CCDC

advancing structural science

Structural databases in drug discovery

Extracting useful information from the CSD and the PDB

Dr Vera Prytkova

Research and Applications Scientist

CCDC

advancing structural science

Why does the CCDC exist?

"A passionate belief that the collective use of data will lead to the discovery of new knowledge, which transcends the results of individual experiments."

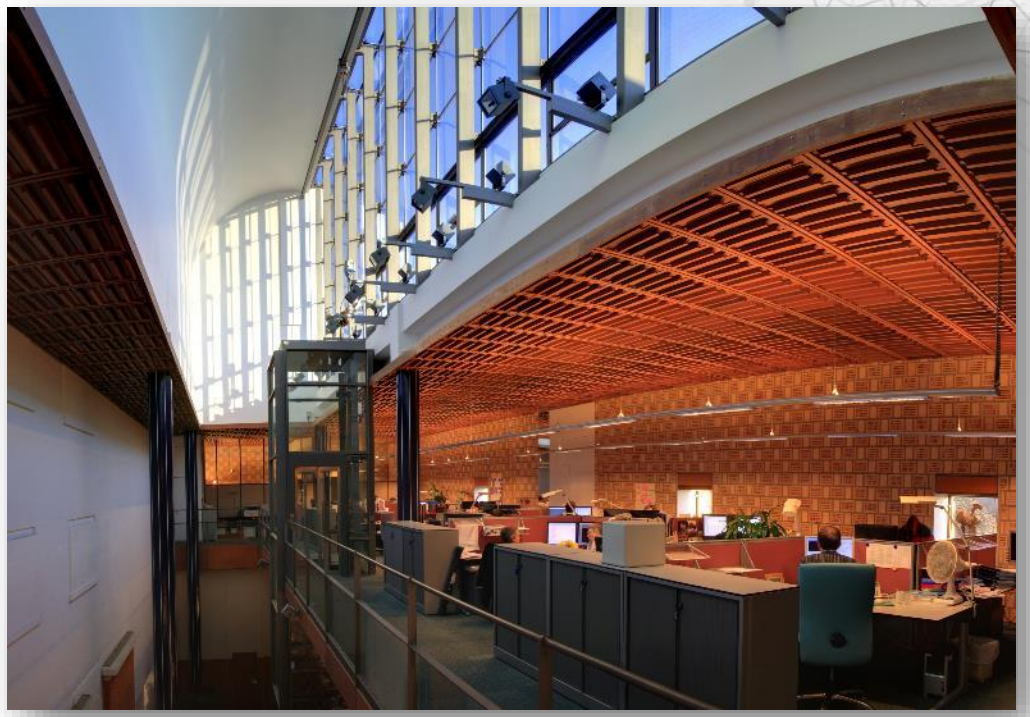
Olga Kennard

What we do

CCDC is a not-for-profit charity and the home of 3D structural data and software for pharmaceutical discovery, materials development, research and education.

We believe that advancing knowledge in structural chemistry is life enhancing:

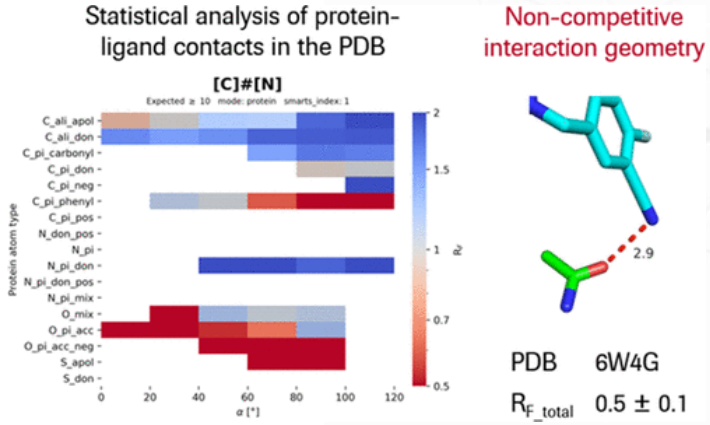
- We deliver knowledge-based solutions that accelerate the discovery and development of new medicines and materials
- We inspire a new generation of structural chemists globally through outreach and education
- We promote research collaboration across academia and industry to advance structural science



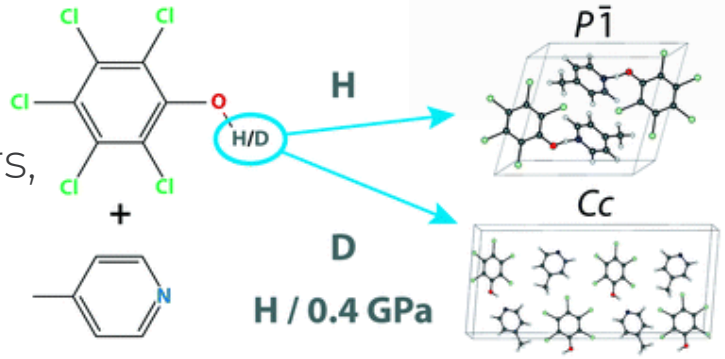
CCDC

Science & research

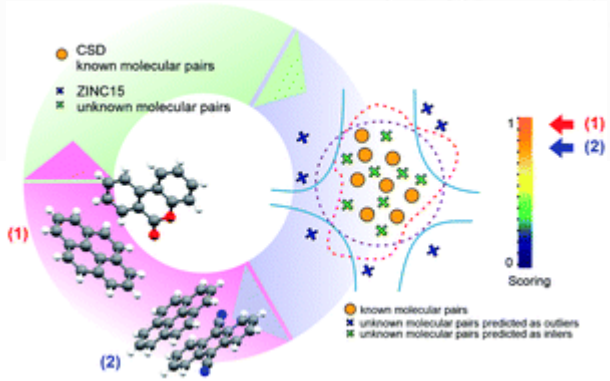
- Expertise in research fields including:
  - Crystal Engineering
  - Drug Design
  - Molecular Structure Analysis
  - Knowledge Mining
- Collaborate with leading researchers, in academia and industry
- Publish scientific reports



"Identification of Noncompetitive Protein-Ligand Interactions for Structural Optimization" A.Tosstorff et al. J.Chem.Inf.Model.2020,60,12,6595-6611



"Suppression of isotropic polymorphism" N.Funnell et al. CrystEngComm, 2021,23,769-776



"One class classification as a practical approach for accelerating π-π co-crystal discovery" A.Vriza et al. Chem.Sci.,2021,12,1702-1219

https://www.ccdc.cam.ac.uk/researchandconsultancy/ccdcresearch/ccdcpublications

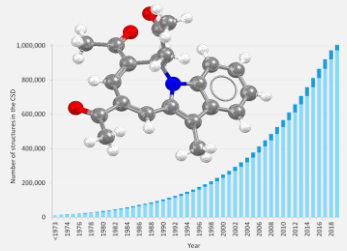
What is the CSD?

