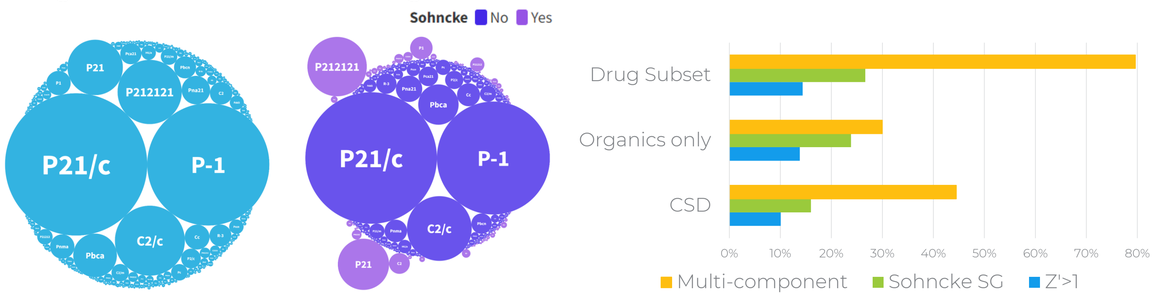
Context is everything: application of CSD-derived knowledge to investigate solid form landscapes

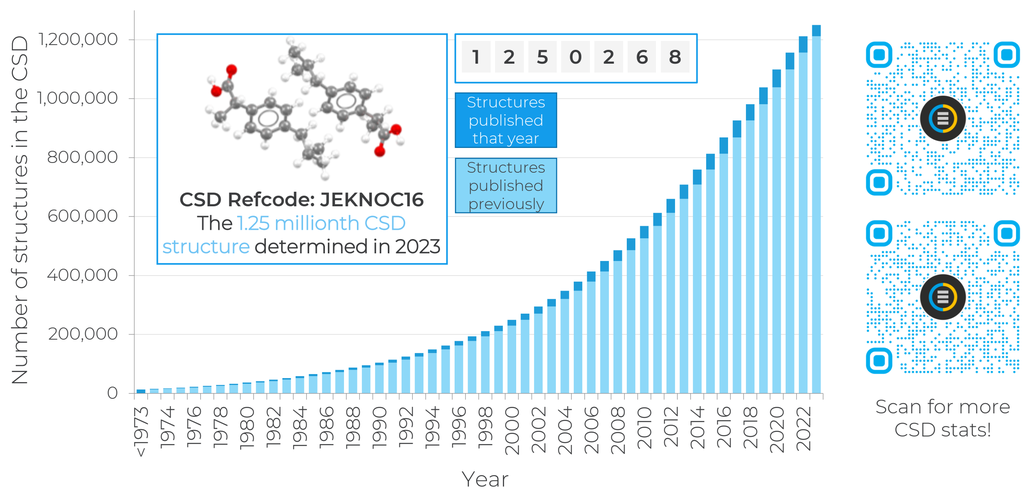
Pablo Martinez-Bulit, Elna Pidcock, Ghazala Sadiq, Joanna S. Stevens
The Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, CB2 1EZ, UK

CCDC advancing structural science

The Cambridge Structural Database

With over 1.25 million curated entries, the Cambridge Structural Database[1] contains crystallographic information of a wide range of chemical compounds capturing both intra- and intermolecular interactions. It is also a source of statistical information for conformational parameters, frequency of occurrence of space groups and high Z' structures, packing coefficients, distributions of chemical motifs, etc.
All these data can be used to perform a quick cheminformatics assessment on any crystalline form of an Active Pharmaceutical Ingredient (API) and contextualise its strengths and weaknesses during solid-form selection.[2]

