Routine *ab initio* structure determination of chlorothiazide by X-ray powder diffraction using optimised data collection and analysis strategies

Kenneth Shankland,* William I. F. David and Devinderjit S. Sivia

ISIS Facility, Rutherford Appleton Laboratory, Chilton, Didcot, Oxon, UK OX11 0QX

The likelihood of solving crystal structures from powder diffraction data is greatly enhanced if data collection and analysis strategies can be designed to effectively remove Bragg peak overlap. In this way, accurate normalised structure factors of essentially single-crystal quality are obtained. Such strategies are illustrated here with the *ab initio* solution, from powder diffraction data using traditional direct methods, of the clinically used diuretic compound chlorothiazide. The structure solution is outstanding in that, despite the non-centrosymmetric, triclinic symmetry, all 17 non-hydrogen atom positions are clearly visible in the *E*-map generated from the top direct methods solution.

The structure solution of moderately complex crystal structures is generally straightforward and routine using single-crystal diffraction data.[1] This is not the case with powder diffraction data, principally because of the inevitable overlap of Bragg reflections that occurs with the compression of the three dimensions of diffraction space onto the one dimension of a powder diffraction pattern. The demand for structural information obtained from powder diffraction data comes from the large number of materials that are only easily synthesised or crystallised in powder form; for example, many zeolites and drug polymorphs. This situation has spurred a significant theoretical[2–6] and experimental effort, leading over the past decade to an increasing catalogue of successes using powder diffraction methods.[7] Many of these have been the crystal structures of small molecules,[8,9] simple inorganic systems[10] or systems in high-symmetry space groups,[11] though a smaller number of much larger systems, which indicate the potential of the powder diffraction technique, have been reported.[12,13] The solution of an inorganic structure with 50 non-H atoms in the asymmetric unit from laboratory X-ray data alone is particularly impressive.[14] However, these more complex systems possess a structural hierarchy of heavy and light atoms that leads to a customised, sequential structure solution process in which the first step involves the detection of part of the overall crystal structure (principally the heavy atoms) by direct methods, with the remainder of the structure being found *via* a series of difference Fourier map operations. Such sequential techniques are less appropriate for the solution of compounds in which heavy atoms are absent and so more reliance is placed upon direct methods to find the bulk of the structure in the initial structure solution.

We show here that a combination of optimised data collection strategies and Bayesian data analysis procedures can turn a complex powder diffraction pattern into effectively a single-crystal data set where the Bragg peak overlap is essentially removed. All of the direct methods algorithms developed for the analysis of single-crystal diffraction data thus become applicable to powder diffraction data without modification or approximation.

The approach is straightforward; conventional direct methods have sufficient power to solve moderately complex crystal structures if accurate structure factor magnitudes, $|F|$, are available. Thus reliable $|F|$ values should be determined whilst recognising that, for direct methods, it is desirable to collect normalised structure factors, $|E|$, with a precision that is independent of the position of the associated Bragg peaks in the powder diffraction pattern. Diffraction data are usually collected in a step-by-step manner at a constant rate across the diffraction pattern. This leads to the familiar fall-off in intensity with increasing angle, that can amount to more than an order of magnitude between low and high angles. Bragg intensities at low angles are measured with great precision whilst high-angle reflections may be scarcely distinguishable from the background. As direct methods utilise $|E|$ values, these should ideally be measured with equal precision across the diffraction pattern. $|E|$ values are related to structure factor magnitudes by

$$|E(\boldsymbol{h})|^2 = |F(\boldsymbol{h})|^2/\varepsilon(\boldsymbol{h}) \sum_{j=1}^{n} f_j^2(\theta)\exp(-2B_j \sin^2\theta/\lambda^2)$$

where $\varepsilon(\boldsymbol{h})$ is the enhancement (epsilon) factor, $f_j(\theta)$ and B_j are the atomic scattering factor and Debye–Waller factor, respectively, for the jth of the n atoms in the unit cell, and λ is the X-ray wavelength. The measured number of photons, $N(\boldsymbol{h})$, is proportional to the structure factor magnitudes, through the relationship

$$N(\boldsymbol{h}) \propto t(\theta)L_{\mathrm{p}}(\theta)m(\boldsymbol{h})|F(\boldsymbol{h})|^2$$

where $t(\theta)$ is the dwell time, $L_{\mathrm{p}}(\theta)$ the Lorentz–polarisation correction and $m(\boldsymbol{h})$ is the reflection multiplicity. The absence of a polarisation correction in a synchrotron radiation measurement gives $L_{\mathrm{p}}(\theta) = 1/\sin\theta \sin 2\theta$.

