Contents

Chapter 1
Software instructions

1.1 Before: obtaining the electronic structure files

Most topological packages read the wavefunction of a molecular system in a format known as .wfn. Besides this, NCIPLOT reads also an XML-like format .wfx, which allows for the use of pseudopotentials. These files are produced by several electronic structure packages like Gaussian (use the keyword output=wfn or output=wfx), GAMESS (use aimpac=.true.) and ORCA (use the !AIM directive). Other codes may write MOLDEN (3) files which may then be transformed to .wfn by the MOLDEN2AIM (4) utility.

Let's see an example for Gaussian. It suffices to set output=wfn in the Route section and give the name of the wfn file at the end of the molecule specification:

P HF/6-311++G(2df,2p) opt output=wfn

tetrahedrane

0 1
C -0.524831 0.524831 0.524831
C 0.524831 -0.524831 0.524831
C -0.524831 -0.524831 -0.524831
C 0.524831 0.524831 -0.524831
H -1.134794 1.134794 1.134794
H 1.134794 -1.134794 1.134794
H -1.134794 -1.134794 -1.134794
H 1.134794 1.134794 -1.134794

tetrahedrane.wfn

The first word in the title (contiguous non blank characters) is used to form the file names in several of the programs we will be illustrating. Hence,

it might be dangerous to use non alpha numerical characters. Notice that there should be a blank line after that containing the name of the .wfn file.

This is valid for a monodeterminantal wavefunction (Hartree-Fock or DFT). Remember that if you are using post Hartree-Fock methods, as well as NaturalOrbitals or NaturalSpinOrbitals, you need to specify that this is density you want to use (density=current):

#CISD/6-31G** popt pop=no density=rhoci output=wfn

H2O

0 1
H
O 1 r
H 2 r 1 a
r=1.0
a=110.0

h2o-no.wfn

In ORCA the input file would looks as follows, where the keyword AIM has been added:

%maxcore 4000
! RKS B3LYP SVP SP AutoAux RIJCOSX AIM TightSCF

%output
print[P_Hirshfeld] 1
print[P_Mayer] 1
end
* xyz 0 1
 C   0.034380593   1.901663427   0.000000000
 H   0.542217494   2.273137971   0.885991138
 H   -0.993020634   2.255906023   -0.000000000
 H   0.542217494   2.273137971   -0.885991138
 H   0.047430638   0.815178301   0.000000000
 O   0.034380593   -1.806674259   -0.000000000
 H   -0.310086665   -2.286973385   -0.758506116
 H   -0.310086665   -2.286973385   0.758506116
*

1.2 NCIPLOT

We will focus on NCIPLOT for the analysis of NCI in molecules. At the time
of writing this book the latest release of NCIPLOT, NCIPLOT4, is publicly
available at:
https://github.com/juliacontrerasgarcia/nciplot.
There is a manual available in the same repository that covers the details of
the current release.

The minimal input is composed by two pieces of information:

- line 1 must contain the number of files to be analyzed
- The following lines must contain as many file names as indicated in the
  first line

The type of calculation to be run (SCF or promolecular) is determined by
the extension of the files to be read (which must all be of the same type):

- name.xyz (Promolecular approximation). It requires a .xyz file. Recommended
  for big systems
- name.wfn or name.wfx wavefunction file in the AIMPAC .wfn format
  (SCF calculation)

The number of files is typically 1 for calculations, where all NCIs within
the given systems are computed:

1
waterdimer.wfn

Once the input constructed, the code is invoked as follows:

nciplot.x < inputfile > outputfile

The main results are collected in four output files:

- name.dat file collects rho vs RDG
- name-grad.cube is a .cube file (3D data) with the RDG data
- name-dens.cube is a .cube file containing $\text{sign}(\lambda_2) \times \rho \times 100$
- name.vmd is a script for the visualization of the results in VMD (1)

