Acta Crystallographica Section B
Structural Science
ISSN 0108-7681

Diffuse scattering study of aspirin forms (I) and (II)

E. J. Chan, T. R. Welberry,* A. P. Heerdegen and D. J. Goossens

Research School of Chemistry, Australian National University, Canberra, ACT 0200, Australia

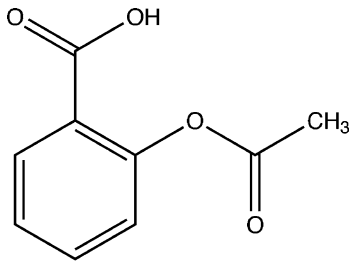
Correspondence e-mail: welberry@rsc.anu.edu.au

Received 24 July 2010
Accepted 16 September 2010

Full three-dimensional diffuse scattering data have been recorded for both polymorphic forms [(I) and (II)] of aspirin and these data have been analysed using Monte Carlo computer modelling. The observed scattering in form (I) is well reproduced by a simple harmonic model of thermally induced displacements. The data for form (II) show, in addition to thermal diffuse scattering (TDS) similar to that in form (I), diffuse streaks originating from stacking fault-like defects as well as other effects that can be attributed to strain induced by these defects. The present study has provided strong evidence that the aspirin form (II) structure is a true polymorph with a structure quite distinct from that of form (I). The diffuse scattering evidence presented shows that crystals of form (II) are essentially composed of large single domains of the form (II) lattice with a relatively small volume fraction of intrinsic planar defects or faults comprising misoriented bilayers of molecular dimers. There is evidence of some local aggregation of these defect bilayers to form small included regions of the form (I) structure. Evidence is also presented that shows that the strain effects arise from the mismatch of molecular packing between the defect region and the surrounding form (II) lattice. This occurs at the edges of the planar defects in the *b* direction only.

1. Introduction

Polymorphism of aspirin crystals has recently become an important issue, following reports of a second polymorph (Vishweshwar *et al.*, 2005), because of the widespread popular use of aspirin as an analgesic. The reported structure appeared to correspond to one of the low-energy structures predicted in the calculations of Ouvrard & Price (2004). Subsequently, Bond *et al.* (2007*a*) cast some doubt on the findings and stated 'form (II) of aspirin as reported ... may just as easily be derived, to the accuracy and precision reported, ... from experimental diffraction data collected from ... a single crystal of the well known form (I)'. In a second paper Bond *et al.* (2007*b*) investigated a number of different crystals that were supposedly form (II). These crystals showed that in addition to the Bragg diffraction peaks that could be indexed on the reciprocal lattice of form (II) there were diffuse streaks extending along a^* in $k =$ odd reciprocal layers. In some samples additional diffuse peaks occurred along these streaks mid-way between the form (II) lattice peaks. The authors concluded that the form (II) crystals consisted of "an intergrowth of two 'polymorphic' domains" and that 'each aspirin crystal is an integral whole in which the domains are intimately connected with each other, with possibly many turnovers of domain within a single crystal'.



Aspirin

In order to test this hypothesis and to clarify the many misconceptions and uncertainties concerning the polymorphism of aspirin we have undertaken a study using diffuse X-ray scattering data obtained at a synchrotron light source from typical form (II) crystals. These data have been analysed using Monte Carlo (MC) simulations of a model crystal. The use of such MC simulations has become a powerful and well accepted technique for aiding the interpretation and analysis of diffuse scattering patterns (Welberry & Butler, 1994; Weber *et al.*, 2001; Weber & Bürgi, 2002; Welberry, 2004; Paściak *et al.*, 2010). The method has been successful in recent studies of polymorphic molecular systems (Chan *et al.*, 2009; Chan & Welberry, 2010) where the accuracy achieved was sufficiently good that precursor effects, foreshadowing the onset of a phase transition, were detected.

In the present work we develop a model for aspirin that involves occupational disorder (or short-range order – SRO), thermal disorder (giving rise to thermal diffuse scattering – TDS) and the effects of local strain in which molecular displacements are coupled to the occupational disorder (analogous to the atomic size effect). By adjusting occupancy variables the model can be made to describe a whole range of structures varying from the pure form (II) to the pure form (I) structure. For all intermediate structures it is necessary to consider strain at the interfaces.

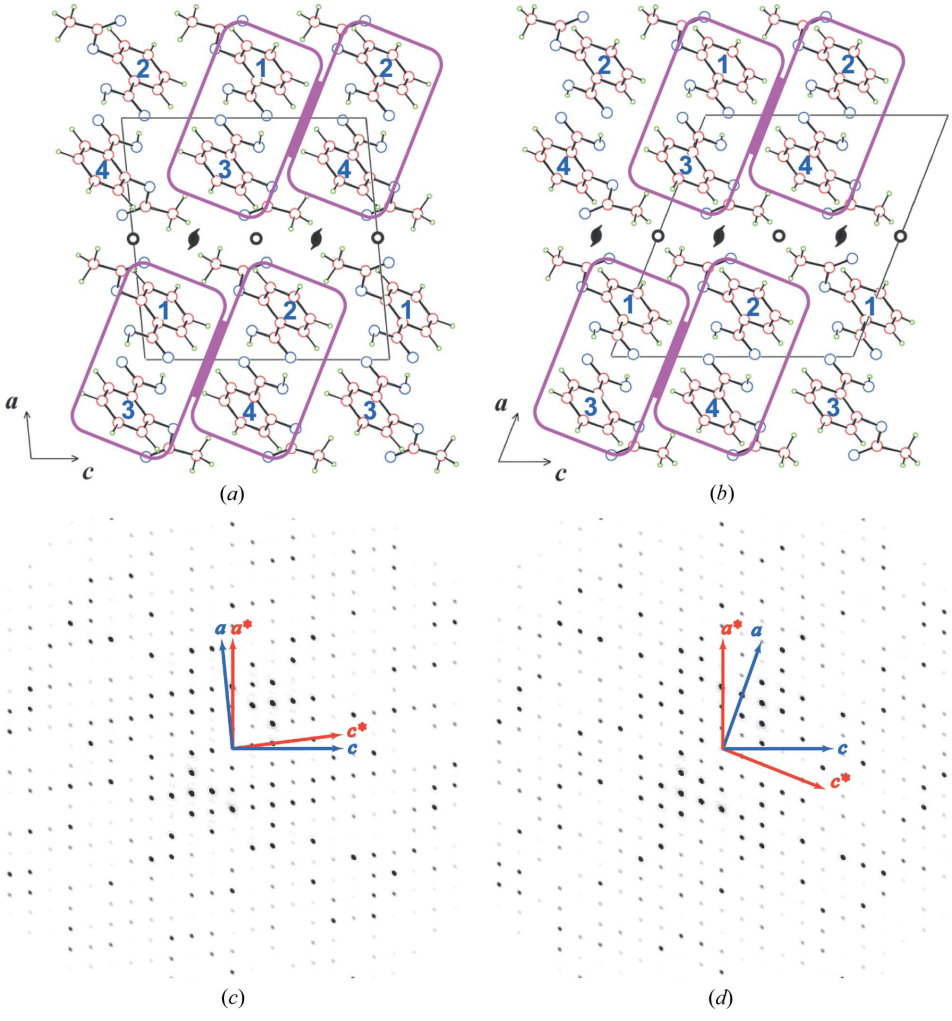


Figure 1
Differences in stacking for molecules of aspirin viewed down the *b* axis in (*a*) form (I) and (*b*) form (II). The numbering of the molecules corresponds to the equivalent position specifiers given in Table 2. The coloured lozenge-shaped envelopes identify the molecular dimers referred to in the text. The corresponding calculated Bragg reflections in the *h*1*l* section are shown in (*c*) for form (I) and (*d*) for form (II).

1.1. Description of form (I) and form (II) structures

Fig. 1 shows plots of the two forms of aspirin viewed down the monoclinic *b* axis. In this projection the structures look identical. The two forms differ in the way that layers stack down **a** so that the centre of symmetry and the 2_1 screw axes occur in different positions. The differences are not apparent in the *h*0*l* reciprocal sections but may be seen in the *h*1*l* sections shown in Figs. 1(*c*) and (*d*). Cell data for the two forms are given in Table 1.

It should be noted that for ease of structural comparison between the two forms, the atomic coordinates for form (II) were placed in a non-standard setting of $P2_1/c$ with an acute β angle. The relationship of this setting to that of Vishweshwar *et al.* (2005) (which has an obtuse β angle) is that $x, y, z \rightarrow \frac{1}{2} - x, -y, z$. The cell setting used here also corresponds to that used by Ouvrard & Price (2004).