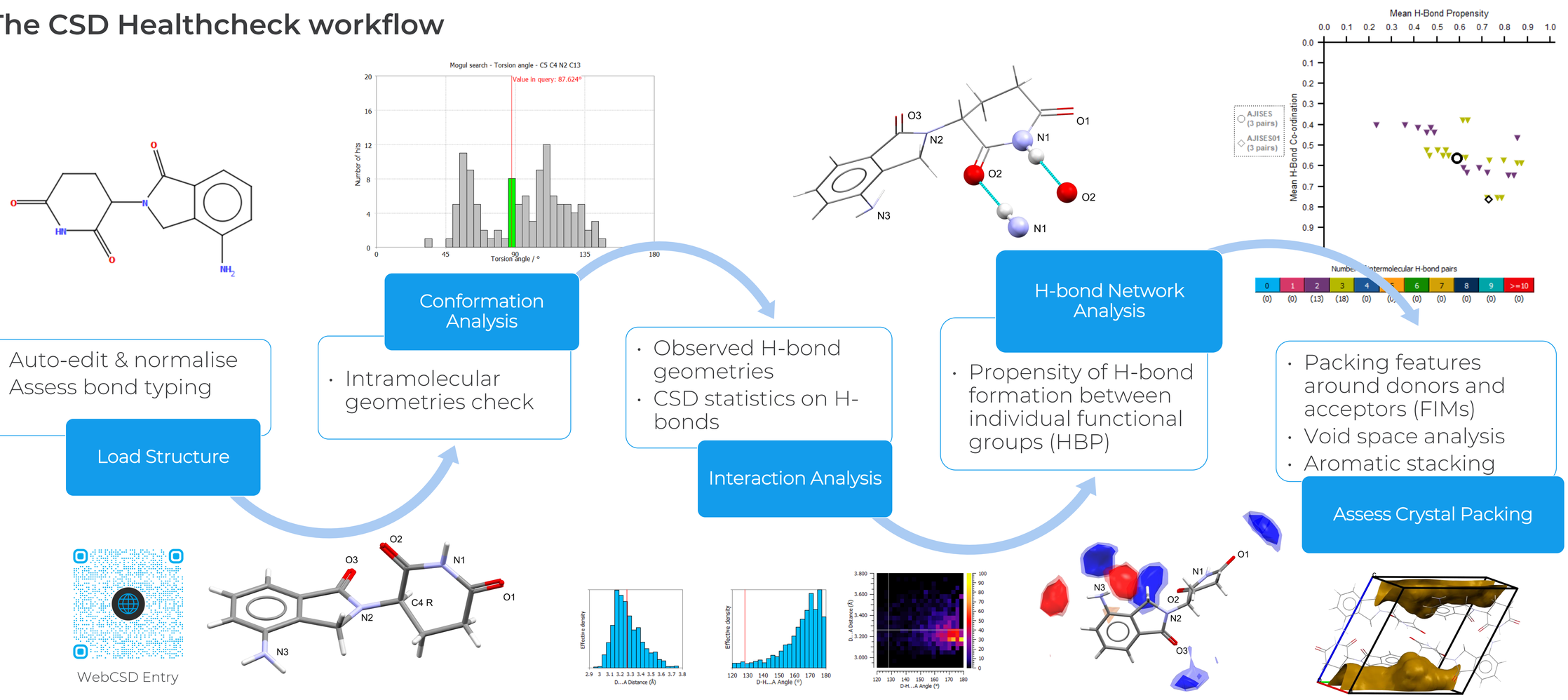




Scan for more CSD stats!

The information in the CSD can be refined to provide maximum relevance by filtering the entries according to features of interest, by excluding entries not relevant to the molecules under study, or by building relevant subsets.

The CSD Healthcheck workflow

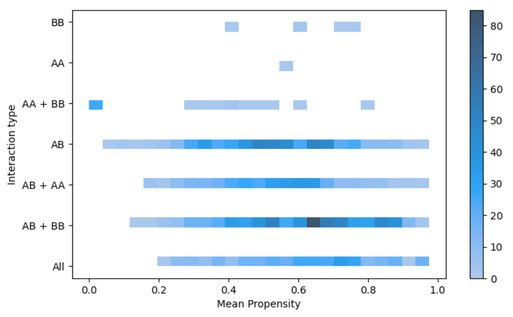


Illustrative workflow of a healthcheck on lenalidomide (CSD Refcode: AJISES), 4th top small molecule by retail sales in 2022.[3] There are eight known crystalline forms, including hydrates and solvates. Every step in this workflow compares the structure of interest to relevant structures in the CSD and can help identify unusual aspects of the crystal forms.

