6th June 2019

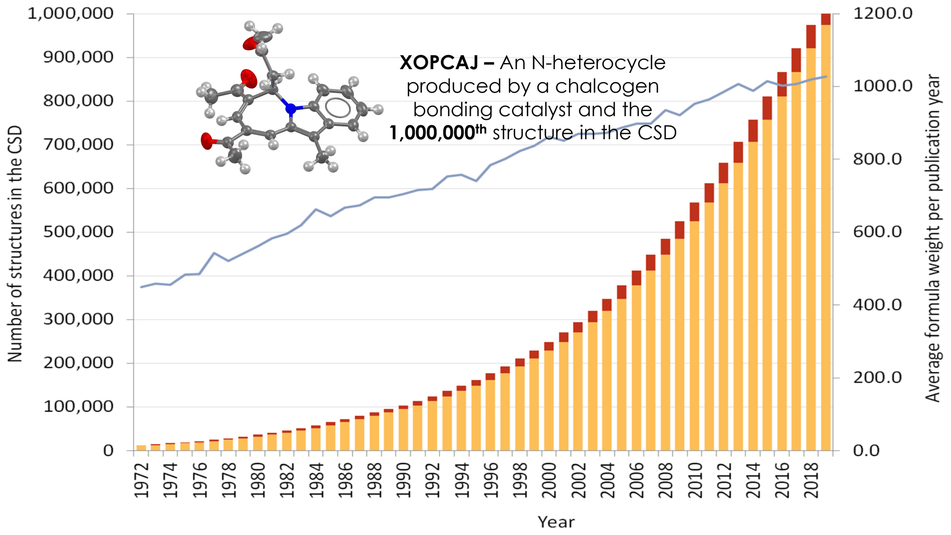
The Cambridge Structural Database: a Powerful Resource in Drug Discovery

Juliette Pradon and Ilenia Giangreco
The Cambridge Crystallographic Data Centre, Cambridge, UK. E-mail: pradon@ccdc.cam.ac.uk; Web: http://www.ccdc.cam.ac.uk

THE CAMBRIDGE STRUCTURAL DATABASE (CSD)

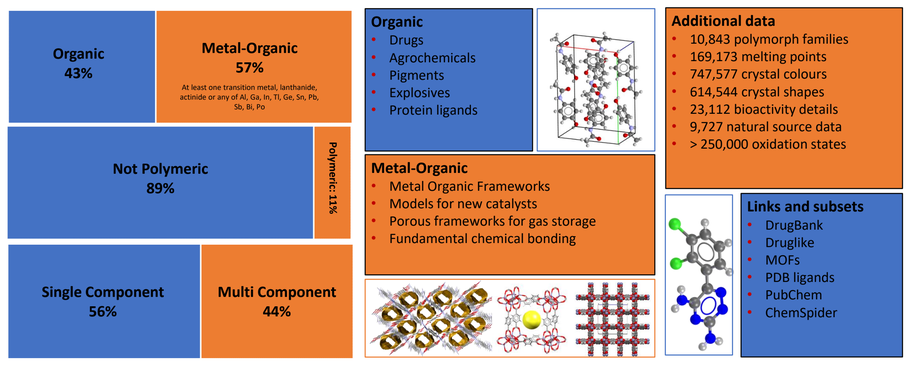
Successful modern drug discovery research makes extensive use of structural data – from target proteins, candidate drug molecules, and complexes of the two. The value of protein-ligand structural information is well accepted, however, knowledge of molecular conformations and interactions derived from small molecule structures alone can have a significant impact in drug discovery. This year commemorates a milestone for structural chemistry as the CSD has reached the addition of its **millionth** structure to its extensive repository of fully curated organic and metal-organic structures.

Number of Entries	Unique Structures	Unique Author Names	Unique Publications
1,011,333	1,000,352	393,226	488,804

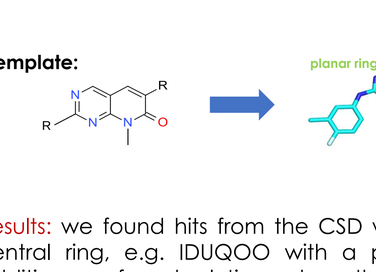


Inside the CSD

The CSD doesn't just contain purely crystallographic data; it also provides a wealth of information on things like melting points, crystal morphology, etc. As well as Crystallographic Information Frameworks (CIFs) destined for the CSD we also make available a number of other CIFs that don't fit the criteria for the CSD through our Access Structures service: https://www.ccdc.cam.ac.uk/structures/. These datasets include calculated structures and can be linked to from a publication.