Millions of molecular geometries, hydrogen bond interactions and pi-pi-stacks, thousands of Z' > 1 structures and multicomponent systems.

Every entry enriched and annotated by experts.

Surpassed **1 million** crystal structures.
Structures available for anyone to download.

Cambridge Structural Database



Mining database

- Interaction patterns
- Scaffold hops
- Substructure searching

Geometry and interaction analysis

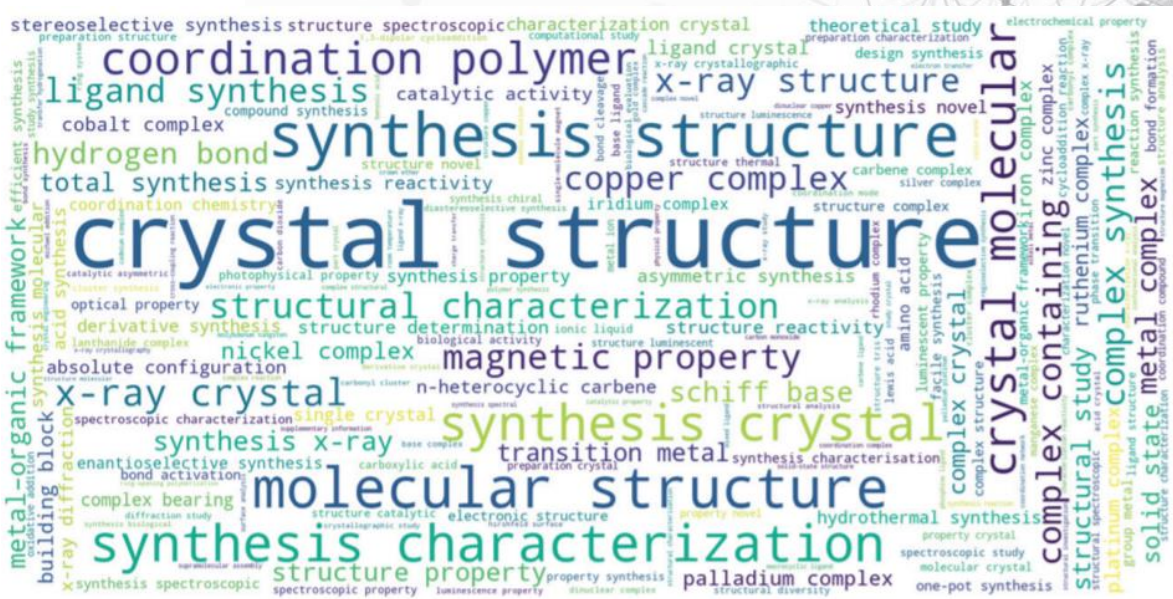
- Intramolecular geometry
- Geometry of interactions
- Likelihood of interactions

CCDC

The CSD in research

The *Web of Science* subject categories that cite the 2006-2016 CCDC standard references most frequently

Subject category	Citations
Crystallography	5256
Chemistry Multidisciplinary	3275
Chemistry Inorganic Nuclear	1931
Chemistry Physical	1148
Materials Science Multidisciplinary	864
Chemistry Organic	568
Biochemistry Molecular Biology	307
Physics Atomic Molecular Chemical	203
Chemistry Medicinal	185
Spectroscopy	181

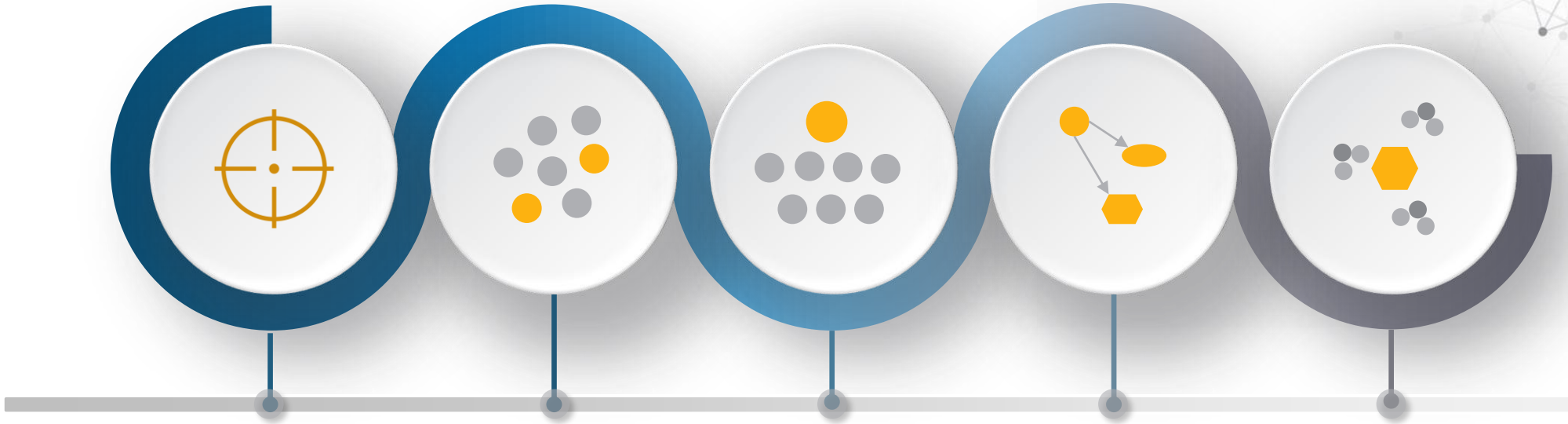


A word cloud of common bigrams in the titles of publications containing CSD-compliant crystal structures.

Willett P. et al, CrystEngComm, 2020,22, 7233-7241 DOI:10.1039/D0CE00045K

CCDC

Power of data in drug discovery



TARGET SELECTION

Assess pocket druggability by analysing intermolecular interaction between the target and the probe and constructing interaction propensity maps.

HIT IDENTIFICATION

Mine databases to identify repeated patterns of interaction or potential scaffold hops.
Fast substructure searching of intermolecular interaction patterns.

HIT TO LEAD

Assess how changes affect protein-ligand interaction using the CSD and the PDB.
Optimize compound geometry.