Greater number of files allow to estimate inter- and intramolecular interactions
separately. For example, 2 files are used for analyzing intermolecular interactions
between a ligand and a protein. This is done with an extra keyword after the
previous lines:

2
ligand.xyz
protein.xyz
INTERMOLECULAR

Additionally, the user can define the cutoffs for ρ and RDG, which will define the boundaries for the search of NCIs. For example, if a rather strong interaction is expected, it is possible to change the default values so that the interaction is not discarded. In this case, the input looks as follows:

1

aceticacid1.wfn
aceticacid2.wfn
INTERMOLECULAR
CUTPLOT 0.16 0.5

The value 0.16 has been chosen so as to include an NCI peak from the 2D plot which lays below $\rho=0.16$.

This change of default values is also needed when interactions with a hole inside the surface appear, revealing that the default value coincides with a peak, leaving a part of it outside of the study range. An example is shown in Fig. 1.1: the cutoff needs to be increased from $\rho=0.05$ to for example 0.08 in order to obtain full disks for the hydrogen interactions.
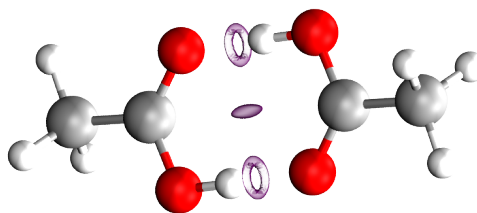

Fig. 1.1 $s(\mathbf{r}) = 0.5$ isosurfaces on the acetic acid dimer with a $\rho(\mathbf{r}) = 0.05$ cutoff.

Depending on the accuracy needed, it is possible to change the size of the step defining the grid. By default, the step is 0.1 (Fig. 1.2-top), so increasing the value allows for a faster (but more pixelated picture) calculation (Fig. 1.2-bottom):

1

linh4.wfn
CUTPLOT 0.16 0.5
INCREMENTS 0.2 0.2 0.2

It is also possible to accelerate the calculation while keeping the accuracy thanks to an adaptive grid. Starting from a coarse grid to quickly explore the whole 3D space around the molecule, the user is prompted to define a succession of progressively finer grids to refine NCI results in regions where NCIs are actually detected. In this way, only NCI relevant points are computed accurately, resulting in a ∼ 10-fold speed-up.
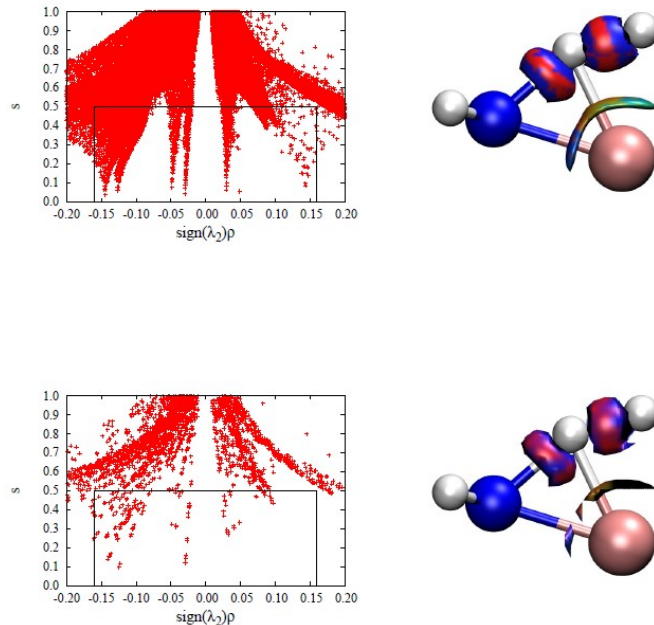
Fig. 1.2 2D (left) and 3D (right) NCI plots with default (top) and bigger INCREMENTS (bottom).

