Routine determination of molecular crystal structures from powder diffraction data

William I. F. David,*[a]† Kenneth Shankland[a] and Norman Shankland[b]

[a] *ISIS Facility, Rutherford Appleton Laboratory, Chilton, Didcot, Oxon, UK OX11 0QX*
[b] *Department of Pharmaceutical Sciences, University of Strathclyde, Glasgow, UK G1 1XW*

The state of the art in determining molecular crystal structures from powder diffraction data using a global optimisation method is illustrated with a fast, automated simulated annealing approach to solving the previously unknown crystal structures of capsaicin, thiothixene and promazine hydrochloride.

The routine structure determination of molecular materials from single-crystal diffraction data is one of the principal triumphs of crystallography. In contrast, structure determination from powder diffraction data is problematic, hampered as it is by the severe loss of information arising from Bragg peak overlap. Here, we present a fast, automated, simulated annealing approach to the problem that provides a significant alternative to direct methods usually employed in attempts to solve structures from powder diffraction data.[1]

X-Ray diffraction data were collected for the compounds illustrated in Fig. 1 at 100 K on BM16 of the European Synchrotron Radiation Facility[2] (λ = 0.6528 Å), and the structures solved as follows. The position, orientation and conformation of an entire molecule in the refined unit cell were postulated and the level of agreement between the trial structure and the experimental diffraction data quantified by:[3]

$$\chi^2 = \sum_h \sum_k [(I_h - c\,|F_h|^2)(V^{-1})_{hk}(I_k - c\,|F_k|^2)]$$

where I_h and I_k are Lorentz-polarisation corrected, extracted integrated intensities from a Pawley refinement[4] of the diffraction pattern, V_{hk} is the covariance matrix from the Pawley refinement, c is a scale factor, and $|F_h|$ and $|F_k|$ are the structure factor magnitudes calculated from the trial structure.

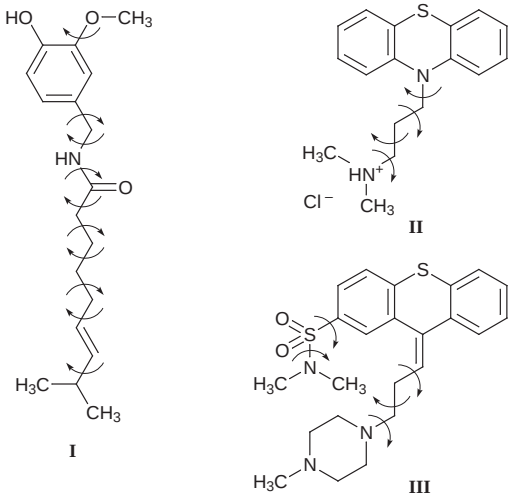

Fig. 1 Capsaicin **I** is the 'hot' component of chilli peppers. Promazine hydrochloride **II** and thiothixene **III** are tranquillisers. Internal degrees of freedom (indicated by the arrows) are 0–360° in every case, except for H–N–C=O in capsaicin, where the range 160–200° was sampled. The configurations around the C=C bonds in capsaicin and thiothixene were fixed as shown, in accordance with prior chemical knowledge, and the piperazine ring fixed in a chair conformation. The molecules have six, nine and five external degrees of freedom respectively.

The trial structure was subjected to a global optimisation in which torsion angles were the only internal degrees of freedom (Fig. 1) and bounds on the external degrees of freedom (three fractional coordinates for position and four quaternions for orientation[5]) were derived from the Euclidean normalisers of the relevant space groups.[6] Finally, the structure solutions obtained at the end of the simulated annealing runs were verified by Rietveld refinements[7] in which only scale and overall temperature factors were refined.