LEAD OPTIMISATION

Check the impact of changes with isosteric replacements or scaffold hops.
Understand how changes affect conformations.

SOLID FORM ANALYSIS

Identify risk of alternate forms, reveal opportunities for intervention or provide reassurance with conformational, interactions and H-bond analysis.

CCDC

Extracting useful information from the CSD and the PDB

1. Mining structural databases to identify *patterns of interaction* or potential *scaffold hops* to design novel motifs and retrieve a diversity of ligand topologies.

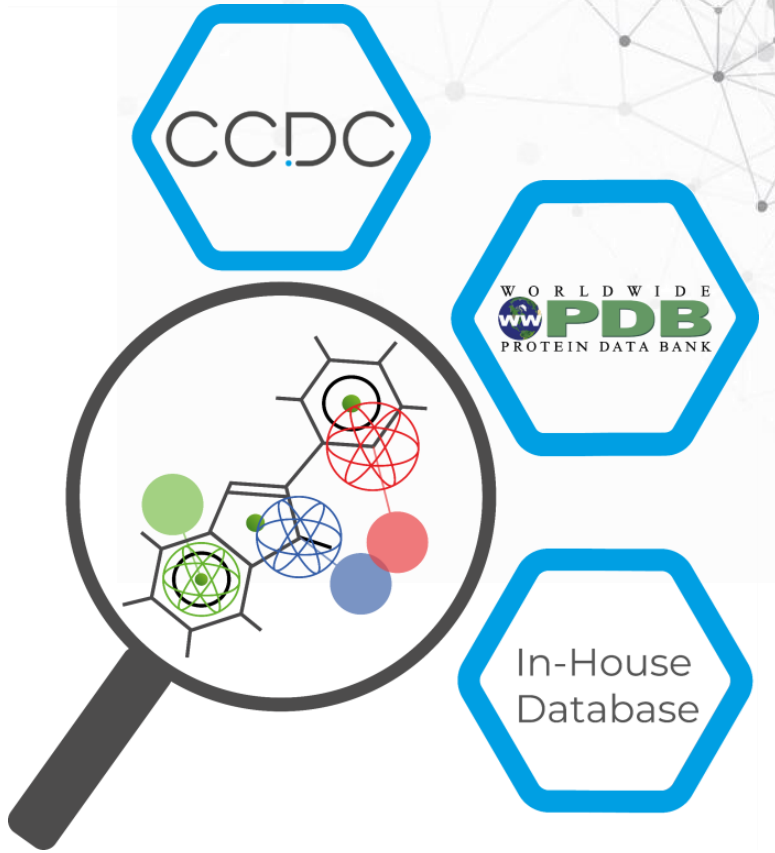
2. Using statistically significant information about *molecular conformations* and *intermolecular interactions* to evaluate the probability of observing a particular conformation of a newly designed drug in the binding site.

3. Performing the conformational analysis of a potential drug candidate. Stability analysis for *solid form development*.

CCDC

Mining structural databases

- 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?

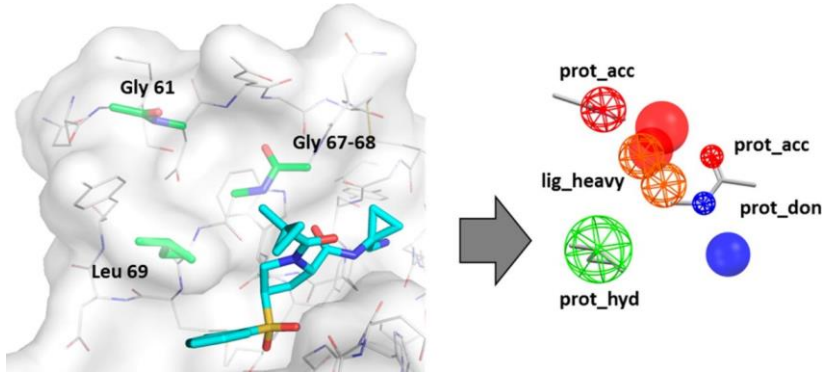


Korb O, et al., (2016) J Med Chem, 59(9):4257, DOI: 10.1021/acs.jmedchem.5b01756

CCDC

Mining structural databases - method

- Structures are annotated with features based on SMARTS patterns
- Features have tolerance radius reflecting the uncertainty in their position
- Features positioned with respect to one another in space form a pharmacophore
- The CSD subset contains the information about the small molecule, the PDB subset – small molecule and the binding site

