovalent region can be approximated without an important loss of information from the sum of atomic densities. This approach, known as promolecular,[13] enables the construction of NCIs in large systems, for which a wave function cannot be generated in a useful time scale. Promolecular densities in NCIPLOT have been obtained from fits to spherically averaged LDA densities of each isolated atom, A, using one exponential function per

https://dx.doi.org/10.1021/acs.jctc.0c00063
J. Chem. Theory Comput. 2020, 16, 4150−4158

Table 1. Comparison of the Computation Time and the Percentage of Nonactive Points (%) between One Grid and Multilevel Grid Methods[a]

		AT2			2V5X			1ACB		
n	α_k	$T_{\rho,s}$	T_{tot}	%	$T_{\rho,s}$	T_{tot}	%	$T_{\rho,s}$	T_{tot}	%
1	1	3.17	8.1	98.14	11.03	17.39	89.11	257.22	264.30	99.80
3	4, 2, 1	0.67	5.45	98.14	3.60	9.94	89.13	7.19	14.06	99.80

[a]Implemented in a MacBook with a 2.5 GHz Intel Core i7 Processor. The computation time includes the time $T_{\rho,s}$ of computing the density and s, and the total running time T_{tot} where the output of the .cube file could be expensive. In these examples, the density and s cutoffs are set to 0.1 and 1.0, respectively. The increments are respectively 0.05 × 0.05 × 0.05, 0.1 × 0.1 × 0.1, and 0.3 × 0.3 × 0.3.

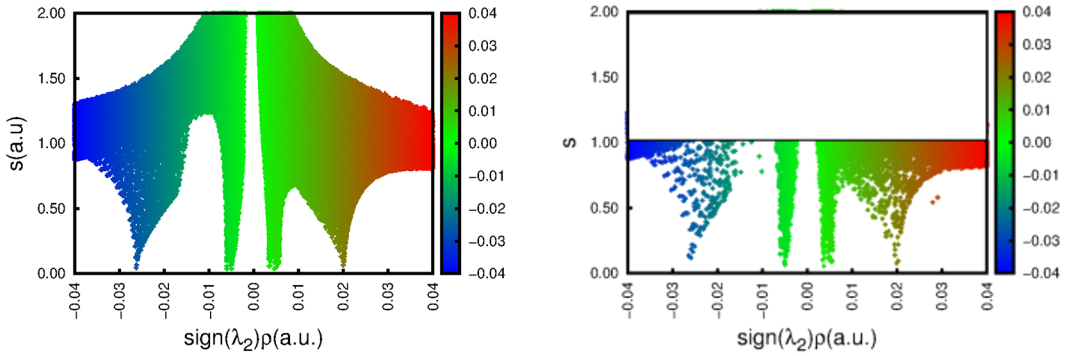


Figure 3. Ω_{NCI} obtained from the integration approach. (left) Full result. (right) Ω_{NCI} defined from the density reconstruction and constant s (eq 2). Parameters used: γ_{ref} = 0.95 (this can be seen from the smaller concentration of points) and s_c = 1.0 (black line).

shell, n: $\rho_A(\mathbf{r}) = \sum_n c_n e^{-\zeta_n \mathbf{r}}$, where c_n and ζ_n are the fitted constants.[10] For metals and charged residues, where the promolecular approach can sometimes fail, the density has been shown to be accurately described when transferred from extremely localized molecular orbitals which take the nearby environment into account.[14]

Focusing on biomolecular applications, an important acceleration of the NCI calculation is needed. Given the fact that high density regions are not needed for the analysis of noncovalent interactions, a multilevel grid is perfectly adapted for the acceleration of the code.

For the multilevel grid method, nested progressively finer grids are computed only in those areas where NCI peaks are found. In order to make sure that grid points are not missed in the coarse grids, the (ρ, s) criteria are made looser in the initial steps until the final (ρ, s) criteria are met in the final fine grid. The same parameter (α_i) that is used to determine the relative size of the grid is used to determine the looseness of the (ρ, s) criteria.

From the algorithmic point of view, it works as follows. We take n levels of grids denoted by $g_1, g_2, ..., g_n$, from coarse to fine with decreasing factors $\alpha_1 > \alpha_2 > ... > \alpha_n = 1$. In this approach, we first compute (ρ, s) at each grid point of the coarsest grid g_1, (ρ_1, s_1), and then check if $\rho_1 > \alpha_1\rho_c$ or $s_1 > \alpha_1 s_c$. If this is the case, this grid point is not interesting from the noncovalent point of view (it has too big density or s). Thus, we call this grid point nonactive. A box is called nonactive if any of its vertices is nonactive. This stringent option has been implemented so as to avoid possible positive (negative) integration values of $\text{sign}(\lambda_2)\rho$ in negative (positive) ranges leading to artificial sign changes. A nonactive box does not contribute to the isosurface and should consequently be ignored. We record the activeness information for all boxes of g_1. Next, for each grid point of g_2, we check whether it belongs to a nonactive box of g_1 or not. If so, it is not necessary to compute (ρ, s) and it can be directly set as nonactive. Otherwise, we compute (ρ_2, s_2) and repeat the procedure, checking whether $\rho_2 > \alpha_2\rho_c$ or $s_2 > \alpha_2 s_c$, to determine the activeness or nonactiveness of the point. Again, we record the activeness information for all boxes of g_2. In general, given any point in the grid g_k with $k \geq 2$, we first check if it is nonactive based on the activeness information on the coarser grid g_{k-1}. This avoids computing the ρ and s quantities many times in a noninteresting region. In the remaining region, (ρ_k, s_k) are computed and compared with $\alpha_k\rho_c$ and $\alpha_k s_c$, to determine the activeness and nonactiveness. Those active boxes of g_k are then recognized and recorded for the following finer grid g_{k+1} analysis. This process is repeated until the final (finest) grid g_n is reached.