The method presented here has two distinctive elements. Firstly, the correlated integrated intensities figure-of-merit is formally equivalent to, but calculated typically two orders of magnitude faster than, the agreement factors used by others[8–10] that operate *via* a point-by-point comparison with the measured diffraction pattern. Secondly, the simulated annealing protocol contains a novel combination of elements that facilitate global searching of the complex multi-dimensional parameter space. These include a temperature reduction that cools more slowly if the χ^2 fluctuations (*cf.* heat capacity) are large, the capacity to generate random mutations (*cf.* a genetic algorithm approach[3]) and an algorithm to generate new parameter sets that samples the neighbouring parameter space efficiently whilst allowing large perturbations with an exponentially decreasing probability.

High quality solutions were obtained with ease for all three structures (Table 1), with remarkably good agreement between the observed and calculated diffraction patterns (Fig. 2). This close agreement arises from the ability of the simulated annealing algorithm to 'fine tune' both the internal and external

Table 1 Refined unit cells, extracted intensity information and final χ^2 for the compounds shown in Fig. 1[a]

	Compound		
	Capsaicin **I**	Promazine HCl **II**	Thiothixene **III**
Space group	$P2_1/c$	$P2_1/c$	$P2_1$
a/Å	12.2234(1)	11.8081(2)	10.1471(4)
b/Å	14.7900(1)	11.4895(3)	8.6883(1)
c/Å	9.4691(1)	13.4270(2)	13.6806(3)
β/°	93.9754(3)	111.722(1)	110.650(1)
2θ Data/°	2.7–22.5	3.0–20.0	2.5–20.4
Res./Å	1.67	1.88	1.84
N_{ref}	379	271	216
t_d/min	40	8	9
χ^2_{Paw}	13.05	3.70	9.43
χ^2_{Riet}	51.19	5.03	63.74
t_s/min	1860	112	205

[a] N_{ref} = Number of reflections in the data range shown. χ^2_{Paw} = Chi-squared for the Pawley type fit to the data range shown. χ^2_{Riet} = Chi-squared for the Rietveld fit of the output simulated annealing model to the data range shown, with only scale and overall temperature factor refined. Res = Maximum spatial resolution of data. t_d = Data collection time over the specified range. t_s = Time to simulated annealing solution (DEC Alphastation 500/500).

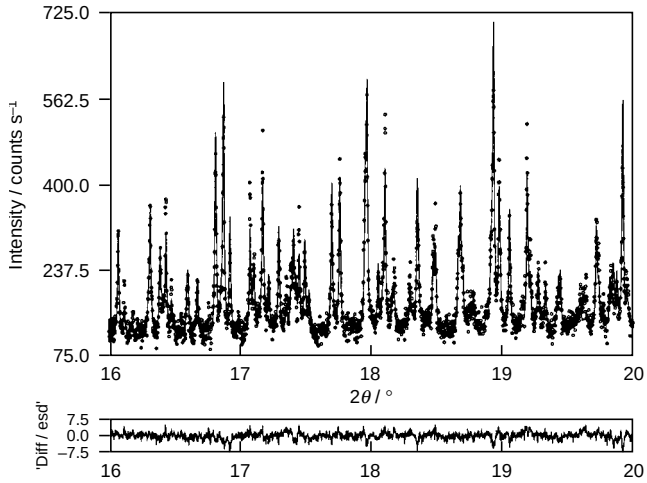

Fig. 2 Calculated (—) and observed (·) diffraction data for promazine hydrochloride **II**, with the crucial higher angle (*i.e.* higher spatial resolution) region shown. The calculated pattern is the result of a Rietveld refinement of the simulated annealing solution in which only the scale factor and overall temperature factor were refined. The atomic coordinates obtained from the simulated annealing algorithm were not refined.

degrees of freedom. Equally remarkable, given the limited spatial resolution of the data, is the agreement between the simulated annealing solution for capsaicin and a 0.7 Å resolution single-crystal structure determined subsequently in order to validate the simulated annealing approach (Fig. 3).

A combination of factors make the simulated annealing approach attractive as a method for structure determination: (*a*) the method works well on relatively low spatial resolution data, opening the way to rapid data collections (Table 1), (*b*) input models are easily constructed in internal coordinates using standard bond lengths and angles, (*c*) the process is entirely structure factor driven and thus no intra- or inter-molecular distance or energy checks are needed, and (*d*) the agreement factor is quick to calculate, leading to the rapid evaluation of the correct solution. In the case of capsaicin, structures were evaluated at a rate of *ca*. 150 per second. On smaller molecules, rates in excess of 1000 structure evaluations per second have been achieved. Ongoing optimisation of the goodness-of-fit χ^2 evaluation code promises to increase this rate still further.