Aspirin molecules arrange themselves in both forms (I) and (II) as bilayers made up of hydrogen-bonded molecular dimers. In Fig. 1 the dimers are identified in projection by the coloured lozenge-shaped envel-

Table 1
Cell data for polymorphs (I) and (II) of aspirin, $C_9H_8O_4$.

Polymorph	Aspirin (I)	Aspirin (II)
Space group	$P2_1/c$ (No. 14)	$P2_1/c$ (No. 14)
a (Å)	11.416 (5)	12.2696 (5)
b (Å)	6.598 (2)	6.5575 (3)
c (Å)	11.483 (5)	11.4960 (4)
β (°)	95.60 (3)	68.163 (2)
V (Å^3)	860.8 (6)	858.58 (6)
Z	4	4

opes. It should be noted that the actual molecular dimers are inclined steeply to the *ac* plane. In form (II) the dimer in one layer [*e.g.* comprising molecules of types (1) and (3)] is related to the corresponding dimer in the next layer along *a* by a centre of symmetry (see Fig. 1*b*). In form (I) the dimers are related instead by a 2_1 screw axis (see Fig. 1*a*).

2. Experimental

2.1. Crystal growth

Suitable crystals of form (II) of aspirin were grown *via* super-cooling a 0.6–0.7 *M* solution of aspirin in acetonitrile to 273 K after an incubation period of 7 d at 308–313 K in a water bath. This was deemed to be sufficient time to completely dissolve any form (I) seeds. All crystallization vessels should be securely sealed.

To obtain good quality diffuse-scattering data large, good quality, single crystals with linear dimensions ~ 0.5 mm were sought. Although numerous crystals were screened for use in the experiment many were discarded because they exhibited signs of non-merohedral twinning or powder streaking and only a small number of suitable good quality specimens were obtained. The sample chosen for the diffuse scattering data collection was a good quality prismatic crystal with well formed facets and edges with no outward signs of any problems in its growth.

Suitable crystals of form (I) are easily grown *via* room temperature evaporation of an acetonitrile solution of the same 0.6–0.7 *M* concentration.

2.2. Bragg experiments

Coordinates obtained from a conventional structure determination using Bragg diffraction data were used as the starting point for disorder modelling. Coordinates for form (I) were taken from a 300 K neutron scattering experiment reported by Wilson (2002). Since the previous characterization for the average structure of form (II) has been the subject of concerns regarding validation (Bond *et al.*, 2007*a*), a Bragg data set was collected. This was obtained from a crystal taken from the same batch of form (II) grown for use in our diffuse scattering experiments. A reliable starting set of coordinates for the disorder modeling was obtained. The CIF and Bragg reflection data are available as supplementary material.[1]

2.3. Diffuse scattering experiments

All data were collected at room temperature using a mar345 image-plate detector. For form (I) data collection was performed on the 11-ID-B beamline at the Advanced Photon Source. Here the photon energy was 58.26 keV [$\lambda = 0.2128$ (1) Å], the detector distance was 543.9 mm and exposure times were 20 s per frame. Data collection for form (II) was performed on the powder diffraction beamline at the Australian Synchrotron. Here the photon energy was 17.71 keV [$\lambda = 0.7000$ (1) Å], the detector distance was 130.0 mm and exposure times were 30 s per frame. Further details of the data collection methods and reciprocal space reconstruction have been described in previous papers (Chan *et al.*, 2009; Beasley *et al.*, 2008).

2.4. Observed diffuse scattering data

Although full three-dimensional data have been collected the majority of the model testing for the current analysis was carried out using data from three plane sections, *hk*0, 0*kl* and *h*1*l*. Some examples are shown in Fig. 2. It should be noted that because the structures of the two forms appear identical when viewed down **b** the *h*0*l* section is not very informative.

Figs. 2(*a*) and (*b*) show the complete *h*1*l* sections for forms (II) and (I). While for form (I) this section shows only Bragg peaks and thermal diffuse scattering (TDS) for form (II) the characteristic streaking along ($l =$ odd) rows that has previously been reported (Bond *et al.*, 2007*b*) is seen. Such streaking along one-dimensional *rods* in reciprocal space is characteristic of the presence of some kind of *planar* defect (normal to the rods) in real space that disrupts the regular stacking of the molecules. It is clear from the complete section views that the streaks in Fig. 2(*a*) contain sharp peaks at the form (II) reciprocal lattice postions, but in addition there are quite strong, though broader, diffuse peaks mid-way between them at the form (I) reciprocal lattice positions (see inset).

This inset pattern in Fig. 2(*a*) is an enlargement of the region indicated by the white rectangle in the main pattern. This region will be used to compare various models later in the paper. Other more intense regions of streaking that might otherwise have been used for this comparison were affected to some extent by the presence of blooming artefacts in the raw data (see Welberry *et al.*, 2005) and were not considered ideal for making the comparisons.

Fig. 2(*c*) shows a comparison of the observed data for the *hk*0 reciprocal sections of forms (I) and (II). It can be seen that overall the two patterns are very similar but, as clearly seen in the inset enlargement of the regions indicated by the white rectangles, there are some significant differences in detail. The most prominent differences appear as continous bands of intensity running in the a^* direction between the $2\bar{2}0$ and $3\bar{2}0$ reciprocal points and similarly between $1\bar{3}0$ and $2\bar{3}0$. Similar less intense examples can be seen on other $k \neq 0$ reciprocal lattice rows. These features appear as though they

[1] Supplementary data for this paper are available from the IUCr electronic archives (Reference: SO5043). Services for accessing these data are described at the back of the journal.

Table 2
List of symmetry operations which describe the equivalent positions for the molecular components used in the disordered model of aspirin.

Locations	Component $\mathcal{A}$	Component $\mathcal{B}$
1	x, y, z	$x, \frac{3}{2}-y, z$
2	$x, \frac{3}{2}-y, \frac{1}{2}+z$	$x, y, z+\frac{1}{2}$
3	$1-x, 1-y, -z$	$1-x, -\frac{1}{2}+y, -z$
4	$1-x, -\frac{1}{2}+y, \frac{1}{2}-z$	$1-x, 1-y, \frac{1}{2}-z$

could be related to the diffuse streaking seen in the $h1l$ section. It is conjectured that they appear here as a result of strain that arises when the planar defects or stacking faults are introduced into the bulk of a form (II) crystal.

3. Model construction

Our aim in this section is to describe the construction of a model crystal in which random numbers are used to define the orientation of the molecule in each molecular site. By suitably ordering these random variables, the crystal may then be made to be all form (I), all form (II) or a disordered combination of the two.

We define two components $\mathcal{A}$ and $\mathcal{B}$. These are defined by the equivalent position specifiers given in Table 2. Component $\mathcal{A}$ comprises the four molecules within a single unit cell shown in Fig. 1(*b*) in their normal form (II) positions and orientations. For component $\mathcal{B}$ each of the molecules of component $\mathcal{A}$ is reflected normal to b and translated by $0.5 \times b$. Used on their own, the component $\mathcal{B}$ specifiers also define a perfect form (II) structure but one that is reflected and translated relative to that described by the component $\mathcal{A}$ specifiers.

In Fig. 3 we show a schematic diagram of a section of structure viewed down b containing both $\mathcal{A}$ and $\mathcal{B}$ components. This is drawn on a square grid for convenience. (Note that there is a 1:1 mapping of the black squares here to the acute angled monoclinic unit cells in Fig. 1*b*). In each of the four molecular sites within a unit cell an $\mathcal{A}$ component is represented by a red ellipse with its major axis horizontal, while a $\mathcal{B}$ component is represented by a blue ellipse with its major axis vertical. If any two neighbouring unit cells in the a (vertical) direction contain the sequence $\mathcal{AAAA}$ or $\mathcal{BBBB}$ this corresponds to the form (II) structure, while a sequence $\mathcal{ABBA}$ or $\mathcal{BAAB}$ corresponds to the form (I) structure. The small region shown is seen to be predominantly the form (II) structure with some small isolated defects that are form (I).