Figure 2 gives a 2D schematic diagram of the above multilevel grid method. In this diagram, three levels of grids are employed with $\alpha_1 = 4$, $\alpha_2 = 2$, and $\alpha_3 = 1$. We can see how at the beginning ρ and s are calculated over a big volume (top left) with a scarce grid (top bottom). In the subsequent computations, finer grids are constructed around the NCI peaks, so that the ρ and s thresholds become smaller and so do the grid volumes. Since only the density and s at the active grid points in the right-most figure need be accurately computed, this approach is much less expensive than dealing with the finest grid directly.

We have tested the performance of this acceleration technique on the computation time and the percentage of the active points in the finest grid. Three examples are illustrated in Table 1, including the adenine−thymine intermolecular interaction (AT2, 30 atoms, Figure 1), the protein−ligand interaction of 2V5X (266 atoms), and the 1ACB protein−inhibitor complex (2289 atoms). It is observed that the multilevel grid method can decrease the time of computing ρ and s by orders of magnitude. This consequently helps to decrease the total run time, especially whenever the computation of (ρ, s) is the most time-consuming step. Moreover, whenever the calculation is performed using

https://dx.doi.org/10.1021/acs.jctc.0c00063
J. Chem. Theory Comput. 2020, 16, 4150−4158

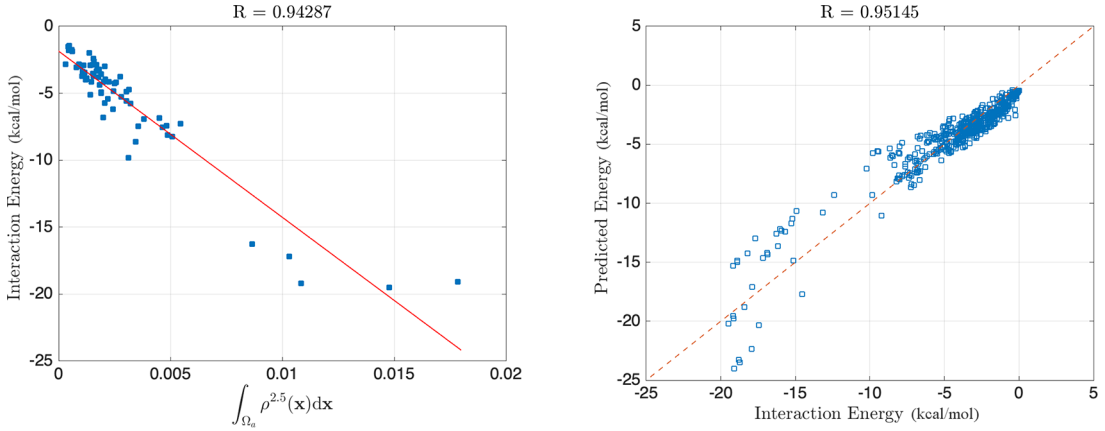


Figure 4. (left) Data fitting of the integral $I_{2.5}$ and the interaction energy for S66 dimers.[15] (right) Predicted interaction energy and exact interaction energy for all S66x8 dimers.[18]

promolecular densities, the whole operation becomes highly parallelizable. Generally speaking, if the NCI region is small compared to the space spawned by the entire system, the multilevel grid method can perform very efficiently.

Quantification. In order to compute chemical properties within the regions highlighted by the NCI method, a clean definition of the interaction region, Ω_{NCI}, is needed. From the promolecular approach exposed above, the total electron density for a complex AB can be expressed as a sum over the fragments A and B: $\rho^{AB}(\mathbf{r}) = \rho^{A}(\mathbf{r}) + \rho^{B}(\mathbf{r})$. Hence, those regions where $\rho \simeq \rho^{A}(\mathbf{r})$ or $\rho \simeq \rho^{B}(\mathbf{r})$ can be considered as noninteracting. Instead, in those regions where neither of these approximations applies, both fragments are contributing to the total density. If, additionally, s is low, the point we are analyzing is part of our interaction region, Ω_{NCI}. In the algorithm:

$$\mathbf{r}_i \in \Omega_{NCI} \begin{cases} s(\mathbf{r}_i) < s_c(\mathbf{r}_i) \\ \rho(\mathbf{r}_i) < \gamma^{ref}\rho^{A}(\mathbf{r}_i) \\ \rho(\mathbf{r}_i) < \gamma^{ref}\rho^{B}(\mathbf{r}_i) \end{cases} \tag{2}$$

where γ^{ref} is a threshold value, typically around 0.95. Stability with respect to γ^{ref} can be softened ensuring a sufficiently low s, i.e., by using a reference $s(\mathbf{r})$ value, $s_c(\mathbf{r}_i)$. Typically a value of s_c = 1.0 encloses all relevant interaction regions: bonds and noncovalent interactions (see Figure 3). On the other hand, an s_c = 1.0 isovalue is high enough to spread from regions of interaction toward atomic tails in certain systems, and this trend is corrected by the threshold value of γ^{ref} multiplied by $\rho^{A}(\mathbf{r}_i)/\rho^{B}(\mathbf{r}_i)$. This can be seen in the depletion of points in the tails when the left and right panels of Figure 3 are compared. The parametric space spawned by both γ^{ref} and s_c has been thoroughly explored (see Figures S4 and S5). Robust and meaningful volumes can be defined by this combination.

Taking the definition of NCI regions introduced above, we have analyzed the correlation between the interaction energy and the integrals of the electron density over the active points Ω_a (the multigrid approximation to the NCI region, Figure 2). The following integrals have been computed for different choices of the s cutoff, s_c:

$$I_n = \int_{\Omega_a} \rho^{n}(\mathbf{r})\,\mathrm{d}r, \qquad n = 0,\ 1,\ \frac{4}{3},\ 1.5,\ \frac{5}{3},\ 2,\ 2.5,\ 3 \tag{3}$$

In order to test the connection between integrals I_n and the energetics, we have first carried out a proof of concept for small systems. The S66 data set[15] is a collection of small noncovalent dimers whose interaction energies have been determined at the golden standard CCSD(T)/CBS level. This constitutes the first step in the benchmarking of noncovalent interactions within the quantum chemistry community.

An analysis of the correlations for the S66 database (see the Supporting Information) has shown that the highest correlation is obtained for I_n with n = 2−2.5 and s_c = 1.0. In general, this cutoff maximizes the correlation between integrals and interaction energies, and it is also more robust with respect to the exponent of I_n. The fit for n = 2.5 is shown in Figure 4, left, where the correlation coefficient R is 0.94 and the mean relative error is 20%. Similar results are also obtained for n = 2, with R = 0.93 (see Figure S1). Since our promolecular approximation uses one exponential per shell, we can compare this result with other quantum mechanical methods and small basis sets. For example, this range of error is the one found with current routine methods such as PBE-D3 when small basis sets are used (for reference, MAE on the S66 set for PBE/6-31+G(d,p) is 53% whereas it goes down to 18% for PBE-D3/6-31+G(d,p)[16]—note that it is less than 8% for bigger basis sets[17]).

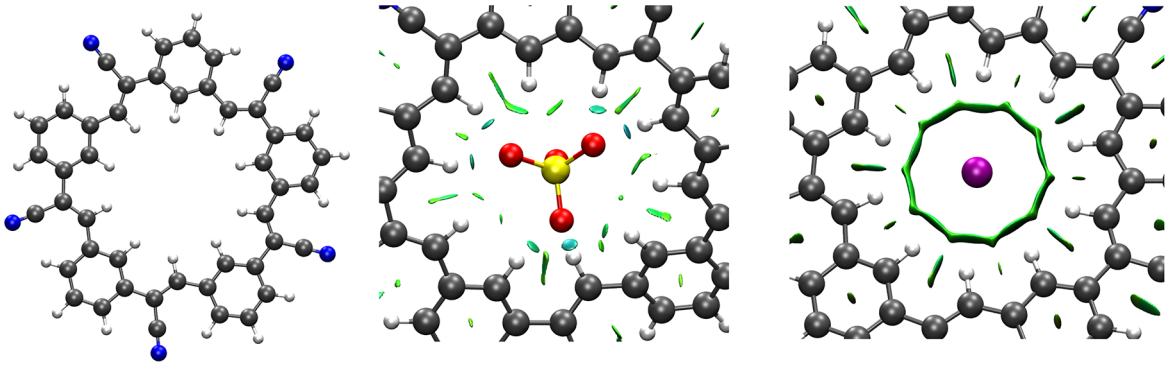
Since intermolecular interactions can span a wide range of distances, in a second step, we have carried out similar analyses for the S66x8 set,[18] leading also to a good description of the involved energetics (see Figure 4 right, and the Supporting Information for more details). This is especially interesting since, on many occasions, the main covalent framework imposes nonequilibrium distances among noncovalently bonded parts of the biomolecule(s).

APPLICATIONS TO LARGE SYSTEMS

In this section, we will focus on the new features available in NCIPLOT4 and illustrate the applicability of estimations of NCI properties and their relationship to structure and energetics in large systems. We will showcase possible active areas of research in which the fast and robust NCIPLOT4 can be useful. We believe this gain in scale to be very important as nearly atomistic experimental control is achieved over larger and more complex structures.