Combining these equations, and noting that $\varepsilon(\boldsymbol{h})m(\boldsymbol{h})$ is the order of the space group, leads to

$$N(\boldsymbol{h}) \propto \left[t(\theta)|E(\boldsymbol{h})|^2 \sum_{j=1}^{n} f_j^2(\theta)\exp(-2B_j \sin^2\theta/\lambda^2) \right] / (\sin\theta \sin 2\theta)$$

Thus $N(\boldsymbol{h})/|E(\boldsymbol{h})|^2$ should be constant in order to collect the same number of photons for a given E magnitude irrespective of its position in the diffraction pattern. Within the levels of approximation in this analysis, the atomic scattering factors and Debye–Waller factors may be represented by average values. This leads to the time spent at a given diffraction angle conforming to the equation

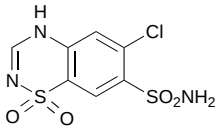
$$t(\theta) \propto (\sin\theta \sin 2\theta)/[f_{\mathrm{av}}^2(\theta)\exp(-2B_{\mathrm{av}} \sin^2\theta/\lambda^2)]$$

where f_{av} is a representative atomic scattering factor (carbon for a typical organic compound) and B_{av} is an estimated overall Debye–Waller factor. The dwell times obtained from this calculation can be scaled to the duration of the experiment, subject to constraints such as a minimum count time below which the overhead incurred in moving a detector makes the counting scheme inefficient. Related protocols[15,16] have been proposed for improving the quality of Rietveld refinements of

structural models, but have not previously been applied to the issue of structure determination.

Even with such a count scheme, the reliability of the $|E|$ estimates for the crucial high-resolution reflections is generally much worse than that of isolated lower resolution reflections, as the increasing extent of reflection overlap with increasing 2θ makes accurate intensity extraction by whole pattern fitting problematic.[6] However, in general, molecular structures undergo anisotropic thermal expansion. Therefore, reflections which are completely overlapped at one temperature are likely to be separated at another (sufficiently different) temperature, and this separation is generally enough to allow direct determination of the reflection intensities in the pattern fitting process. More reliable intensity estimates can thus be obtained by collecting diffraction data at more than one temperature, and combining the data sets to give effectively a single-crystal quality data set, without resort to computational schemes such as equipartitioning or permutation of intensities.

The methods described above were used in the collection of synchrotron X-ray diffraction data from a powder sample of the drug substance chlorothiazide **I** for which only the unit-cell dimensions and space group were previously known.[17]



I

Experimental

A sample of chlorothiazide powder obtained from the Sigma Chemical Co. was recrystallised from ethanol to ensure that only a single powder phase was present. The sample was loaded into a 1 mm capillary and placed inside a cryostat on station 9.1 at the Daresbury Synchrotron Radiation Source. Using an incident wavelength of 1.0985 Å and a temperature of 130 K, diffraction data were collected using the variable count time scheme shown in Fig. 1(*a*). The scheme was calculated for the range 10–60° 2θ in 0.01° steps using $\lambda = 1.1$ Å, $B = 1.0$, a total count time of 6 h and a minimum count time of 1.0 s. To simplify implementation within the confines of the instrument control software, the original continuously variable scheme was approximated by 33 ranges of constant count time with a 0.5 s difference between each adjacent range. The raw powder diffraction data obtained are shown in Fig. 1(*b*). The pattern represents the summation of two 6 h data sets collected using the count scheme described. In addition, but unrelated to the count scheme, data were collected in the direction 60°→10° 2θ so that the fall in incident synchrotron flux with time after beam injection was somewhat offset by the increasingly strong Bragg reflections encountered at lower angles. A 10 h data set was also collected at 160 K.