3.1. Use of random variables

With the possibility of any individual molecular site being occupied by a component $\mathcal{A}$ molecule or a component $\mathcal{B}$ molecule the most general description of the system that allows for disorder of these would be to assign binary random variables, $\xi_{i,j,k,L}$, to each molecular site, where $\xi_{i,j,k,L} = -1$ corresponds to component $\mathcal{A}$ and $\xi_{i,j,k,L} = +1$ corresponds to component $\mathcal{B}$. Here i, j, k are indices labelling unit cells in the a, b and c crystal directions and L is the equivalent position specifier within the unit cell as defined in Table 2.

Although this is the most general formulation that would allow the system to be explored by modelling there are two

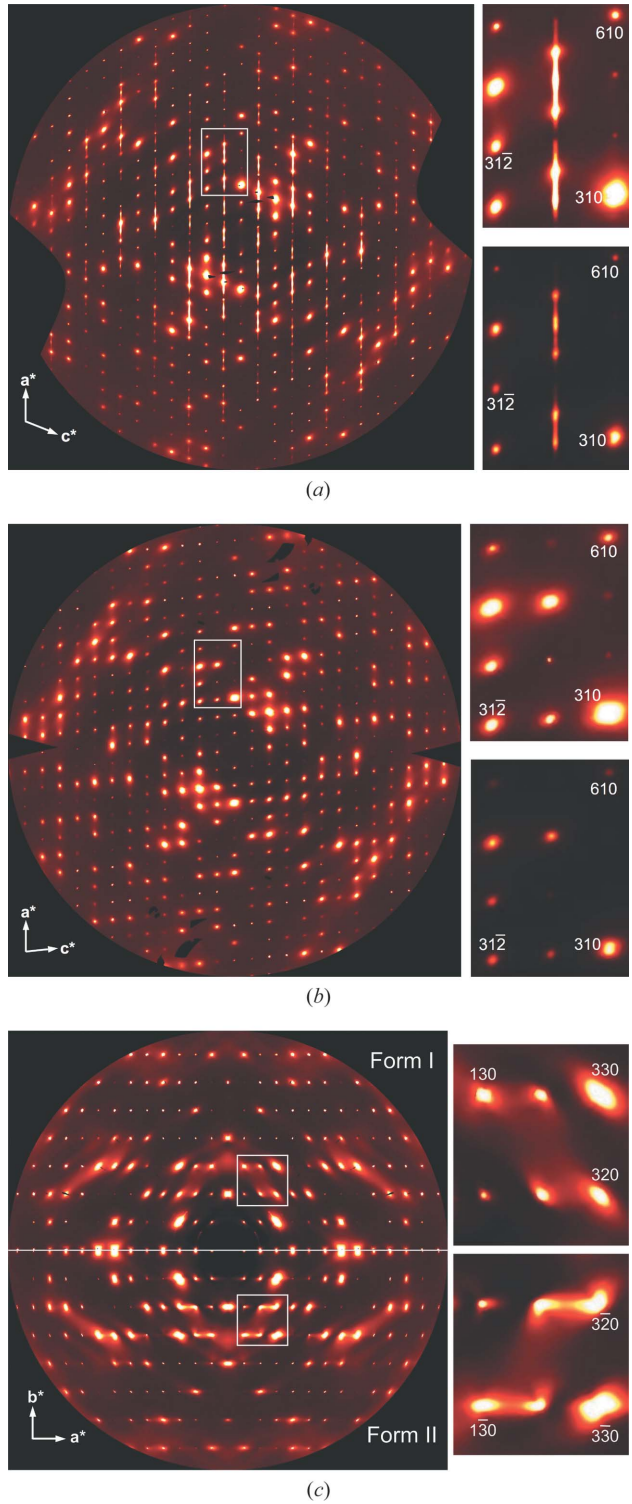


Figure 2
(*a*) The $h1l$ reciprocal layer of aspirin form (II) showing the diffuse streaking along l = odd rows. Inset are magnified images, on two different intensity scales, of the region outlined by the white rectangle. (*b*) A comparable $h1l$ reciprocal layer of aspirin form (I). (*c*) A comparison of the $hk0$ reciprocal layers for forms (I) and (II). Inset are magnified images of the regions outlined by the white rectangles. See text for details.

considerations that allow a significant simplification to be made. First, a detailed inspection of the average structures of both forms indicates that the molecular dimers indicated by the coloured lozenge-shaped envelopes in Fig. 1 are strong hydrogen-bonded units that are likely to remain intact irrespective of any disorder, *i.e.* any break in the pefect crystal sequence is not likely to involve disruption of these strongly hydrogen-bonded dimers. This means that a single random variable can be used to represent both individuals of the pair of sites such as **13** or **24** in the bilayer.

For an occupationally disordered system the diffuse scattering is proportional to the difference in the scattering factors for the two disordered species (Welberry, 1985). Since a single pair of sites such as **13** in the bilayer does not contain information about the 2_1 screw axis, neither does the difference, $|F_{\mathcal{AA}} - F_{\mathcal{BB}}|$, between the molecular scattering factors for an $\mathcal{AA}$ pair and a $\mathcal{BB}$ pair. In some preliminary experiments it was found that simply using the individual molecular pairs as the basic units resulted in models which gave quite strong (and structured) scattering in the neighbourhood of the 010, 030 and 050 reciprocal lattice positions. This can be avoided by using as the basic unit of disorder two neighbouring pairs **13** and **24** shown by the linked coloured lozenge-shape envelopes in Fig. 1. The pairs **13** and **24** are related by the 2_1 screw if all four sites are occupied by $\mathcal{A}$ or all four are occupied by $\mathcal{B}$.

As a result of these two considerations, for the simulations described in this paper we simply use a set of random variables, $\sigma_{i,j,k}$, where each variable defines the occupants of the four sites of the screw-related pairs. As before, the indices i, j, k refer to the unit cell translations. $\sigma_{i,j,k} = -1$ corresponds to the case when all four molecular sites are occupied by component $\mathcal{A}$ and $\sigma_{i,j,k} = +1$ corresponds to the case when all four molecular sites are occupied by component $\mathcal{B}$. Such quartets of sites may be seen in Fig. 3. The quartet highlighted by the magenta-coloured motif illustrates a typical target site to be used in the Monte Carlo simulations described in §4.2. The four surrounding quartets, highlighted by the green coloured motifs, are neighbouring sites with which the target site is assumed to interact. For the remainder of the paper we use the symbols **A** to represent a quartet of sites all occupied by component $\mathcal{A}$ and **B** for a quartet of sites all occupied by component $\mathcal{B}$, *i.e.*

$$\mathbf{A} = \begin{matrix} \mathcal{A} & \mathcal{A} \\ \mathcal{A} & \mathcal{A} \end{matrix}; \quad \mathbf{B} = \begin{matrix} \mathcal{B} & \mathcal{B} \\ \mathcal{B} & \mathcal{B} \end{matrix}. \tag{1}$$

We will use the term 'component' interchangeably when referring to either a single site or to a quartet of sites.

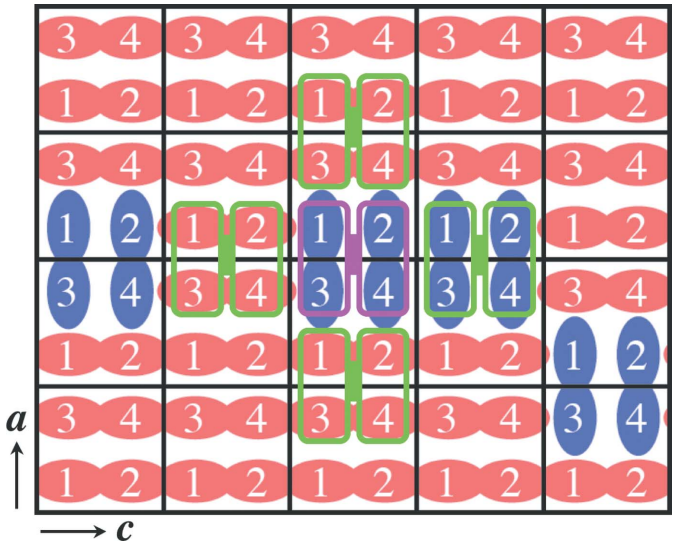


Figure 3
Schematic representation of the *ac* plane of the model structure. The monoclinic lattice shown in Fig. 1 is represented here as a square grid. The numbering of molecular sites within the unit cell corresponds to that used in Fig. 1 and Table 2. The figure shows the disposition of a typical target quartet of molecular sites (magenta motif) together with its four neighbouring quartets (green motifs) which are used in the MC ordering process. The two differently coloured and oriented types of ellipse correspond to the two different components $\mathcal{A}$ and $\mathcal{B}$ described in the text.