Analysis of NCIs in Supramolecular Systems. Supramolecular chemistry is an expanding area of research in which NCIs are often the driving force.[19] Mechanically interlocked

https://dx.doi.org/10.1021/acs.jctc.0c00063
J. Chem. Theory Comput. 2020, 16, 4150−4158



a) Pentacyanostar **A** b) Pentacyanostar **A** complexed with PO_4^{3-} c) Pentacyanostar **A** complexed with I^-

Figure 5. NCIPLOT analysis of pentacyanostar **A** interacting with different anions, with carbon, hydrogen, nitrogen, oxygen, phosphorus, and iodine atoms colored gray, white, blue, red, yellow, and magenta, respectively. $s(\mathbf{r}) = 0.3$ isosurfaces in (b) and (c).

systems, molecular machines, and anion capture systems are brilliant examples of this fruitful area.[20]

In many cases, quantum mechanical treatment of the full system is achievable with density functional theory. NCIPLOT4 enables the rapid analysis of computational electron densities and provides researchers with an understanding of the NCIs at play. In this brief example, we will focus on novel anion capture systems in which key interactions are purportedly of the CH···X^- type.[21,22] The specificity and characteristics of the NCIs in such systems are still relatively unknown.

As an example, a novel pentacyanostar **A** (Figure 5a), reported by Lee and co-workers,[21] was treated at the B3LYP-D3(BJ)/def2SVP level.[23,24] **A** has a cavity that can bind anions. In this case, complexation of PO_4^{3-} and I^- was studied by optimizing the structures to a minimum. NCIPLOT4 was used to process the calculated electron densities for the optimized geometries, which leads to the characterization and understanding of the interactions in the system.

Results are collected in Figure 5b,c. After running NCIPLOT4, visual inspection of the isosurfaces in the standard coloring scheme instantly reveals whether the CH···X^- interaction is plausibly a hydrogen bond and therefore directional and specific, or van der Waals.

In the case of PO_4^{3-}, the interaction type, as reflected locally by $\rho(\mathbf{r})$ and $s(\mathbf{r})$, is clearly akin to a hydrogen bond: small, lenticular, bluish surfaces. As the interactions target the oxygen atoms specifically, it can be inferred that the spherically averaged interaction is suboptimal, as one or more electron-rich oxygen moieties always remain outside the twisted star. Hence, it will favor shared capture by a sandwich star complex, sharing oxygen moieties between two stars as experimentally observed. Out of the 10 available hydrogen atoms, only six are interacting strongly in the single-star structure.

The complex with I^-, while still symmetric, does not show any strong hydrogen bond interaction. In this case, having the anion interact dispersively with the whole crown is more energetically favored than forming specific interactions. This signals that dispersive interactions are relatively strong for this anion, while hydrogen bonds are comparatively less relevant. This has the important consequence that the planarity of the star is far less altered, and accordingly little to no improvement would be expected upon dimerization or stacking of the receptor.

Note that the cost of performing such an analysis is remarkably low thanks to the improvements in the code, but the information it provides can be used for rationalization and thus reverse engineering.

Dissecting Interaction Energies in Biosystems. In order to validate our approach for larger systems, we will now analyze a peptide model where a quantum chemical reference is accessible. This will allow us to identify the relative weights of attractive and repulsive NCI contributions. Note that this is not feasible without overfitting for the small S66 set.

We have carried out a classical molecular dynamics (MD) simulation with a simple peptide model and extracted a series of frames to be used as the training set for a linear regression (Ridge Regression) to determine how NCI quantities are related to attractive, van der Waals, and repulsive contributions to the interaction energy. Our system is the dimer of the peptide NYNYN, composed of asparagine and tyrosine residues. These amino acids are rich in polar groups and aromatic rings (tyrosine residues), allowing us to explore a wide range of noncovalent interactions in a compact system. Each monomer has an acetylated N-terminus and an amidated C-terminus, so the overall structure is neutral. We carried out a 10 ns classical MD simulation, selecting 1000 equally spaced frames.

In order to have a good reference interaction energy, each frame was then computed at the BSSE-corrected PBE-D3 level with the 6-31G+(d,p) basis set.[25,26] The DFT calculation yields a wide range of values, going from −24.08 to −76.32 kcal/mol, which enables exploration of the NCI capabilities. Further details about these calculations are reported in the Supporting Information.

The NCIPLOT integrals have been split into different ranges of interaction to cover the different interaction types. We have used ranges of $\text{sign}(\lambda_2)\rho$ from −0.1 to −0.02 corresponding to attractive interactions, from −0.02 to 0.02 for van der Waals, and from 0.02 to 0.1 assigned to repulsive interactions (see the Supporting Information for full data). Stability with respect to these ranges has also been verified.