In conclusion, we have demonstrated a very powerful method for solving molecular crystal structures from powder diffraction data. Although the experimental data were collected at a synchrotron source, the ability to solve structures reliably with data extending to resolutions of only $\approx$1.8 Å means that the approach will also be applicable to data collected on laboratory-based X-ray diffractometers. Approximately 77% of organic structures reported in the Cambridge Structural Database[11] have fewer atoms than thiothixene. It follows that the approach should have broad applicability in molecular crystal structure analysis.

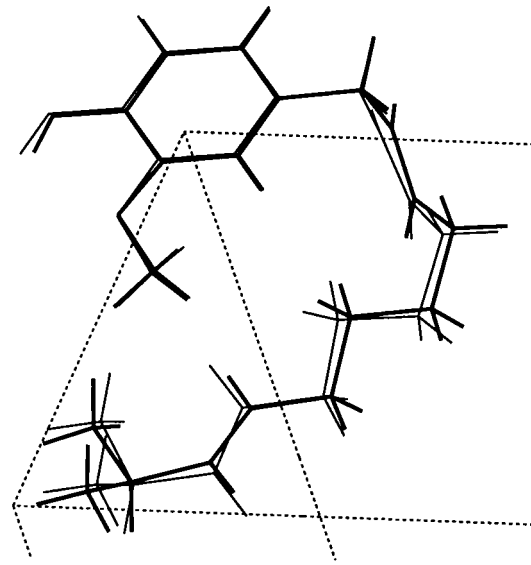

Fig. 3 The solved crystal structure of capsaicin **I** output directly from the simulated annealing algorithm, with the single-crystal solution shown overlaid in bold. The minimum, maximum and mean distances between pairs of corresponding capsaicin atoms are 0.045, 0.418 and 0.150 Å, respectively.

We thank Dr Andy Fitch of the ESRF for setting up BM16, Dr Chris Frampton of Roche Drug Discovery for determining the single-crystal structure of capsaicin, and Dr Joel Oliver of Procter and Gamble for helpful discussions regarding the structure of capsaicin.

Notes and References

† E-mail: wifd@isise.rl.ac.uk

1 D. M. Poojary and A. Clearfield, *Acc. Chem. Res.*, 1997, **30**, 414.
2 A. N. Fitch, *Nucl. Instrum. Methods Phys. Res. B*, 1995, **97**, 63.
3 K. Shankland, W. I. F. David and T. Csoka, *Z. Kristallogr.*, 1997, **212**, 550.
4 G. S. Pawley, *J. Appl. Crystallogr.*, 1981, **14**, 357.
5 J. Foley, A. Van Dam, S. Feiner and J. Hughes, *Computer Graphics, Principles and Practice*, Addison Wesley, Reading, MA, 2nd edn., 1993, ch. 21.
6 *International Tables for X-ray Crystallography Vol. A* (*Space Group Symmetry*), ed. T. Hahn, Kluwer Academic Publishers, Dordrecht, 1989.
7 H. M. Rietveld, *J. Appl. Crystallogr.*, 1969, **2**, 65.
8 K. D. M. Harris and M. Tremayne, *Chem. Mater.*, 1996, **8**, 2554.
9 Y. G. Andreev, P. Lightfoot and P. G. Bruce, *J. Appl. Crystallogr.*, 1997, **30**, 294.
10 B. M. Kariuki, H. Serrano-Gonzalez, R. L. Johnson and K. D. M. Harris, *Chem. Phys. Lett.*, 1997, **280**, 189.
11 F. H. Allen and O. Kennard, *Chem. Des. Automat. News*, 1993, **8**, 1.

Received in Cambridge, UK, 2nd February 1998; 8/00855H