4. Monte Carlo simulations

4.1. MC simulation of molecular motions

The observed scattering patterns of form (II) of aspirin (Fig. 2) not only show the features that are directly attributable to the planar defects or stacking faults, but show extensive diffuse scattering that derives from the molecular motions. In numerous studies (*e.g.* Beasley *et al.*, 2008; Chan *et al.*, 2009) we have used a method of modelling such molecular motions by using a small number of suitably chosen harmonic (Hooke's law) springs which connect the molecules and provide 'effective' intermolecular interactions. With the advent and availability of increasingly fast computing it has recently become viable to include such springs on all interatomic atom–atom vectors shorter than a conveniently chosen upper limit, *e.g.* 4 Å. Moreover, a simple empirical formula, involving only three adjustable parameters, has been developed to specify spring constants and provide an objective means for limiting the number of springs that need to be used (Chan *et al.*, 2010).

For the present study such a harmonic spring model was used to fit to the form (I) diffuse scattering data and then the same parameter values were tranferred to the form (II) structure model.

4.2. MC ordering of components

In this section we describe how Monte Carlo simulation is used to create an occupancy distribution for the quartet sites described above. The aim is to create in the computer a model that mimics the effects that occur in the real crystal. In the case of aspirin it is readily apparent from the observed patterns that the key features are planar defects or stacking faults. In the real systems such planar defects may extend over many lattice spacings, whereas in our simulation techniques we are limited to a crystal size of typically $48 \times 48 \times 48$ unit cells. Even a plane comprising the maximum of 48×48 unit cells is rather small compared with what might occur in practice, but even more restrictive is the fact that the 48 unit cells in the third dimension allow very little scope to adequately represent

such things as stacking fault probabilities. Some compromises need to be made and although patterns resembling the observations can be simulated within the confines of this rather small simulation cell the differences in scale compared with the real system need to be borne in mind.

4.2.1. MC energy. Using the random variables $\sigma_{i,j,k}$, defined above, the MC energy used is given by the simple Ising model expression:

$$E_{\rm occ} = \sum_{i,j,k} \sigma_{i,j,k}\{J_x[\sigma_{i+1,j,k} + \sigma_{i-1,j,k}] + J_y[\sigma_{i,j+1,k} + \sigma_{i,j-1,k}] + J_z[\sigma_{i,j,k+1} + \sigma_{i,j,k-1}] + J_{x2}[\sigma_{i+2,j,k} + \sigma_{i-2,j,k}]\}. \quad (2)$$

Here J_x, J_y, J_z, J_{x2} are interaction parameters that are used to induce correlation along the three crystal directions. J_{x2} is an interaction between second-nearest neighbours in the x direction. These quantities are generally unknown but can be adjusted to induce the required correlation effects. We have generally found it most convenient to use an iterative feedback procedure in order to obtain particular target correlation values. Starting with initial arbitrary values for the J_i values the induced correlations are measured after each MC cycle. These are compared with the target values and then the J_i values are adjusted accordingly by an amount proportional to the discrepancy. For $\sigma = \pm 1$ variables the nearest-neighbour correlation coefficent is given by

$$C - x = \frac{P_{11} - m_{\rm B}^2}{m_{\rm B}(1 - m_{\rm B})}. \quad (3)$$

Here P_{11} is the joint probability, $P(\sigma_{i,j,k} = 1, \sigma_{i-1,j,k} = 1)$ and $m_{\rm B}$ is the concentration of component **B**. It should be noted that use of equations such as (2) to achieve distributions with predetermined correlation values can be problematical since it is well known that even the simplest nearest-neighbour Ising model undergoes a phase transition at a particular value of the interaction parameter J (Onsager, 1944). For the present purposes it should be considered simply as a means to an end in order to generate distributions that demonstrate particular effects.

4.2.2. Occupancy simulations. In the simulations described here it was also convenient to choose *a priori* values for the concentrations, $m_{\rm A}$ and $m_{\rm B}$, of the two components **A** and **B**. In order to achieve this $\sigma_{i,j,k}$ were first set up randomly so that any quartet of molecular sites had a probability $m_{\rm A}$ of containing **A** and $1 - m_{\rm A}$ of containing **B**. In order to maintain these fractions during the MC iteration, at each MC step the variables on two different randomly chosen sites were interchanged, with the change in the system energy being computed using both sites. 5000 cycles of iteration were carried out, by which time the values of J_i and the corresponding C_i had stabilized at their target values. Note that a cycle is defined as that number of individual MC steps required to visit each site once on average.

4.2.3. Different models. Two qualitatively different models have been investigated, although equation (2) has been used to obtain realisations of both of them. For Model 1 it is assumed that the concentrations of the two components **A** and **B** are equal, *i.e.* $m_{\rm A} = 0.5$. The planar nature of the disordered domains is induced by applying large positive values for the C_y and C_z correlations. Keeping these in-plane correlations constant the effect of different stacking sequences is investigated by varying the correlations along a normal to the planes by adjusting the first- and second-nearest neighbour interactions, J_x and J_{x2}.

For Model 2 it is assumed that the crystal is essentially composed of a single domain of one component (**A**) and that a relatively small number of planar (**B**) defects occur within it. These, as before, are induced by applying large positive values for the C_y and C_z correlations. In the examples shown in this paper a value of $m_{\rm B} = 0.1$ was used. With a concentration of this magnitude the highest negative correlation that is possible, which occurs when $P_{11} = 0$ in equation (3), is $C_x = -(0.1)^2 \div (0.1 \times 0.9) = -0.111$. By insisting that the nearest-neighbour correlation is -0.111 the possibility of neighbouring layers both containing **B** is zero. By adjusting the second-neighbour correlation small regions of **BABA**... ordering, corresponding to the aspirin form (I) structure, can be induced. It should be noted that a value of $m_{\rm B} = 0.1$ was chosen as a conveniently low value. With this concentration it was possible to produce realisations that were essentially a single domain of one component, but at the same time still provided a sufficient number of defects in the simulated crystal to allow the demonstration of the effects of defect clustering.

5. Results

5.1. MC simulation of molecular motions

Fig. 4 shows observed and calculated diffraction patterns for the $hk0$ sections of both forms (I) and (II) of aspirin. The calculated patterns for both of these were produced using the same spring parameters derived from a fit of the experimental diffuse scattering data of form (I). The model for form (I) was comprised of 36 different types of linear spring that had non-zero force constants. In addition four torsional springs were used on the dihedral angles that define the internal rotations around single bonds.

Spring force constants were defined in terms of the interatomic distances according to the equation (Chan *et al.*, 2010)

$$K_i = A \exp B(d_0 - R_{\rm vdw}) - C, \quad (4)$$

where A, B and C are the coefficients for an exponential decay curve, K_i is the force constant and $R_{\rm vdw}$ is the sum of the van der Waals radii for the two interacting atoms. d_0 is the equilibrium separation of a pair of atoms. Springs were not considered for the model beyond a threshold of $(d_0 - R_{\rm vdw}) + 0.3$ Å. The van der Waals radii were taken from Bondi (1964). It was found after consecutive trials that values of $A = 11$, $B = -0.4$, $C = 8$, $G_t = 25$ rad^{-1} gave calculated diffuse diffraction patterns for form (I) that best fitted the observed data. A single global torsional constant, G_t, was used for the four torsional springs. These values were very similar to

those refined for paracetamol (see Chan *et al.*, 2010, for further details of the method.

The model for form (II) was comprised of 46 different types of linear spring that had non-zero force constants for interactions between $\mathcal{A}$–$\mathcal{A}$ or $\mathcal{B}$–$\mathcal{B}$ components. In addition, there were another 40 types of spring required for interactions between $\mathcal{A}$–$\mathcal{B}$ or $\mathcal{B}$–$\mathcal{A}$ components. Of these only four were involved in the interaction between neighbouring layers in the *a* stacking direction, while the other 36 were involved in the relatively infrequently occurring interactions at the edges of the domains in the *bc* planes. The values of the former 46 $\mathcal{A}$–$\mathcal{A}$ and the latter four $\mathcal{A}$–$\mathcal{B}$ force constants were assigned using equation (4) with the same *A*, *B* and *C* coefficients as for form (I). However, the remaining 36 $\mathcal{A}$–$\mathcal{B}$ force constants involved in the relatively infrequently occurring interactions at the edges of the domains in the *bc* planes were assigned a single constant value.

