JCTC Journal of Chemical Theory and Computation

ARTICLE

pubs.acs.org/JCTC

NCIPLOT: A Program for Plotting Noncovalent Interaction Regions

Julia Contreras-García,[†] Erin R. Johnson,[‡] Shahar Keinan,[†] Robin Chaudret,[§,∥] Jean-Philip Piquemal,[§,∥] David N. Beratan,[†] and Weitao Yang*[,†]

[†]Department of Chemistry, Duke University, Durham, North Carolina 27708

[‡]School of Natural Sciences, University of California, Merced, 5200 North Lake Road, Merced, California 95343

[§]UPMC Univ Paris 06, UMR 7616, Laboratoire de Chimie Théorique, case courrier 137, 4 place Jussieu, F-75005, Paris, France

[∥]CNRS, UMR 7616, Laboratoire de Chimie Théorique, case courrier 137, 4 place Jussieu, F-75005, Paris, France

Supporting Information

ABSTRACT: Noncovalent interactions hold the key to understanding many chemical, biological, and technological problems. Describing these noncovalent interactions accurately, including their positions in real space, constitutes a first step in the process of decoupling the complex balance of forces that define noncovalent interactions. Because of the size of macromolecules, the most common approach has been to assign van der Waals interactions (vdW), steric clashes (SC), and hydrogen bonds (HBs) based on pairwise distances between atoms according to their vdW radii. We recently developed an alternative perspective, derived from the electronic density: the non-covalent interactions (NCI) index [*J. Am. Chem. Soc.* **2010**, *132*, 6498]. This index has the dual advantages of being generally transferable to diverse chemical applications and being very fast to compute, since it can be calculated from promolecular densities. Thus, NCI analysis is applicable to large systems, including proteins and DNA, where analysis of noncovalent interactions is of great potential value. Here, we describe the NCI computational algorithms and their implementation for the analysis and visualization of weak interactions, using both self-consistent fully quantum-mechanical as well as promolecular densities. A wide range of options for tuning the range of interactions to be plotted is also presented. To demonstrate the capabilities of our approach, several examples are given from organic, inorganic, solid state, and macromolecular chemistry, including cases where NCI analysis gives insight into unconventional chemical bonding. The NCI code and its manual are available for download at http://www.chem.duke.edu/~yang/software.htm.

I. INTRODUCTION

Noncovalent interactions are of critical importance in many chemical, biological, and technological systems. Protein−ligand interactions,[1] the self-assembly of nanomaterials,[2] the folding of proteins,[3] and the packing of molecular crystals[4] are controlled by a delicate balance of numerous, weak noncovalent interactions. The large sizes of the materials and structures of interest makes understanding noncovalent forces particularly difficult.

Accurately describing noncovalent interactions, including their spatial characteristics, constitutes the first step in the process of decomposing the complex balance of chemical forces. However, even this spatial characterization of interactions has generated controversy.[5−8] Because of the size of many systems of interest, the most common approach to examining noncovalent interactions has been to assign van der Waals interactions (vdW), steric clashes (SC), and hydrogen bonds (HBs) in terms of pairwise distances between atoms based on their vdW radii.[9] This procedure, although lacking generality and giving rise to systematic errors (e.g., the length of HBs is generally overestimated), enables the rapid enumeration of noncovalent interactions.

More elaborate algorithms for mapping and analyzing noncovalent interactions can be derived from the electronic and kinetic-energy densities. Indeed, the theory of atoms in molecules[10] has been used to understand and to quantify weak interactions on the basis of the electron density.[10−12] This approach relies on the fact that critical points of the density ($\nabla\rho = 0$) arise when atoms interact. If the interaction is bonding, the point is expected to be a first order saddle point. Many properties, including the density itself, its Laplacian ($\nabla^2\rho$), and the kinetic-energy density, have been found to correlate with the interaction energy in vdW and HB complexes for families of related compounds.[11,12] A parallel approach is based on the analysis of the electron localization function (ELF).[13,14] This is a function of the electron density and the kinetic-energy density that was developed to highlight regions of electron localization (i.e., covalent bonds, lone pairs, etc.). HBs have also been studied using this approach, since the ELF value is proportional to the strength of the HB.[15,16]

Recently, we introduced a new approach to visualize noncovalent interactions, based on the analysis of the electron densities and their reduced gradients, $s(\vec{r})$.[17] This approach was chosen for its ability to highlight interactions in the low-density regime (*vide infra*). The non-covalent interactions (NCI) index identifies interactions in a chemical system solely on the basis of the electron density and its derivatives. Indeed, a similar approach was recently introduced to analyze covalent bonds, highlighting the ability of this type of function to identify all types

Received: November 8, 2010
Published: January 25, 2011

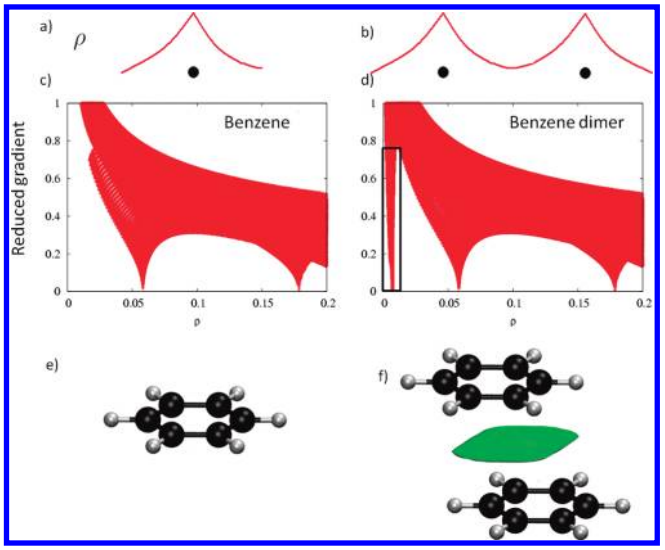

Figure 1. (a) Representative behavior of atomic density. (b) Appearance of a $s(\rho(\vec{r}))$ singularity when two atomic densities approach each other. (c, d) Comparison of the reduced density behavior for the benzene monomer and dimer; a singularity in s appears at low density values in the dimer case. (e) Benzene monomer. (f) Appearance of an intermolecular interaction surface in the benzene dimer, associated with the additional singularity in the $s(\rho)$ plot. The isosurface was generated for $s = 0.7$ au and $\rho < 0.01$ au.

of bonding situations.[18] Here, we introduce a practical strategy for NCI analysis and visualization of weak interactions.