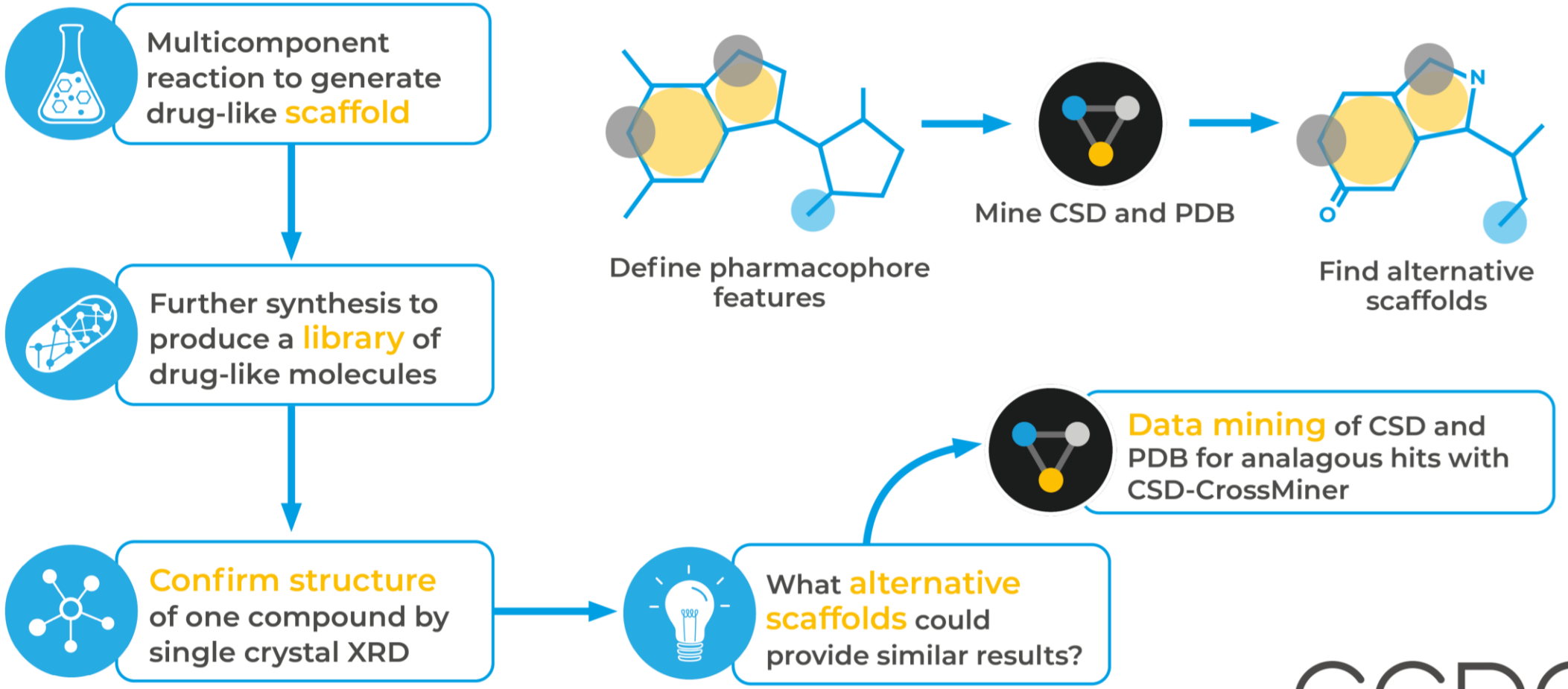


- Either database is searched using a relevant pharmacophore, pre-built with features describing the small molecule drug, the binding site and/or their interaction

Korb O, et al., (2016) J Med Chem, 59(9):4257, DOI: 10.1021/acs.jmedchem.5b01756

CCDC

Exploring molecular scaffolds

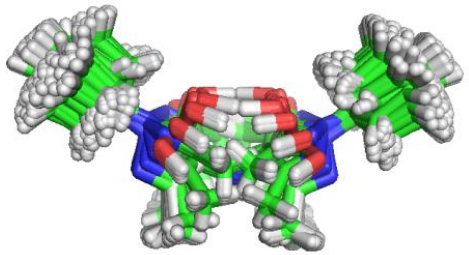


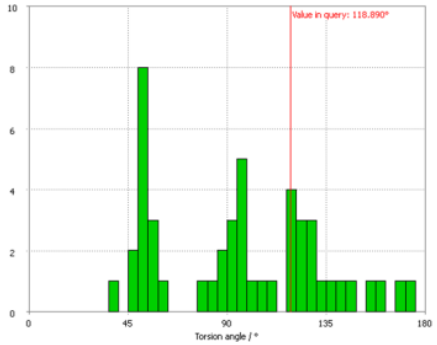
CCDC
advancing structural science

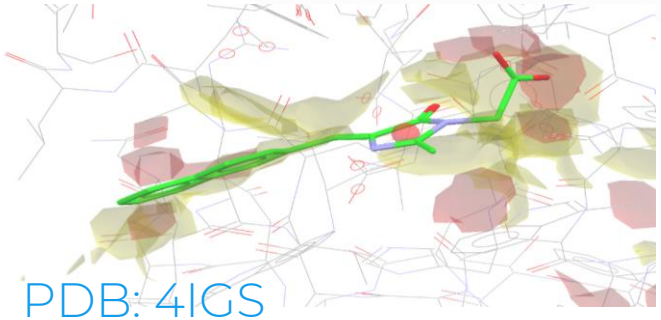
Konstantinidou M., et al., (2020) EurJOC, 34:5601, DOI:10.1002/ejoc.202000933

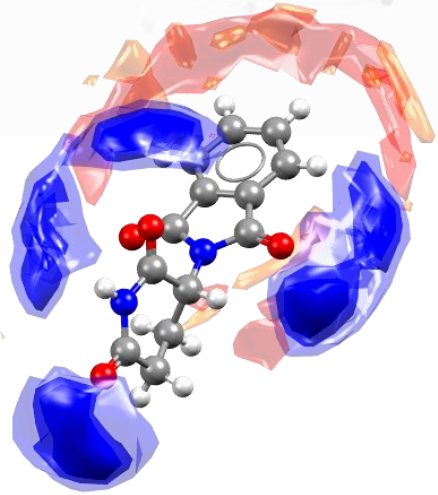
The CSD - molecular conformations and intermolecular interactions

- How likely is my small molecule drug conformation?
- Can I generate a realistic molecular conformation?
- What is the interaction propensity for this small molecule drug?
- Where in the binding site will the small molecule drug interact?





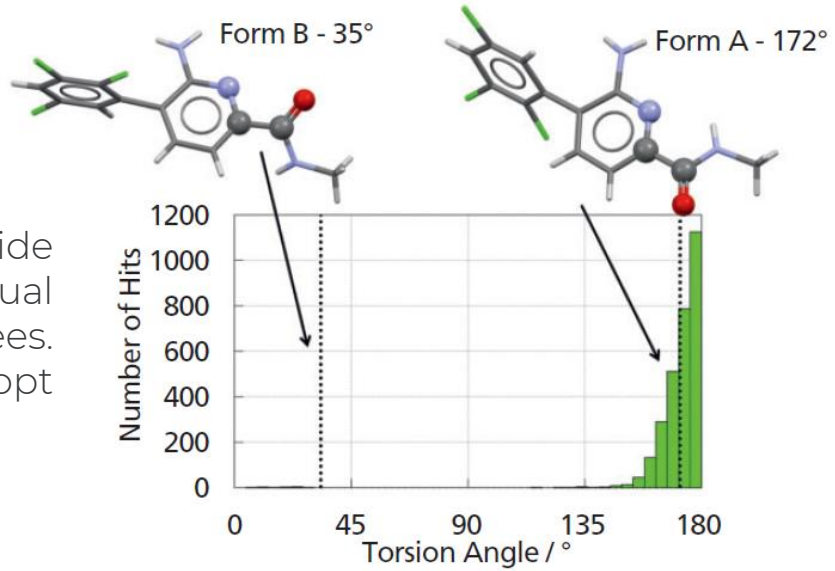




CCDC

Intramolecular Geometry

- Millions of fragments of molecules abstracted from the CSD and data for bond lengths, bond angles and torsion angles stored in an easily searchable format. These data are mined to check the intramolecular geometry of candidates.
- Unusual torsion angles