Multi-component HBP

The Hydrogen Bond Propensity[4] (HBP) tool ranks the likelihood of hydrogen bond interactions between functional groups in a molecule. By programmatically running HBP on a cocrystal dataset, we can study hydrogen bonding trends in multi-component systems.

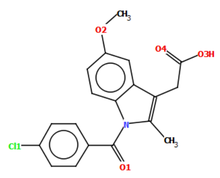
Interactions	Self A:A - Self B:B ≤ -0.2	Self A:A - Self B:B ≤ -0.2	Self A:A - Self B:B > 0.2
A:A	0%	0%	0%
A:A + B:B	1%	3%	3%
A:B	27%	24%	33%
A:B + A:A	5%	20%	47%
A:B + B:B	52%	27%	6%
A:A + A:B + B:B	15%	25%	11%
B:B	0%	0%	0%



A dataset of ~5,000 cocrystals was analysed, showing that the observed hydrogen bonding networks vary depending on the differences in self-association propensities for the active (A) and the coformer (B).

Interaction Analysis

Databases for specific contacts can be constructed by mining the CSD. With a database capturing popular H-bonding interactions, an API can be screened against coformers to rank them based on the overall potential number of interactions that occur frequently between functional groups in the API and the coformer.



Indomethacin
INDMET

H-bond Donor	Top H-bond Acceptors	F. Occ. (%)	Hits	Coformer acceptor matches
		95	459	3-methylpyridine, isonicotinamide, nicotinamide, pyrazine
		75	1118	L-aspartic acid, L-glutamic acid
		66	249	
		56	53	
		55	78	
		54	126	Folic acid

H-bond Acceptor	Top H-Bond Donors	F. Occ. (%)	Hits	Coformer donor matches
		6	54	
		40	58	imidazole, theophylline, xanthine
		27	76	4-hydroxybenzoic acid, apigenin, capsaicin, ethylparaben, gentisic acid, hesperetin, L-tyrosine, methylparaben, pamoic acid, propylparaben, t-butylhydroxyanisole
		98	92	
		82	181	isonicotinamide, L-glutamine, nicotinamide, phthalamide
		57	454	L-aspartic acid, L-glutamic acid
		20	300	

Conclusions

The solid-form informatics workflows and tools developed at the CCDC allow scientists to quickly screen crystalline forms and identify suboptimal characteristics during the solid-form selection process. They can also help guide the experimental screenings for polymorphs, solvates, hydrates, or cocrystals by identifying trends in the Cambridge Structural Database and its subsets, and providing relevant context for the molecules under study.

Acknowledgements

We would like to thank the Crystal Form Consortium members for their continuous support towards the development of these tools.

References

1. Groom C.R., Bruno, I.J, Lightfoot, M.P., Ward, S.C., *Acta Cryst*, 2016, *B72, 171-179*
2. Feeder, N., Pidcock, E., Reilly, A.M., Sadiq G., Doherty, C.L., Back, K.R., Meenan, P., Docherty, R., *J Pharm Pharmacol*, 2015, *67(6)*, 857–868.
3. McGrath, N.A., Brichacek, M., Njardasson, J.T., *J Chem Ed*, 2010, *87(12), 1348-1349.*
4. Galek, P.T.A., Fabian, L., Motherwell, W.D.S., Allen, F.H., Feeder, N., *Acta Cryst*, 2007, *B63, 768-782*

www.ccdc.cam.ac.uk
pbulit@ccdc.cam.ac.uk