Structure determination from powder data: Mogul and CASTEP

A. J. Florence[1,*], J. Bardin[1], B. Johnston[1], N. Shankland[1], T. A. N. Griffin[2], K. Shankland[2]

[1] Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, Glasgow G4 0NR, UK
[2] School of Pharmacy, Whiteknights, PO Box 224, Reading RG6 6AD, UK
[*] alastair.florence@strath.ac.uk

Keywords: structure determination from powder data, simulated annealing, torsion angle constraints, Mogul, DFT, geometry optimisation

Abstract. When solving the crystal structure of complex molecules from powder data, accurately locating the global minimum can be challenging, particularly where the number of internal degrees of freedom is large. The program Mogul provides a convenient means to access typical torsion angle ranges for fragments related to the molecule of interest. The impact that the application of modal torsion angle constraints has on the structure determination process of two structure solution attempts using DASH is presented. Once solved, accurate refinement of a molecular structure against powder data can also present challenges. Geometry optimisation using density functional theory in CASTEP is shown to be an effective means to locate hydrogen atom positions reliably and return a more accurate description of molecular conformation and intermolecular interactions than global optimisation and Rietveld refinement alone.

Introduction

Structure determination from powder diffraction data (SDPD) using simulated annealing (SA) has become a widely used tool for structural analysis in the absence of single crystals. It remains to be established precisely where the limit of effectiveness of SA actually lies for complex structures. In this context, the term 'complexity' has two connotations, molecular (internal Degrees of Freedom, DoF; torsion angles) and crystallographic (external DoF; position and orientation). Whilst SA successes in dealing with relatively large external DoF have been reported (36 external DoF; [1]) it has also been shown that structures containing 13 internal DoF can pose a significant challenge to SA [2]. It is therefore of interest to explore strategies that maximise the chances of success with structures containing large numbers of internal DoF. Here we demonstrate the application of prior chemical knowledge, in the form of modal torsion angle constraints extracted from the CSD [3] using Mogul [4], to increase the frequency with which the global minimum is located in SDPD using DASH [5].

Once the global minimum has been located and the structure solved, Rietveld refinement [6] is typically carried out to obtain a final structure that is both chemically reasonable and gives the best possible fit to the data. When dealing with structures of flexible molecules, it is often necessary to confirm that weakly-scattering hydrogen atoms are correctly located in the SA structure prior to refinement. Although manual location of hydrogen atom positions between nearby donor and acceptor atoms is often facile for rigid groups, this process can be non-trivial for flexible hydrogen-bond donor or acceptor groups. Density functional theory (DFT) can be applied as a complementary tool to SDPD for structure verification [7] and refinement [8] and here we utilise geometry optimisation of the SDPD structure with CASTEP [9] to accurately locate hydrogen atom positions in a molecular crystal structure solved by SDPD.

Mogul-assisted structure determination

The impact of Mogul-derived torsion angle preferences on the SA approach to structure solution is considered here, noting that the same principles may be applicable to other global optimisation algorithms. The Mogul program has been developed for ease of use in retrieving molecular geometry data from the CSD and is therefore easily implemented within the SDPD process.

Application to verapamil hydrochloride

Data (2-70° 2θ; step size 0.017°; variable count time scheme; 298 K) were collected from a capillary-mounted sample using a Bruker-AXS D8 diffractometer (Cu $K\alpha_1$; Bruker Lynxeye), yielding χ^2_{Pawley} =1.56 for a DASH fit to the data range 3-42° 2θ. In a previously reported systematic study of the relationship between search space complexity and frequency of success in SDPD [2], verapamil hydrochloride (VHCl; triclinic, $P\bar{1}$, CSD refcode CURHOM, Z' = 1) had the highest total number of DoF, with 13 internal (figure 1) plus 9 external DoF. As such, it constitutes a good starting point to assess the impact of Mogul-derived search space restrictions (table 1).