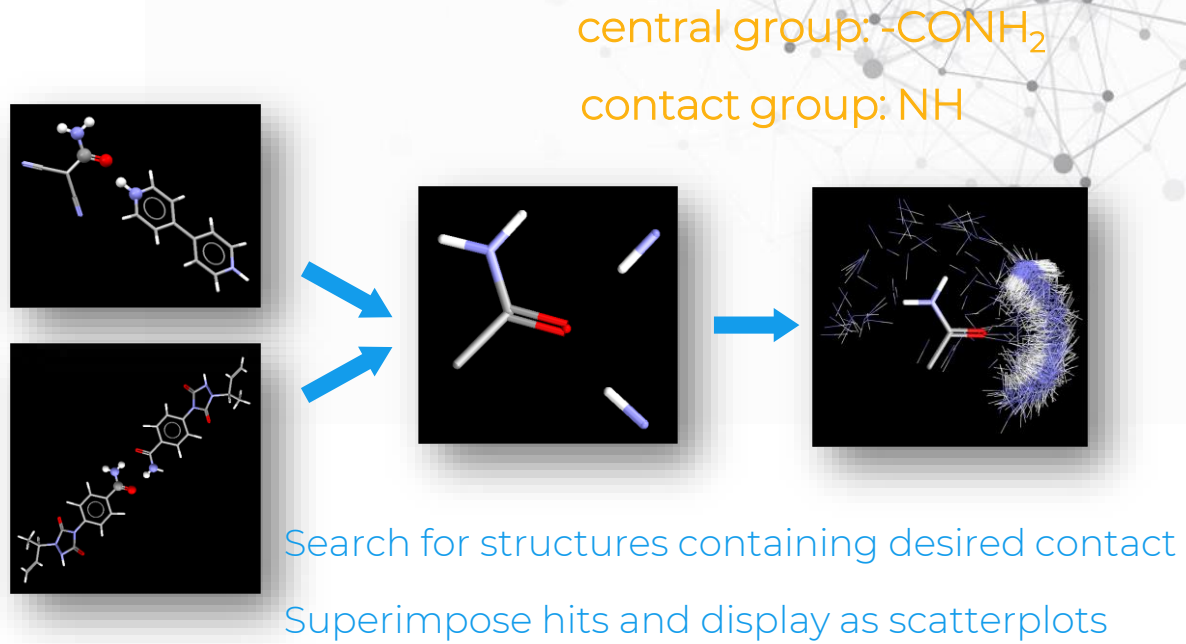
Distribution of torsion angles observed for the amide carbonyl around 180 degrees. Form B shows an unusual orientation of this group with a value of 35 degrees. Barely visible on the scale used, six structures adopt torsion values of less than 45 degrees.

Feeder N., et al., (2015) JPP, 67:857, DOI:10.1111/jphp.12394

CCDC

Intermolecular interaction geometries

- A knowledge base of *intermolecular interactions* for individual functional group pairs
- Experimental data taken from:
  - Cambridge Structural Database
  - Protein Data Bank (protein-ligand complexes only)
- Interaction distributions displayed as scatterplots or contour surfaces
- Contours calculated with respect to “random” density based on number of central and contact groups in unit cell



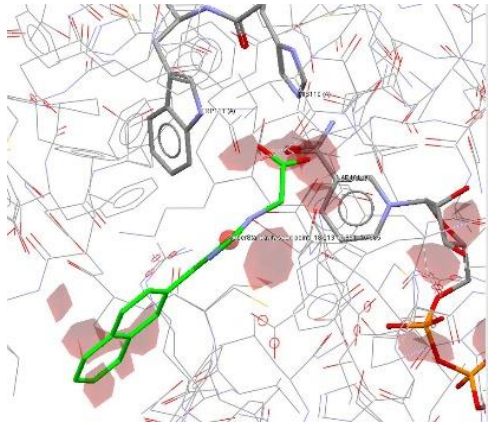
Verdonk M. et al.(1999) J. Mol. Biol., 289:1093, DOI: 10.1006/jmbi.1999.2809

CCDC

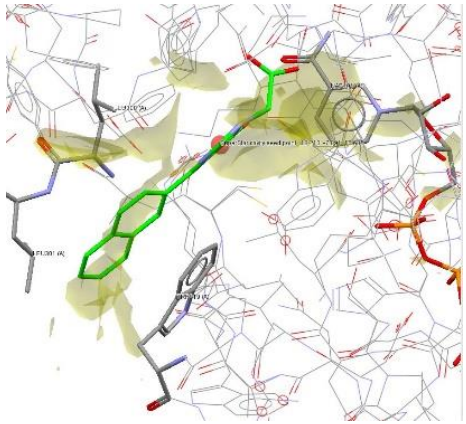
Binding site analysis

- Experimental information about *non-bonded interactions* occurring in small-molecule crystal structures is used to show the experimentally observed distribution of *functional groups of probes*.
- A *protein binding site* is broken down into structure fragments and the scatterplots, showing the distribution of a chosen probe around these structure fragments.
- *Propensity maps* are used to aid in manual docking and structure-based drug design.

H-bond acceptor propensity map



hydrophobic propensity map

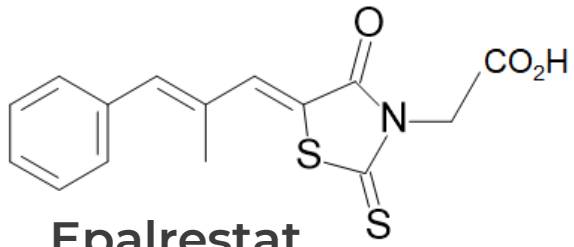


Verdonk M. et al.(1999) J. Mol. Biol., 289:1093, DOI: 10.1006/jmbi.1999.2809

CCDC

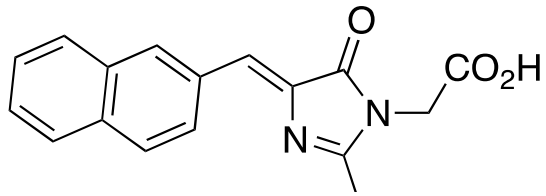
Rationalising activity

- Analyse the protein-inhibitor interactions within the binding site
  - Carboxylate group and the naphthyl ring in the ligand correlates well with the areas of interactions predicted.
- Rationalise the activity of the docked pose of the new compounds
  - Activity linked to an explored lipophilic pocket that is also explored by known compound epalrestat.



