TECHNICAL ARTICLE

Utilizing organic and organometallic structural data in powder diffraction

Jason C. Cole,[a] Elena A. Kabova, and Kenneth Shankland
School of Pharmacy, University of Reading, Whiteknights, Reading RG6 6AD, UK

(Received 16 June 2014; accepted 26 August 2014)

The Cambridge Structural Database (CSD) is a database of small molecule organic and organometallic crystal structures elucidated using X-Ray and neutron crystallography. The CSD is distributed alongside a system of software (the Cambridge Structural Database System) to academic and industrial users. The system contains a number of applications (in particular DASH, ConQuest, and Mogul) that can be used to aid crystallographers in the solution and refinement of crystal structures from powder diffraction data, and in the interpretation of crystal structure models (in particular, Mercury). This publication uses a racemic form of ornidazole ($Z' = 3$) to illustrate the efficacy of DASH in the crystal structure solution from powder diffraction data. Furthermore, numerous features in Mogul and Mercury that aid crystal structure solution and interpretation of crystallographic data are revised. Finally, a review of a new method for using database-derived geometric information directly in structural solution is presented. © *2014 International Centre for Diffraction Data.*
[doi:10.1017/S0885715614000827]

Key words: CSD, Mogul, DASH, Mercury, powder diffraction, structure determination

I. INTRODUCTION

The Cambridge Structural Database (CSD) is a database of organic and organometallic crystal structures that have been published in the academic literature, patents or received through direct private communications from crystallographers. The majority of entries in the CSD contain an accurate description of the chemical connectivity and the full three-dimensional (3D) coordinates of the associated structure along with bibliographic information extracted from the original publication. The Cambridge Structural Database System (CSDS) allows users to search and retrieve crystallographic entries and perform the data analysis on the wealth of information contained within the database. It is used widely across the scientific community in, for example, drug design and development, crystal engineering, protein–ligand structure solution and refinement.

The Cambridge Crystallographic Data Centre (CCDC) is a not-for-profit organization dedicated to the curation and redistribution of crystallographic information from small molecule crystal structures. The CCDC's primary responsibility is the creation of the CSD. CCDC is funded by annual contributions from its user community.

CCDC acts as a custodian of crystallographic information to the wider chemical community. Individual structures are available for download free-of-charge from the CCDC's web site, and, in 2013, over 100 000 structures were accessed via this portal. The complete database is available along with access software to financial contributors. Key parts of the CSDS are summarized in Figure 1. The software applications, ConQuest and Mercury (Bruno *et al.*, 2002) provide the means to search, retrieve and analyze the structures. The knowledgebases, Mogul (Bruno *et al.*, 2004) and IsoStar (Bruno *et al.*, 1997) form the next tier of the CSDS. These are databases of *derived* information: Mogul captures data from CSD entries regarding intra-molecular geometric preferences; IsoStar presents distributions of interaction preferences for a wide variety of chemical groups. Finally, there are software applications that utilize CSDS data in their operation. Of particular relevance to powder diffraction is DASH, a system for solving crystal structures from powder diffraction data.

Applications in the CSDS are interlinked; for example, Mercury and ConQuest allow a user to perform database searches, while Mercury can cross-link to Mogul and IsoStar for structural analyses and DASH can make direct use of Mogul for assessing conformational preferences in structure solution.

Parts of the CSDS can also be usefully integrated into third party applications. For example, Mogul is ideally suited to the generation of restraints that can be utilized in crystal structure refinement. Indeed, research groups in the field of protein structure refinement now use Mogul routinely for generating ligand dictionaries in macro-molecular structure refinement (Smart *et al.*, 2001). Small molecule crystallographers can use Mogul via a link in the CRYSTALS refinement package (Betteridge *et al.*, 2003). Most recently, Rigaku have developed a link to Mogul that allows users to restrain bond lengths and angles in structural models during Rietveld refinement to the means of those suggested by Mogul (Rigaku, 2013).

In this paper, the structure solution of a racemic form of ornidazole from powder diffraction data is reported. This structure and several other CSD structures are used to illustrate how the CSDS can benefit structural solution and refinement from powders and to show how the CSDS can aid in the analysis of the structural information obtained.

A. Growth of Data in the CSD

Editors at the CCDC incorporate all organic and organometallic crystal structures published in the chemical literature

[a]Author to whom correspondence should be addressed. Electronic mail: cole@ccdc.cam.ac.uk

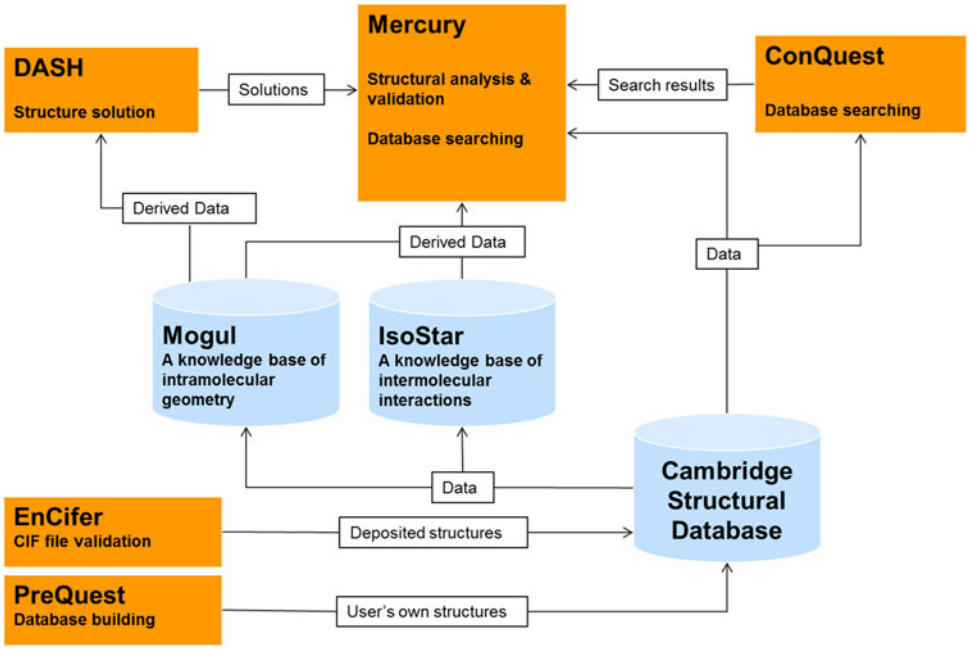

Figure 1. (Color online) A summary of the key parts of the CSDS that are used in structural studies. The annotations summarize the flow of information through the system from a user's perspective. Databases are shown in as cylinders in blue while applications are shown as larger rectangles in orange.

into the CSD. In addition to published structures, they encourage researchers to deposit with the CCDC previously unpublished structural data, or new crystal structures that are not intended for wider publication. This strategy makes the CSD a more valuable resource to the community.

In Figure 2, the growth of crystallographic output is shown. The growth appears to approximate to the exponential phase of a sigmoidal growth curve [see Figure 2(a)]. The ever increasing volume of data presents a significant challenge for data curation organizations, driving the need for process automation.

A recent article on the use of structure determination by powder diffraction (SDPD) methods for solving small-molecule crystal structures of pharmaceutical interest summarized the growth of molecular crystal structures in the CSD (Shankland *et al.*, 2013). A similar analysis that considers