The paper first summarizes NCI theory. Next, the algorithms underlying NCI analysis are discussed. Finally, we apply the approach to chemical examples that span a diverse set of interaction types and system sizes. Emphasis is placed on the ability of the new approach to describe the bonding in systems currently under debate.

Details of the computations are found in the Supporting Information. The software and manual may be downloaded at http://www.chem.duke.edu/~yang/software.htm

II. THEORETICAL BACKGROUND

The NCI analysis provides an index, based on the electron density and its derivatives, that enables identification of noncovalent interactions.[17] The NCI index is based on a 2D plot of the reduced density gradient, s, and the electron density, ρ, where

$$s = \frac{1}{2(3\pi^2)^{1/3}} \frac{|\nabla\rho|}{\rho^{4/3}} \quad (1)$$

When a weak inter- or intramolecular interaction is present, there is a crucial change in the reduced gradient between the interacting atoms, producing density critical points between interacting fragments (Figure 1a,b). Troughs appear in $s(\rho)$ associated with each critical point. Since the behavior of s at low densities is dominated by ρ, s tends to diverge except in the regions around a density critical point, where $\nabla\rho$ dominates, and s approaches zero. This fact is highlighted in Figure 1c,d, which shows that the main difference in the $s(\rho)$ plots between a monomer and a dimer is the steep trough at low density. When we search for the points in real space giving rise to this feature, the noncovalent region clearly appears in the (supra)molecular complex (green isosurface in Figure 1f).

Further analysis of the electron density in the troughs is required to assign the origins of these troughs (steric interactions, hydrogen bonds, etc.). The electron density values within the troughs are an indicator of the interaction strength. However, both attractive and repulsive interactions (i.e., hydrogen-bonding and steric repulsion) appear in the same region of density/reduced gradient space. To distinguish between attractive and repulsive interactions, we examine the second derivatives of the density along the main axis of variation.

On the basis of the divergence theorem,[19] the sign of the Laplacian ($\nabla^2\rho$) of the density indicates whether the net gradient flux of density is entering ($\nabla^2\rho < 0$) or leaving ($\nabla^2\rho > 0$) an infinitesimal volume around a reference point. Hence, the sign of $\nabla^2\rho$ determines whether the density is concentrated or depleted at that point, relative to the surroundings. To distinguish between different types of weak interactions, one cannot use the sign of the Laplacian itself, because the sign is dominated by negative contributions from the nuclei.[20] Instead, contributions to the Laplacian along the axes of its maximal variation must be analyzed. These contributions are the eigenvalues λ_i of the electron-density Hessian (second derivative) matrix, such that $\nabla^2\rho = \lambda_1 + \lambda_2 + \lambda_3$, $(\lambda_1 < \lambda_2 < \lambda_3)$. At the nuclei, all of the eigenvalues are negative, while away from them, $\lambda_3 > 0$. In molecules, the λ_3 values vary along the internuclear direction, while λ_1 and λ_2 report the variation of density in the plane normal to the λ_3 eigenvector. Interestingly, the second eigenvalue (λ_2) can be either positive or negative, depending on the interaction type. On the one hand, bonding interactions, such as hydrogen bonds, are characterized by an accumulation of density perpendicular to the bond, and $\lambda_2 < 0$. Nonbonded interactions, such as steric repulsion, produce density depletion, such that $\lambda_2 > 0$. Finally, vdW interactions are characterized by a negligible density overlap that gives $\lambda_2 \lesssim 0$. Thus, analysis of the sign of λ_2 enables us to distinguish different types of weak interactions, while the density itself enables us to assess the interaction strength.

The dependence of $s(\vec{r})$ on $\text{sign}(\lambda_2)\rho$ is shown in Figure 2a, which uses a modification of our earlier reduced gradient and density plots. The 3D plots resulting from using $\text{sign}(\lambda_2)\rho$ are shown in Figure 2, bottom. The low-density, low-reduced gradient trough in the hydrogen-bonded water dimer now lies at negative values, indicative of an attractive interaction. Conversely, the low-density, low-gradient trough for the sterically crowded bicyclo[2,2,2]octane molecule remains at positive values, indicating a repulsive interaction. Finally, the low-density, low-gradient trough for the dispersion-bound methane dimer is very near zero, with slightly negative $\text{sign}(\lambda_2) \times \rho$ negative values of -0.0025 au, indicative of weak attraction.

III. ALGORITHMIC DETAILS

Figure 3 shows the protocol for visualizing noncovalent interactions in NCIPLOT. A detailed summary of the tasks associated with each routine is shown in Table 1 (cube construction, properties, visualization, and I/O flow).

The key data-flow details are highlighted in Figure 3. The input of data is shown in red, and its processing is indicated in black. Two basic types of data constitute the input: the density information (based on wave functions or molecular geometries) and the analysis options, which determine the noncovalent interactions to be plotted. Four algorithms analyze the data: (i) the selection of interactions (through the input), (ii) the construction of the cube and the grid (in CUBE, see Table 1),