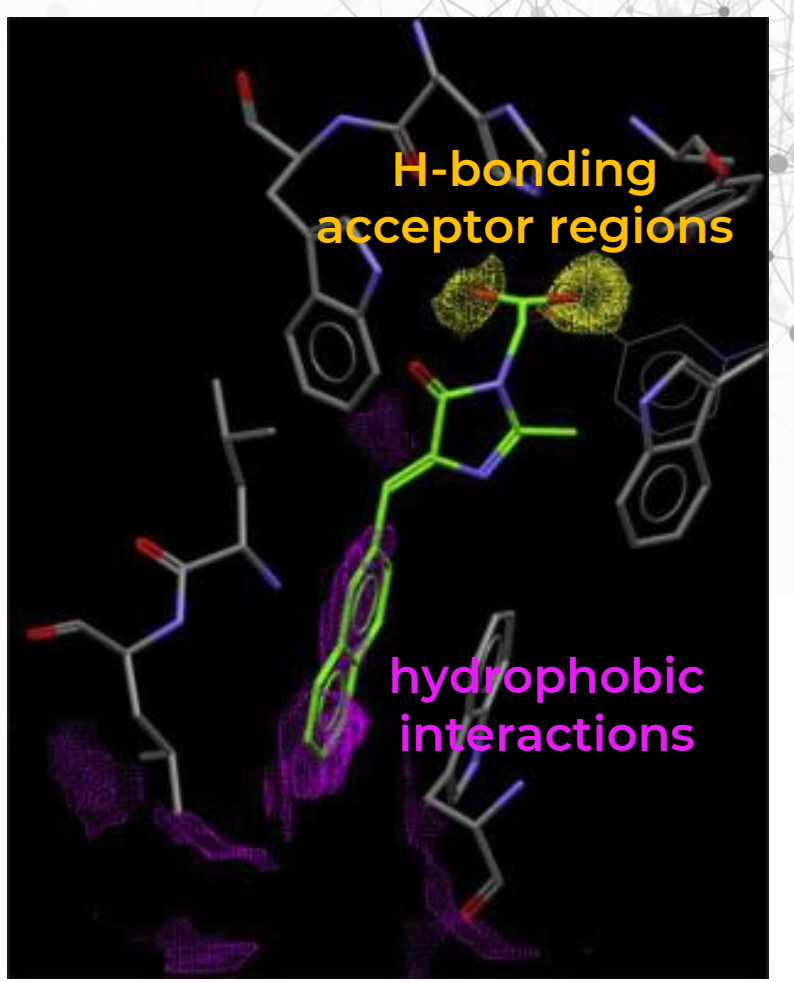
Epalrestat

IC_{50} = 0.085 µM



New inhibitor

IC_{50} = 0.10 µM



Saito et al., (2017) Eur. J. Med. Chem. 125: 965. DOI: 10.1016/j.ejmech.2016.10.016

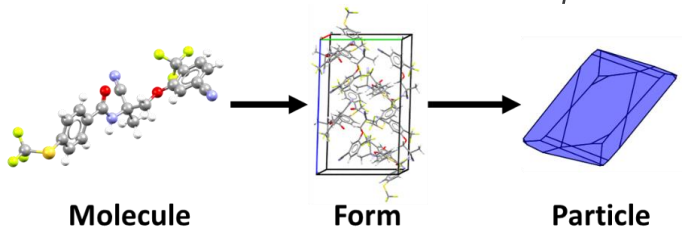
CCDC

Solid form informatics

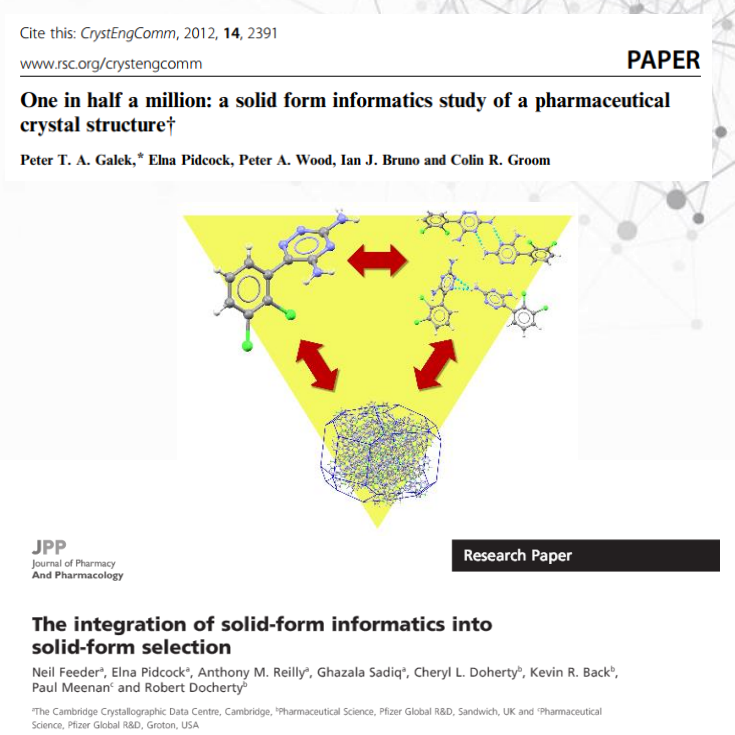
- Is this the most stable solid form of a potential drug?
- Are there weaknesses which may relate to risk of alternate forms?

The term "**solid form informatics**" first introduced in mid-2000s by Bob Docherty (Pfizer):

Use of structural knowledge to inform key decisions in pharmaceutical development



Galek P. et al.,(2012), CrystEngComm, 14: 2391, DOI: 10.1039/C2CE06362J



Cite this: *CrystEngComm*, 2012, **14**, 2391

www.rsc.org/crystengcomm **PAPER**

One in half a million: a solid form informatics study of a pharmaceutical crystal structure†

Peter T. A. Galek,* Elna Pidcock, Peter A. Wood, Ian J. Bruno and Colin R. Groom

JPP Journal of Pharmacy And Pharmacology

Research Paper

The integration of solid-form informatics into solid-form selection

Neil Feeder[a], Elna Pidcock[a], Anthony M. Reilly[a], Ghazala Sadiq[a], Cheryl L. Doherty[b], Kevin R. Back[b], Paul Meenan[c] and Robert Docherty[b]

[a]The Cambridge Crystallographic Data Centre, Cambridge, [b]Pharmaceutical Science, Pfizer Global R&D, Sandwich, UK and [c]Pharmaceutical Science, Pfizer Global R&D, Groton, USA

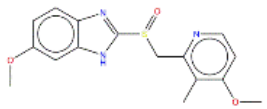
CCDC

Solid form risk assessment

- Crystal structure
  - Space group, chirality of molecule, Z', voids, disorder
- Intramolecular geometry
  - Bond lengths, bond angles, torsion angles
- Intermolecular interactions
  - Likelihood of donor acceptor pairings
  - Likelihood of functional group coordination scenario
  - Geometry of intermolecular interactions
- Statistics on all of these aspects: helps with the recognition of "unusual" and identifies opportunities. Feeds into risk management and design of experiment.