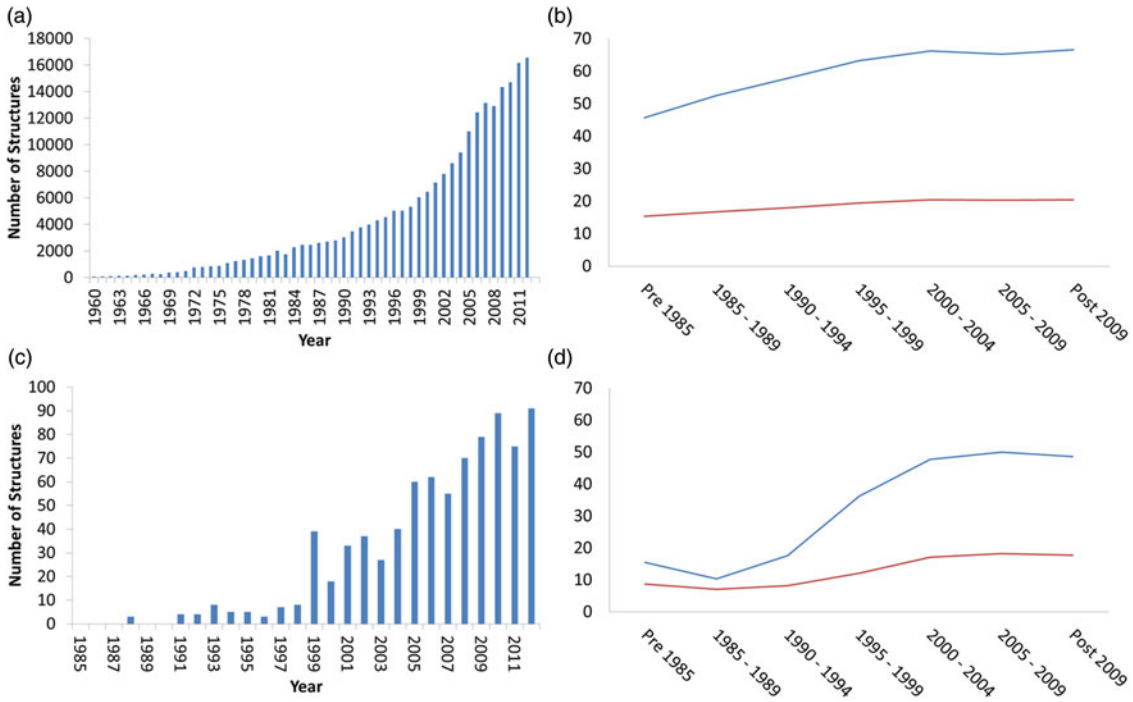

Figure 2. (Color online) (a) Growth in the number of organic crystal structures in the CSD. (b) Growth in the average complexity of organic systems in the CSD. The mean of the number of atoms is in blue and number of DoF is in red (see supplementary information for details of how these values are calculated) for structures published since 1985. (c) Growth in the number of organic crystal structures in the CSD solved from powder data. (d) Growth in the average complexity of organic crystal structures in the CSD from powder data. The data shown are derived from structures published up to the end of 2012.

https://doi.org/10.1017/S0885715614000827 Published online by Cambridge University Press

all structures (Shankland *et al.* considered only structures with $Z' = 1$) is presented here and shown in Figure 2(b). While the variance in the data points is typically quite large, viewed over time we can see a broad trend that suggests that the complexity of molecular organic crystal structures deposited into the CSD has, unsurprisingly, increased up until the year 2000.

Figures 2(c) and (d) show equivalent plots, but this time only considering those structures in the CSD derived from powder diffraction data. The total numbers of structures are tiny by comparison with the full database; just 0.5% of the total in 2012, however, these structures show the same underlying trends as the full database: a significant growth in volume of data received, and an increase in overall structural complexity. It is interesting to note that the structural complexity of SDPD studies is now, on average, comparable to that of the single-crystal organic structures added to the CSD. This reinforces the notion that crystal structure determination of pharmaceutically relevant compounds using powder diffraction data alone is not only routine (David *et al.*, 1998), but returns structures that are of high quality (Karki *et al.*, 2007; Lapidus *et al.*, 2010).

B. Powder diffraction studies in the CSD

A key application area for the CSD is determining trends for specific chemical features (such as bond lengths) in the specific chemical systems. Typically these studies will rely on the atomic coordinates established from single-crystal studies of well-ordered structures that yield relatively accurate coordinate positions, as a consequence of the high data-to-parameter ratio in the least-squares structure refinement. Powder diffraction refinements suffer from a much lower data-to-parameter ratio and generally rely upon either the use of rigid-bodies or soft restraints on structural parameters in order to maintain chemical sense during the refinement stage. Nevertheless, they are of great import in situations where no single crystal can be obtained and where, for example, variable pressure/temperature/humidity experiments are required. Lapidus *et al.*, (2010) have compared the structural quality of compounds that were determined by powder diffraction studies and single-crystal studies. They conclude that if carefully performed, powder diffraction studies of co-crystals are of comparable accuracy with single-crystal studies, but naturally the precision of the results is lower. As such, the CSDS currently retains a discriminatory flag on powder studies. If structural data are abundant for a given feature of interest, the user can choose to omit powder studies from the results, but even with 700 000 structures to choose from, users can find that the number of crystal structures relevant to their search are limited; in such cases a key criterion for using data is to understand its relative precision, and interpret results accordingly.

II. USING THE CSDS TO AID POWDER DIFFRACTION STUDIES

A. Structure solution using DASH and Mogul

DASH (Florence *et al.*, 2005; David *et al.*, 2006b) utilizes simulated annealing (David *et al.*, 1998; Shankland *et al.*, 2002b) to minimize the differences between observed powder diffraction data and data calculated from a structural model in direct space. One of the key features of DASH is the speed with which candidate structures are evaluated; rather than using the time-consuming method of calculating a full diffraction pattern for every candidate structure (of which there are typically millions in a crystal structure determination) it uses reflection intensity data and the correlations between reflection intensities obtained for overlapping reflections in the powder X-ray diffraction (PXRD) pattern. Combined with a simulated annealing algorithm whose parameters are set automatically by the program, this leads to a fast and effective protocol for the crystal structure solution, as evidenced by many structures solved using it. A few recent examples include trospium chloride with 14 degrees of freedom (DoF: see supplementary information for a full definition) (Skorepova *et al.*, 2013), mebendazole with 11 DoF (Ferreira *et al.*, 2010), nimustine hydrochloride with 16 DoF (Bekoe *et al.*, 2012), indomethacin–nicotinamide co-crystal with 18 DoF (Majumder *et al.*, 2011), and naltrexone hydrochloride with 12 DoF (Guguta *et al.*, 2009). Some more complex examples of crystal structures include AR-C69457CC with 26 DoF (Johnston *et al.*, 2004), chlorothiazide N,N-dimethylformamide solvate with 42 DoF (Fernandes *et al.*, 2006), carbamazepine form II and cyheptamide form II with 28 DoF (Fernandes *et al.*, 2007) and docetaxel monohydrate with 23 DoF (Vella-Zarb *et al.*, 2013), confirming that DASH is capable of solving structures with DoF greater than the average shown in Figure 2(d).

By way of an illustrative example of a structure with significantly more DoF than the average, we include here the solution from powder data of the antifungal agent ornidazole. Synchrotron powder diffraction data were collected at 130 K on the high-resolution powder diffractometer (BM16) of the European Synchrotron Radiation Facility (ESRF) by Shankland and David ($\lambda = 0.652\,78$ Å, personal communication, 1997) and the pattern indexed to a triclinic unit cell, $a = 13.636\,01$ Å, $b = 14.055\,59$ Å, $c = 8.930\,88$ Å, $\alpha = 71.6038°$, $\beta = 78.5696°$, $\gamma = 64.8111°$, $V = 1465.72$ Å^3 with space group $P\bar{1}$, suggesting $Z' = 3$ and total of 30 DoF (nine positional, nine rotational, and 12 torsional). Repeated attempts to solve the structure using DASH met with failure and even the publication of a single-crystal structure (Shin *et al.*, 1995) which then enabled the use of correctly folded starting conformations (thus eliminating 12 DoF) failed to yield a result. Recently, the data were revisited as part of a systematic study of the performance of DASH across a wide range of structural complexities. The model used in this latest study was derived from the fully ordered component of the CSD entry NETRUZ01 (Anderson *et al.*, 2009) and the crystal structure solved by DASH using 500 simulated annealing runs each utilizing 5×10^7 SA moves for the 30 DoF, i.e. all three independent molecules fully flexible around their four torsion angles. The best DASH solution obtained has a very favorable $\chi^2_{SA}/\chi^2_{Pawley}$ ratio of 2.74, strongly suggesting that the structure has been solved. A scale-factor-only Rietveld refinement in TOPAS (Coelho, 2003) gave an R_{wp} value of 11.47 (Figure 3), which compares favorably with the Pawley R_{wp} value of 6.68. The DASH solution is in excellent agreement with that of NETRUZ01; crystal packing similarity using Mercury returns an RMSD of 0.11 Å for 15/15 molecules and an overlay of one of the independent molecules is shown in Figure 4. Of the 500 runs performed in this work, which took a total run time of 480 CPU hours on a 12 core 2.6 GHz Xeon(R) processor, only four reached the global

https://doi.org/10.1017/S0885715614000827 Published online by Cambridge University Press