It is clear from Fig. 4(*a*) that the spring model is able to reproduce very well the diffuse scattering pattern for form (I). However, while the same spring model gives a good description of many of the features of the form (II) pattern there are a number of features, the strongest of which is shown in the inset to Fig. 4(*b*), that are not reproduced at all. It should be noted that the model from which the calculated pattern of Fig. 4(*b*) was obtained had a strongly faulted structure, as described in §5.2. Fig. 4(*b*) shows that for a purely thermal model the *hk*0 section shows no sign of the faulting. For the defects in the structure to result in observable effects in this section it is necessary to introduce strain at the boundaries of the faulted regions. This is described in §5.4.

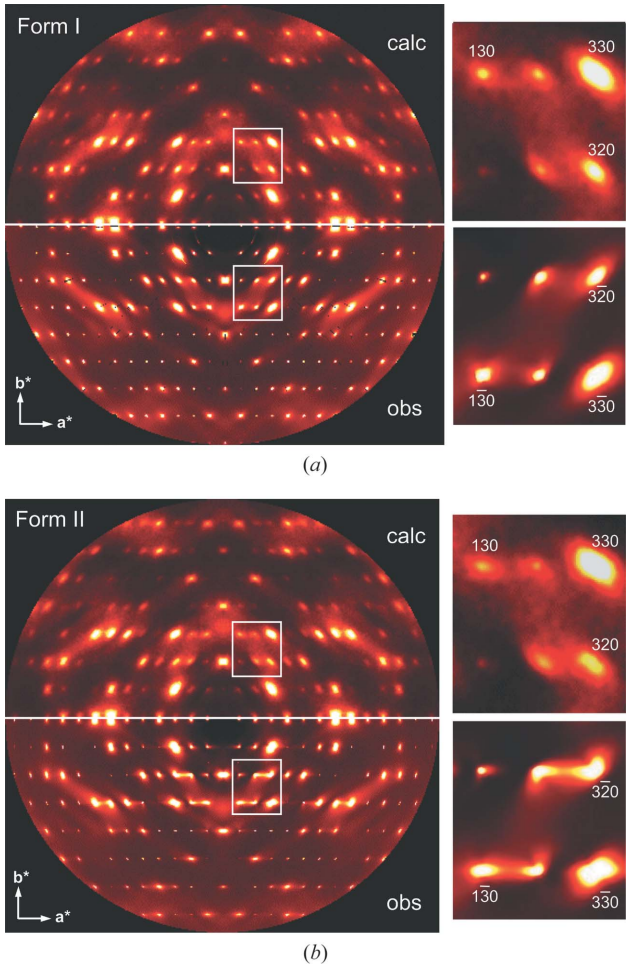


Figure 4
A comparison of observed and calculated *hk*0 intensities for the purely thermal simulations of (*a*) aspirin form (I) and (*b*) aspirin form (II). Inset are magnified images of the regions outlined by the white rectangles. See text for details.

5.2. Ordering models for the faulted form (II) structure

5.2.1. Model 1. Fig. 5 shows example realisations produced using equation (2). The central plots show a complete single layer from the three-dimensional crystal. In all five examples the values of the in-plane correlations were set at $C_y = C_z = 0.7$ and the concentrations of **A** (red) and **B** (blue) components both kept to 50%. The different examples show the variation from a high positive correlation, $C_x = 0.7$, to a high negative correlation, $C_x = -0.7$. Alongside each main figure is a strip showing a simple planar stacking sequence of the same correlation value. To the right of each figure is the diffraction pattern calculated for the region of reciprocal space corresponding to the inset rectangular region of the *h*1*l* data shown in Fig. 2(*a*).

These diffraction patterns have been calculated directly from the simulations. It has been our usual practice in making such calculations to subtract the average lattice so that only the diffuse scattering remains. In this case the average was not subtracted and so the patterns contain both the Bragg peaks and the diffuse scattering. The reason for doing this is to show that such disordered sequences of equal numbers of the **A** and **B** components do not give Bragg peaks on the *h*1*l*, l = odd rows. As the correlations approach their limiting values of ± 1 the diffuse streaks coalesce into sharp peaks at the form (I) position ($C_x = -1.0$) or the form (II) position ($C_x = +1.0$).

Although this shows that there is a continuous pathway between the form (I) structure and a (twinned) form (II) structure it also demonstrates that nowhere along this sequence is the model consistent with the observed pattern (Fig. 2*a*). Form (II) Bragg peaks are present on the *h*1*l*, l = odd rows in addition to the diffuse streaking.

5.2.2. Model 2. Fig. 6 shows three example realisations, also produced using equation (2), in which the concentration of component **B** was fixed at 0.1. For all three examples the in-plane correlations were $C_y = C_z = 0.9$. The nearest-neighbour correlation in the stacking direction was fixed at $C_x = -0.11$, which is the maximum negative correlation that can be obtained for $m_B = 0.1$. This ensures that there are no occurrences of neighbouring **BB** pairs of layers. The examples show that as the second neighbour correlation C_{x2} increases there is a tendency to form small clusters in which the **A** and **B** layers alternate. These correspond to small domains of the form (I) structure. Note that for Fig. 6(*a*) the diffuse streaking is continuous with little sign of an incipient form (I) peak, but for Fig. 6(*c*) the diffuse streak has largely coalesced into a form (I) peak mid-way between the form (II) Bragg peaks. The

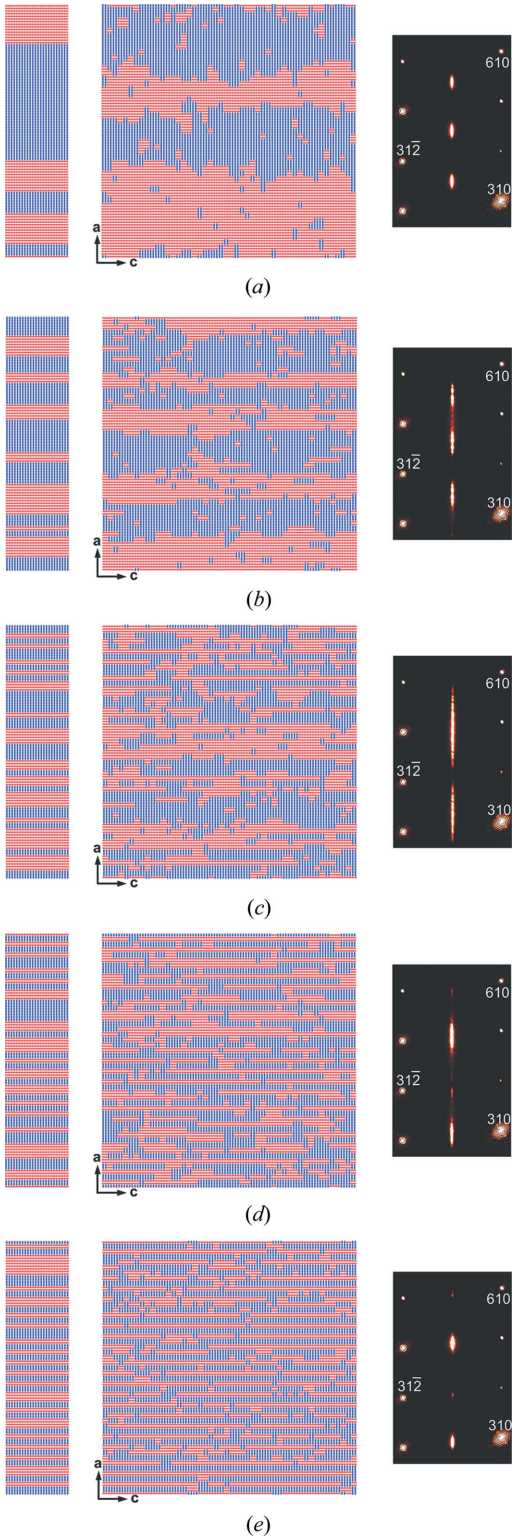


Figure 5
The distribution of the **A** (red) and **B** (blue) components in the *ac* plane for five Model 1 examples for which the occupancies were $m_A = m_B = 0.5$ and the in-plane correlations were $C_y = C_z = 0.7$. The correlations, C_x, in the layer-stacking direction were: (*a*) +0.7; (*b*) +0.35; (*c*) 0.0; (*d*) −0.35; (*e*) −0.7. On the right are corresponding calculated diffraction patterns. The region of reciprocal space shown corresponds to the inset rectangular region of the *h*1*l* data in Fig. 2(*a*). On the left are shown, for comparison, simple planar one-dimensional stacking sequences with the same correlation values.

intermediate example shows an incipient form (I) peak and this pattern qualitatively closely resembles the observed pattern shown in Fig. 2(*a*).