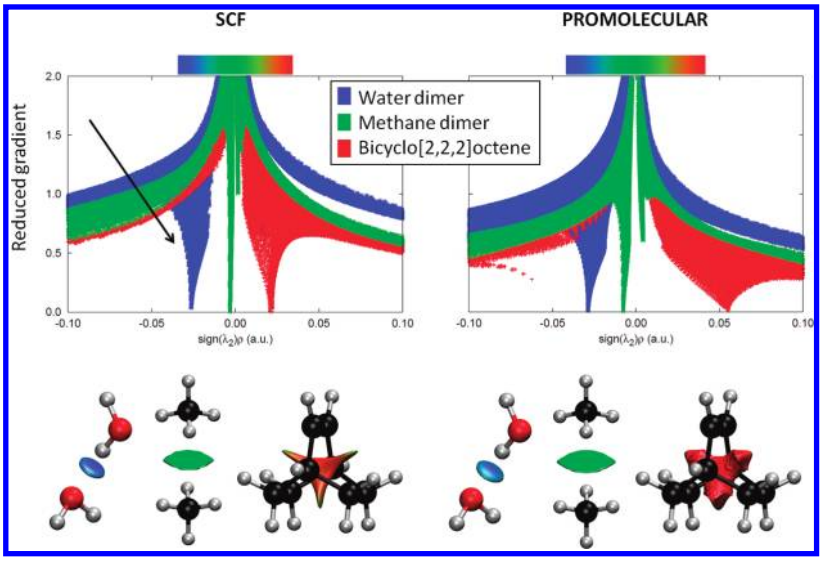

Figure 2. Top: Overlapping troughs in $s(\rho)$ plots, which can be distinguished when $\mathrm{sign}(\lambda_2)\rho$ is used as the ordinate. Favorable interactions appear on the left, unfavorable on the right, and vdW near zero. The same $s(\rho)$ features are obtained using self-consistent (left) and promolecular (right) calculations, with a shift toward negative (stabilizing) regimes. Bottom: Taking the shift in troughs into account (i.e., changing the cutoff), the isosurface shapes remain qualitatively unaltered for selected small molecules. Figures are shown for both SCF (left) and promolecular densities (right). NCI surfaces correspond to $s = 0.6$ au and a color scale of $-0.03 < \rho < 0.03$ au for SCF densities. For promolecular densities, $s = 0.5$ au (water and methane dimers) or $s = 0.35$ au (bicyclo[2,2,2]octene), and the color scale is $-0.04 < \rho < 0.04$ au.

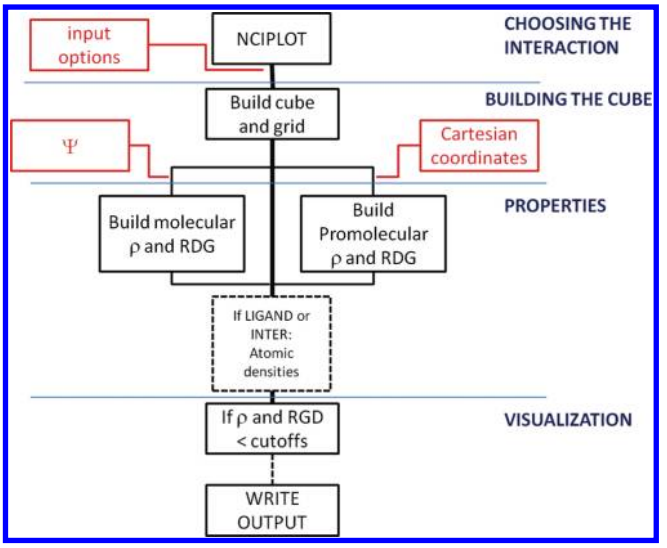

Figure 3. Flowchart for program routines for noncovalent interactions' visualization in NCIPLOT. Red labels highlight the information that can be input by the user, whereas black labels show the internal flow of information. The flow is divided into four main algorithmic parts: input, cube construction, properties, and visualization.

(iii) the calculation of properties at each point (utilizing a number of routines), and (iv) the calculation of visualization data (carried by the main routine, NCIPLOT). Since the input is keyword oriented, the program includes a number of parsing routines. These main features are discussed in the following sections.

III.A. Building the Cube. Interaction analysis is based on examination of local properties on a cubic grid constructed within the program. This procedure was found to be extremely efficient for computing stable properties (as is the case of NCI, see section III.B).[21] Furthermore, this approach enables us to discard contributions from high-density points in the construction of isosurfaces. The spatial region to be analyzed is determined, by default, in terms of the molecular geometry. Unless otherwise noted, a cube is constructed from the outermost x,y,z coordinates for all of the molecules in the input. An extra radial threshold in each direction is added to ensure that the isosurfaces are contained within the cube (no intermolecular interactions are expected in those regions, but isosurfaces can spread beyond the atoms). A practical threshold was defined as 2 Å:

$$x_i(0) = \min[x_i] - 2\ \text{Å} \quad (2)$$

$$x_i(1) = \max[x_i] + 2\ \text{Å} \quad (3)$$

where $x_i = x, y, z$. This step also eliminates spurious symmetry-related cancellations (e.g., if benzene were analyzed without this extra threshold, there would be no points along the C_6 axis).

It can be useful to construct a user-defined cube or to analyze the interactions only around one point or molecule (*vide infra*). All of these options are implemented in the program and are described in the Supporting Information.

III.B. Properties. Density, reduced gradient, and λ_2 values are calculated at each point on the grid. Densities can be obtained from quantum-mechanical calculations or from promolecular estimates. Topological features of the electron density in the weak interaction region are very stable with respect to the calculation method, to such an extent that these features are already contained in the sum of atomic densities (ρ_i^{at}).[22,23] The molecular density computed from the sum of atomic contributions, also known as promolecular density (ρ^{pro}), is

$$\rho^{\mathrm{pro}} = \sum_i \rho_i^{\mathrm{at}} \quad (4)$$

Promolecular densities lack the relaxation introduced in a SCF Hartree−Fock or DFT calculation; however, the promolecular densities are very useful for describing large biomolecular

Table 1. NCIPLOT Tasks and Details of the Input and Output[a]

MODULE	ROUTINE	TASK	INPUT	OUTPUT
MAIN	NCIPLOT	main routine		visualization
CUBE	CUBE	constructs cube and grid	geometry	cube, grid
I/O	TIMER	accumulates run times	process and resetting	elapsed times
	GETARGS	get command	arguments	argument count
	GETDATE	extracts time and date	(operatingsystem)	date and time
	ZATGUESS	provides atomic #s	atomic symbol	atomic #
	RWFN	stores wfn information	wfn file	wfn information
	RPROM	stores xyz information	xyz file	xyz information
PROPS	PROPPROM	calculates properties at $\vec{r}$ (PROM)	$\vec{r}$	ρ, $s(\rho)$
	CALCHESS	calculates λ_2 at $\vec{r}$ (PROM)	$\vec{r}$	λ_2
	PROPWFN	calculates properties at $\vec{r}$ (SCF)	$\vec{r}$	ρ,s (ρ), λ_2
	F012	derivatives at $\vec{r}$ (SCF)	$\vec{r}$	derivatives
	PHI012	summation over primitives	$\vec{r}$	summation
	PRI012	construction of primitives	$\vec{r}$	primitives
	INDEX0	angular momentum assignation	primitive type	L_i
	RS	matrix diagonalization (eispack)	matrix	eigenvalues