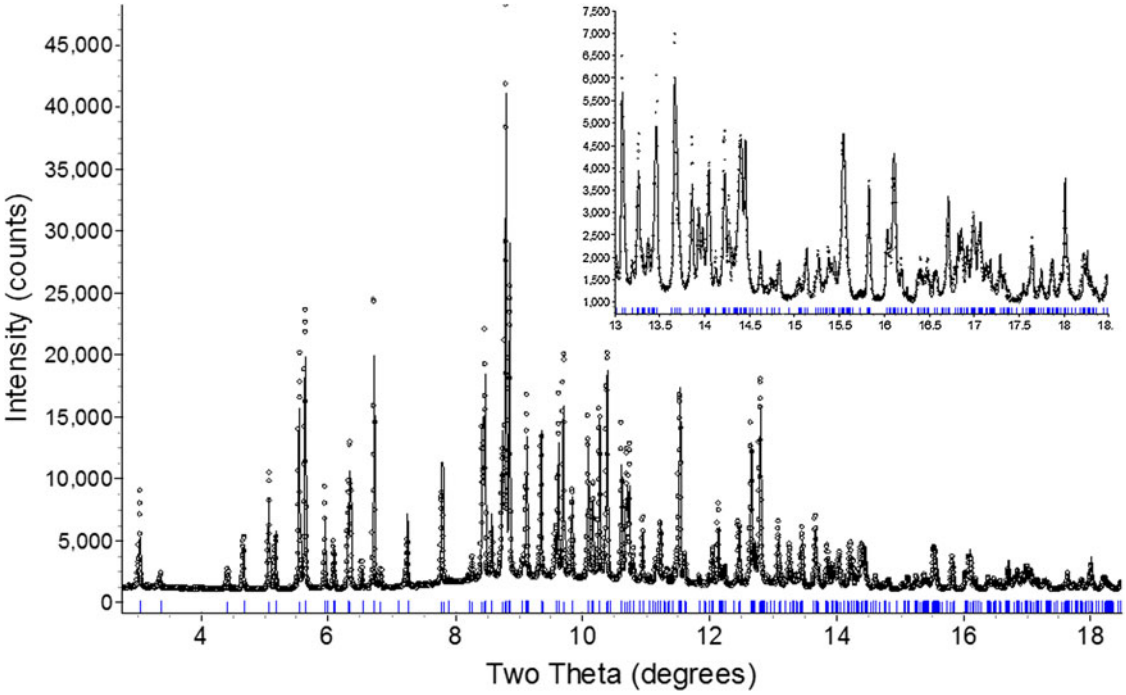

Figure 3. (Color online) A scale-factor-only Rietveld refinement of the best DASH solution for racemic ornidazole.

minimum. Given that the core SA algorithm in DASH has not varied substantially over the years, it is clear in retrospect that the ornidazole crystal structure *could* have been solved when the data were collected, had sufficient computational power been available. It is not clear why this particular structure requires so many long SA runs in order to reliably locate the global minimum; other complex structures with $Z' > 1$ have been solved with much greater success rates using DASH, e.g. carbamazepine form II and cyheptamide form II, both of which have $Z' = 4$.

The size of the search space that needs to be explored by the SA algorithm scales exponentially with the number of DoF in the structure under study, making it (in general) more difficult and more time-consuming to reliably locate the global minimum in that space (Shankland *et al.*, 2002a). Utilizing prior structural knowledge can however, dramatically extend the applicability of such algorithms. In a normal DASH run, the input structural model already utilizes a wealth of structural information, in the form of well-known bond lengths, angles, and fixed torsion angles; it is only those (flexible) torsion angles that are not known in advance that are treated as parameters to be determined as part of the simulated annealing process. Of course, one also has prior information on the values likely to be adopted by some of the torsional parameters in the molecule and this information can be used to restrict search space to these regions of higher probability. This additional information can be extracted from relevant structures in the CSD using Conquest, or more conveniently using the Mogul knowledge-base, which contains conformational preferences of molecules in the CSD in the form of distributions of bond lengths, angles, and torsion angles. Mogul has previously been used (Florence *et al.*, 2005) to extend the scope of structure solutions using DASH; the approach taken was to use Mogul to aid an investigation in assessing the likely conformational preferences of a given torsion angle. The investigator then made a judgment, based on the prior structural evidence available, about the angular limits applied to a given flexible torsion angle in the molecule under study, and in so doing, cut down the amount of space that needs to be explored. This method, while effective, has the drawback in its reliance on the expertise of the user and so a more automated method has been developed in which SA moves for torsion angles are biased toward regions that are likely to be observed in such a molecule, as predicted by Mogul. The effectiveness of this Mogul directional biasing (MDB; David *et al.*, 2006a) has yet to be fully evaluated, but some encouraging results have been obtained on a number of structures and these are summarized in Table I. It is clear that the original Mogul approach improves results across the board and that MDB is effective in the cases of famotidine and capsaicin.

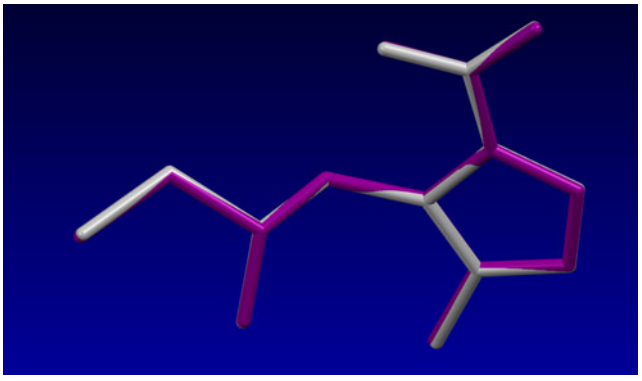

Figure 4. (Color online) A representative molecular overlay, generated in Mercury 3.3, of the best DASH solution and NETRUZ01.

The result for verapamil hydrochloride is poor when using MDB. We include it to illustrate that MDB is not a panacea. The result can be explained by consideration of the structure and the applied method. The correct structure in verapamil hydrochloride contains a torsion angle that resides in a smaller peak of the Mogul distribution; it therefore has a less likely

https://doi.org/10.1017/S0885715614000827 Published online by Cambridge University Press

TABLE I. Success rates for three standard test cases using different Mogul approaches. The success rate reported is determined based on 50 SA runs of 10^7 moves. All datasets taken from Florence *et al.* (2005).

	Success rate (%)				
	DoF external	DoF internal	Fully flexible	Mogul	MDB
Famotidine	6	7	44	64	72
Capsaicin	6	10	10	22	18
Verapamil HCl	9	13	4	26	2

conformation than the majority of similar structures in the CSD that possess a C_{10}–C_9–N_1–C_8-type torsion. In the conventional Mogul algorithm, the relative peak height is ignored; all Mogul peaks are treated as regions in which sampling should be equally biased. This is not the case in the MDB algorithm: here the minor peak regions are down-weighted, so MDB can occasionally actively bias a structural solution away from the correct answer. The success rate with MDB achieved with verapamil HCl can be improved to 22% if the C_{10}–C_9–N_1–C_8 torsion angle is allowed to rotate freely. Having such a torsion angle is not always problematic; indeed, in verapamil HCl, the adjacent torsion angle (C_{11}–C_{10}–C_9–N_1) also resides in a minor peak in Mogul but seems to have limited influence on the success rate achieved (see Figure 5 and Figures S1 and S2 in the supplementary information). It is worth noting that the extent of the contribution made by MDB to torsion angle sampling has not as yet been extensively trained. Further work in this area may improve the achievable success rates using this approach in such cases.

B. Crystal structure analysis and validation using Mercury

The CSDS contains the structural visualizer, Mercury. Mercury is a useful application for analyzing the results of structural analyses and is freely available for download from CCDC's website, although its more advanced features are available only to contributors. It allows for extensive analysis of molecular geometry and interactions above and beyond basic contact analysis. An additional module (the Solid Form module) allows scientists to quantify similarities and differences between polymorphs, hydrates, and solvates, identify regions of structural similarity, and understand the strengths and weaknesses of structures by searching for extended functional group interaction motifs or general packing features, quickly and easily. Particularly relevant features of Mercury are summarized in Table II.

The program provides built-in functionality for analyzing hydrogen bonds and short contacts, but also the ability to create custom definitions of contact types; within Mercury, the user can create a definition which allows the display of only short contacts between user-specified atom types. The definition of short can be either based on absolute distance or can be normalized to account for the VdW radii of participating atoms.

An example is shown in Figure 6 for the CSD entry BAGCET (Snegaroff *et al.*, 2011); the structure of 3,5-dichloro-2-iodopyridine. In Figure 6(a), all close contacts less than VdW radii are shown, whereas in Figure 6(b) multiple different contact definitions (one for halogen bonds, one for Cl—Cl contacts, and one for H—Cl contacts) were used to highlight a more interpretable network. The network is fully interactive: the user can click on contacts to expand the network, or use a contact dialog to control the contacts on view.

The investigator can easily generate molecular shells based both on interatomic distances, VdW's corrected distances or through the use of energy calculations based on the UNI force field (Gavezzotti, 1994; Gavezzotti and Filippini, 1994). Hydrogen bonds can be analyzed using graph set analyses as described by Bernstein (Bernstein *et al.*, 1995; Motherwell *et al.*, 2000; Lemmerer *et al.*, 2011). The available structures of ornidazole provide a good opportunity to highlight the application of some of these features.