It should be noted that although both Model 1 and Model 2 examples are derived from the same equation (2) they are qualitatively quite different. The Model 1 examples shown in Fig. 5 are composed of an intimate mixture of both **A** and **B** components. In contrast, the Model 2 examples in Fig. 6 all contain a single domain of component **A** with a relatively low concentration of included defects. While the former model gives Bragg peaks on the *h*1*l*, $l =$ odd, rows only for the limiting end-members of the sequence, for the latter form (II) peaks are present throughout.

5.3. Further evidence supporting Model 2

Further evidence that the Model 2 type of ordering is a more viable model for describing the real aspirin (II) crystal is given by a comparision of the 0*kl* section of scattering for the two types of model. Fig. 7 shows a comparison of observed and calculated patterns for examples taken from the two series. Fig. 7(*a*) corresponds to the model shown in Fig. 5(*c*), while Fig. 7(*b*) corresponds to the model shown in Fig. 6(*b*).

While the outer regions of both patterns are quite similar to the observed data the very strong diffuse peaks indicated by the arrows in Fig. 7(*a*) represent a clear discrepancy from the observed data. For the Model 2 example these peaks are much reduced in intensity and in fact match the observed peaks quite well with the possible exception of the lowest-order peaks (close to the origin) which still appear to calculate with some excess intensity. It appears likely that this results from the fact that the concentration of defects (10%) in the simulation models is rather higher than in the actual aspirin (II) crystal.

5.4. Strain effects

All of the simulation results described above have been based on coordinates derived from Bragg scattering measurements of the average crystal structures for form (I) and form (II). These have been used to define molecular components $\mathcal{A}$ and $\mathcal{B}$ that define alternative positions and orientations for molecules in each molecular site. In addition models have been used:

(i) to describe the distribution of the $\mathcal{A}$ and $\mathcal{B}$ components over the sites of the crystal;

(ii) to describe atomic displacements owing to normal thermal motion.

Although together these have been able to account for most features of the observed diffraction patterns, the features observed in the *hk*0 section of form (II) (Fig. 4*b* inset) are still unaccounted for. It has been conjectured that these arise from local strain induced by the misfit of the $\mathcal{A}$ and $\mathcal{B}$ components at domain interfaces. In this section we describe some additional simulations that have been carried out to test this hypothesis.

The $\mathcal{A}$–$\mathcal{B}$ interface between adjacent domains in the *a* crystal direction is not expected to result in strain effects since

such an interface is part of the normal packing in aspirin form (I). However, $\mathcal{A}$–$\mathcal{B}$ interfaces in both the b and c directions are not part of the normal packing sequence for either form (I) or an ideal form (II) structure. Hence, such interfaces are potentially the source of strain that might give rise to the kind of features observed in Fig. 4(b). In fact, the interface in the c direction shows no unusual contact distances and the molecular packing is still quite good. However, simulations were carried out to test what effects any strain induced in this direction would have on the diffraction pattern. As a result of these tests it was concluded that there was no evidence in the observed patterns that could be attributed to strain at the $\mathcal{A}$–$\mathcal{B}$ interface in the c direction.

In the b direction the situation is quite different with some contact distances being anomalously short and others too long. Fig. 8 shows a molecular drawing of this interface. The arrows indicate the direction in which molecules might be assumed to have to move to alleviate the stress caused by the poor molecular packing at this interface.

In order to test this as a mechanism for the observed strain effects, some additional (artificial) interaction vectors were introduced into the MC model to allow such local displacements to be induced. These are indicated by the horizontal blue and orange dotted lines in Fig. 8. Along each of these vectors an additional *size-effect* term was added to the MC energy

$$E_{\rm size} = \sum_{\substack{\rm extra\\ \rm vectors}} K_{m,n}(d_{m,n} - d_0 - \zeta_{m,n})^2. \qquad (5)$$

Here $K_{m,n}$ is a force constant, $\zeta_{m,n}$ is a size-effect parameter, $d_{m,n}$ is the instantaneous length of the vector between atoms m and n, and d_0 is the length of the vector as defined by the equilibrium positions of the $\mathcal{A}$ and $\mathcal{B}$ components. A positive value of $\zeta_{m,n}$ will tend to push the molecules apart, while a negative value will tend to pull them together. For the examples discussed here $K_{m,n}$ was set to a value of ~ 2.5 times the average value of the springs used in the normal intermolecular interactions and the value of $|\zeta_{m,n}|$ used was 1 Å.

Three different cases have been tested. In the first example $\zeta_{m,n}$ was set to be positive for the blue vectors and negative for the orange vectors, thereby producing the pattern of shifts shown in Fig. 8. For the second example $\zeta_{m,n}$ was set to be

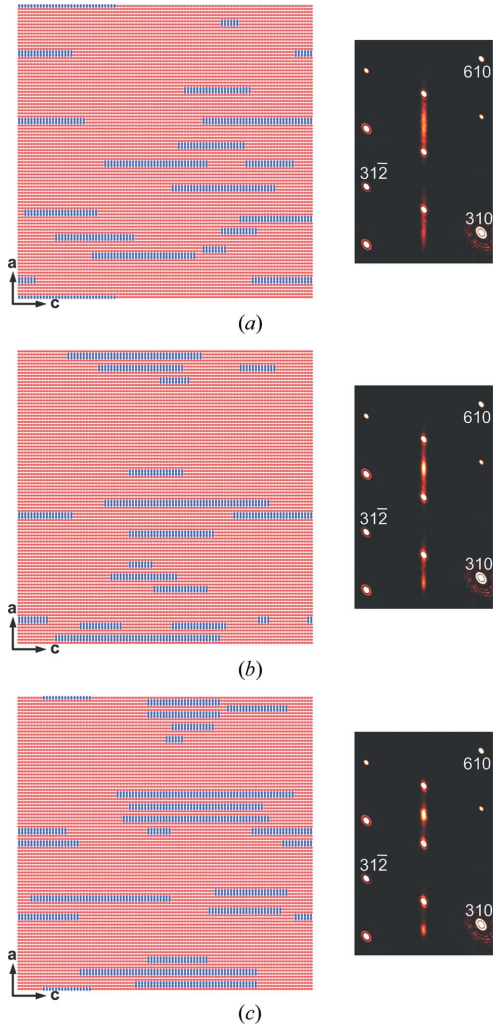


Figure 6
Three Model 2 examples obtained for different values of the second-neighbour correlation, C_{x2}. For all three examples $m_{\rm B} = 0.1$, $C_x = -0.11$ and $C_y = C_z = 0.9$. (*a*) $C_{x2} = 0.01$; (*b*) $C_{x2} = 0.2$; (*c*) $C_{x2} = 0.4$. Note that as C_{x2} increases small domains of the form (I) structure appear (alternating layers **ABAB** . . . *etc.*).

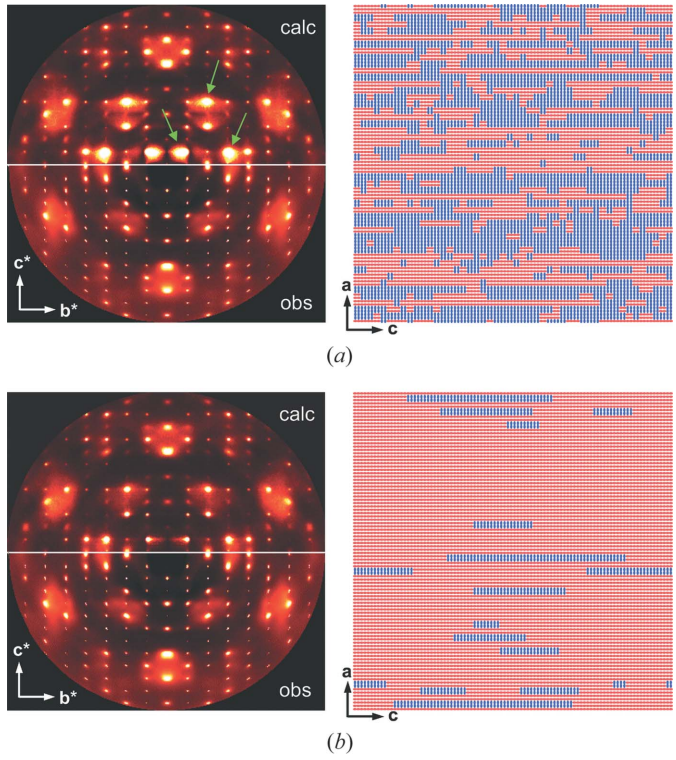


Figure 7
A comparison of calculated (top halves) and observed (lower halves) 0*kl* intensities for two different models. (*a*) Model has random layer stacking of components **A** (red) and **B** (blue). (*b*) Model is a single domain of **A** with relatively isolated **B** defects. Note that the distributions shown are of the *ac* plane, as in Fig. 5. Each horizontal line defect in (*b*) corresponds to a planar domain in *bc*.

positive for both the blue vectors and the orange vectors. For the third example $\zeta_{m,n}$ was set to be negative for both.