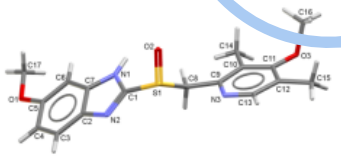
CCDC

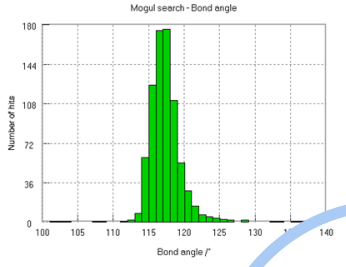
Solid form risk assessment



- Auto-edit & normalise
- Assess bond typing

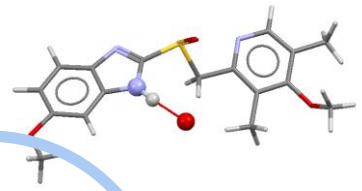
Load Structure





Conformation Analysis

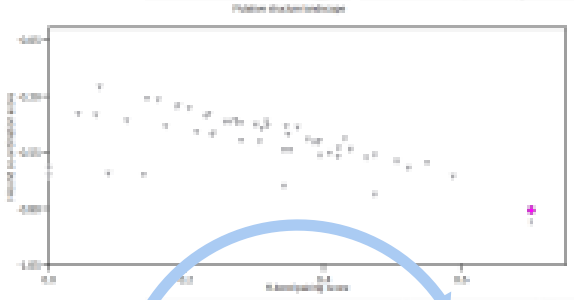
- Geometry check



- H-bond geometry
- Compare with CSD

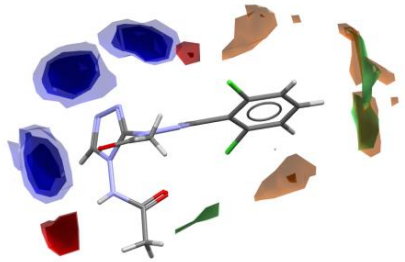
Interactions Analysis

D	A	H...D (Å)	D...A (Å)	Mean (CSD)	D-H...A (°)	Mean (CSD)
N1	O2	1.892	2.743	2.86 ± 0.13	175.03	162.6 ± 10.3



- HBP analysis
- Full HBP & coordination chart

H-bond Network Analysis



- Analyse FIMs
- Assess void space

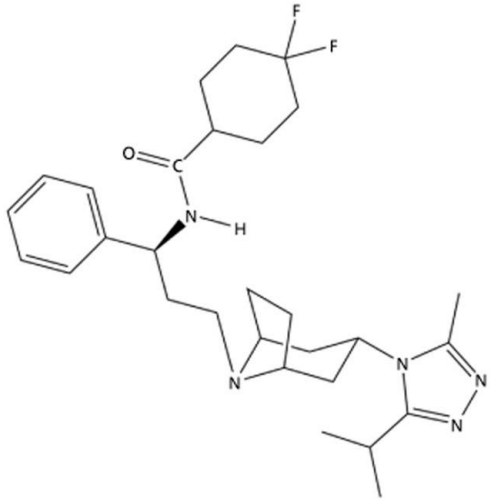
Assess Crystal Packing

✓ / ✗

	Compound B Form I	CSD Statistics
Crystal System	Triclinic	CSD Freq.: 24.7 %
Space Group	P -1	CSD Freq.: 23.8 %
Cell Lengths	a: 9.6421 b: 10.3865 c: 10.1539	N/A
Cell Angles	α: 89.929 β: 110.939 γ: 116.937	N/A
Cell Volume	830.784	N/A
Z: Z'	2: 1	N/A
R-factor (%)	5.66	N/A
Temperature (K)	100	N/A
Density (g cm⁻³)	1.381	Mean: 1.386 ± 0.239

CCDC

Solid form risk assessment - Example



Maraviroc - viral entry inhibitor

Analysis for stability and polymorphism

Informatics based **Healthcheck assessment**

+

Experimental screening of solid form

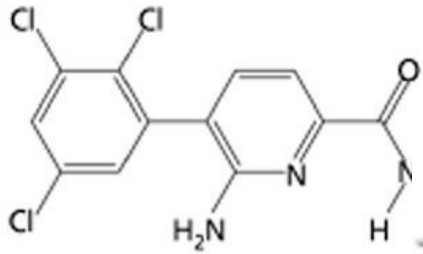
Healthcheck **suggested a more stable form** possible: confirmed by further experimental work.

CSD-Materials tools to assess solid form stability, reducing development risk.

CCDC
advancing structural science

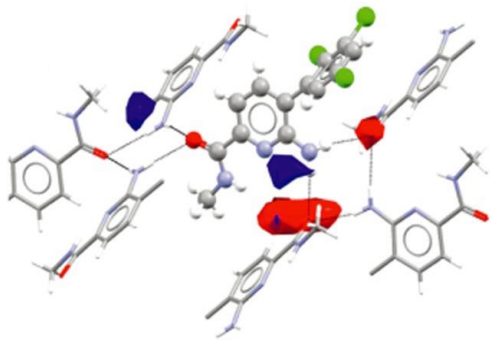
Feeder N., et al., (2015) JPP, 67:857, DOI:10.1111/jphp.12394