Five related ornidazole (single-crystal) structures are available in the CSD, the oldest dating from 1995. In this structure (the racemate; NETRUZ), there are three molecules

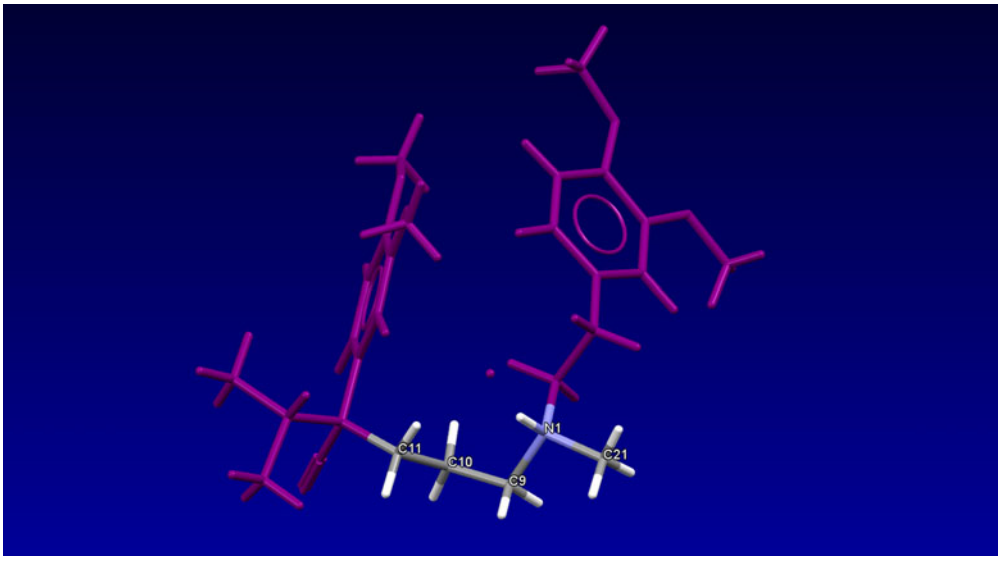

Figure 5. (Color online) The connectivity of verapamil HCl. The atoms involved in the C_{11}–C_{10}–C_9–N_1 and C_{10}–C_9–N_1–C_8 torsion angles have been labeled.

https://doi.org/10.1017/S0885715614000827 Published online by Cambridge University Press

TABLE II. Relevant features in Mercury for crystal structure analysis.

Feature Name	Description
Standard 3D visualization	Capacity for visualizing crystal structures
Visualization style management	Ability to define/save styles to facilitate creation of publication quality images
Geometric measurement	Ability to measure distances, angles torsion angles between atoms or centroids. Current measurements can also be tabulated and exported to comma separated files
Packing and slicing	Sophisticated manipulation of packing box of unit cells and the ability to show molecular slices around planes
Molecular shell generation	Ability to generate molecular shells based on interatomic distances
Symmetry visualization	Ability to visualize symmetry elements and color molecules based on symmetry elements
Powder pattern generation	Ability to simulate a powder pattern from the loaded structure
UNI force field contact analysis	Ability to generate molecular shells based on intermolecular energies from the UNI force field
Graph set analysis	Ability to generate graph set (Bernstein *et al.*, 1995; Motherwell *et al.*, 2000; Lemmerer *et al.*, 2011) representations of hydrogen bond-mediated paths within structures
Void detection	Ability to detect void and channels within a crystal structure
IsoStar contact lookup	Ability to cross-reference interactions to CSD-based distributions of non-bonded interactions
Structural similarity	Ability to compare the packing of crystal structures of identical or chemically similar molecules
Motif and Packing Searching	Ability to detect common motifs or molecular synthons in CSD structures (Desiraju, 1995)
Hydrogen bond propensity analysis	Ability to build predictive models to assess the relative likelihood of the formation of given hydrogen bonds within a structure (Allen *et al.*, 2013)

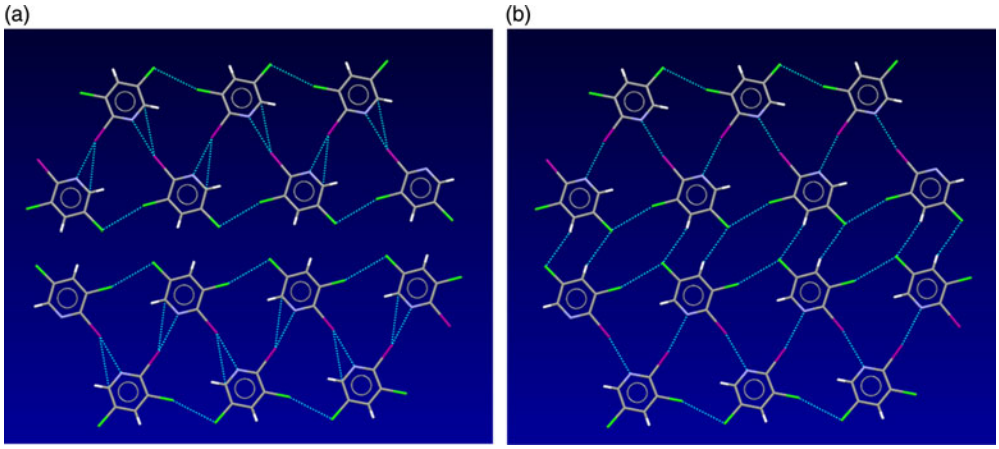

Figure 6. (Color online) An illustration of the benefit of using multiple contact definitions to aid comprehensibility of packing in CSD entry BAGCET: (a) Standard close contacts using all atom–atom contacts within VdW radii, (b) Multiple contacts showing halogen bonds (sum of VdW radii between a halogen and a nitrogen), Cl—Cl interactions (sum of VdW radii) and Cl—H contacts (VdW + 0.1).

in the asymmetric unit. A chirally pure form of ornidazole (NOBVEF; Skupin *et al.*, 1997) was determined soon afterwords and more recently a redetermination of the racemic structure at a lower temperature (120 K) and a new co-crystal with 4-nitrobenzoic acid (NETRUZ01 and CUBBEH) were published by Anderson *et al.* (2009). Deng *et al.* (2007) have also published a hemi-hydrate structure (WINKUA).

In the original racemate structure, Mercury highlights the presence of five strong interactions (-8 kcal mol^{-1} or stronger) to the lattice energy according to the UNI force field. Two of these (the strongest) interactions are stacking relationships between imidazole rings related by inversion symmetry [see Figure 7(a)]. A further two of these interactions represent a large ring mediated by hydrogen bonds. Graph set analysis clarifies the nature of the hydrogen bonding. It shows the formation of a complex R_4^4 (18) ring [see Figure 7(b)]. The fifth strong interaction is due to two symmetry independent ornidazole molecules mediated by an additional hydrogen bond.

In the chirally pure structure, a different pattern is observed; because of the lack of inversion symmetry, the packing is instead dominated by herring bone packing mediated by an OH—N hydrogen bond [see Figure 7(c)].

One can study the prevalence of particular structural arrangements using the crystal packing feature in Mercury. This feature allows the user to select one or more functional groups and search for the same spatial arrangements of the groups in the CSD. By way of example, the pair of imidazole rings forming the strongest stacking interaction in NETRUZ01 [see Figure 7(a)] were selected in Mercury and searched against the CSD (v 5.35) directly from within Mercury (the query and the result summary file are included in the supplementary material). 21% of structures that could form such a stack do, although interestingly the stack is not present in any of the other forms of ornidazole (CUBBEH, NOBVEF, WINKUA) reported in the CSD.

The hemi-hydrate structure of ornidazole (WINKUA) allows demonstration of the benefit of void detection for structural analysis. Void detection operates by analyzing the accessibility of grid points for a given probe radius, and then generating a contoured surface that envelopes the accessible points.

https://doi.org/10.1017/S0885715614000827 Published online by Cambridge University Press

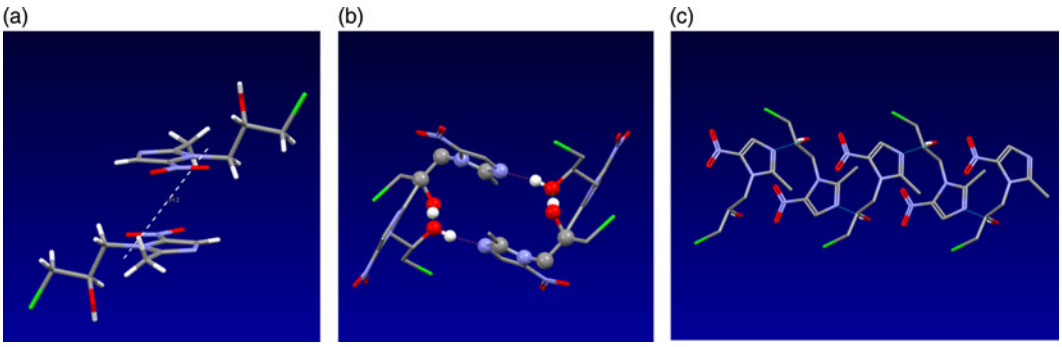

Figure 7. (Color online) Interactions in some of the ornidazole structures in the CSD: (a) Stacking in racemic ornidazole; the intermolecular energy (in kcal mol^{-1}) as calculated by the UNI force field is shown. (b) An R_4^4 (18) ring in racemic ornidazole. (c) Herring bone interactions in chirally resolved ornidazole.