Fig. 9 shows the result for the second of these examples, *i.e.* one in which both blue and orange vectors increased in length, thereby tending to force apart the $\mathcal{A}$ and $\mathcal{B}$ domains. This example shows good qualitative agreement with the observed pattern. Not only have the diffuse bands between the Bragg peaks (in the inset enlargement) been induced but numerous other weaker examples also appear in the correct positions (indicated by the arrows). The asymmetric feature labelled 'P' is a particularly important indicator since it results directly from the fact that the $\mathcal{A}$ and $\mathcal{B}$ domains are being forced apart. For the third example in which the domains are pulled together the diffuse bands were similar, but the feature 'P' was bent upwards rather than downwards. For the first example in which the forces tended to induce the pattern of shifts shown in Fig. 8, the resulting diffuse scattering did not show any of these effects. Rather there was a tendency for an extra peak to be formed as a result of the rows of molecules seen in Fig. 8 being displaced alternately to the left and the right.

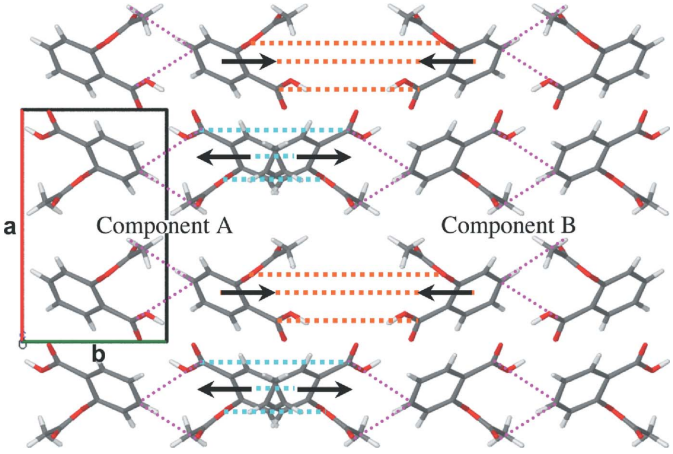


Figure 8
A plot of the $\mathcal{A}$–$\mathcal{B}$ interface between adjacent domains along the *b* crystal direction. The arrows have been inserted to indicate the direction in which molecules might have to move to alleviate the stress. The horizontal blue and orange dotted vectors indicate artificial interactions that were inserted to induce such movements. The pink dotted vectors represent some of the normal interactions within the domains.

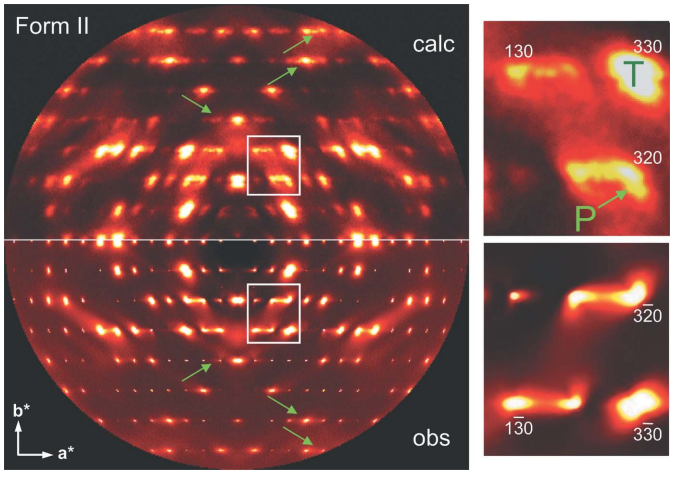


Figure 9
Comparison of observed and calculated *hk*0 intensities for a Model 2 simulation in which strain at the $\mathcal{A}$–$\mathcal{B}$ interface is taken into account. Inset are magnified images of the white outlined region exhibiting the strong scattering effects induced by the distortion. Arrows indicate similar (though weaker) features.

The example shown in Fig. 9 convincingly demonstrates that it is this kind of local distortion that is occurring in the real system. However, inducing the effect by the additional vectors shown in Fig. 8 is very artificial and this has had a detrimental effect on the normal thermally induced peaks such as that labelled 'T'. There is clearly much scope for attempting to improve the modelling of strain, but this would be very difficult to achieve until it becomes possible to work with bigger systems.

6. Conclusions

Full three-dimensional diffuse scattering data have been recorded for both polymorphic forms [(I) and (II)] of aspirin and these data have been analysed using MC computer modelling. Although the modelling was focused mainly on the key sections *hk*0, *h*1*l* and 0*kl*, once appropriate modelling parameters were obtained the calculated patterns for other sections were checked and also found to agree satisfactorily with the observed data.

The limited size of the computer models that it was possible to use imposed some limitations on how true a representation could be made of the aspirin II structure. Some aspects are better represented than others. The linear dimension of 48 unit cells imposes an upper limit on the lateral extent (in the *b* and *c* directions) of the planar defects. The width of the diffuse streaks is reciprocally related to this dimension. This width does appear qualitatively similar in the simulated patterns (Fig. 6*a*) to that of the observed streaks (Fig. 2*a*), so this aspect appears to be reasonably well modeled. Similarly the limit of 48 unit cells in the stacking direction (along **a**) limits the statistical accuracy that can be achieved in the faulting parameters. This affects the distribution of intensity along the streaks. Again, the models presented have been able to reproduce these effects quite convincingly. What is rather less satisfactory is that it was not possible to use defect concentrations lower than $\sim 10\%$, as for example in Fig. 6. This is significantly higher than the concentration envisioned to exist in the real crystals. This means that the diffuse streaking produced by the models, though having the same form, is too strong relative to the thermal diffuse scattering. The images showing the real-space distributions of components given in Fig. 6 should therefore be considered to be indicative only.

The observed scattering in form (I) is well reproduced by a simple harmonic model of thermally induced displacements. This model used the recently reported strategy for modelling diffuse scattering from molecular crystals (Chan *et al.*, 2010) in which a simple empirical formula [see (4)] is used to specify

the spring constants of the inter-atomic springs used to represent intermolecular interactions and to provide an objective means for limiting their number.

The data for form (II) exhibited, in addition to thermal diffuse scattering similar to that in form (I), diffuse streaks originating from stacking fault-like defects as well as other effects attributable to strain induced by these defects. An MC model has been developed which, within the limitations dictated by the finite simulation size currently feasible, satisfactorily accounts for all these effects. This model for form (II) incorporated a description of the thermal displacements carried over directly from the form (I) analysis and used the same parameter values.

The modelling of the defects used a description for the structure in which each molecular site contained either an $\mathcal{A}$ or $\mathcal{B}$ molecular component (see Table 2). A crystal comprising all $\mathcal{A}$ components corresponds to a perfect form (II) crystal, while one consisting of all $\mathcal{B}$ components corresponds to the same form (II) structure reflected in the *ac* plane and translated by $0.5 \times b$. For perfect *bc* planes of molecules the form (I) structure consists of a sequence along *a* corresponding to $\mathcal{AA}\,\mathcal{BB}\mathcal{AA}\,\mathcal{BB}\mathcal{AA}\ldots$ *etc*. Here the $\mathcal{AA}$ and $\mathcal{BB}$ pairs correspond to bilayers of strongly hydrogen-bonded molecular dimers. In the model described it is assumed that these molecular dimers remain intact throughout and faulting of the sequence can only occur at the interface between two such bilayers. In the model the extent of the order in the *bc* planes is induced by imposing correlations between components in the *b* and *c* directions, but although these were moderately large given the finite simulation size, they can only be considered as an approximation to the much longer range of order in the real crystals.

