Table of Contents

IsoStar

2020.0 CSD Release

CCDC
advancing structural science

Introduction

Entries in the CSD and PDB contain thousands of interactions, between all kinds of different functional groups. Extracting these out and mapping the space of preferred interactions around functional groups can lend to predictive properties of query molecules. This can help with ligand design, understanding of interactions, identifying bioisoteres, and many other components along the drug discovery pipeline.

IsoStar is a library of information about the intermolecular interactions formed by a wide variety of chemical groups. It contains data on intermolecular interactions in small molecule crystal structures taken from the CSD, as well as protein-ligand interactions in X-ray structures from the PDB. The library is presented in a web application for download based on the query molecule and probe type. These files are then viewed and analysed in the IsoStar client, which is included with the CSD Suite.
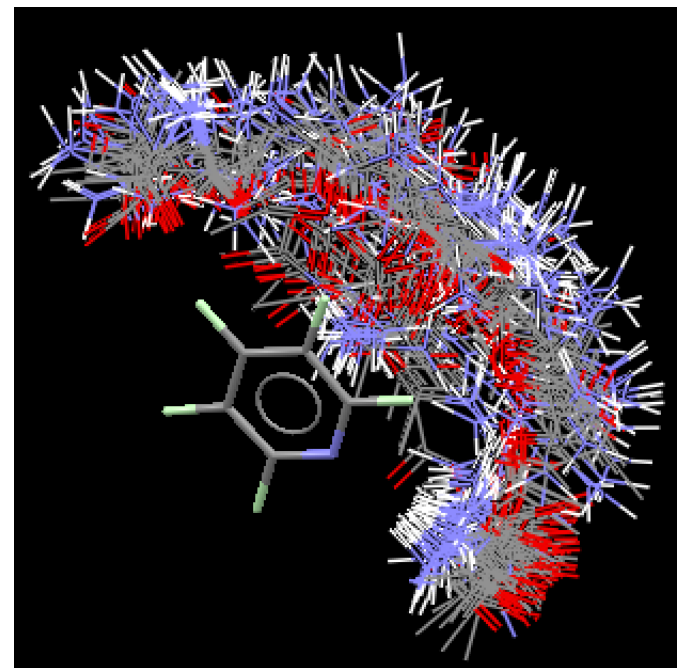
IsoStar

Crystallographic information in IsoStar is presented as 3D scatterplots, or contoured density surfaces. Each data point in these scatterplots has been calculated by searching the CSD or PDB for nonbonded interactions between a pair of functional groups A and B, and then least-square superimposed.

This example will walk you through downloading information from the IsoStar server, and then using it to compare the interaction preferences of an uncharged pyridine in the CSD vs. the PDB. The context of both of these is quite different, even though the query molecule is the same.

1. On the IsoStar webpage (isostar.ccdc.cam.ac.uk), under *Ring systems,* click *N,C,H only.*

2. About halfway down the page, click *uncharged pyridine.* This takes you to a page presenting the interacting groups for uncharged pyridines. Note that the tables are broken into element-based functional groups, with guides at the top of each table.

While most entries have statistics associated with them, some also have entries in the *Theory* column. These are a series of calculations for various energies contributions, all done at the 6-31G** basis set level using CADPAC. More information can be found by following the *Theory* link and clicking on *Methodology*.

3. In the **C,H only** table, click the entry for *phenyl*, under *CSD.* This downloads an IsoStar file (.istr).

4. Click the downloaded IsoStar file to launch the IsoStar client. The IsoStar window has visualisers at the bottom of the screen, which by default shows all entries. Click the **vdW Overlaps** button to hide all contact groups except for those within van der Waal contact of any target atom in the central group.