The WINKUA structure was solved using single-crystal diffraction where, owing to the high data resolution, the water position would have been seen in electron density maps. In some structures, however, solvent can be disordered and so not so easily resolved. In a powder study, the lack of resolution is exacerbated by reflection overlap, making the characterization of solvates challenging. Void analysis can show where channels and holes exist within a structure, so that the investigator can identify possible locations where solvent may reside. In Figure 8, we show a void analysis (with a probe radius of 1.0 Å and a grid spacing of 0.4 Å) of the WINKUA structure with the half water molecule removed. Unsurprisingly, a hole is detected that is exactly coincident with the solvent location; one can see that performing such an analysis on a structure in its early stages of refinement might indicate when the possibility of the inclusion of a solvent molecule in the structural model may be warranted.

Void analysis can also be useful for rationalizing structural changes. A recent study (Fujii *et al.*, 2012) uses void detection to illustrate critical differences between a sequence of hydrates of Lisinopril, leading to an understanding of the dehydration and rehydration mechanism. In Figure 9, four images are shown which illustrate the formation of a large channel on solvent loss in the dihydrate to the monohydrate, and then the structural reorganization with the loss of the second solvent. It is interesting that the larger channel is retained in the anhydrous structure. Fujii *et al.* explain this by considering the differences between the hydrogen bonding patterns in the dihydrate and the monohydrate.

C. Intramolecular geometric analysis and validation using Mercury and Mogul

The advanced features in Mercury also provide tools for studying intra-molecular geometry. These features are summarized in Table III.

Mercury contains cross-links to other parts of the CSDS. For example, Mogul can be launched directly from within Mercury to generate a structure report of detailed geometric parameters lying in regions that one would deem unusual in comparison with distributions of similar features in the CSD.

Mogul reports were generated for the racemate structures deposited in the CSD (NETRUZ and NETRUZ01). In NETRUZ, some features were shown to be unusual. These are summarized in Table IV (see also Figure 10).

A noticeable feature from the Mogul report of the first structure is that it highlights a number of bond lengths and angles that are deemed "unusual". A standard Mogul report tells the user the underlying z-score for an observation. This score gives the number of standard deviations away from the mean that a given parameter lies; thus a z-score of 2.0 corresponds to an observation that lies two standard deviations from the mean and is therefore somewhat unusual. The length of the bond between C4 and C5 in NETRUZ is marked as unusual with a

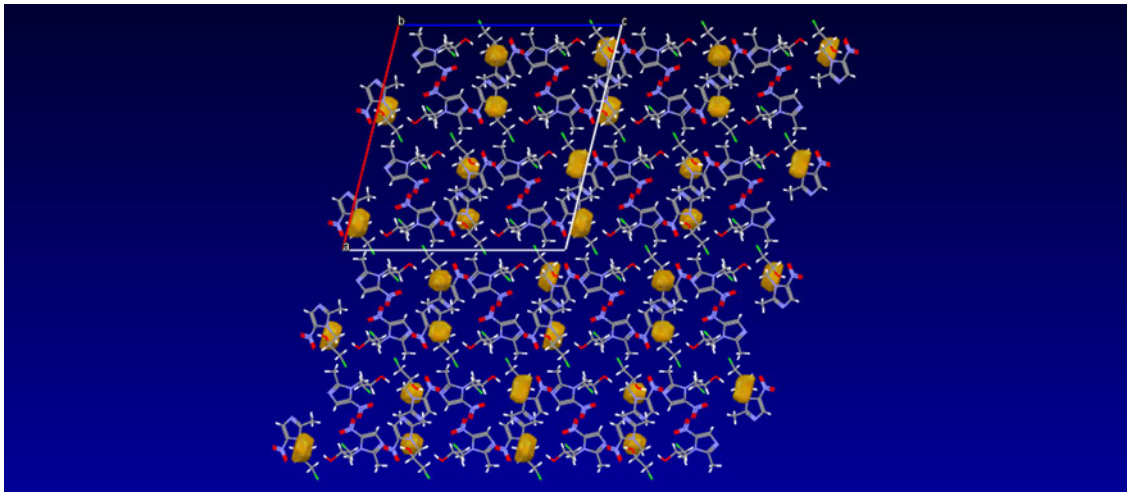

Figure 8. (Color online) Void analysis of ornidazole hemi-hydrate where the half water has been removed from the structure.

https://doi.org/10.1017/S0885715614000827 Published online by Cambridge University Press

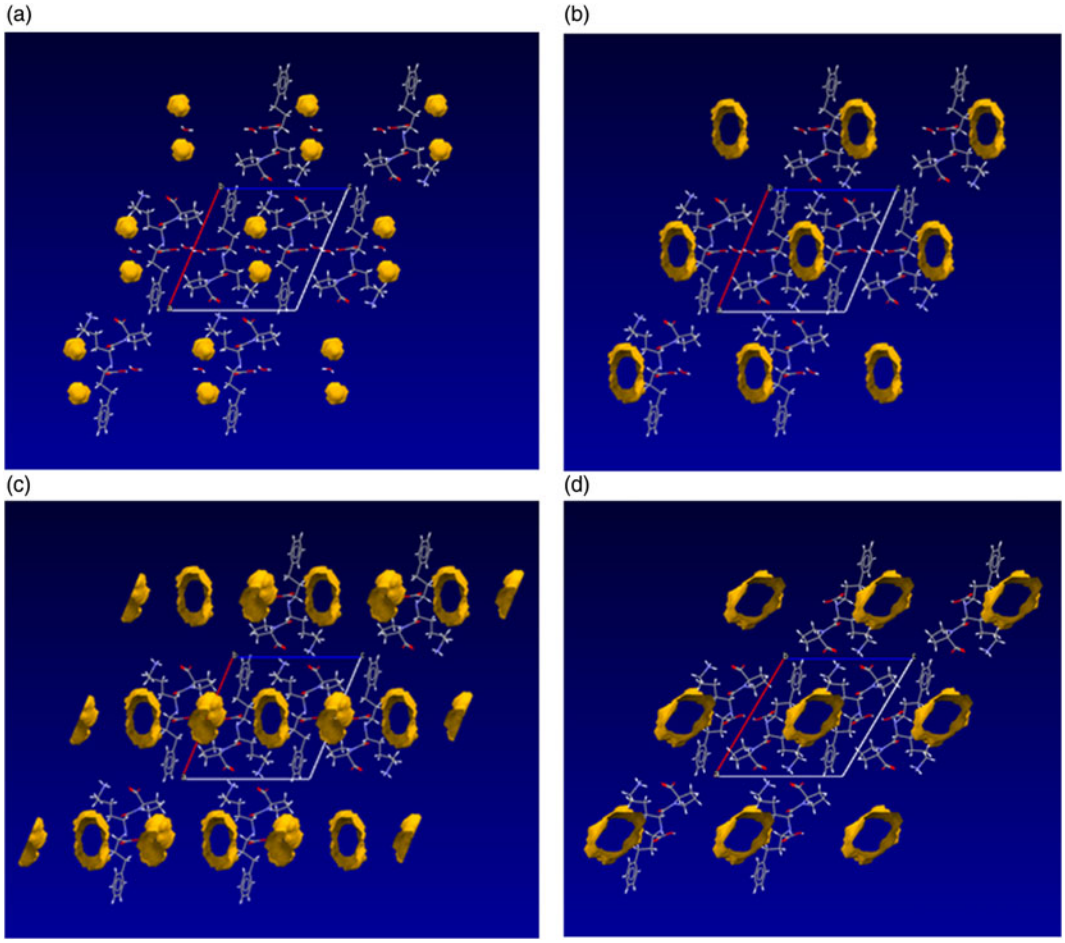

Figure 9. (Color online) A comparison of voids in Lisinopril hydrate structures, all viewed down the b axis: (a) Small voids in the dihydrate structure. (b) The larger channel in the monohydrate structure, caused by the loss of single water from the dehydrate. (c) The monohydrate structure, but with the remaining water molecule removed manually to show the additional void space created. (d) The anhydrous structure which actually results from the reorganization that occurs upon removal of the remaining water molecule.