Results

The first 20 lines of the 130 K diffraction pattern [Fig. 1(*b*)] were indexed using the DICVOL91[18] program which returned a unit cell with dimensions $a = 6.372$ Å, $b = 8.915$ Å, $c = 4.853$ Å, $\alpha = 96.12°$, $\beta = 99.47°$, $\gamma = 74.40°$, $V = 261.3$ Å^3 (implying a space group of $P1$, given the estimated molecular volume), in good agreement with the previous cell determination. The unit-cell dimensions, reflection intensities, zero point and peak shape (Voigt, with an asymmetry correction for axial divergence[19,20]) were refined using the SR15LS[21] program to give a 130 K cell of dimensions $a = 6.3720(1)$ Å, $b = 8.9159(1)$ Å, $c = 4.8554(1)$ Å, $\alpha = 96.1321(3)°$, $\beta = 99.4760(4)°$, $\gamma = 74.4119(5)°$, with a χ^2 of 3.71 and an R_p of 3.38% for the fit.

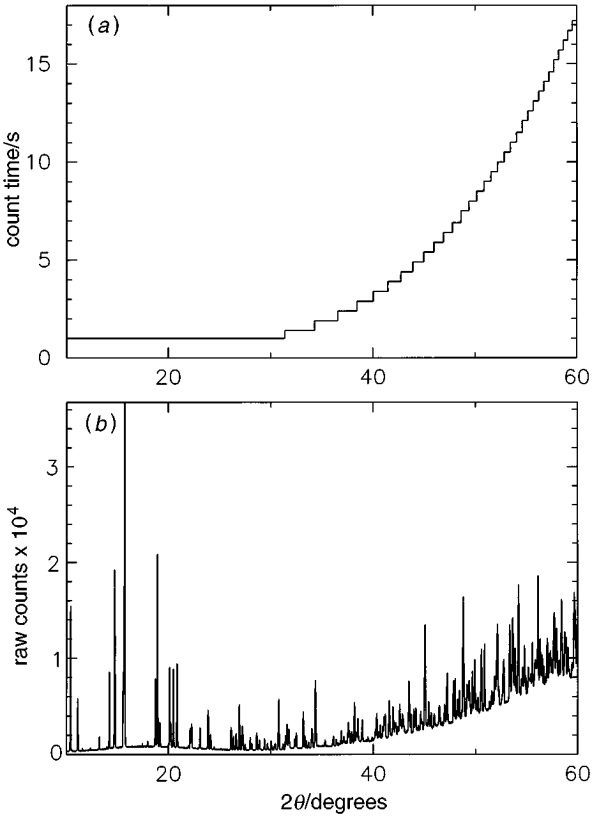


Fig. 1 (*a*) The variable count time scheme used in the collection of diffraction data from chlorothiazide. (*b*) Raw powder diffraction data obtained on station 9.1 at the Daresbury Synchrotron Radiation Source from a powder sample of chlorothiazide at 130 K.

Fig. 2 shows the effect over a small region of the diffraction pattern of the anisotropic thermal expansion experienced by chlorothiazide upon changing the temperature from 130 to 160 K, whilst Table 1 lists the positions and extracted intensities for the corresponding reflections. At 130 K, the $2\,\bar{1}\,0$ and $\bar{1}\,2\,1$ reflections are so closely spaced in 2θ that they are treated as a single reflection in the extraction process, and the resultant intensity is equipartioned between the two contributors. At 160 K, the doublet is sufficiently resolved to allow the reflections to be treated as separate entities in the extraction. This behaviour is reflected in other overlapped Bragg peaks throughout the diffraction pattern, *i.e.* 22 out of the 30 overlap sets present at 130 K are split at 160 K. The estimated intensity ratio of *ca.* 1.2:1 for the $2\,\bar{1}\,0/-\bar{1}\,2\,1$ pair is in good agreement with the calculated ratio of *ca.* 1.3:1 obtained from the refined 130 K structural model.