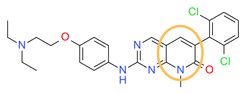
CSD-CROSSMINER



- Which structural motifs bind similar protein binding sites?
- Which ligand motifs have similar protein interaction patterns?
- Which ligand modifications and scaffold hops are tolerated in a protein binding site?

CSD-CrossMiner[1] provides the ability to search structural databases in terms of pharmacophore queries allowing to design molecular mimics of established lead compounds, shed some light on the cross-pharmacology between protein targets, as well as on the selectivity of bioactive small molecules.

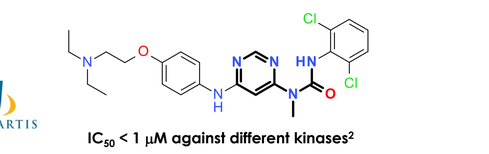
Kinase Inhibitor Scaffold Hopping Based on Ligand Features



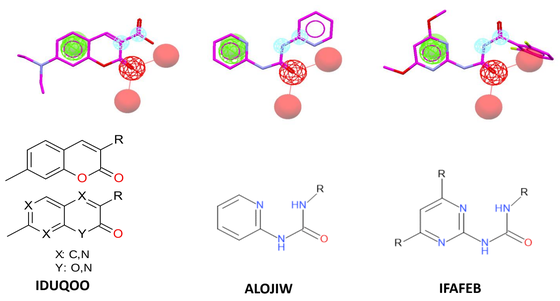
PD166285 tyrosine kinase inhibitor

Aim: replace the pyrimidone ring to find new potent tyrosine kinase inhibitors.

Template:



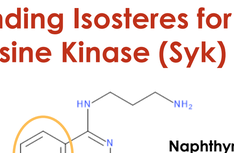
Results: we found hits from the CSD with a different central ring, e.g. IDUQOO with a pyrane ring. In addition, we found solutions where the central ring is replaced by an urea moiety able to form an intramolecular H-bond.



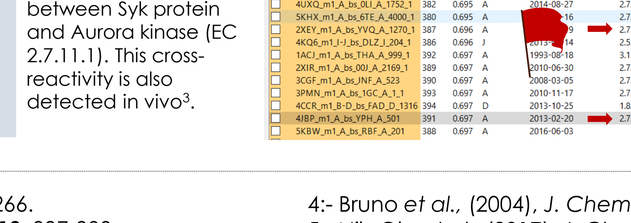
NOVARTIS

IC_{50} < 1 μM against different kinases[2]

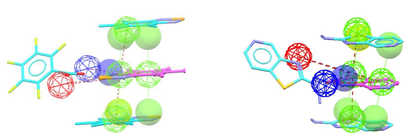
Finding Isosteres for Spleen Tyrosine Kinase (Syk) Inhibitors



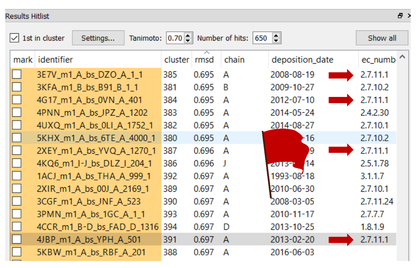
Aim: find isosteres of the 1,6-naphthyridine ring to obtain more potent and selective analogs while preserving the molecular interaction pattern.



Results: we found hits from the CSD with an isothiazole as an isostere of the 1,6-naphthyridine ring.