TABLE III. Features relevant for analysis of molecular geometry.

Feature Name	Description
Molecule overlay	An automatic tool to overlay molecules for conformer comparison
Mogul geometry check	Report generation identifying intramolecular geometric features lying outside the areas that would be predicted by the CSD
Structural editing	Automated and manual methods for detecting bond types within structures and editing structures.
MOPAC interface	Interface to the semi-empirical molecular optimization software MOPAC (Stewart, 2012)

z-score of 2.5. Inspecting the underlying distribution further suggests that this bond is rather shorter than might be expected by comparison with other related structures in the CSD (see Figure 11).

The report highlights a number of other unusual features in the first structure. The unusual bonds suggest that there may be residual issues in the structure, which need further attention. NETRUZ01 shows no such problems. However, in this structure a disordered chloromethyl group was resolved. It should be noted that the bond lengths and angles that Mogul identifies as unusual in the original structure are relatively remote from this problematic group; Mogul is not pointing to the problematic region, but experienced crystallographers understand that refinement of a structure is a holistic process. Missing elements of a model can lead to other regions of a model being compromised in refinement, and thus when unusual geometry is observed, the structural model deserves some scrutiny. Additional examples of the use of the CSD in Rietveld analysis can be found in the paper of Kaduk (2007).

Use of Mogul in validating starting models for SDPD

When solving a structure with a global optimization-based program such as DASH, a 3D description of the molecule under study is a prerequisite. Such descriptions, typically in internal coordinate format, can be derived from models created using a wide range of molecular modeling computer programs, models taken directly from reported crystal structures in the CSD (e.g. in the case of polymorphic systems) or models that are based upon closely related structures. Regardless of how the initial model is constructed, it is advisable to check

https://doi.org/10.1017/S0885715614000827 Published online by Cambridge University Press

TABLE IV. Unusual features in structure NETRUZ, and their equivalent in NETRUZ01. The *z*-score, minimum and maximum values reflect the underlying distributions. In Mogul, as released, any feature with a *z*-score >2.0 is deemed unusual. The minimum and maximum are the highest and lowest values observed in the underlying distribution.

	NETRUZ		NETRUZ01		Query value	
	Query value	*z*-score	Query value	*z*-score	Min.	Max.
Bonds	(Å)		(Å)		(Å)	(Å)
C_1–N_1	1.337	2.5	1.359	0.7	1.337	1.399
C_4–C_5	1.481	2.2	1.529	0.3	1.465	1.623
C_3–N_1	1.401	2.1	1.385	0.4	1.355	1.402
Angles	(°)		(°)		(°)	(°)
Fragment 1						
N_1–C_1–N_2	113.4	2.6	111.9	0.2	110.3	113.4
C_5 –C_4 –N_1	116.1	2.4	112.3	0.1	109.9	116.1
Fragment 2						
C_{11}–N_4–C_{10}	131.0	2.9	128.4	1.3	128.4	131.0
C_{14}–C_8–N_4	126.4	2.8	123.9	0.2	121.0	127.4
C_{11}–N_4–C_8	124.2	2.2	126.5	1.2	124.2	127.0
C_9 –C_{10}–N_4	108.3	2.2	107.6	0.7	105.7	108.3
Fragment 3						
C_{16}–C_{17}–N_9	129.5	2.6	126.7	1.2	126.5	130.4
O_9–N_9–C_{17}	122.3	2.3	119.6	0.9	109.8	124.4

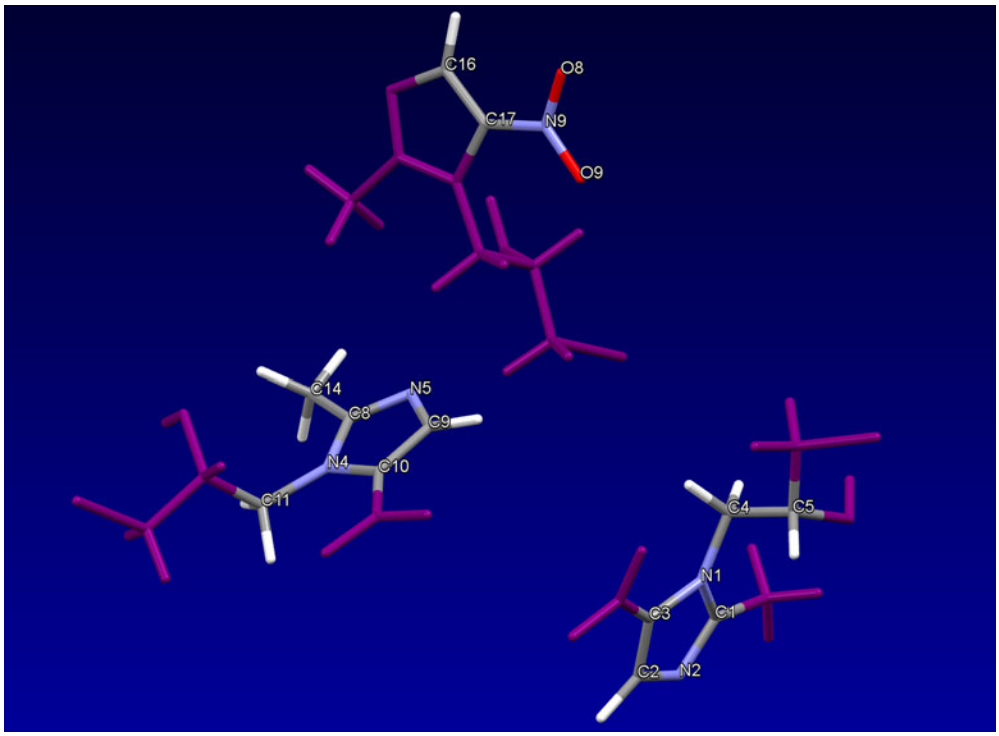

Figure 10. (Color online) The connectivity of ornidazole in NETRUZ. The labeled atoms are those involved in the bonds and angles listed in Table IV.

that its basic molecular geometry (i.e. bond lengths, bond angles and non-rotatable torsions) is chemically reasonable and Mogul provides a quick and easy way of identifying structural elements that deviate substantially from CSD-derived expectation values. All identified issues with the model flagged up by Mogul should be addressed *before* any global optimization is performed, for two reasons: (a) the more accurate the starting model, the more likely it is that one will locate the global minimum in the structure solution space; and (b) one can then refine the solved crystal structure as a series of connected rigid bodies, secure in the knowledge that bond lengths and angles in the crystal structure are already chemically reasonable and do not necessarily need to be refined. This second point greatly simplifies the refinement process and helps prevent situations where improvements in the fit to the diffraction data come at the expense of chemical sense.

D. New developments: full interaction maps (FIMS)

The CSDS includes a component, IsoStar (Bruno *et al.*, 1997), which produces interaction maps between functional groups based on the CSDS. Each interaction map reflects how pairs of functional groups tend to orient themselves with respect to each other in the CSD. In addition the

https://doi.org/10.1017/S0885715614000827 Published online by Cambridge University Press

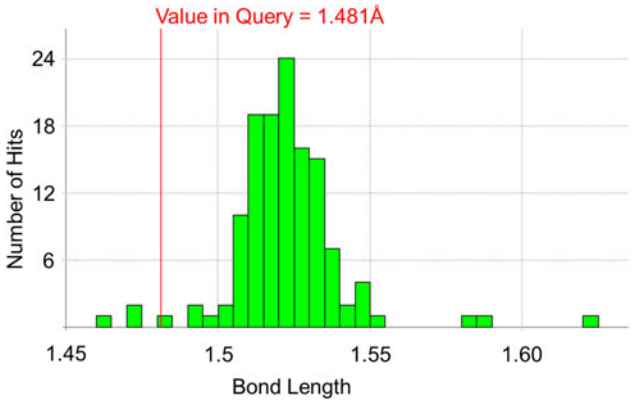

Figure 11. (Color online) Mogul distribution for the C_4–C_5 bond length in the racemic structure of ornidazole (NETRUZ).

"propensity" for a given interaction can be assessed by observing the frequency of the interaction in the CSD.

These individual interaction maps can now be combined into full interaction maps (Wood *et al.*, 2013). A full interaction map is a contoured map that shows the relative propensity for one or more probes around a molecule. This tool allows the user to assess a structure rapidly to check if the interactions formed are comparable to those that we observe in the CSDS.