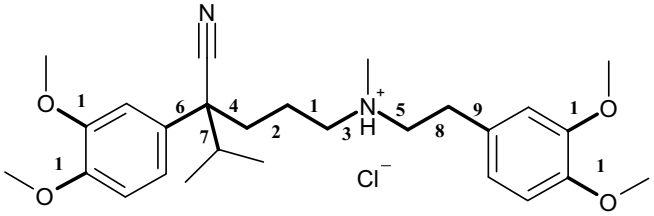


Figure 1. Molecular structure of VHCl, highlighting the 13 internal DoF.

To assess the impact of the restrictions on the stochastic search, a large number of runs are required 'with' and 'without' the restrictions applied. Executions of DASH runs on the Grid MP system at the STFC Rutherford Appleton Laboratory allowed these large numbers of runs to be performed overnight (SA control parameters: 2×10^7 moves, 0.01 cooling rate and

final simplex refinement). The frequency of success was taken to be the number of SA runs in each batch of 800 reaching the global minimum, defined as $\chi^2_{Profile} / \chi^2_{Pawley} \leq 10$ (table 2).

Table 1. Allowed ranges for the 13 optimisable torsion angles (τ) identified in figure 1.

τ	Allowed torsion angle search space			
1, 2, 3, 4, 5, 7, 8	Trimodal	30° → 90°	150° → 210°	270° → 330°
9	Bimodal	45° → 135°	225° → 315°	
6	Unimodal	0° → 360°		
10, 11, 12, 13	Unimodal	-45° → 45°		

Table 2. Frequency of success in solving the crystal structure of VHCl.

Internal DoF	Allowed torsion angle search space	Frequency of success
0	All torsions fixed at values in CUR-HOM	800 / 800
13	0° → 360° (all angles)	27 / 800
13	As table 1	29 / 800

The complexity of the overall search space is reflected in the greatly reduced frequency of success accompanying the introduction of internal DoF. Interestingly, the torsion angle search space restrictions brought no significant benefit in this particular case. Thus, with VHCl, the implementation of the SA algorithm in DASH is equally effective in full (0° → 360°) or reduced (table 1) torsion angle search space.

Application to CSD refcode XELBEV

When looking for more challenging examples than VHCl, it is logical to look towards structures with $Z' > 1$. XELBEV (monoclinic, $P2_1$), for example, poses a significant increase in complexity, with a total of 18 internal plus 12 external DoF (figure 2).

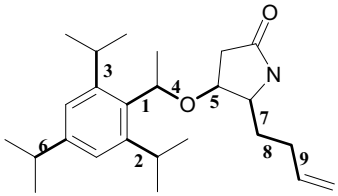


Figure 2. The molecular structure of CSD refcode XELBEV, Z' = 2, highlighting the 9 internal DoF.

A DASH Pawley fit to a XELBEV pattern simulated using Mercury CSD 2.0 (λ = 1.54056 Å; range 5-30° 2θ; step size 0.017°) yielded χ^2_{Pawley} =1.35. Torsion angle search space restrictions were derived using Mogul (table 3) and frequency of success (table 4) determined as described for VHCl. The frequency of success with fixed conformations is lower than for VHCl, reflecting the increased complexity due to the additional external DoF. In contrast to

the unrestricted batch (0 successes, lowest $\chi^2_{Profile}$ = 27), the batch that operated within the constrained torsion angle space returned 3 successes. Thus, Mogul plus DASH combined effectively to solve the XELBEV crystal structure. It is worth noting that Mogul-derived search space restrictions only apply during the SA part of the search - they do not apply to the post-SA simplex refinement executed upon reaching (in this case) 2×10^7 moves. Thus, torsion angles that lie beyond the imposed search space can still be accessed at the simplex stage.

Table 3. Allowed ranges for the 9 optimisable torsion angles (τ) identified in figure 2 (identical ranges were imposed on each of the 2 molecules in the asymmetric unit).