With this model it has been demonstrated that in principle a continuous sequence of different disordered structures can be generated from the perfect form (I) crystal at one extreme to a perfect form (II) crystal at the other. Through this progression the diffuse streaks that occur on the $h1l$, $l =$ odd, reciprocal rows change continuously from a sharp peak falling on the form (I) lattice at one extreme, through an intermediate continuous streak showing no peaking, to a sharp peak falling on the form (II) reciprocal lattice at the other.

The intermediate structures in this sequence are consistent with the description given by Bond *et al.* (2007*b*) that 'each aspirin crystal is an integral whole in which the domains are intimately connected with each other, with possibly many turnovers of domain within a single crystal'. However, we have shown that when the frequencies with which the two components $\mathcal{A}$ and $\mathcal{B}$ occur are equal there are no Bragg peaks on the $h1l$, $l =$ odd, reciprocal lattice rows except at the limiting extremes (when the streaking also disappears). Moreover, for intermediate disordered models the interference of the scattering from the intimately intermingled $\mathcal{A}$ and $\mathcal{B}$ domains gives strong diffuse peaks in the $0kl$ section and these are clearly not present in the observed scattering (see Fig. 7*a*).

These considerations led us to the conclusion that a crystal of form (II) is essentially comprised of a single domain of one component ($\mathcal{A}$, say) with isolated planar defects comprised of bilayers of the other component ($\mathcal{B}$). These defects may be essentially randomly occurring, in which case the diffuse streaks are continuous and show no peaking or they may cluster to form small regions of the form (I) structure as in Fig. 6(*b*). In this case a diffuse peak appears at the position of the form (I) reciprocal lattice on the $h1l$, $l =$ odd, reciprocal lattice rows. For this model, sharp form (II) Bragg peaks are present in the $h1l$, $l =$ odd, rows throughout. Such sharp Bragg peaks are clearly present in the data (Fig. 2*a*), and moreover the diffuse peaks in the $0kl$ section are now much reduced in intensity and are in much better agreement with the observed data (see Fig. 7*b*).

Finally it has been shown that various diffuse scattering features that occur in the $hk0$ section of the form (II) data that are absent from the form (I) data (see Fig. 2*c*) are due to strain induced by the presence of the defects. The results of modelling these effects have shown that they arise from the mismatch of the packing of the $\mathcal{A}$ and $\mathcal{B}$ components at the edges of the planar defects in the *b* direction. The sign of the size-effect distortion indicates that the spacing of an $\mathcal{A}$–$\mathcal{B}$ interface is increased relative to the normal $\mathcal{A}$–$\mathcal{A}$ or $\mathcal{B}$–$\mathcal{B}$ spacing. No evidence for similar mismatching in the *c* direction was found. With the current limitation on the size of the model system that we can use it is not possible to represent the domain structure completely realistically. Consequently, it is difficult to be sure of the lateral extent of the planar faults and of the overall volume fraction of the faulted material (*i.e.* the concentration of component $\mathcal{B}$). However, from the present study it is clear that strain effects are highly significant and since they only occur at the *b* edges of the faulted domains this implies that these must occur reasonably frequently in the real material.

Ouvrard & Price (2004) noted that aspirin form (II) should have 'a low shear elastic constant, implying that it is so readily deformed that there may be problems in its growth'. This same point was picked up by Bond *et al.* (2007*b*). The sample used for the present diffuse scattering study was a good quality prismatic crystal with well formed facets and edges with no outward signs of any problems in its growth. It was also significantly bigger than samples normally used for Bragg structure analysis. We consider that what the present study has revealed is that the kind of planar defects described are intrinsic to the form (II) stucture. It may well be that these arise as a result of 'a low shear elastic constant', but there is clear evidence that their inclusion results in significant strain to the form (II) lattice. This implies that there must be a balance. The defects may be easy to form but the strain that is induced results in a significant energy penalty that limits their number.

Bond *et al.* (2007*b*) have also been concerned with trying to assess the composition of the crystals in terms of the relative fractions of forms (I) and (II). This is not completely straightforward. The example shown in Fig. 6(*b*), which we consider to be qualitativly a good representation of what occurs in the real structure, has 10% of **B** bilayer defects. Since what comprises the form (I) structure is an alternating sequence of **BABA**... it might be considered that 10% of **B**

bilayer defects would correspond to 20% of form (I). However, a good fraction of the defects are composed of single isolated **B** bilayers. For these the neighbouring **A** bilayers are still part of the form (II) structure so cannot be counted as part of the form (I) structure.

The present results are based on data derived from a single aspirin (II) crystal grown *via* super-cooling a 0.6–0.7 *M* solution of aspirin in acetonitrile. Although no other crystals were examined in the same amount of detail, all of the samples that were surveyed showed similar streaking along ($h1l$, l = odd) rows, with the form (II) peaks prominent and with a relatively minor presence of more diffuse form (I) peaks. Whether form (II) crystals with different degrees of faulting can be made by controlling the growth conditions remains open to conjecture.

In summary, the present study has provided strong evidence that the aspirin form (II) structure is a true polymorph with a structure quite distinct from that of form (I). The diffuse scattering evidence we have presented shows that crystals are essentially composed of large single domains of the form (II) lattice with a relatively small volume fraction of intrinsic planar defects or faults comprising misoriented bilayers of molecular dimers. There is evidence of some local aggregation of these defect bilayers to form small included regions of the form (I) structure. Evidence has also been presented that shows that strain effects arise from the mismatch of molecular packing between the defect region and the surrounding form (II) lattice. This occurs at the edges of the planar defects only in the *b* direction.

Diffuse scattering data for aspirin form (I) were collected on the 11-ID-B beamline at the Advanced Photon Source, Argonne, IL, USA. Diffuse scattering data for aspirin form (II) were collected on the powder diffraction beamline at the Australian Synchrotron, Victoria, Australia. The support of the Australian Research Council, the Australian Synchrotron Research Program and the NCI National Facility at the ANU is gratefully acknowledged. DJG gratefully acknowledges support of the Australian Institute of Nuclear Science and Engineering. We would also like to thank Dr Peter Chupas and Dr Karena Chapman of the Advanced Photon Source, and Dr Kia Wallwork from the powder diffraction beamline of the Australian Synchrotron for assistance with the collection of the diffuse scattering data. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357.

References

Beasley, A. G., Welberry, T. R., Goossens, D. J. & Heerdegen, A. P. (2008). *Acta Cryst.* B**64**, 633–643.
Bond, A. D., Boese, R. & Desiraju, G. R. (2007*a*). *Angew. Chem. Int. Ed.* **46**, 615–617.
Bond, A. D., Boese, R. & Desiraju, G. R. (2007*b*). *Angew. Chem. Int. Ed.* **46**, 618–622.
Bondi, A. (1964). *J. Phys. Chem.* **68**, 441–451.
Chan, E. J. & Welberry, T. R. (2010). *Acta Cryst.* B**66**, 260–270.
Chan, E. J., Welberry, T. R., Goossens, D. J. & Heerdegen, A. P. (2010). *J. Appl. Cryst.* **43**, 913–915.
Chan, E. J., Welberry, T. R., Goossens, D. J., Heerdegen, A. P., Beasley, A. G. & Chupas, P. J. (2009). *Acta Cryst.* B**65**, 382–392.
Onsager, L. (1944). *Phys. Rev.* **65**, 117–149.
Ouvrard, C. & Price, S. L. (2004). *Cryst. Growth Des.* **4**, 1119–1127.
Paściak, M., Wołcyrz, M., Pietraszko, A. & Leoni, S. (2010). *Phys. Rev. B*, **81**, 014107.
Vishweshwar, P., McMahon, J., Oliveira, M., Peterson, M. L. & Zaworotko, M. J. (2005). *J. Am. Chem. Soc.* **127**, 16802–16803.
Weber, T. & Bürgi, H.-B. (2002). *Acta Cryst.* A**58**, 526–540.
Weber, T., Estermann, M. A. & Bürgi, H.-B. (2001). *Acta Cryst.* B**57**, 579–590.
Welberry, T. R. (1985). *Rep. Prog. Phys.* **48**, 1543–1593.
Welberry, T. R. (2004). *Diffuse X-ray Scattering and Models of Disorder.* IUCr Monographs on Crystallography. Oxford University Press.
Welberry, T. R. & Butler, B. D. (1994). *J. Appl. Cryst.* **27**, 205–231.
Welberry, T. R., Goossens, D. J., Heerdegen, A. P. & Lee, P. L. (2005). *Z. Kristallogr.* **222**, 1052–1058.
Wilson, C. C. (2002). *New. J. Chem.* **26**, 1733–1739.