Solid form risk assessment - Example

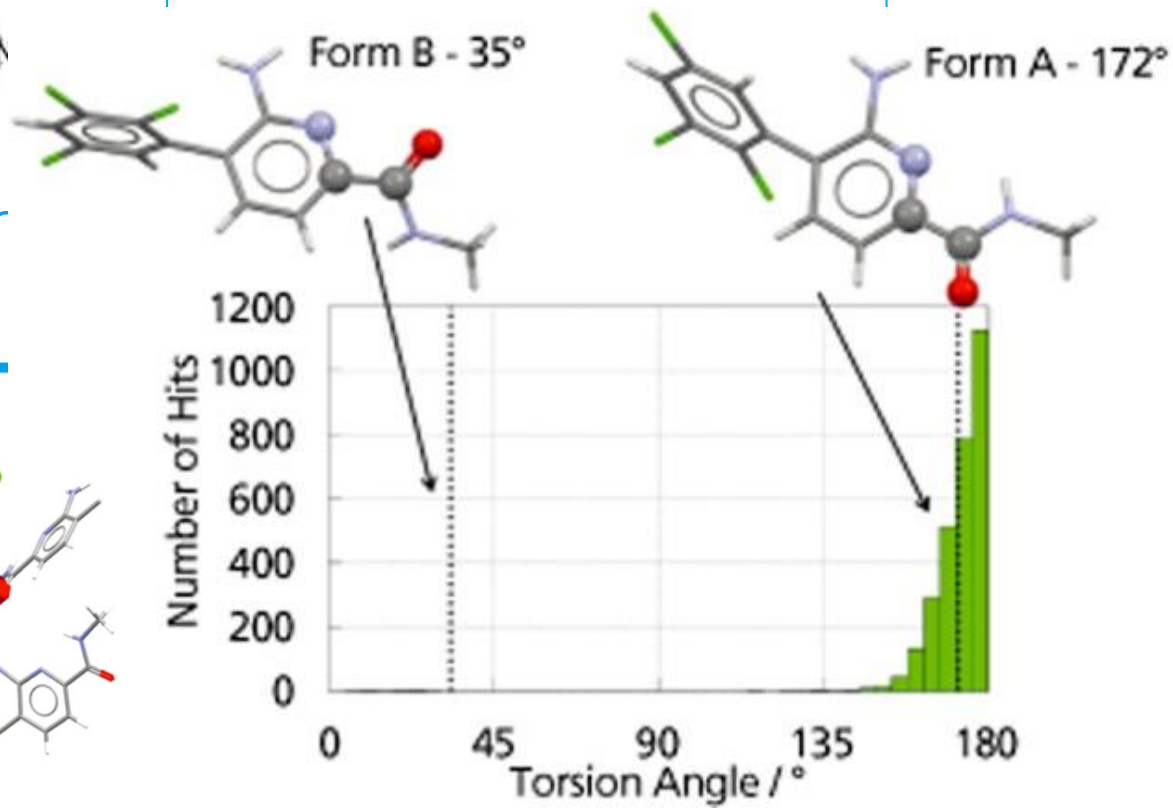


Metastable forr

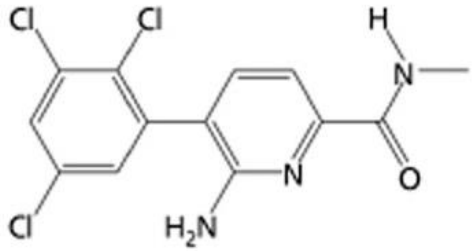
Pain candidate drug



HB donor
HB acceptor

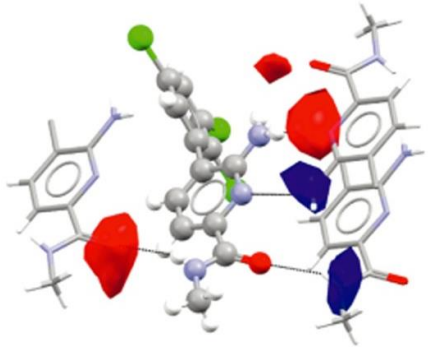


CSD-Materials tools to assess solid form stability, reducing development risk.



Stable form B

Healthcheck confirmed which of 2 forms was most stable.



CCDC
advancing structural science

Feeder N., et al., (2015) JPP, 67:857, DOI:10.1111/jphp.12394

Solid form risk assessment - Example

Crizotinib - anti-cancer drug

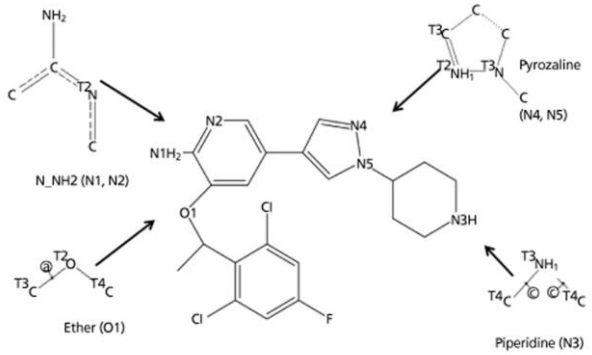
Analysis for stability and polymorphism

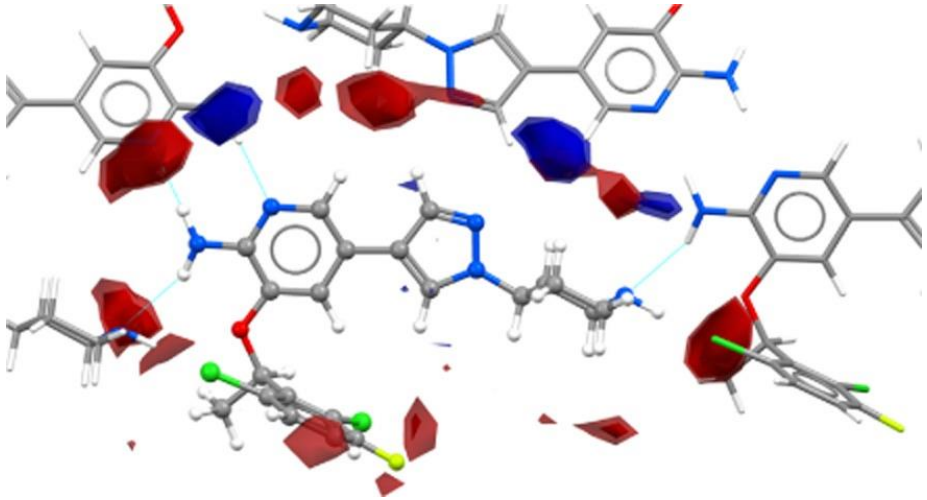
Informatics based **Healthcheck assessment**

+

Experimental screening of solid form

Monomorphic - **agreement** from both analyses.





CSD-Materials tools to assess solid form stability, reducing development risk.

CCDC
advancing structural science

Feeder N., et al., (2015) JPP, 67:857, DOI:10.1111/jphp.12394

Summary

- The data stored in *structural databases*, such as the CSD and the PDB is valuable in drug discovery research
- The databases can be searched to identify *interaction patterns* or potential *scaffold hops*.
- The information about *molecular conformations* and *intermolecular interactions* provides statistical guidance of observing a conformation of a drug in the binding site.
- The collection of statistical data on crystal structure helps with the recognition of “unusual” and feeds into *risk management*.

CCDC

Thank you

Vera Prytkova,
Research and Applications Scientist
vprytkova@ccdc.cam.ac.uk

CCDC

CCDC

advancing structural science