In Figure 12, a FIMS plot is shown for the uncharged NH probe (a surrogate for a hydrogen bond donor) and the carbonyl oxygen (a surrogate for a hydrogen bond acceptor) around one of the three ornidazole molecules in NETRUZ01. Superimposed on the plot are the locations of hydrogen bonds and Cl—O contacts. The plot shows that the OH—N hydrogen bond lies in a region that would be deemed highly likely by the CSD; the nitrogen atom lies in the middle of a red contoured region representing the carbonyl probe. The carbonyl probe also highlights the likelihood of carbonyl-like oxygen being located relatively close to chlorine atom in the molecule, and this is indeed observed in the structure with a halogen bond between the chlorine and a nitro-oxygen. Finally, the imidazole ring nitrogen is not satisfied via a strong hydrogen bond in this molecule; the CSD would show that such an interaction is often formed (cyan region), and indeed in the other two symmetry independent molecules in the structure do form OH—N hydrogen bonds to their imidazole nitrogen. In the molecule shown a weaker CH—N hydrogen bond does exist.

Analysis of this kind can help to rationalize a structure. The FIMS evidence here is easily generated and it gives a rapid qualitative picture of the likelihood of interactions formed and not formed by a molecule based on CSD evidence. The evidence above would have provided tentative support for the hypothesis that it may be possible to form a hydrate of this compound, as we can see that certain strong acceptors are unsatisfied in the crystal structure.

III. CONCLUSIONS

Powder diffraction as a technique for crystal structure determination is becoming more prevalent in the community as method development and increasing compute power makes its application more routine. We have illustrated how a user can take advantage of the CSDS when solving, refining and critically assessing small molecule crystal structures derived from PXRD data.

The use of prior structural information in SDPD can help compensate for the low information content of the majority of PXRD patterns, by significantly reducing the size of the search space for global optimization algorithms; the CSDS is an ideal basis for biasing search space in this way.

As SDPD methods improve and user confidence in them increases, we see increasingly complex structures being solved from PXRD data alone, but users must pay particular attention to the chemical and crystallographic sense of their refined crystal structures. The CSDS is a valuable resource for the critical assessment of refined structures, both in terms of

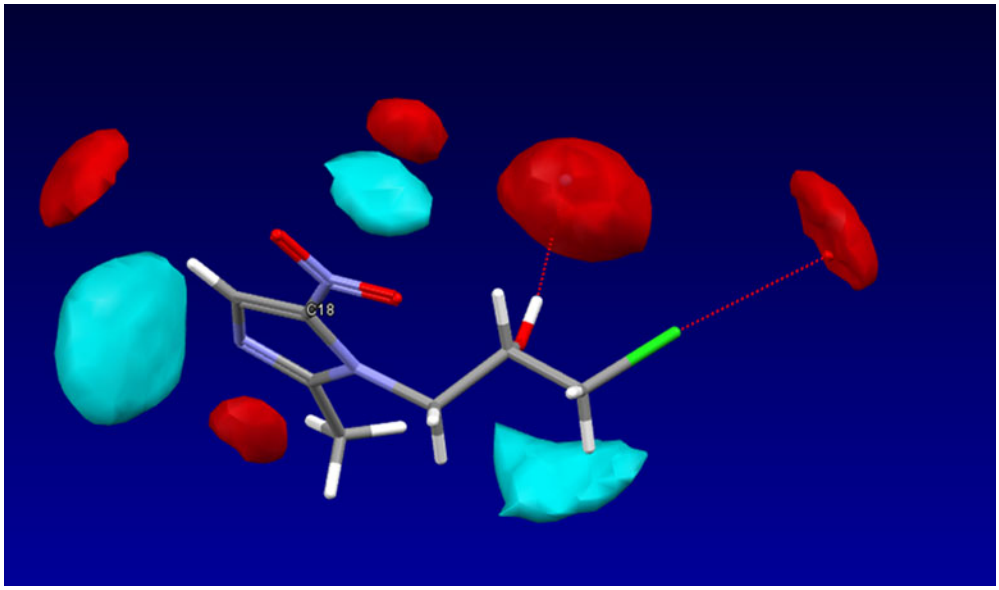

Figure 12. (Color online) A full interaction map for one of the three independent molecules present in racemic ornidazole (NETRUZ01). Favorable regions for the uncharged NH probe are shown in cyan. Favorable regions for the carbonyl oxygen probe are shown in red. A Cl—O contact in the actual structure and a hydrogen bond are shown, illustrating the predictive power of FIMS.

https://doi.org/10.1017/S0885715614000827 Published online by Cambridge University Press

molecular geometry and intermolecular interactions, especially hydrogen bonds. Alongside other resources, such as EnCifer (Allen *et al.*, 2004), PLATON (Spek, 2009), and checkCIF (IUCr, 2014), crystallographers are well placed to solve, refine, and critically assess structures derived from powder diffraction.

The incorporation of computational chemistry tools (e.g. MOPAC, UNI force field) into the CSDS mirrors the trend toward further verification of crystal structures using periodic, dispersion-corrected, density functional theory calculations (Bruening *et al.*, 2011), in that often the "fine detail" of the structure cannot be determined from the PXRD data alone.

ACKNOWLEDGEMENT

The ornidazole data were collected on the high-resolution powder diffractometer (BM16) at the ESRF in 1997 by Prof. Bill David (STFC ISIS Facility), Dr. Kenneth Shankland and Dr. Norman Shankland (CrystallografX Ltd.). We gratefully acknowledge support from the beamline staff during data collection. E.K. gratefully acknowledges financial support from both the CCDC and the University of Reading.

Supplementary Materials and Methods

The supplementary material for this article can be found at http://www.journals.cambridge.org/PDJ

Allen, F. H., Johnson, O., Shields, G. P., Smith, B. R., and Towler, M. (**2004**). "CIF applications. XV. enCIFer: a program for viewing, editing and visualizing CIFs," J. Appl. Crystallogr. **37**, 335–338.

Allen, F. H., Wood, P. A., and Galek, P. T. A. (**2013**). "Role of chloroform and dichloromethane solvent molecules in crystal packing: an interaction propensity study," Acta Crystallogr. B-Struct. Sci. **69**, 379–388.

Anderson, K. M., Probert, M. R., Whiteley, C. N., Rowland, A. M., Goeta, A. E., and Steed, J. W. (**2009**). "Designing co-crystals of pharmaceutically relevant compounds that crystallize with Z' > 1," Crystal Growth Des. **9**, 1082–1087.

Bekoe, S. L., Urmann, D., Lakatos, A., Glaubitz, C., and Schmidt, M. U. (**2012**). "Nimustine hydrochloride: the first crystal structure determination of a 2-chloroethyl-N-nitrosourea hydrochloride derivative by X-ray powder diffraction and solid-state NMR," Acta Crystallogr. C – Crystal Struct. Commun. **68**, O144–O148.

Bernstein, J., Davis, R. E., Shimoni, L., and Chang, N. L. (**1995**). "Patterns in hydrogen bonding – functionality and graph set analysis in crystals," Angew. Chem. Int. Ed. **34**, 1555–1573.

Betteridge, P. W., Carruthers, J. R., Cooper, R. I., Prout, K., and Watkin, D. J. (**2003**). "CRYSTALS version 12: software for guided crystal structure analysis," J. Appl. Crystallogr. **36**, 1487–1487.

Bruening, J., Alig, E., van de Streek, J., and Schmidt, M. U. (**2011**). "The use of dispersion-corrected DFT calculations to prevent an incorrect structure determination from powder data: the case of acetolone, C11H11N3O3," Z. Kristallogr. **226**, 476–482.

Bruno, I. J., Cole, J. C., Lommerse, J. P. M., Rowland, R. S., Taylor, R., and Verdonk, M. L. (**1997**). "IsoStar: A library of information about nonbonded interactions," J. Comput. Aided Mol. Des. **11**, 525–537.

Bruno, I. J., Cole, J. C., Edgington, P. R., Kessler, M., Macrae, C. F., McCabe, P., Pearson, J., and Taylor, R. (**2002**). "New software for searching the Cambridge Structural Database and visualizing crystal structures," Acta Crystallogr. B-Struct. Sci. **58**, 389–397.

Bruno, I. J., Cole, J. C., Kessler, M., Luo, J., Motherwell, W. D. S., Purkis, L. H., Smith, B. R., Taylor, R., Cooper, R. I., Harris, S. E., and Orpen, A. G. (**2004**). "Retrieval of crystallographically-derived molecular geometry information," J. Chem. Inf. Comput. Sci. **44**, 2133–2144.

Coelho, A. (**2003**). Topas user manual. Version v3.1. Bruker AXS GmbH, Karlsruhe, Germany.