The data collection strategies described only partly address the problem of collecting precise $|E|$ values. The reflection intensities extracted by the above refinement were transformed to an optimal estimate of the structure factor amplitudes using a Bayesian procedure which entails a formal method of propagating errors that circumvents the problems associated with negative intensities and minimises the effects of Bragg peak overlap, enabling the best statistical decorrelation of neigh-

Table 1 Extracted structure factor intensities in the range 23.6–24° 2θ for chlorothiazide at 130 and 160 K

	130 K			160 K		
$h\,k\,l$	2θ/degrees	I	$\sigma(I)$	2θ/degrees	I	$\sigma(I)$
$\bar{2}\,\bar{2}\,1$	23.834	4.30	0.28	23.793	4.04	0.20
$2\,\bar{1}\,0$	23.865	9.39	0.18	23.850	9.80	0.97
$\bar{1}\,2\,1$	23.869	9.39	0.18	23.859	8.40	0.93

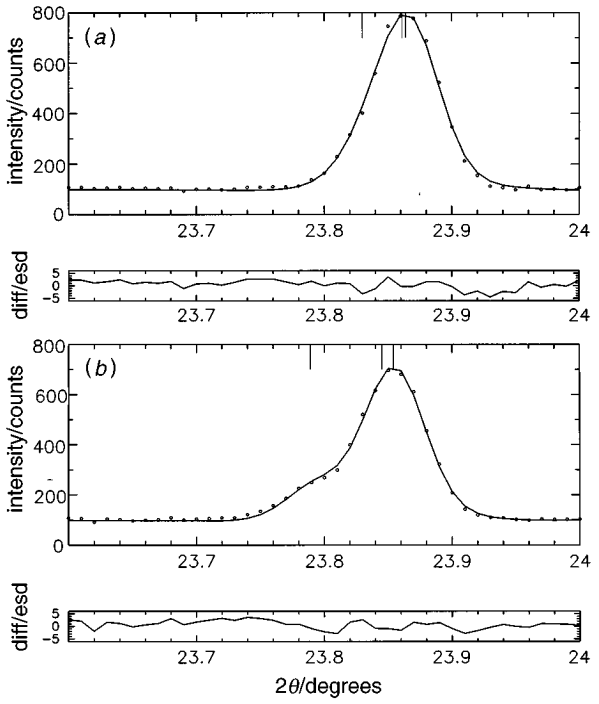


Fig. 2 Cell and intensity least squares fits to a selected data region at (*a*) 130 K and (*b*) 160 K. The unit cell at 160 K is $a=6.3802(1)$ Å, $b=8.9270(1)$ Å, $c=4.8594(1)$ Å, $\alpha=96.1818(3)^\circ$, $\beta=99.4932(4)^\circ$, $\gamma=74.3422(5)^\circ$. The tick marks, corrected for zero-point error, show the positions (in order of increasing 2θ) of the $\bar{2}\,\bar{2}\,1$, $2\,\bar{1}\,0$, and $\bar{1}\,2\,1$ reflections in both cases.

bouring $|E|$ value estimates.[6] Using only the 130 K data, a total of 420 structure factors were obtained and used as input to the MITHRIL94[22] direct methods package. 1542 Triplets and 1605 negative quartets were generated from the top 157 and top 140 $|E|$ values respectively; 15 possible solutions were subsequently developed by tangent refinement. The solution with the highest combined figure of merit (CFOM = 2.48, next highest 2.23, then 1.77) produced an *E*-map in which the positions of all the non-H atoms in the structure were found by an automatic peak search (Table 2). An initial Rietveld refinement against the 130 K pattern in which only the scale factor was varied gave a reasonable fit to the data, showing the determined structure to be substantially correct. Hydrogen atoms were added in calculated positions ($C_x-H=0.95$ Å, $N_x-H=0.9$ Å) with fixed B_{iso} values of 3.0 and allowed to ride on their parent atoms. The final cycle of least squares had 65 variables and gave agreement factors of $R_B=6.33\%$, $R_p=5.10\%$, $R_{wp}=5.46\%$, $R_e=1.77\%$ and $\chi^2=9.54$. Refined atomic coordinates are listed in Table 2 and the corresponding profile fit is shown in Fig. 3. Fig. 4 shows four adjacent unit cells of the refined structure in projection down the *c* axis. The intermolecular hydrogen-bond distances shown are all within the ranges observed for organic molecular crystals containing these functional groups.[23]