[a] The table is organized following the main algorithmic subdivisions of the flowchart (Figure 3): cube construction, properties, and visualization, as well as I/O flow. wfn file stands for the wavefunction file (wfn extension) commonly used for post-SCF analyses.

systems. In these systems, noncovalent interactions are crucial for describing the interplay of structure and reactivity.[3] Because electron density calculations for these large systems are extremely expensive computationally, use of the promolecular density is an attractive option. When relaxed densities are compared to promolecular ones, a shift in the $s(\rho)$ troughs is observed toward bonding regimes (see Figure 2 top). Specifically, a large shift toward smaller density values is observed in the troughs corresponding to regions of nonbonded overlap, introducing less repulsion and greater stability. However, once this shift is taken into account (by changing the density cutoff), results at the self-consistent and promolecular levels are qualitatively equivalent for all of the cases considered (see Figure 2, bottom).

SCF densities are constructed from the wave function information stored in the wfn file, whereas promolecular densities are constructed from the atomic positions stored in the xyz coordinate file(s). In order to store atomic densities, fully numerical LSDA[24] free-atomic densities were generated for the neutral atoms H to Ar, spherically averaged over space and summed over spins. Because atomic densities are piece-wise exponentially decaying for each shell of electrons, they were then fit to one (H, He), two (Li−Ne), or three (Na−Ar) Slater-type functions of the form $\rho^{at} = \sum_j c_j e^{-\mathbf{r}/\zeta_j}$. Once these densities are written as simple sums of exponential functions, the NCI surfaces can be calculated very efficiently for each (supra)molecule, since all of the necessary data (ρ, s, λ_2) can be obtained analytically.

III.C. Visualization: The Cutoffs. The ρ,s coordinates of the density troughs define the appropriate cutoffs for the noncovalent interactions. For example, a cutoff of $\rho < 0.05$ au is appropriate for recovering the noncovalent interactions in the benzene dimer, including the nonbonding regions at the center of each ring (Figure 1c,d). All points giving rise to ρ values above this threshold need to have their s values set to a large value. This enables the user to recover only the noncovalent interactions when $s = S$ is plotted (for some isosurface value S) because this rescaling eliminates all points in density/reduced gradient space with greater density values and $s < S$ (i.e., points with $\rho > 0.05$ au and $s < 0.5$ au would also appear in the isosurface representation

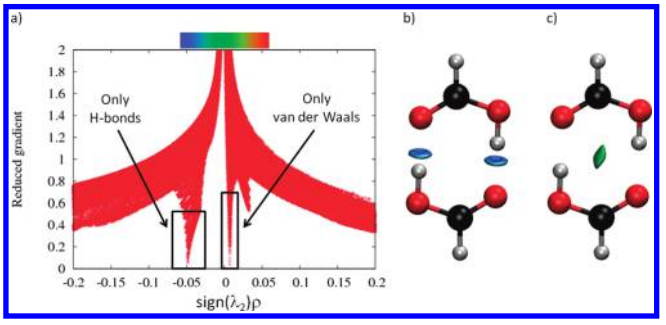

Figure 4. NCI analysis of formic acid dimer. (a) $s(\rho)$ plot for the SCF density. Peaks appear at $\rho \simeq 0.01$ au for vdW and $\rho \simeq 0.05$ au for hydrogen bonds. (b) If the cutoffs are set at $s = 0.7$ au and $\rho < 0.02$ au, the isosurface only recovers the vdW interactions in the system. (c) If the cutoffs are set at $s = 0.5$ au and $0.02 < \rho < 0.06$ au, only the hydrogen bonds are displayed. The NCI color scale is $-0.06 < \rho < 0.06$ au.

in Figure 1). Another example is provided in Figure 4. The formic acid dimer troughs appear at $\rho = 0.01$ au for vdW contacts, and $\rho = 0.05$ au for hydrogen bonds. If the cutoff is set to $\rho = 0.02$ au, the isosurface will only recover the vdW interactions in the system (Figure 4). Furthermore, placing a threshold for the interval $\rho = [0.02-0.06]$ au enables the user to isolate the hydrogen bonds in a similar manner.

It is convenient, therefore, to perform a preliminary run, where only $s(\rho)$ values are produced, and the user can use these data to determine optimal cutoffs. A second run can subsequently target the noncovalent interactions in a given molecule with no interference from other density regions. For this reason, the current implementation enables the user to decide which file types are to be output.

III.D. Input: Selectively Displaying the Interactions. In order to plot a certain interaction selectively, there are several constraints that can be applied to display the weak interactions of interest. Criteria for filtering the interactions include the strength of the interaction, its localization in space, and its nature (whether it is intermolecular or intramolecular). The details for each of these cases are described below.

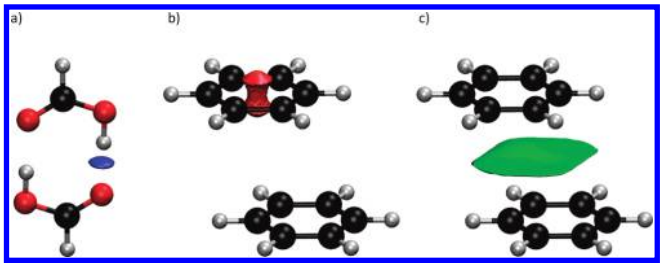

Figure 5. Results of several input options for the selective representation of a given interaction based on (a) its localization in 3D space, defined by the cube, (b) its localization in 3D space, defined by a radial threshold around a point (in this case, the center of the top benzene was chosen), (c) its inter-/intramolecular nature (in this case only intermolecular interactions are shown). NCI surfaces correspond to $s = 0.4$ au and a color scale of $-0.04 < \rho < 0.04$ au, using promolecular densities.