David, W. I. F., Shankland, K., and Shankland, N. (**1998**). "Routine determination of molecular crystal structures from powder diffraction data," *Chem. Commun.*, 931–932.

David, W. I. F., Shankland, K., Cole, J. C., Maginn, S., Motherwell, W. D. S., and Taylor, R. (**2006a**). *DASH User Manual Guide* (Cambridge Crystallographic Data Centre, Cambridge, UK).

David, W. I. F., Shankland, K., van de Streek, J., Pidcock, E., Motherwell, W. D. S., and Cole, J. C. (**2006b**). "DASH: a program for crystal structure determination from powder diffraction data," J. Appl. Crystallogr. **39**, 910–915.

Deng, L., Wang, W., and Lv, J. (**2007**). "Ornidazole hemihydrate," Acta Crystallogr. E-Struct. Rep. Online **63**, O4204–U2347.

Desiraju, G. R. (**1995**). "Supramolecular synthons in crystal engineering – a new organic-synthesis," Angew. Chemie, Int. Ed. Engl. **34**, 2311–2327.

Fernandes, P., Florence, A. J., Shankland, K., Shankland, N., and Johnston, A. (**2006**). "Powder study of chlorothiazide N,N-dimethyl-formamide solvate," Acta Crystallogr. E, Struct. Rep. Online **62**, O2216–O2218.

Fernandes, P., Shankland, K., Florence, A. J., Shankland, N., and Johnston, A. (**2007**). "Solving molecular crystal structures from X-ray powder diffraction data: the challenges posed by gamma-carbamazepine and chlorothiazide N, N,-dimethylformamide (1/2) solvate," J. Pharm. Sci. **96**, 1192–1202.

Ferreira, F. F., Antoni, S. G., Pires Rosa, P. C., and Paiva-Santos, C. D. O. (**2010**). "Crystal structure determination of mebendazole form a using high-resolution synchrotron X-Ray powder diffraction data," J. Pharm. Sci. **99**, 1734–1744.

Florence, A. J., Shankland, N., Shankland, K., David, W. I. F., Pidcock, E., Xu, X. L., Johnston, A., Kennedy, A. R., Cox, P. J., Evans, J. S. O., Steele, G., Cosgrove, S. D., and Frampton, C. S. (**2005**). "Solving molecular crystal structures from laboratory X-ray powder diffraction data with DASH: the state of the art and challenges," J. Appl. Crystallogr. **38**, 249–259.

Fujii, K., Uekusa, H., Itoda, N., Yonemochi, E., and Terada, K. (**2012**). "Mechanism of dehydration-hydration processes of lisinopril dihydrate investigated by ab Initio powder X-ray diffraction analysis," Crystal Growth Des. **12**, 6165–6172.

Gavezzotti, A. (**1994**). "Are crystal structures predictable?," Acc. Chem. Res. **27**, 309–314.

Gavezzotti, A. and Filippini, G. (**1994**). "Geometry of the intermolecular X-H. . .Y (X, Y = N, O) hydrogen-bond and the calibration of empirical hydrogen-bond potentials," J. Phys. Chem. **98**, 4831–4837.

Guguta, C., van Eck, E. R. H., and de Gelder, R. (**2009**). "Structural insight into the dehydration and hydration behavior of naltrexone and naloxone hydrochloride. Dehydration-induced expansion versus contraction," Crystal Growth Des. **9**, 3384–3395.

IUCr (**2014**). checkCIF – a service of the International Union of Crystallography.

Johnston, A., Florence, A. J., Shankland, K., Markvardsen, A., Shankland, N., Steele, G., and Cosgrove, S. D. (**2004**). "Powder study of N- 2-(4-hydroxy-2-oxo-2,3-dihydro-1,3-benzothiazol-7-yl)ethyl-3- 2-(2-n aphthalen-1-ylethoxy)ethylsulfonyl propylaminium benzoate," Acta Crystallogr. E, Struct. Rep. Online **60**, O1751–O1753.

Kaduk, J. A. (**2007**). "Chemical reasonableness in Rietveld analysis; organics," Powder Diffr. **22**, 74–82.

Karki, S., Fabian, L., Friscic, T., and Jones, W. (**2007**). "Powder x-ray diffraction as an emerging method to structurally characterize organic solids," Org. Lett. **9**, 3133–3136.

Lapidus, S. H., Stephens, P. W., Arora, K. K., Shattock, T. R., and Zaworotko, M. J. (**2010**). "A comparison of cocrystal structure solutions from powder and single crystal techniques," Crystal Growth Des. **10**, 4630–4637.

Lemmerer, A., Adsmond, D. A. and Bernstein, J. (**2011**). "An investigation of the hydrogen-bond preferences and co-crystallization behavior of three didonor compounds," Crystal Growth Des. **11**, 2011–2019.

Majumder, M., Buckton, G., Rawlinson-Malone, C., Williams, A. C., Spillman, M. J., Shankland, N., and Shankland, K. (**2011**). "A carbamazepine-indomethacin (1:1) cocrystal produced by milling," Crystengcomm **13**, 6327–6328.

Motherwell, W. D. S., Shields, G. P., and Allen, F. H. (**2000**). "Automated assignment of graph-set descriptors for crystallographically symmetric molecules," Acta Crystallogr. B, Struct. Sci. **56**, 466–473.

Rigaku (**2013**). Integrated X-ray Powder Diffraction Software PDXL 2.2: Structure Analysis User Manual, Rigaku, Distributed with the Rigaku PDXL software (v2.2).

https://doi.org/10.1017/S0885715614000827 Published online by Cambridge University Press

Shankland, K., David, W. I. F., McCusker, L. B., and Baerlocher, C. (eds) (**2002a**). *Structure Determination from Powder Diffraction Data* (Oxford University Press, USA).

Shankland, K., McBride, L., David, W. I. F., Shankland, N., and Steele, G. (**2002b**). "Molecular, crystallographic and algorithmic factors in structure determination from powder diffraction data by simulated annealing," J. Appl. Crystallogr. **35**, 443–454.

Shankland, K., Spillman, M. J., Kabova, E. A., Edgeley, D. S., and Shankland, N. (**2013**). "The principles underlying the use of powder diffraction data in solving pharmaceutical crystal structures," Acta Crystallogr. C, Crystal Struct. Commun. **69**, 1251–1259.

Shin, H. S., Song, H., Kim, E., and Chung, K. B. (**1995**). "The crystal and molecular-structure of 1-(3-chloro-2-hydroxypropyl)-2-methyl-5-nitroimidazole (ornidazole)," Bull. Korean Chem. Soc. **16**, 912–915.

Skorepova, E., Cejka, J., Husak, M., Eigner, V., Rohlicek, J., Sturc, A., and Kratochvil, B. (**2013**). "Trospium chloride: unusual example of polymorphism based on structure disorder," Crystal Growth Des. **13**, 5193–5203.

Skupin, R., Cooper, T. G., Frohlich, R., Prigge, J., and Haufe, G. (**1997**). "Lipase-catalyzed resolution of both enantiomers of omidazole and some analogues," Tetrahedron-Asymmetry **8**, 2453–2464.

Smart, O. S., Womack, T. S., Sharff, A., Flensburg, C., Keller, P., Paciorek, W., Vonrhein, C., and Bricogne, G. (**2001**). *Grade* (Global Phasing Ltd., Cambridge, UK). http://www.globalphasing.com

Snegaroff, K., Tan Tai, N., Marquise, N., Halauko, Y. S., Harford, P. J., Roisnel, T., Matulis, V. E., Ivashkevich, O. A., Chevallier, F., Wheatley, A. E. H., Gros, P. C., and Mongin, F. (**2011**). "Deprotonative metalation of chloro- and bromopyridines using amido-based bimetallic species and regioselectivity-computed CH acidity relationships," Chem. Eur. J. **17**, 13284–13297.

Spek, A. L. (**2009**). "Structure validation in chemical crystallography," Acta Crystallogr. D, Biol. Crystallogr. **65**, 148–155.

Stewart, J. J. P. (**2012**). *MOPAC2012, Stewart Computational Chemistry* (Colorado Springs, CO, USA).

Vella-Zarb, L., Dinnebier, R. E., and Baisch, U. (**2013**). "The devil is in the detail: a rare H-Bonding Motif in new forms of Docetaxel," Crystal Growth and Des. **13**, 4402–4410.

Wood, P. A., Olsson, T. S. G., Cole, J. C., Cottrell, S. J., Feeder, N., Galek, P. T. A., Groom, C. R., and Pidcock, E. (**2013**). "Evaluation of molecular crystal structures using Full Interaction Maps," Cryst. Eng. Comm. **15**, 65–72.

https://doi.org/10.1017/S0885715614000827 Published online by Cambridge University Press