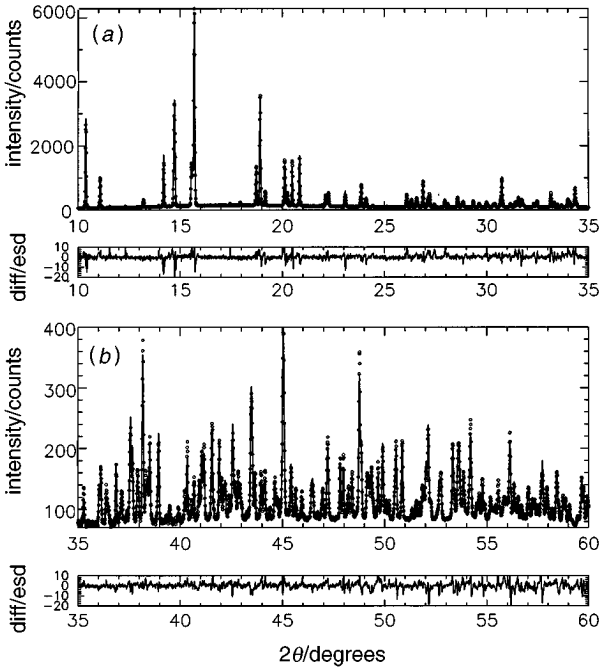


Fig. 3 Final observed, calculated and $(y_{obs}-y_{calc})/\sigma(y_{obs})$ plots for chlorothiazide at 130 K. The fitted data are those shown in Fig. 1(*b*), normalised to take account of the variation in dwell times across the pattern.

Table 2 Peak numbers, heights and atom assignments for the *E*-map corresponding to the top MITHRIL94 solution. Refined fractional atomic coordinates and B_{iso} values for the assigned atoms are also given. Atom S(2) was used to fix the location of the molecule in the unit cell. Δ_{ref} is the distance between an initial atomic position determined from the *E*-map and a corresponding final refined position

peak	height	atom	*x*/*a*	*y*/*b*	*z*/*c*	B_{iso}/Å^2	Δ_{ref}/Å
1	4418	S(2)	0.4164	0.2481	0.0214	0.13(7)	—
2	3323	S(1)	0.2559(7)	0.8019(4)	0.6364(7)	0.56(8)	0.159
3	2719	Cl(1)	0.7575(5)	0.6876(4)	0.9701(8)	1.55(8)	0.115
4	1840	C(6)	0.383(2)	0.522(1)	0.346(3)	0.1(3)	0.344
5	1518	N(3)	0.336(2)	0.938(1)	0.545(2)	1.4(2)	0.366
6	1284	C(2)	0.735(2)	0.336(1)	0.403(2)	0.4(2)	0.260
7	1272	C(1)	0.820(2)	0.102(1)	0.117(2)	0.5(2)	0.381
8	1267	N(2)	0.880(1)	0.202(1)	0.345(2)	0.5(2)	0.336
9	1245	C(3)	0.808(3)	0.433(2)	0.639(3)	2.7(4)	0.429
10	1215	C(5)	0.446(2)	0.631(1)	0.569(2)	1.1(3)	0.222
11	1107	C(4)	0.653(2)	0.575(2)	0.700(3)	1.8(3)	0.110
13	1004	O(3)	0.294(2)	0.177(1)	0.138(2)	2.2(2)	0.255
16	974	O(2)	0.262(1)	0.8213(8)	0.942(2)	1.6(2)	0.289
17	937	C(7)	0.514(2)	0.382(1)	0.266(3)	0.4(2)	0.121
20	878	O(4)	0.309(1)	0.3297(8)	−0.222(2)	1.0(1)	0.313
21	856	O(1)	0.0467(1)	0.7991(8)	0.468(2)	0.9(2)	0.094
23	739	N(1)	0.638(1)	0.107(1)	−0.044(2)	1.9(2)	0.322
		H(1)	0.237	0.547	0.250	3.00	—
		H(4)	0.957	0.393	0.715	3.00	—
		H(2)	0.934	0.019	0.056	3.00	—
		H(3)	1.006	0.173	0.464	3.00	—
		H(5A)	0.470	0.939	0.635	3.00	—
		H(5B)	0.340	0.924	0.359	3.00	—