τ	Allowed torsion angle search space			
4, 7, 8	Trimodal	30° → 90°	150° → 210°	270° → 330°
1	Bimodal	45° → 135°	225° → 315°	
9	Bimodal	90° → 135°	225° → 270°	
6	Unimodal	0° → 360°		
5	Unimodal	45° → 315°		
2,3	Unimodal	190° → 290°		

Table 4. Frequency of success in solving the crystal structure of XELBEV.

Internal DoF	Allowed torsion angle search space	Frequency of success
0	Torsions fixed at values in XELBEV	579 / 800
18	0° → 360° (all angles)	0 / 800
18	As table 3	3 / 800

Geometry optimisation of HCT form II using CASTEP

In this section, DFT geometry optimisation is applied to determine accurate hydrogen atom positions in the crystal structure of a polymorph (form II) of the thiazide diuretic, hydrochlorothiazide (HCT; figure 3). The crystal structure of form II [10] was solved by SA using DASH from lab powder data collected at 298 K to 1.76 Å resolution. The positions of hydrogen atoms H1 and H2 (figure 3) were manually located in the global minimum structure by rotating the NH_2 group around the N-S bond to obtain chemically sensible intermolecular contacts. Subsequent Rietveld refinement of this structure yielded an R_{wp} of 3.75%.

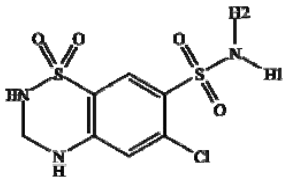
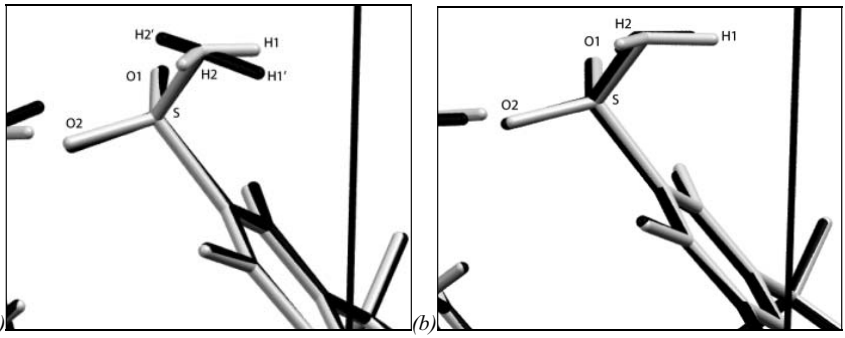


Figure 3. HCT molecule with labelling of H1 and H2 shown.

A recent single-crystal determination of form II HCT [11] confirms the accuracy of the SDPD structure with respect to the position of non-hydrogen atom positions. However, the positions of H1 and H2 differ significantly in the two structures, as demonstrated by the torsion angle τ_1 (H2-N-S-O2) describing the NH_2 orientation. The manually located NH_2 group is rotated by 40.8° from the position determined in the single-crystal structure [τ_1 = -8.0° (*SDPD*) and 32.7° (*single-crystal*); figure 4(a)].



(a) *(b)*

Figure 4. Overlays of (a) HCT form II SDPD (black) and single-crystal (grey) structures with H1 and H2 labelled and (b) CASTEP optimised SDPD (grey) and single-crystal (black) structures.

Geometry optimization of crystal structures using CASTEP

First-principles DFT calculations were performed on the experimental SDPD and single-crystal structures using CASTEP v.3.2. The unit cell parameters were fixed at their experimental values for all calculations. Preliminary optimizations in space group *P*1 confirmed the structures were close to minima and subsequent optimizations were carried out in $P2_1/c$ to facilitate comparison of the optimized structures with the experimental data. A generalised gradient approximation (GGA-PBE) was used to describe the exchange-correlation potential [12] with a plane-wave basis set cut-off of 435eV and a single k-point for the BZ sampling. MEDIUM convergence criteria were used (energy per atom $2x10^{-5}$ eV, forces 0.05 eV.Å^{-1}, displacement $2x10^{-3}$Å). Calculations were performed on the Scientific Computing Application Resource for Facilities (SCARF) cluster at the SFTC Rutherford Appleton Laboratory and both optimizations converged satisfactorily within the default number of iterations and utilised 2.7 h of total computing time per structure.