1. **Strength:** It is possible to select the interaction in terms of its strength by the choice of cutoff parameters. Figure 4 uses the information in section III.C to plot only the vdW contact region in the formic acid dimer, while avoiding the hydrogen bonds. It is also possible to define an interval range, as in the formic acid dimer case, to identify only hydrogen bonds and exclude vdW interactions. An interval of the kind $\rho = [0.02-0.06]$ au produces this result.
2. **Geometry:** There are two possibilities when choosing a given interaction from its location in 3D space:
   (i) An appropriate choice of the cube boundaries enables the selection of individual interactions (CUBE keyword in the Supporting Information). The cube in Figure 5a captures only one of the hydrogen bonds in the formic acid dimer. This option is especially suitable if there are several interactions of the same type, but only one of them is of interest. In this case, the strength criterion would not differentiate the interactions, and the geometric selection should be used.
   (ii) An alternate implementation for the geometric criterion consists of defining the center of the cube instead of its boundaries (RADIUS keyword in the Supporting Information). This option uses the origin and length of the box sides as input, rather than the Cartesian coordinates themselves. Results for the benzene dimer are shown in Figure 5b, where the steric clashes within a single ring are highlighted by searching for noncovalent interactions near the center of the upper benzene molecule.
3. **Pure Intermolecular Interactions:** All of the interactions with at least a specified fraction (e.g., $f = 0.9$) of the density from a single molecule are turned off:

$$\frac{\rho_{\text{molec}}}{\rho_{\text{tot}}} = \begin{cases} \geq f & \text{intramolecular} \\ < f & \text{intermolecular} \end{cases} \quad (5)$$

This choice causes only intermolecular interactions to be plotted, screening out the intramolecular interactions. Figure 5c shows results for the benzene dimer, where the internal steric repulsions are removed, since most of their density is obtained from a single benzene contribution. In order to construct molecular densities, atoms need to be assigned to each monomer. This is readily automated if each monomer is uploaded in a different file. This procedure enables the characterization of monomers and the construction of ρ_{molec}.

4. **Centered around a Given Molecule:** If the radial and the intermolecular options are used together, the intermolecular interactions around just one of the molecules will be highlighted (implemented via the LIGAND keyword, described in the Supporting Information). In this case, the desired cube is set around one entire molecule rather than around a given point. This option is particularly useful to study inclusion complexes and protein−ligand interactions, where a small molecule binds in a cavity and we wish to describe the interactions at this active site. In these cases, the interactions of interest are not only intermolecular but are localized around the smaller of the two partners (see protein−ligand interactions for HDAC8 protein in section IV.B.2).

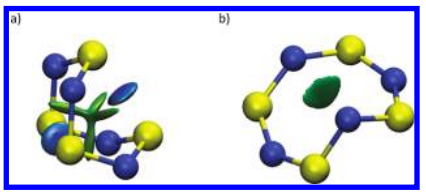

Figure 6. S_4N_4 main conformations. (a) Boat conformation b) Cage conformation. The boat conformation is more stable due to the bridging S−S bond. NCI surfaces correspond to $s = 0.4$ au and a color scale of $-0.05 < \rho < 0.05$ au.

IV. APPLICATIONS

Here, we review some examples where the applicability of NCIPLOT is illustrated for a range of systems (large and small molecules, inorganic complexes, and biological systems). Examples are chosen to highlight the capabilities of NCI and to shed light on challenging open issues as to the nature of specific interactions.

IV.A. Coordination Chemistry. *IV.A.1. S−S Bond in S_4N_4.* Tetrasulfur tetranitride, S_4N_4, is a textbook example that illustrates structure, bonding, and reactivity in main group inorganic compounds. Although discovered in 1835, S_4N_4 remains a subject of intensive study. S_4N_4 exists as an eight-membered ring. The most stable conformation is a D_{2d} "boat", in which the ring folds back on itself to give two close S−S contacts bridging the ring.[25] Another possible but less-stable conformer is the C_s "cage" structure. Both geometries were found in previous DFT calculations[26] and are shown in Figure 6. The greater stability of the boat conformation was explained by the presence of S−S bonding interactions.[27] These bonding interactions are clearly revealed by the NCI analysis and are predicted to be fairly strong, on the order of HB strengths, with $s(\rho)$ troughs at $\rho \sim 0.044$ au. There is also a nonbonding interaction through the center of the boat. Conversely, there is only a weak vdW interaction at the center of the more open cage structure. Thus, the NCI representation explains the greater stability for the boat conformation of S_4N_4.

IV.A.2. Hg Complexes. Understanding the solvation of ions and their interactions with ligands is of prime relevance for rationalizing their bioactivity. Since metals play a decisive role in many protein active sites as cofactors, it would be useful to have simple means of describing metal−ligand interactions.

Mercury(II) is a heavy metal cation especially challenging for quantum-mechanical treatment, as both correlation and relati-

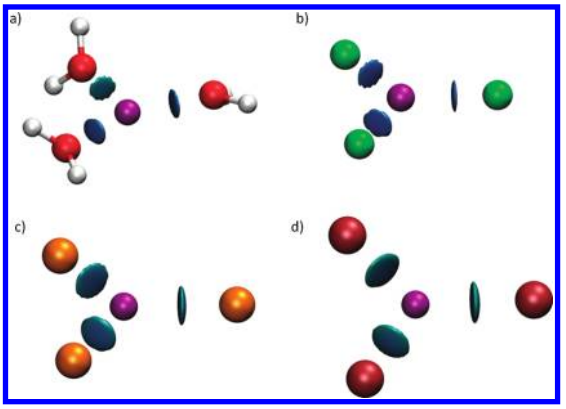

Figure 7. Complexes of Hg. (a) $[Hg(H_2O)_3]^{2+}$, (b) $[Hg(F)_3]^-$, (c) $[Hg(Cl)_3]^-$, (d) $[Hg(Br)_3]^-$. NCI surfaces correspond to $s = 0.3$ au and a color scale of $-0.1 < \rho < 0.1$ au.

