Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

Analysis of an industrial production suspension of *Bacillus lentus* subtilisin crystals by powder diffraction: a powerful quality-control tool

Christian G. Frankaer,[a]‡ Olga V. Moroz,[b]‡ Johan P. Turkenburg,[b] Stein I. Aspmo,[c] Majbritt Thymark,[c] Esben P. Friis,[c] Kenny Stahl,[a] Jens E. Nielsen,[c] Keith S. Wilson[b]* and Pernille Harris[a]*

[a]Department of Chemistry, Technical University of Denmark, Kemitorvet 207, DK-2800 Kgs. Lyngby, Denmark, [b]Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York YO10 5DD, England, and [c]Novozymes A/S, Krogshøjvej 36, DK-2880 Bagsvaerd, Denmark

‡ Christian G. Frankaer and Olga V. Moroz contributed equally to this work.

Correspondence e-mail: keith.wilson@york.ac.uk, ph@kemi.dtu.dk

Received 2 December 2013
Accepted 21 January 2014

PDB references: *Bacillus lentus* subtilisin, 4cfy; 4cfz; 4cg0

A microcrystalline suspension of *Bacillus lentus* subtilisin (Savinase) produced during industrial large-scale production was analysed by X-ray powder diffraction (XRPD) and X-ray single-crystal diffraction (MX). XRPD established that the bulk microcrystal sample representative of the entire production suspension corresponded to space group $P2_12_12_1$, with unit-cell parameters a = 47.65, b = 62.43, c = 75.74 Å, equivalent to those for a known orthorhombic crystal form (PDB entry 1ndq). MX using synchrotron beamlines at the Diamond Light Source with beam dimensions of 20 × 20 µm was subsequently used to study the largest crystals present in the suspension, with diffraction data being collected from two single crystals (∼20 × 20 × 60 µm) to resolutions of 1.40 and 1.57 Å, respectively. Both structures also belonged to space group $P2_12_12_1$, but were quite distinct from the dominant form identified by XRPD, with unit-cell parameters a = 53.04, b = 57.55, c = 71.37 Å and a = 52.72, b = 57.13, c = 65.86 Å, respectively, and refined to R = 10.8% and $R_{\rm free}$ = 15.5% and to R = 14.1% and $R_{\rm free}$ = 18.0%, respectively. They are also different from any of the forms previously reported in the PDB. A controlled crystallization experiment with a highly purified Savinase sample allowed the growth of single crystals of the form identified by XRPD; their structure was solved and refined to a resolution of 1.17 Å with an R of 9.2% and an $R_{\rm free}$ of 11.8%. Thus, there are at least three polymorphs present in the production suspension, albeit with the 1ndq-like microcrystals predominating. It is shown how the two techniques can provide invaluable and complementary information for such a production suspension and it is proposed that XRPD provides an excellent quality-control tool for such suspensions.

1. Introduction

Single-crystal macromolecular crystallography (MX) is well established as the prime technique for determining the three-dimensional structure of proteins and is responsible for close to 90% of the structures deposited in the Protein Data Bank (PDB). The rapid growth of its success depended on developments in two key areas, firstly gene technologies allowing the cloning and overexpression of target proteins and secondly highly efficient systems for recording diffraction data, especially synchrotron sources and two-dimensional detectors. These were accompanied in the last decade by high-throughput robotic approaches for crystallization screening, in part arising from structural genomics projects. Recent developments in synchrotron sources, in particular the use of microfocus beams, have resulted in structures being solved from very small single crystals. In one example, a 1 µm beam at the ESRF ID13 beamline was used to solve the structure of

xylanase II at 1.5 Å resolution from a crystal with a smallest dimension of 20 µm (Moukhametzianov *et al.*, 2008). In a second example, the Diamond Light Source microfocus I24 beamline was used to record data on the structure of CPV15 polyhedrin (space group *I*23; a = 103 Å), with a final data set being assembled from about 40 crystals, each of about 1 µm (Wagner *et al.*, 2013). Obtaining meaningful diffraction data from crystals down to 1 µm is now on the horizon, with a number of synchrotron facilities targeting such crystals with the next generation of microfocus beamlines.

In spite of these advances, single microcrystals of less than 1 µm in the maximum dimension pose a formidable challenge using MX and remain too small for single-crystal analysis, although we note that the advent of free-electron laser crystallography has the potential to revolutionize this field in the near future (Chapman *et al.*, 2011). One solution to the analysis of submicrometre crystals is the use of X-ray powder diffraction (XRPD). This technique is widely used for small-molecule structure analysis, both with synchrotron radiation as well as home laboratory X-ray sources. The pioneering work on protein powder diffraction (von Dreele, 1999; von Dreele *et al.*, 2000) established the potential of structure solution and refinement of small protein structures from high-resolution synchrotron experiments. Recent years have seen further developments of protein XRPD using either synchrotron radiation (Margiolaki & Wright, 2008; Margiolaki *et al.*, 2013) or home laboratory X-ray sources (Hartmann *et al.*, 2010).