2

C,H only				
● Links to statistical data	● Links to theoretical energy data			
Contact Group	*CSD*	*PDB*	*Stats*	*Theory*
any alkyl C-H	4994	5336	●	
methylene	1997	1326	●	
methyl	1997	2189	●	
any aromatic C-H	4932	2750	●	
substituted aromatic carbon	2495	454	●	
phenyl	1426	521	●	

5
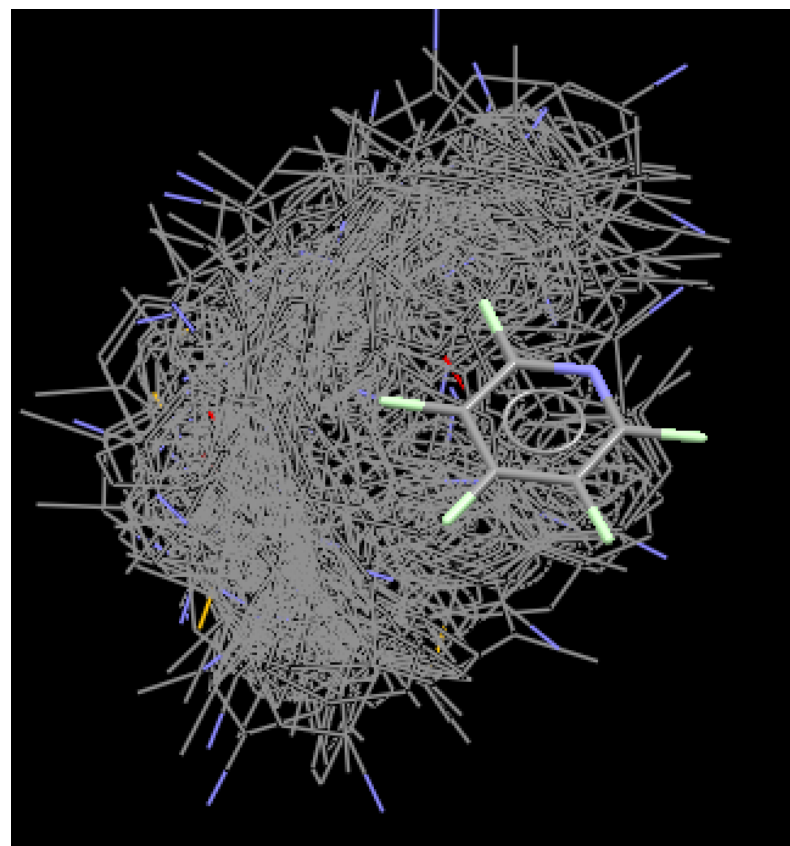
5. The visualiser in IsoStar is the same as Mercury, so the functionality is the same. Take a moment to rotate and zoom the scatterplot to get an appreciation for the pi-pi stacking and edge interactions around the uncharged pyridine. In the image on the right, each atom of the pyridine was selected, and then displayed as capped sticks by right clicking *Styles >> Capped Sticks*.

6. The scatterplot is made of 1426 entries at time of publication. While this lends vital information about interaction preferences, it is important to be able to drill down into each case to see the interaction. To enable this ability, click the tick box next to *Hyperlink*. Now click on any interaction in the scatterplot to see the individual CSD entry in the previously empty quadrant.
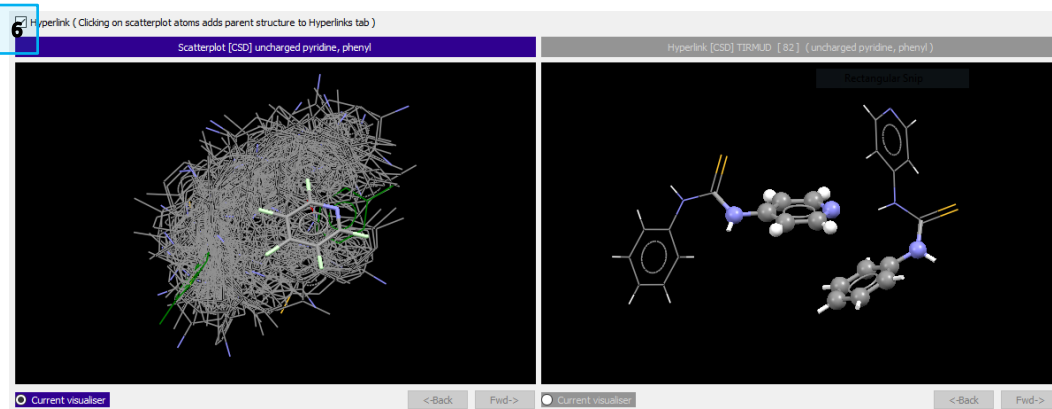
7. Scatterplots can be converted to contoured density surfaces, which provides richer information that looking at the probes overlaid. Click the **Create/Edit** button next to *Contour Surfaces* to launch the **Contour Surfaces** window.