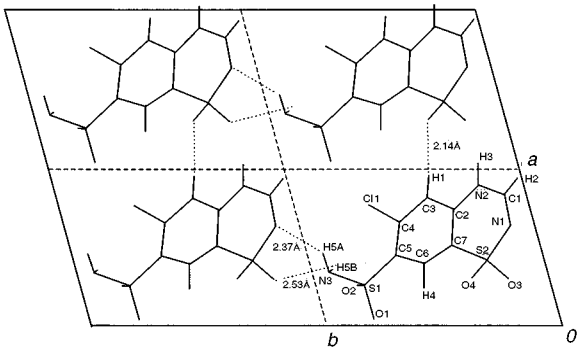


Fig. 4 The refined structure of chlorothiazide at 130 K, shown as a projection down the *c* axis of four adjacent unit cells. Intermolecular hydrogen bonds, including a weak C—H···O interaction, are shown by the finely dotted lines. An additional 2.17 Å intermolecular hydrogen bond running from N(3)—H(5B)···O(2) (in molecule at $x, y, z-1$) is not shown.

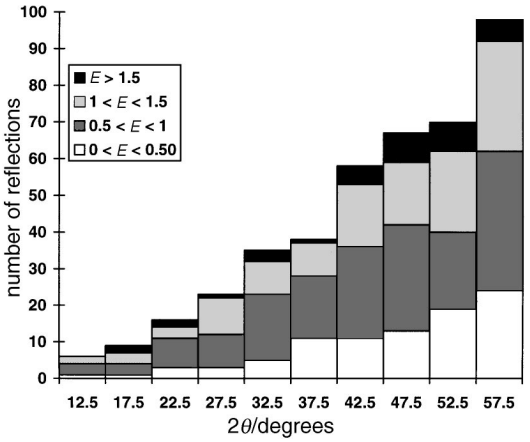


Fig. 5 Distribution of *E* magnitudes for the 420 reflections extracted from the 130 K data set

Discussion

It is clear that the combination of an improved data collection strategy and an optimal structure factor extraction procedure has resulted in a very high quality, routine structure solution, both in terms of the ease with which it was identified from its CFOM and the proximity of the initial atomic positions to their final refined positions. It is significant that 80% of the data collection time was spent in the range 40–60° and that 75% of strong $|E|$ magnitudes ($|E|>1.5$) lie in this range (Fig. 5). This underlines the fundamental point that the strategy has ensured accurate estimates of these crucial strong $|E|$ magnitudes. Furthermore, the superiority of the Bayesian method of extracting of structure factors over the traditional method of cell and intensity least squares (CAILS, commonly known as the Pawley method) is demonstrated when structure solutions that neglect the Bayesian analysis are attempted. All positive intensities extracted by the CAILS procedure (about 90% of the total) were used, and structure factor magnitudes calculated using $|F|=\sqrt{F^2}$ and $\sigma(F)=\sigma(F^2)/2|F|$. The *E*-map generated from the top MITHRIL94 solution (CFOM = 2.56) did not contain the correct structure. The second and fifth ranked solutions did generate *E*-maps with almost the entire structure visible, but only when peak connectivity was extended down to include the top 32 peaks in the *E*-map, *cf*. Table 2, which shows a solution with fewer significant spurious peaks.