vistic effects play a crucial role in its bonding and electronic structure. Figure 7a−d show how NCI is used to quickly and efficiently assess Hg complexation sites and discern the strength of binding between Hg^{2+} and its ligands. The DFT optimized geometry of the $[Hg(H_2O)_3]^{2+}$ complex indicates that the three waters bound to Hg are not equivalent, with one of the waters further away than the other (2.2 Å vs 2.3 Å, see Figure 7a). In order to analyze the nature of the stabilization, we decompose the interaction energy of the complex using the RVS[28] (reduced variational space) procedure. Both polarization and charge transfer are significantly weaker for one water molecule than for the other two. While two water molecules show a polarization energy of −15 to −16 kcal/mol and a charge transfer energy of −10 kcal/mol, the third water shows stabilization due to polarization and charge transfer of only −12.7 kcal/mol and −6.2 kcal/mol, respectively. Specific details of the calculation are found in the Supporting Information. The weaker binding of one of the water molecules is clearly revealed by the NCI representation.

NCI also recovers the ordering of binding energies when different ligand series are analyzed. Figure 7b−d show $[Hg(X)_3]$ complexes (X = F, Cl, or Br). F is more strongly bound to Hg than Cl than Br. This result is in agreement with the $[Hg(X)_3]^-$ binding energies, which are −632.8 (F), −571.0 (Cl), and −562.6 kcal/mol (Br) (see calculation details in the Supporting Information).

IV.A.3. BH_3NH_3 and the Dihydrogen Bond. Recently, the term "dihydrogen bond" was coined to describe an interaction of the type D−H···H−E, where D is a typical hydrogen donor (such as N or O).[40] The novelty of this bond is that the acceptor atom is also a hydrogen. Thus, the accepting hydrogen atom must be negatively charged, and E is an atom capable of accommodating a hydridic hydrogen. Transition metal and boron atoms are known examples of atoms serving as E atoms.

BH_3NH_3 is a widely studied example among dihydrogen-bonded complexes.[8,35,36] Crabtree et al.[37] placed the NH···HB contacts at the upper end of the energy range for hydrogen bonds. Popelier, instead, assigned these interactions to the range of normal H-bond strengths.[35] Morrison and Siddick[36] assigned the energies to the lower end of the hydrogen-bond-strength spectrum. The most relevant point made by the later authors is that there is a large conformational difference between the native crystal structure and the gas-phase minima considered by Crabtree et al.[37] and Popelier.[35]

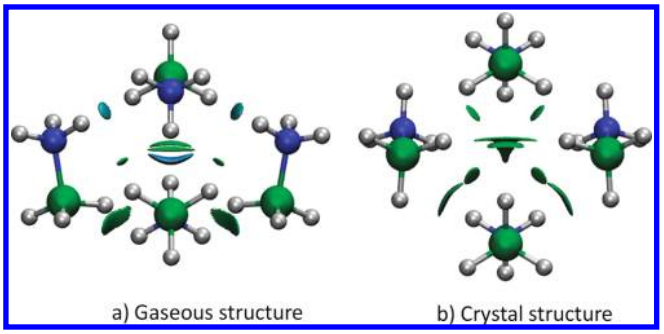

Figure 8. Dihydrogen interactions in a BH_3NH_3 tetramer in (a) the fully optimized gas-phase geometry and (b) the solid-state geometry. NCI surfaces correspond to $s = 0.4$ au and a color scale of $-0.03 < \rho < 0.03$ au.

Figure 8 shows the NCI results for two different $(BH_3NH_3)_4$ complexes. In Figure 8a, the structure was completely optimized in the gas phase, whereas in Figure 8b, the tetramer geometry was derived from the crystalline structure (with the hydrogen positions optimized to ensure correct description of the H−H contacts).[38] Calculations were performed using the B3LYP functional and support qualitatively different interaction types in each case. Whereas the gas-phase tetramer gives rise to a highly negative interaction energy, the crystalline structure is not stable with this density functional. This difference suggests that the solid structure is only stabilized by dispersion interactions (which are completely neglected using this functional[39]); stronger interactions are present in the fully optimized gas-phase complex.

The NCI description is able to recover the different nature of the interactions in both tetramers, supporting Morrison and Siddick's conclusion.[36] Whereas the crystalline tetramer only shows vdW contacts, the completely optimized structure gives rise to stronger dihydrogen bonds (with $\rho \simeq 0.02$ au), which are closer in energy to normal hydrogen bonds.

IV.B. Large Systems. *IV.B.1. High Affinity Host−Guest Complexes.* Host−guest complexation, like protein−ligand binding, depends upon a balance of stabilizing and destabilizing noncovalent interactions that contribute to the net thermodynamics. These systems are well-known exemplars for understanding noncovalent interactions in more complex biomolecular systems. Recent experimental results show that synthetic host−guest systems can achieve binding affinities that rival those of the tightest-bound protein−ligand complexes. For example, the seven-unit cucurbitural host (CB[7], Figure 9) binds cationic adamantyl,[29] ferrocene derivatives,[30] and bicyclo[2,2,2]octane derivatives[31−33] with binding constants of 10^9 to 10^{13} M^{-1}. Thus, a key test of our NCI method is to understand the magnitude of noncovalent interactions in host−guest complexes.