The positions of H1 and H2 in the optimized SDPD structure show a much improved correspondence to the experimental single-crystal structure, with $\Delta\tau_1$ falling from 40.8° (*SDPD* vs. *single-crystal* structure) to 8.8° (*CASTEP SDPD* vs. *single-crystal;* figure 4(b)). A Rietveld fit of the optimised SDPD structure returned a final R_{wp} of 3.72 %, unchanged from the original published structure. Thus, the combination of SA, quantum chemical geometry optimization and Rietveld refinement has yielded a structure for form II HCT that shows both excellent fit to the experimental data and more accurate hydrogen atom positions than from SA and Rietveld refinement alone.

Concluding remarks

Accurate crystal structures are of vital importance to those concerned with studying the basic science underpinning physical form diversity in the organic solid-state. In the absence of suitable single-crystal samples, methods for SDPD are often invaluable. Where there is a requirement to solve crystal structures with large numbers of internal DoF, limiting the search space using experimentally derived torsion angle constraints can be an effective tool for increasing the probability of successful structure solution. These methods add no computational overhead to the global optimization method and, particularly when used in combination with a distributed computing global optimization implementation, can enable structures with significant complexity to be tackled efficiently and with confidence. Once a global minimum structure has been obtained, DFT calculations can be recommended where there is a particular interest in the determination of accurate hydrogen atom positions.

References

1. Fernandes, P., Shankland, K., Florence, A.J., Shankland, N. & Johnston, A., 2007, *J. Pharm. Sci.*, **96**, 1192.
2. Florence, A.J., Shankland, N., Shankland, K., David, W.I.F., Pidcock, E., Xu, X., Johnston, A., Kennedy, A.R., Cox, P.J., Evans, J.S.O., Steele, G., Cosgrove, S.D. & Frampton, C.S., 2005, *J. Appl. Crystallogr.*, **38**, 249.
3. Allen, F.H., 2002, *Acta Crystallogr. Sect. B-Struct. Sci.*, **58**, 380.
4. 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. & Orpen, A.G., 2004, *J. Chem. Inf. Comp. Sci.*, **44**, 2133.
5. David, W.I.F., Shankland, K., van de Streek, J., Pidcock, E., Motherwell, W.D.S. & Cole, J.C., 2006, *J. Appl. Crystallogr.*, **39**, 910.
6. Rietveld, H.M., 1969, *J. Appl. Crystallogr.*, **2**, 65.
7. Neumann, M.A., Tedesco, C., Destri, S., Ferro, D.R. & Porzio, W., 2002, *J. Appl. Crystallogr.*, **35**, 296.
8. Smrcok, L., Jorik, V., Scholtzova, E. & Milata, V., 2007, *Acta Crystallogr. Sect. B-Struct. Sci.*, **63**, 477.
9. Clark, S., Segall, M.D., Pickard, C.J., Hasnip, P.J., Probert, M.I.J., Refson, K. & Payne, M.C., 2005, *Z. Kristallogr.*, **220**, 567.
10. Florence, A., Johnston, A., Fernandes, P., Shankland, K., Stevens, H.N.E., Osmundsen, S. & Mullen, A.B., 2005, *Acta Crystallogr. Sect. E-Struct. Rep. Onl.*, **61**, o2798.
11. Kennedy, A.R., Johnston, A. & Florence, A.J., 2007, *unpublished result*.
12. Perdew, J.P., Burke, K. & Ernzerhof, M., 1996, *Phys. Rev. Lett.*, **77**, 3865.

Acknowledgements. The authors thank STFC's e-Science facility for access to computing resources and the University of Strathclyde for funding JB.