The data set was split into a training set used for fitting with a Ridge Regression (700 structures) and then the linear model was used to predict the interaction energies on the remaining 300 structures.[27] Results are shown in Figure 6. The average error of prediction is 9%, with a training score of 0.62 and a test score of 0.59 (more details about the fitting procedure are

https://dx.doi.org/10.1021/acs.jctc.0c00063
J. Chem. Theory Comput. 2020, 16, 4150−4158

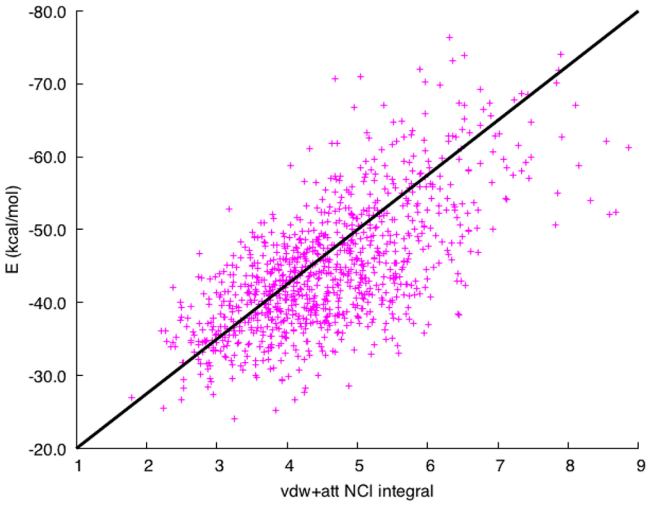


Figure 6. BSSE-corrected interaction energies plotted against the sum of attractive and van der Waals integrals (n = 1) for the 1000 frames extracted from the simulation of the NYNYN dimer. The standard deviation is 1.287, and the mean absolute percentage error is 11.68%. When Ridge Regression is used, the average error of prediction on the test set goes down to 9%.

reported in the Supporting Information). These results indicate that there is a good correlation between NCI integrals and the interaction energy. Further analysis on the relationship between NCI integrals and interaction energies in biomolecules is discussed in ref 28.

On top of the charges, volumes were also considered. The linear coefficients obtained in the regression are the following: for the attractive, van der Waals, and repulsive charges, respectively, we obtained −8.24, 8.16, and 1.88, while for the volumes we obtained −0.23, −0.12, and 0.46 for the three terms in the same order. Since the volume integral is defined by $I_{n=0}$, these coefficients confirm the results from the previous section, i.e., the relevance of the charge over the volume of the region. Moreover, for the peptide sequence, the magnitude of the coefficients shows that the attractive and van der Waals terms are more relevant than the repulsive one. This is due to the fact that in absolute terms the repulsive integral is significantly lower than the other two. This points to the fact that the system strives to minimize repulsive interactions.

Most importantly, these preliminary results suggest that the correlation between DFT results and NCI properties (which can be easily calculated for large systems using the promolecular density) is good enough to envisage the development of scoring functions for noncovalently bonded complexes where a good computational speed−accuracy is needed. Indeed, preliminary results of some of us have shown that the inclusion of NCI information in the parametrization of affinity functions enables them to outperform existing comparable protein−protein affinity estimation methods.[29]

Localization of the Energetic Contributions. Probably the greatest advantage up to now of the NCI approach was its structural information. As opposed to delocalized wave function methods, NCI provides information on which atoms are contributing to the (de)stabilization. It is important to verify that the energetic contributions are related to the changes in the structure. Due to their relevance, conformational changes in biosystems constitute a good example of this structure−energy relationship.

We have checked the local energy connection in a Aβ40 peptide. A 200 ns classical simulation in aqueous solution has been run for a single peptide extracted from PDB 2M4J,[30] a mature amyloid fibril. The aim was to study how the secondary structure adapts from the mature fibril β architecture to the aqueous medium.

The initial and final geometries of the Aβ monomer are reported in Figure 7; after 200 ns, an α-helix is spontaneously formed in the C-terminus region, encompassing residues 28−37 (KGAIIGLMVG).

Figure 8a shows the evolution of the Aβ40 secondary structure along the 200 ns trajectory, where the formation of the α-helix between residues 28 and 37 takes place after approximately 100 ns. The plot shows the color coded secondary structure for each type of residue along the trajectory frames. The evolution over the same time frame of the integral of the noncovalent interactions density in the range $-0.02 \leq \mathrm{sign}(\lambda_2)\rho \leq 0.02$ au in residues 28−37, corresponding to the weakest noncovalent contacts, is shown in Figure 8b.

Comparing parts a and b of Figure 8, we can observe that the formation of the α-helix motif (in red in Figure 8a) is accompanied by a significant increase of the integral of ρ in Figure 8b. The value of this integral mirrors the tighter packing of KGAIIGLMVG as the secondary structure evolves from the fibril-like geometry to the equilibrium one for the monomer in solution, which has a high α-helix content.[31] It should be noted that, when the full system is subjected to the same analysis, the local information is blurred by the other residues (see Figure S8).