In order to use this option, the number of adaptive grids and their relative size (finishing with "1"-i.e. the basic increment) are defined after the keyword CG2FG (Coarse Grid To Fine Grid). By default, the CG2FG option is set to 1 1 using only one-level grid, i.e. without acceleration. In practice, we recommend to set CG2FG to 3 4 2 1 or 4 8 4 2 1 with 3- or 4-level grids for acceleration. This leads to an input looking as follows:

1
bigmolecule.xyz
CG2FG 4 8 4 2 1

It is possible to focus on the visualization of a given interaction type following various criteria:

1. Strength. By selecting electron density and reduced density gradient cutoffs, single peaks in the NCI diagram can be isolated and visualized.
2. Geometry. Contrary to the previous criterion, chemical interactions can be isolated not from the NCI diagram, but from their spatial location. For that, the computation of ρ and s is restricted to a cubic box defined by the user. Three algorithms exist to defined the cubic box:

   - ATCUBE keyword. The cubic box is built around a selected list of atoms.
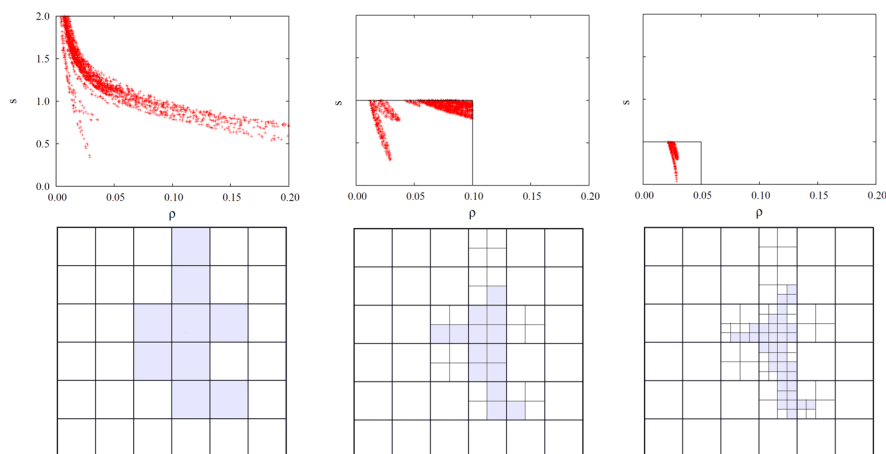
Fig. 1.3 2D schematic diagram of multi-level grids, from coarse (left) to fine grids (right) with $4 : 2 : 1$ increments and $\rho_c = 0.05$ and $s_c = 1.0$. The density and the s values are computed at the active points of the coarse, medium and fine grids (from left to right). Top: cutoff on (ρ,s). Bottom: size of the boxes. The small blue boxes are the active boxes of the finest grid, which really contribute to form the isosurface. Adapted with permission from (2). Copyright 2020 American Chemical Society.

- RADIUS keyword. The cubic box is defined by the origin and the length of its axes.
- CUBE keyword. The cubic box is defined by its boundaries.

3. Pure intermolecular interactions Promolecular density can be divided into molecular or fragment contributions. The ratio between molecular densities (ρ_{mol}) and the total density (ρ_{total}) is used as a criterion to differentiate intermolecular from intramolecular interactions. This filter is only applicable within the promolecular approach.

$$\frac{\rho_{mol}}{\rho_{total}} = \begin{cases} f < 0.9 & \textbf{Intramolecular} \\ f \geq 0.9 & \textbf{Intermolecular} \end{cases} \tag{1.1}$$

4. Centered around a given molecule. In some complicated cases such as a enzyme-substrate complexes, it is fundamental to analyze all the interactions (intra and intermolecular) of the substrate and the active site. This is automatically done by the keyword LIGAND, which builds a cubic box not around a single point, but around a whole molecule.