Mining minima algorithms (M2) have been used to dissect the enthalpic and entropic contributions to high binding affinities.[33] Despite the evident electrostatic complementarity of the cationic guests and the electronegative, carbonyl portions of CB[7], electrostatic interactions do not provide a significant net driving force for ligand binding. This is because the strong Coulombic attractions between the guests and CB[7] are offset by the energetic cost of removing solvating water from the cationic guests and the polar hosts upon binding. Instead, two other factors are found to be crucial for stabilizing these complexes. The entropy penalty is unusually small in relation to the enthalpy of binding. This is likely the case because of the rigidity of both

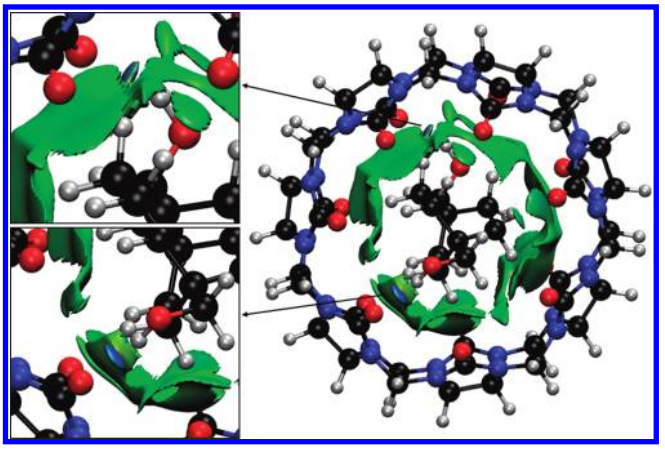

Figure 9. Cucurbit[7]uril-bicyclo[2,2,2]octane derivative inclusion complex. Anchor posts are highlighted in the insets. NCI surfaces show only intermolecular interactions. The gradient cutoff is $s = 0.5$ au, and the color scale is $-0.04 < \rho < 0.04$ au.

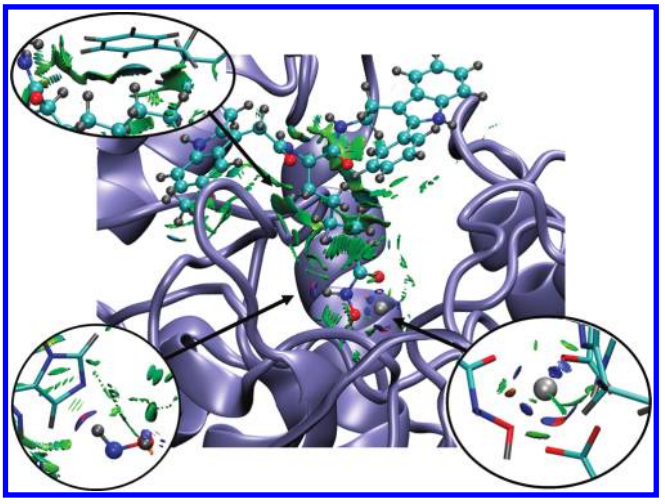

Figure 10. NCI surface around a V5X ligand in the active site of HDAC8 protein. Specific interactions are enlarged in the insets. NCI surfaces show only intermolecular interactions. The gradient cutoff is $s = 0.35$ au, and the color scale is $-0.04 < \rho < 0.02$ au.

the host and the guest structures.[31,33] Furthermore, the host and guest molecules studied are complementary in their preferred conformations. These characteristics give rise to favorable enthalpy−entropy compensation that enhances the binding affinities.

NCI analysis allows an assessment of host−guest complementarity and the extent to which weak enthalpic interactions stabilize a complex (NCI would not identify stabilization due to purely Coulombic forces). Figure 9 shows the binding between a bicyclo[2.2.2]octane derivative and CB[7]. In agreement with M2 results, vdW interactions are established throughout the cavity. The binding of the guest is further favored by the two hydroxyl anchors: the hydroxyl substituents on the guest establish strong hydrogen bonds with two carbonyls of the CB7 host (see Figure 9 insets).

IV.B.2. Protein−Ligand Interactions. Figure 10 shows the promolecular NCI surface around a V5X ligand in the active site of the HDAC8 protein (obtained from the protein data bank, as 2v5x.pdb, ref 34), with the details of the NCI surfaces enlarged. Specific parameters for the Zn^{2+} contribution to the promolecular density were calculated as described in section III.B. The NCI analysis reveals a set of complex interactions between the ligand and the protein, which arise from a combination of specific atom−atom interactions (i.e., hydrogen bonds) as well as broad surfaces indicative of stabilizing vdW interactions. On the lower right-hand side of the figure, we show the stabilization of the Zn^{2+} ion by the protein and ligand. At the lower left, we show the hydrogen bond between the ligand and His140. The top left of the figure shows the vdW surface between the ligand and the Phe139 phenyl ring. NCI analysis clearly highlights how a ligand "fits" the geometry of the active site, and the many small contributions that combine to determine the interaction energy between the ligand and protein.

V. CONCLUSIONS

Algorithms for the analysis of noncovalent interactions based on the NCI index were described. An efficient and flexible implementation was established for both SCF and promolecular densities, allowing for the analysis of noncovalent interactions in both small molecules and macromolecules. The utility of the method was illustrated with examples from organic and inorganic chemistry. Special emphasis was given to the ability of the new NCI index to provide insight into open issues in bonding.

The NCI code and a manual may be downloaded at http://www.chem.duke.edu/~yang/software.htm

■ ASSOCIATED CONTENT

Supporting Information. Input files for selected examples and the programming details are available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: weitao.yang@duke.edu.

■ ACKNOWLEDGMENT

This work is supported by the National Science Foundation (CHE-09-11119 & CHE-1012357), and the National Institute of Health through the UPCMLD project (P50GM067082) and ROI-GM-061870. J.C.G. thanks the Spanish MALTA-Consolider Ingenio-2010 program under project CSD2007-00045, and E.R.J. thanks the Natural Sciences and Engineering Research Council of Canada. Support from the French National Research Agency (ANR) is acknowledged, Grant ANR-08-BLAN-0158 (R.C. and J.P.P.). Contributions from H. L. Schmider (wave function properties), A. Martín Pendás (parsing routines), E.J. Toone and Y. Wang (CB[7] data), and J. Li (S_4N_4 data) are gratefully acknowledged.

■ REFERENCES

(1) (a) Chandler, D. *Nature* **2005**, *437*, 640. (b) Panigrahi, S. K.; Desiraju, G. R. *Proteins: Struct., Funct., Bioinf.* **2007**, *67*, 128.

(2) (a) Fenniri, H.; Packiarajan, M.; Vidale, K. L.; Sherman, D. M.; Hallenga, K.; Wood, K. V.; Stowell, J. G. *J. Am. Chem. Soc.* **2001**, *123*, 3854. (b) Kruse, P.; Johnson, E. R.; DiLabio, G. A.; Wolkow, R. A. *Nano Lett.* **2002**, *2*, 807.