This example proves that the integrals of the promolecular density in the noncovalent interaction regions carry compact

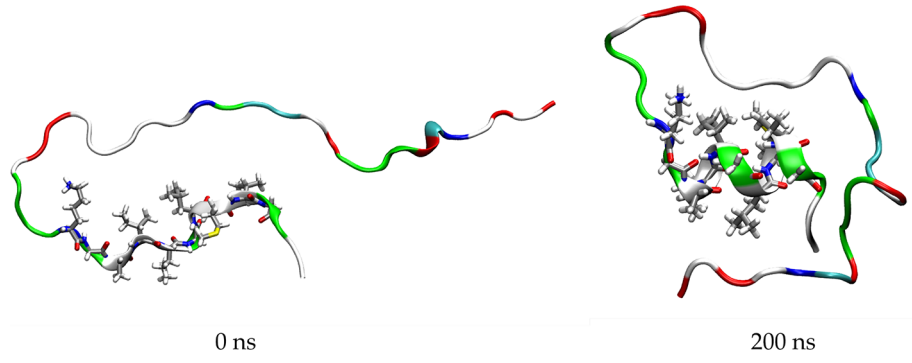
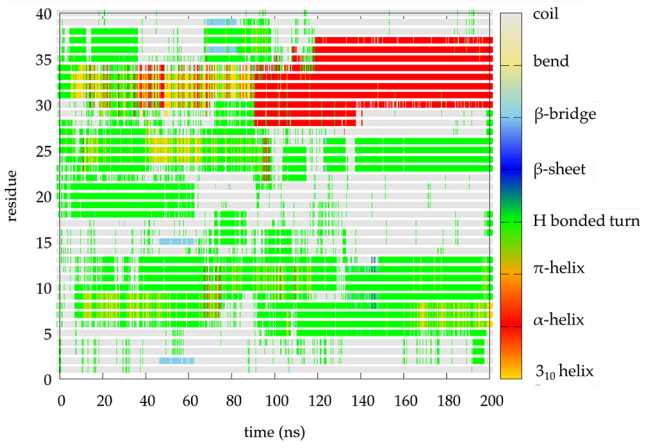
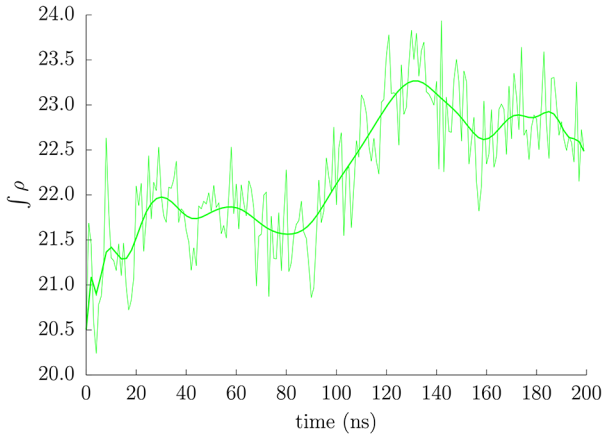


Figure 7. Aβ40 monomer at the initial geometry (PDB) 2M4J and after 200 ns of classical dynamics.

https://dx.doi.org/10.1021/acs.jctc.0c00063
J. Chem. Theory Comput. 2020, 16, 4150−4158



(a) secondary structure



(b) integral of ρ in $-0.02 \leq \mathrm{sign}(\lambda_2)\rho \leq 0.02$ a.u. for residues 28 to 37.

Figure 8. Evolution of noncovalent interactions along the 200 ns simulation of the Aβ monomer.

information on the temporal evolution of the system and can be linked to both structural and energetic features. Indeed, the increase of α-helix content is reflected in the stabilization of the system, as shown in Figure 9.

Overall, this new local energy analysis should enable analysis of the evolution of weak interactions in biological systems by providing information on attractive, repulsive, and van der Waals interactions on both the energy and the position in a fast and efficient manner that goes well beyond distance analysis.

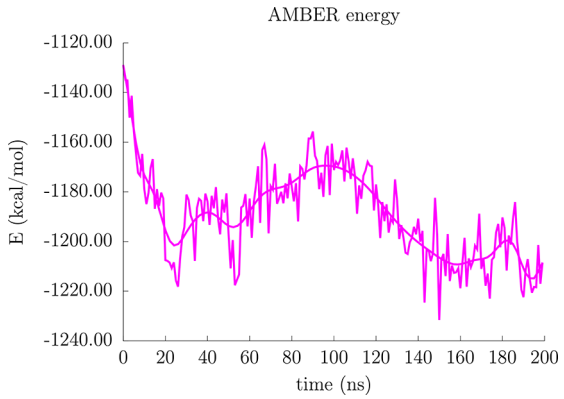


Figure 9. Evolution of the AMBER energy along the 200 ns simulation of the Aβ monomer.

Special fields of applications are protein−protein and protein−ligand interactions and conformational analysis, which are at the heart of biological and pharmacological developments.

CONCLUSIONS AND OUTLOOK

We have presented the new general structure and capabilities of the new release of NCIPLOT, NCIPLOT4, which is meant to provide access to a quantification of NCI properties in systems ranging from a few atoms to several hundred atoms. Two main implementations have been carried out thereto: integrals over promolecular densities and an acceleration through a multilevel grid and parallelization. The advantages of the new version allow locally following structural and energetic changes in large systems.