8. Note that the *Probe* is by default set to *Ring_Centre*. This means that the centroid of a ring will be taken as the representative point. Click **Create** to create the surface. The default values create a density surface with red being the highest concentration, to blue, and then green for lower threshold. Each of these thresholds can be edited. Click **Close** to close the window.

9. The new contour surface is showed overlaid with all of the probes. Hide the probes by clicking the *Hide All* tick box in the **Show contacts in van der Waals (vdW) corrected distance range** box. Inspect the contour surface. Note the hotspots (red areas) and the uneven distribution of them around the ring. There is also a higher density of probes underneath the aromatic ring, for the pi-pi stacking interactions.

10. In IsoStar, scatterplots are symmetrised. This means that all contacts are reflected into an area of space based on the symmetry of the central group (e.g. only one quarter of an aromatic iodine group with its contact group is shown since the central group has two planes of mirror symmetry, whereas

half of all contacts to an azide group are shown since it has one plane of mirror symmetry). The symmetry can be expanded and contracted as desired by clicking on **Expand** and **Contract** next to *Scatterplot Symmetry*. Expand the symmetry. While this does show all of the information, it can get visually cluttered.
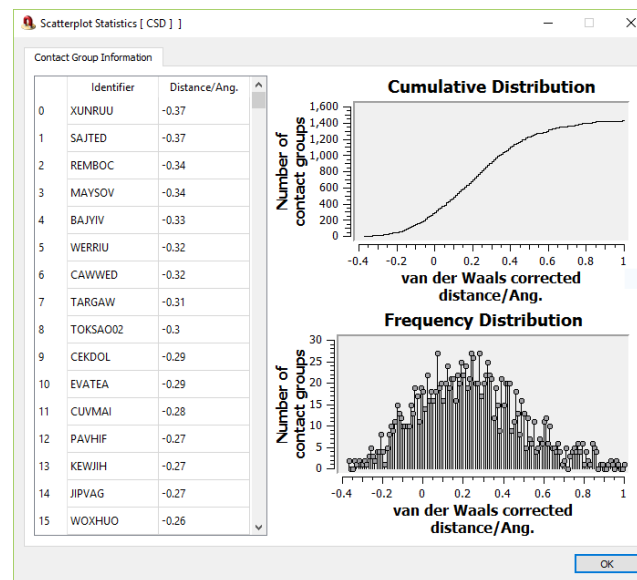
11. Click the **Overview** button to see a drill down of all contact group information. Note the bell curve type distribution of the vdW corrected distances between the pyridine and contact groups.

Remember that this data is based only on small molecule interactions. How would this compare with protein-ligand data from the PDB?

12. Go back to the IsoStar website, and download the 'phenyl' group information from the PDB (521 entries). Open this in IsoStar as you did previously. Opt to **Load in Other** to load in the second visualiser.

13. Click **vdW Overlaps** to show only contacts within vdW distances. Create contour surfaces again by clicking **Create/Edit**, and clicking **Create**. You may be warned that this will result in a statistically insignificant data set. Proceed by clicking **Yes**. Close the window, and click *Hide All* to hide the probes.

14. Note that the hotspots are exclusively at pi-pi stacking regions, showing a strong preference for this interaction in the PDB. Given that the PDB is full of protein-based interactions, and that pi-pi stacking is a common bioactive motif, it is not surprising to see the difference in CSD vs. PDB hotspot distributions for uncharged pyridine. Compare this contour to the contour of the CSD from step 9 above. What other PDB vs. CSD reasons might account for this difference?

Although IsoStar contains data for many different functional groups, users may wish to generate their own scatterplots for contacts not currently included in the library. This may be done by performing a nonbonded search of the CSD with ConQuest, and then converting the output to a scatterplot using IsoGen. IsoGen is a linux-only program, and is not covered in this tutorial. However, the IsoStar documentation explains how to do this.