(3) (a) Fiedler, S.; Broecker, J.; Keller, S. *Cell. Mol. Life Sci.* **2010**, *67*, 1779. (b) Dill, K. A. *Biochemistry* **1990**, *29*, 7133.

(4) Silvi, B. *Phys. Rev. Lett.* **1994**, *73*, 842.

(5) Bickelhaupt, F. M.; Baerends, E. *Angew. Chem., Int. Ed.* **2003**, *42*, 4183.
(6) Weinhold, F. *Angew. Chem., Int. Ed.* **2003**, *42*, 4188.
(7) Cioslowski, J.; Mixon, S. T. *Can. J. Chem.* **1992**, *70*, 443.
(8) Matta, C. F.; Hernández-Trujillo, J.; Tang, T.; Bader, R. F. W. *Chem.—Eur. J.* **2003**, *9*, 1940.
(9) Klein, R. A. *Chem. Phys. Lett.* **2006**, *425*, 128.
(10) Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; Oxford University Press: Oxford, U.K., 1990.
(11) Espinosa, E.; Souhassou, M.; Lachekar, H.; Lecomte, C. *Acta Crystallogr., Sect. B* **1999**, *55*, 563.
(12) Grabowski, S. J. *J. Phys. Chem. A* **2001**, *105*, 10739.
(13) Becke, A. D.; Edgecombe, K. E. *J. Chem. Phys.* **1990**, *92*, 5397.
(14) Silvi, B.; Savin, A. *Nature* **1994**, *371*, 683.
(15) Contreras-García, J.; Recio, J. M. *Theor. Chem. Acc.* DOI: 10.1007/s00214-010-0828-1.
(16) Alikhani, M. E.; Fuster, F.; Silvi, B. *Struct. Chem.* **2005**, *16*, 203.
(17) Johnson, E. R.; Keinan, S.; Mori-Sánchez, P.; Contreras-García, J.; Cohen, A. J.; Yang, W. *J. Am. Chem. Soc.* **2010**, *132*, 6498.
(18) Bohórquez, H. J.; Boyd, R. J. *Theor. Chem. Acc.* **2010**, *127*, 393.
(19) Arfken, G. *Mathematical Methods for Physicists*; Academic Press: Orlando, FL, 1985.
(20) Bader, R. F. W.; Essén, H. *J. Chem. Phys.* **1984**, *80*, 1943.
(21) Noury, S.; Krokidis, X.; Fuster, F.; Silvi, B. *Comput. Chem.* **1999**, *23*, 597.
(22) Spackman, M. A.; Maslen, E. N. *J. Phys. Chem.* **1986**, *90*, 2020.
(23) Pendás, A. M.; Luaña, V.; Pueyo, L.; Francisco, E.; Mori-Sánchez, P. *J. Chem. Phys.* **2002**, *117*, 1017.
(24) Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992**, *45*, 13244.
(25) DeLucia, M. L.; Coppens, P. *Inorg. Chem.* **1978**, *17*, 2336.
(26) Pritchina, E. A.; Gritsan, N. P.; Zibarev, A. V.; Bally, T. *Inorg. Chem.* **2009**, *48*, 4075.
(27) Gleiter, R. *J. Chem. Soc. A* **1970**, 3174. Salahub, D. R.; Messmer, R. P. *J. Chem. Phys.* **1976**, *64*, 2039.
(28) Stevens, W. J.; Fink, W. H. *Chem. Phys. Lett.* **1987**, *139*, 15.
(29) Liu, S.; Rupic, C.; Mukhpadhayay, P.; Chakrabarti, S.; Zavalij, P. Y.; Isaacs, L. *J. Am. Chem. Soc.* **2005**, *127*, 1595.
(30) Jeon, W. S.; Moon, K.; Park, S. H.; Chun, H.; Ko, Y. H.; Lee, J. Y.; Samal, S.; Selvapalam, N.; Rekharsky, M. V.; Sindelar, V.; Sobransingh, D.; Inoue, Y.; Kaifer, A. E.; Kim, K. *J. Am. Chem. Soc.* **2005**, *127*, 12984.
(31) Rekharsky, M. V.; Mori, T.; Yang, C.; Ko, Y. H.; Selvapalam, N.; Kim, H.; Sobransingh, D.; Kaifer, A. E.; Liu, S.; Isaacs, L.; Chen, W.; Moghaddam, S.; K. Gilson, M. K.; Kim, K.; Inoue, Y. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 20737.
(32) Liu, S.; Ruspic, C.; Mukhopadhyay, P.; Chakrabarti, S.; Zavalij, P. Y.; Isaacs, L. *J. Am. Chem. Soc.* **2005**, *127*, 15959.
(33) Moghaddam, S.; Inoue, S.; Gilson, M. K. *J. Am. Chem. Soc.* **2009**, *131*, 4012.
(34) Vannini, A.; Volpari, C.; Gallinari, P.; Jones, P.; Mattu, M.; Carfi, A.; Defrancesco, R.; Steinkuhler, C.; Di Marco, S. *EMBO Rep.* **2007**, *8*, 879.
(35) Popelier, P. L. A. *J. Phys. Chem. A* **1998**, *102*, 1873.
(36) Morrison, C. A.; Siddick, M. M. *Angew. Chem., Int. Ed. Engl.* **2004**, *116*, 4884.
(37) Richardson, T. B.; de Gala, S.; Crabtree, R. H.; Siegbahn, P. E. M. *J. Am. Chem. Soc.* **1995**, *117*, 12875.
(38) Klooster, W. T.; Koetzle, T. F.; Siegbahn, P. E. M.; Richardson, T. B.; Crabtree, R. H. *J. Am. Chem. Soc.* **1999**, *121*, 6337.
(39) Johnson, E. R.; Wolkow, R. A.; DiLabio, G. A. *Chem. Phys. Lett.* **2004**, *394*, 334.
(40) Brown, M. P.; Heseltine, R. W. *Chem. Commun. (London)* **1968**, *23*, 1551.

■ NOTE ADDED AFTER ASAP PUBLICATION

This article was published ASAP on January 25, 2011. The name of the first author in ref 26 has been corrected. The correct version was published on February 2, 2011.