After a proof of concept on small systems, two examples on the potentiality of the new code for larger systems have been presented. We have showcased how large structures, including molecular dynamics results, can be analyzed in a fast and visual manner thanks to the new code. Given the use of promolecular densities, rather good correlations are obtained between the attractive (hydrogen bond and van der Waals) NCI integrals used in small systems and our reference DFT energies computed also with small basis sets. The remarkable affordability of this descriptor allows us to treat increasingly large systems and supports previous observations of qualitative NCI−energetic connections.[7,32] It enables identification of the charge within the NCI region as the main source of information. The optimized exponent is closed to the GGA exchange term. However, further physical insight for the analytical connection needs to be established.

Going toward bigger systems, we show the advantage of simultaneously having access to the energetics and the position of the interactions. We exploit this information coupled to the time coordinate, showing how the NCIPLOT4 implementation enables identification of the formation of new secondary structures: when it appears and which residues are involved.

Overall, the combined use of promolecular densities and multiple integration ranges makes NCI properties an extremely flexible tool for the study of complex systems: fragments for NCI calculation can be defined picking any combination of atoms, which allows the user to focus on any subsystem within a larger one, and the flexibility in choosing the integration ranges allows the user to focus on a particular subset of interaction strength. The combination of the two allows for the effective decomposition of any complex interaction network at will and for the calculation of the contribution of each NCI component to the overall network.

The program is open source, available under a GNU General Public License at https://github.com/juliacontrerasgarcia/nciplot.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.0c00063.

Two-dimensional plots along the dynamics trajectories; adaptative grid I/O example; integral I/O examples; data for plots benchmark integrations (PDF)

This video shows the temporal evolution of NCIs ligand and receptor of the Thioflavin-T/Aβ40 and NIAD-4/Aβ40 complexes along a molecular dynamics simulation. Despite being accommodated in a narrow binding

https://dx.doi.org/10.1021/acs.jctc.0c00063
J. Chem. Theory Comput. 2020, 16, 4150−4158

channel, the two ligands undergo significant fluctuations of the NCIs contacts along the trajectory. For this reason, direct comparison of the 2D NCI plots along the trajectory does not convey useful information on the strength or stability of ligand-receptor NCIs. Conversely, integration analysis allows one to distinguish between the two ligands, assessing whether their network of NCIs with the receptor is stable or not (see main text) (MOV)

AUTHOR INFORMATION

Corresponding Author

Julia Contreras-García − *Laboratoire de Chimie Théorique (LCT), Sorbonne Université, CNRS, 75005 Paris, France;* orcid.org/0000-0002-8947-9526; Email: contrera@lct.jussieu.fr

Authors

Roberto A. Boto − *Laboratoire de Chimie Théorique (LCT), Sorbonne Université, CNRS, 75005 Paris, France; Materials Physics Center, CSIC-UPV/EHU, 20018 Donostia-San Sebastián, Spain;* orcid.org/0000-0002-8012-1499

Francesca Peccati − *Laboratoire de Chimie Théorique (LCT), Sorbonne Université, CNRS, 75005 Paris, France;* orcid.org/0000-0002-7813-8216

Rubén Laplaza − *Laboratoire de Chimie Théorique (LCT), Sorbonne Université, CNRS, 75005 Paris, France; Departamento de Química Física, Universidad de Zaragoza, 50009 Zaragoza, Spain;* orcid.org/0000-0001-6315-4398

Chaoyu Quan − *SUSTech International Center for Mathematics, and Department of Mathematics, Southern University of Science and Technology, 518055 Shenzhen, China; Institut des Sciences du Calcul et des Données (ISCD), Sorbonne Université, 75005 Paris, France*

Alessandra Carbone − *Laboratoire de Biologie Computationnelle et Quantitative (LCQB), Sorbonne Université, CNRS, IBPS, 75005 Paris, France; Institut Universitaire de France, 75005 Paris, France;* orcid.org/0000-0003-2098-5743

Jean-Philip Piquemal − *Laboratoire de Chimie Théorique (LCT), Sorbonne Université, CNRS, 75005 Paris, France; Institut Universitaire de France, 75005 Paris, France;* orcid.org/0000-0001-6615-9426

Yvon Maday − *Laboratoire Jacques-Louis Lions (LJLL), Sorbonne Université, Université Paris-Diderot SPC, CNRS, F-75005 Paris, France; Institut Universitaire de France, 75005 Paris, France*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.0c00063

Funding

All authors express their gratefulness to LabEx CALSIMLAB (public grant ANR-11-LABX-0037-01 constituting a part of the "Investissements d'Avenir" program, reference ANR-11-IDEX-0004-02). This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 810367), project EMC2 (J.-P.P. and Y.M.). A.C., J.-P.P., and Y.M. acknowledge funding from the Institut Universitaire de France. R.L. thanks ED 388 for a Ph.D. grant. Y.M. and C.Q. acknowledge funding from PICS-CNRS and PHC PROCOPE 2017 (Project No. 37855ZK).