Quantification with NCI integrals can be done also following the above Strength criterion. In order to define integration ranges, the keyword RANGE is invoked, and the $sign(\lambda_2)\rho$ intervals subsequently defined:

1
bigsystem.xyz
RANGE 3
-0.1 -0.02
-0.02 0.02
 0.02 0.1

This provides both the total integrals in the NCI region as well as by ranges. The total integrals look as follows:

```
--------------------------------------------------------------------
Integration  over the volumes of rho^n
--------------------------------------------------------------------
n=1.0           :       3.71920693
n=1.5           :       0.80392978
n=2.0           :       0.17992719
n=2.5           :       0.04164749
n=3.0           :       0.00995114
n=4/3           :        1.33570670
n=5/3           :        0.48721346
Volume          :       88.37619958
rho-sum_i rho_i :        0.00178346


--------------------------------------------------------------------
Integration  over the volumes of sign(lambda2)(rho)^n
--------------------------------------------------------------------
n=1.0           :       -3.71920693
n=1.5           :       -0.80392978
n=2.0           :       -0.17992719
n=2.5           :       -0.04164749
n=3.0           :       -0.00995114
n=4/3           :        -1.33570670
n=5/3           :        -0.48721346
--------------------------------------------------------------------
```

The first set of results corresponds to the integrals $\int \rho^n d\boldsymbol{r}$, and the second one to $\int \text{sign}(\lambda_2)\rho^n d\boldsymbol{r}$.

On top of this, three ranges are defined in the input file: [-0.1,-0.02], [-0.0, 0.02] and [0.02,0.1], to differenciate hydrogen bonds, van der Waals and steric crowding, respectively. For example if we are interested in strongly attractive non-covalent interactions we will look a the output of the [-0.1,0.02] range:

```
--------------------------------------------------------------------
 Interval        :      -0.10000000      -0.02000000
--------------------------------------------------------------------
Integration  over the volumes of rho^n
--------------------------------------------------------------------
```

n=1.0   :   3.71920693
n=1.5   :   0.80392978
n=2.0   :   0.17992719
n=2.5   :   0.04164749
n=3.0   :   0.00995114
n=4/3   :   1.33570670
n=5/3   :   0.48721346
Volume   :   88.37619958
rho-sum_i rho_i :   0.00178346

By comparing the total and the range integration we see that in this example all the non-covalent interactions are of in the strongly attractive regime.

1.3 After: Visualization with vmd

NCIPlot produces a vmd script for visualization. Alternatively you can also follow these instructions to render your NCI visualization:

1. Loading the cube files
   File > New Molecule>Browse[1]

   a. You choose the name-dens.cube file
   b. On the same window (do not repeat the File>New Molecule process) you load the name-grad.cube file

2. Drawing the NCI surface
   Graphics > Representations > Create Rep

   a. Drawing Method : Isosurface
   b. Isovalue: 0.3 (+ Enter)
   c. Coloring Method: Volume
   d. In Trajectory (3rd tab): Color Range Data Range: -3 to 3

3. Other settings

   - Changing the background color to white: Graphics>Colors>Display (in Categories)> 8 white (in Colors)
   - Getting rid of the axes: Display>Axes>Off
   - Not showing perspective: Display>Orthographic

4. Saving your picture File > Render > Filename: Yourname.bmp
   Start Rendering

   You can also save your state for ulterior use: File > Nom du fichier: yourname.vmd

---

[1] You can define your default directory on the vmd command terminal. Go to the files location at the command vmd terminal:
>cd dir (dir=your working directory)
For example:
>cd (to go to your home)
>cd Desktop (to go to your Desktop)
>dir (to see what is inside the repository)

References

[1] VMD Software. https://www.ks.uiuc.edu/Research/vmd/

[2] Boto, R.A., , Peccati, F., Laplaza, R., Quan, C., Carbone, A., Piquemal, J.P., Maday, Y., Contreras-García, J.: J. Chem. Theory Comput. (2020)

[3] Schaftenaar, G.: The molden program. The code is avalaible at http://www3.cmbi.umcn.nl/molden/

[4] Zou, W.: The molden2aim code (2020). The code is avalaible at github.com/zorkzou/Molden2AIM