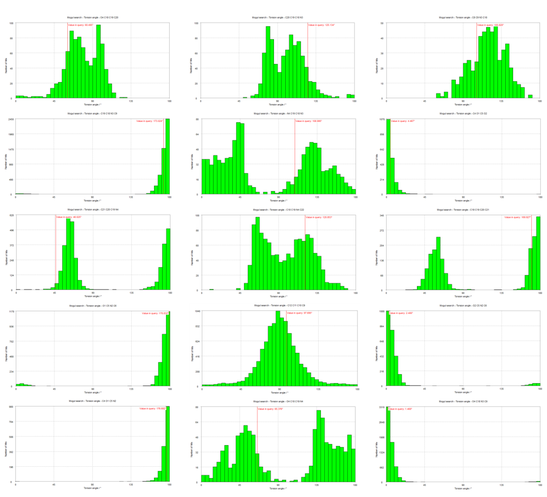
By searching the PDB we found that there is a risk of cross-reactivity between Syk protein and Aurora kinase (EC 2.7.11.1). This cross-reactivity is also detected in vivo[3].



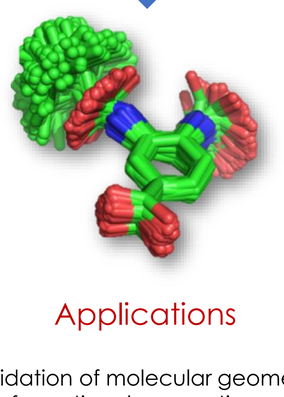
FROM DATA TO KNOWLEDGE

Here are two examples showing the value that can be obtained from the fully curated database and consider the intelligent software required to extract powerful insights that can inform the design, development and identification of new and better pharmaceutical products.

Molecular Shape



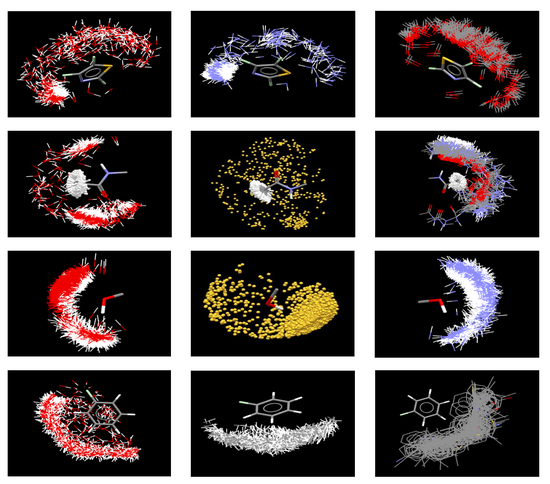
Pre-computed libraries of bond lengths, valence angles, torsion angles and ring conformations derived from the CSD[4].



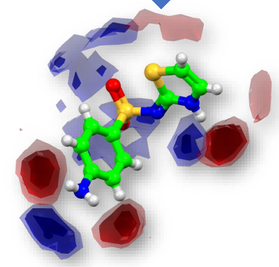
Applications

- Validation of molecular geometries.
- Conformational generation and validation for identification of low energy conformations[5,6].
- Creation of restraint data/ligand dictionaries for protein crystallographers and structural biologists.

Molecular Interactions



Pre-computed libraries of intermolecular interactions derived from the CSD and the PDB[7]. The interaction distributions are displayed as scatterplots or contour surfaces.



Applications

- Analyse the interaction preferences of molecules[8].
- Obtain evaluation of intermolecular packing of a crystal structure, e.g. evaluate the stability of polymorphic structures.
- Obtain interaction preferences for cavities in proteins[9].

1:- Korb *et al.,* (2016), *J. Med. Chem.*, ***59***, 4257-4266.
2:- Furet *et al.,* (2008), *Bioorg. Med. Chem. Lett.*, ***18***, 897-900.
3:- Thoma *et al.,* (2015), *Bioorg. Med. Chem. Lett.*, ***25***, 4642-4647.
4:- Bruno *et al.,* (2004), *J. Chem. Comput. Sci.*, ***44***, 2133-2144.
5:- Nils-Ole *et al.,* (2017), *J. Chem. Inf. Model.*, ***57***, 2719-2728.
6:- Cole *et al.,* (2018), *J. Chem. Inf. Model.*, ***58***, 615-629.
7:- Bruno *et al.,* (1997), *J. Comput.-Aided Mol. Des.*, ***11***, 525-537.
8:- Wood *et al.,* (2013), *CrystEngComm*, ***15***, 65-72.
9:- Verdonk *et al.,* (1999), *J. Mol. Biol.*, ***289***, 1093-1108.