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Lu, W.; Chan, 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.

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

(3) Bader, R. F. *Atoms in Molecules*; Wiley Online Library: 1990.

(4) Becke, A. D.; Edgecombe, K. E. A simple measure of electron localization in atomic and molecular systems. *J. Chem. Phys.* **1990**, *92*, 5397−5403.

(5) Kohout, M. A measure of electron localizability. *Int. J. Quantum Chem.* **2004**, *97*, 651−658.

(6) 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.

(7) Contreras-García, J.; Johnson, E. R.; Yang, W. Analysis of hydrogen-bond interaction potentials from the electron density: integration of noncovalent interaction regions. *J. Phys. Chem. A* **2011**, *115*, 12983−12990.

(8) Saleh, G.; Gatti, C.; Lo Presti, L. Energetics of non-covalent interactions from electron and energy density distributions. *Comput. Theor. Chem.* **2015**, *1053*, 53−59.

(9) Armstrong, A.; Boto, R. A.; Dingwall, P.; Contreras-Garcia, J.; Harvey, M. J.; Mason, N.; Rzepa, H. S. The Houk-List Transition states for organocatalytic mechanism revisited. *Chem. Science* **2014**, *5*, 2057.

(10) 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.

(11) Contreras-Garcia, 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.

(12) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865.

(13) Spackman, M. A.; Maslen, E. N. Chemical properties from the promolecule. *J. Phys. Chem.* **1986**, *90*, 2020.

(14) Meyer, B.; Guillot, B.; Ruiz-Lopez, M. F.; Genoni, A. Are Bond Critical Points Really Critical for Hydrogen Bonding? *J. Chem. Theory Comput.* **2016**, *12*, 1052.

(15) Rezác, 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.

(16) Peccati, F.; Desmedt, E.; Contreras-García, J. A Regression Approach to Accurate Interaction Energies Using Topological Descriptors. *Comput. Theor. Chem.* **2019**, *1159*, 23−26.

(17) Goerigk, L.; Hansen, A.; Bauer, C. A.; Ehrlich, S.; Najibi, A.; Grimme, S. A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Phys. Chem. Chem. Phys.* **2017**, *19*, 32184−32215.

(18) Brauer, B.; Kesharwani, M. K.; Kozuch, S.; Martin, J. M. The S66x8 benchmark for noncovalent interactions revisited: Explicitly correlated ab initio methods and density functional theory. *Phys. Chem. Chem. Phys.* **2016**, *18*, 20905−20925.

(19) Kolesnichenko, I. V.; Anslyn, E. V. Practical applications of supramolecular chemistry. *Chem. Soc. Rev.* **2017**, *46*, 2385−2390.

(20) Stoddart, J. F. The chemistry of the mechanical bond. *Chem. Soc. Rev.* **2009**, *38*, 1802−1820.

(21) Lee, S.; Chen, C.-H.; Flood, A. H. A pentagonal cyanostar macrocycle with cyanostilbene CH donors binds anions and forms dialkylphosphate (3) rotaxanes. *Nat. Chem.* **2013**, *5*, 704−710.

(22) Liu, Y.; Zhao, W.; Chen, C.-H.; Flood, A. H. Chloride capture using a C−H hydrogen-bonding cage. *Science* **2019**, *365*, 159−161.
(23) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297−3305.
(24) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456−1465.
(25) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(26) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
(27) Pedregosa, F.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.
(28) Peccati, F. The NCIPLOT4 Guide for Biomolecules: an Analysis Tool for Non-Covalent Interactions. *J. Chem. Inf. Model.* **2020**, *60*, 6−10.
(29) Raucci, R.; Laine, E.; Carbone, A. Local Interaction Signal Analysis Predicts Protein Protein Binding Affinity. *Structure* **2018**, *26*, 905−915.
(30) Lu, J.-X.; Qiang, W.; Yau, W.-M.; Schwieters, C.; Meredith, S.; Tycko, R. Molecular Structure of β-Amyloid Fibrils in Alzheimer's Disease Brain Tissue. *Cell* **2013**, *154*, 1257−1268.
(31) Sticht, H.; Bayer, P.; Willbold, D.; Dames, S.; Hilbich, C.; Beyreuther, K.; Frank, R. W.; Rösch, P. Structure of Amyloid A4-(1−40)-Peptide of Alzheimer's Disease. *Eur. J. Biochem.* **1995**, *233*, 293−298.
(32) Alonso, M.; Woller, T.; Martín-Martínez, F. J.; Contreras-García, J.; Geerlings, P.; De Proft, F. Understanding the fundamental role of the pi/pi sigma/sigma and sigma/pi dispersion interaction in shaping carbon-based materials. *Chem. - Eur. J.* **2014**, *20*, 4931−4941.

https://dx.doi.org/10.1021/acs.jctc.0c00063
J. Chem. Theory Comput. 2020, 16, 4150−4158