Conclusions

The approach to structure solution from powder data presented here permits accurate structure factor magnitudes to be extracted across a complete diffraction pattern, a crucial step if conventional direct methods are to be employed effectively in solving large structures. Applicable to both synchrotron and laboratory X-ray powder diffraction experiments, the approach also enhances the widely used sequential method of structure solution, by allowing larger initial fragments to be located with greater accuracy for subsequent use in difference Fourier calculations. Other methods, based on fitting trial models to diffraction data[24,25] and using an agreement factor (such as R_{wp}) for discrimination, will also benefit from the improved signal to background ratio at high 2θ obtained when an optimised count time scheme is used.

We thank G. Bushnell-Wye for access to, and calibration of, station 9.1. We also thank S. W. Love for his help during the data collection.

References

1 G. H. Stout and L. H. Jensen, *X-ray Structure Determination*, Wiley Interscience, New York, 1989.
2 W. I. F. David, *J. Appl. Crystallogr.*, 1987, **20**, 316.
3 G. Bricogne, *Acta Crystallogr., Sect. A*, 1991, **47**, 803.
4 J. Jansen, R. Peschar and H. Schenk, *J. Appl. Crystallogr.*, 1992, **25**, 231.
5 M. A. Estermann and V. Gramlich, *J. Appl. Crystallogr.*, 1993, **26**, 396.
6 D. S. Sivia and W. I. F. David, *Acta Crystallogr., Sect. A*, 1994, **50**, 703.
7 J. I. Langford and D. Louër, *Rep. Prog. Phys.*, 1996, **59**, 131.
8 P. Lightfoot, M. Tremayne, C. Glidewell and K. D. M. Harris, *J. Chem. Soc., Perkin Trans. 2*, 1993, 1625.
9 R. G. Delaplane, W. I. F. David, R. M. Ibberson and C. C. Wilson, *Chem. Phys. Lett.*, 1993, **201**, 75.
10 H. Fjellvag and P. Karen, *Inorg. Chem.*, 1992, **31**, 3260.
11 D. M. Poojary, R. B. Borade, F. L. Campbell and A. Clearfield, *J. Solid State Chem.*, 1994, **112**, 106.
12 R. E. Morris, W. T. A. Harrison, J. M. Nicol, A. P. Wilkinson and A. K. Cheetham, *Nature (London)*, 1992, **359**, 519.
13 R. E. Morris, J. J. Owen, J. K. Stalick and A. K. Cheetham, *J. Solid State Chem.*, 1994, **111**, 52.
14 D. M. Poojary, A. Cabeza, M. A. G. Aranda, S. Bruque and A. Clearfield, *Inorg. Chem.*, 1996, **35**, 1468.
15 W. I. F. David, *Accuracy in Powder Diffraction*, NIST Special Publication no. 846, 1992, 210.
16 I. C. Madsen and R. J. Hill, *J. Appl. Crystallogr.*, 1994, **27**, 385.
17 P. L. Dupont and O. Dideberg, *Acta Crystallogr., Sect. B*, 1970, **26**, 1884.
18 A. Boultif and D. Louer, *J. Appl. Crystallogr.*, 1991, **24**, 987.
19 M. M. Eddy, A. K. Cheetham and W. I. F. David, *Zeolites*, 1986, **6**, 449.
20 L. W. Finger, D. E. Cox and A. P. Jephcoat, *J. Appl. Crystallogr.*, 1994, **27**, 892.
21 W. I. F. David, R. M. Ibberson and J. C. Matthewman, Rutherford Appleton Laboratory Report RAL-92-032, 1992.
22 C. J. Gilmore and S. R. Brown, *J. Appl. Crystallogr.*, 1988, **21**, 571.
23 G. A. Jeffrey and W. Saenger, *Hydrogen Bonding in Biological Structures*, Springer-Verlag, Berlin, 1991.
24 M. Tremayne, B. M. Kariuki and K. D. M. Harris, *J. Mater. Chem.*, 1996, **6**, 1601.
25 Y. G. Andreev, P. Lightfoot and P. G. Bruce, *Chem. Commun.*, 1996, 2169.

Paper 6/06998C; *Received* 14*th October*, 1996