Using a home laboratory source, XRPD offers a method of characterizing a crystalline protein suspension (Hartmann *et al.*, 2011; Ståhl *et al.*, 2013). For well diffracting samples, XRPD data can be collected within 30 min. XRPD is especially useful for identifying which single-crystal components from a known library of structures are present in a powder or crystalline suspension. This is a much easier task than *ab initio* structure solution, which suffers from a relative paucity of data, highlighting in part the substantial overlap of reflections in the scattered images. This approach does of course require that a known single-crystal form or set of alternative forms are known with which to fit the observed XRPD pattern.

It is precisely this approach which we apply here. As the efficiency of industrial production techniques continues to grow, enzymes and proteins are produced and handled at higher concentrations. However, the efficiency of production often leads to a new problem: the precipitation of proteins as crystals in the production pipeline, which can in turn cause handling and yield problems for the processes with crystalline proteins and potentially poses problems of re-solubilization. The sizes and morphologies of the crystals present in such production suspensions varies considerably, with the samples often being largely composed of very small crystallites of less than 1 µm in the maximum dimension. Knowledge of the three-dimensional structures of the microcrystals in such samples is of considerable interest for further optimization of production.

However, structure solution poses a real challenge for the reasons given above. Here, we report how the application of a combined XRPD and MX approach demonstrates the presence of three different microcrystalline forms in a suspension from a large-scale industrial production of the widely used *Bacillus lentus* subtilisin (Savinase). Savinase is widely used as a protease in the detergent industry, and a considerable number of three-dimensional structures of the wild-type enzyme have been deposited in the PDB, together with those of mutants and complexes. The enzyme is a member of the subtilisin family, with two metal ions contributing to its structure and activity. We show how MX microdiffraction and XRPD can complement one another for studying such samples containing small crystals as well as microcrystals. XRPD provided the primary information, clearly identifying a dominant crystal form in the production suspension. This was complemented by a MX analysis of some larger single crystals present in the sample, which proved to be in different crystal forms from that identified as dominant by XRPD, suggesting the presence of significant polymorphism in the sample. The ability to quickly characterize the major crystal forms encountered during the recovery process using XRPD has the potential to greatly help in troubleshooting the production process and to provide valuable information for further refining the enzyme/protein manufacturing.

2. Experimental

2.1. Savinase samples

The crystal suspension from the industrial enzyme production and the highly purified Savinase solution used for the controlled crystallization experiment were donated by Novozymes A/S. In brief, the Savinase was produced by a *Bacillus* host and the suspension was treated with 3% benzalkonium chloride for 48 h to sterilize it. The bulk of the biomass was removed by centrifugation and the remaining suspension is a complex mixture of Savinase, metabolites, salts and biomass remnants.

2.2. X-ray powder diffraction

The crystallites in the production suspension were recovered by filtration using an Ultrafree-MC centrifuge filter (Amicon/Millipore) with a pore size of 0.22 µm. Approximately 5 µl of wetted protein powder samples were mounted in a specially designed sample holder for XRPD (Frankaer *et al.*, 2011), in which the sample was prevented from drying out. An X-ray powder pattern was recorded at room temperature for 2 h on a Huber G670 diffractometer using Cu $K\alpha_1$ radiation (λ = 1.5406 Å). The signal-to-noise ratio was further enhanced by extending the exposure time to 16 h. No decay in intensity was observed during data collection for this system.

The subsequent analysis of the XRPD data included an identification of the crystal form followed by verification in which optimized unit-cell parameters are extracted (Ståhl *et al.*, 2013). For identification of the crystal form, powder diffraction patterns were computed from coordinate files using *PROTPOW* (Hartmann *et al.*, 2010). The calculated patterns include the bulk-solvent contribution using average values of the solvent parameters k_{sol} and B_{sol} of 0.35 e Å^{-3} and 46 Å^2

Table 1
Crystallographic statistics.

Crystal form	A1	A2	1ndq-like
Data collection			
PDB code	4cg0	4cfz	4cfy
Beamline	Diamond I03	Diamond I04	Diamond I04
Wavelength (Å)	0.976	0.980	0.979
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (Å)	a = 53.04, b = 57.55, c = 71.37	a = 52.72, b = 57.13, c = 65.86	a = 47.23, b = 61.82, c = 75.25
Resolution range (Å)	17.8–1.36	23.94–1.57	23.88–1.17
No. of reflections	279430	187599	326612
Unique reflections	46289	28009	58833
Monomers in asymmetric unit	1	1	1
Completeness† (%)	97.3 (78.3)	98.4 (86.1)	78.2 (14.8)
$\langle I/\sigma(I)\rangle$†	14.9 (2.2)	16.0 (2.3)	39.8 (5.1)
$CC_{1/2}$†	0.999 (0.609)	0.998 (0.749)	0.998 (0.916)
Average multiplicity†	6.0 (4.1)	6.7 (4.3)	5.8 (1.2)
R_{merge}†‡ (%)	6.8 (67.0)	8.4 (52.8)	4.2 (15.2)
Refinement statistics			
Percentage of R_{free} reflections (%)	5.07	5.03	5.06
R_{cryst}§ (%)	10.8	14.1	9.2
R_{free} (%)	15.5	18.0	11.8
R.m.s. deviations from ideal geometry¶			
Bond distances (Å)	0.018 (0.019)	0.018 (0.019)	0.019 (0.019)
Bond angles (°)	1.719 (1.933)	1.866 (1.948)	1.807 (1.934)
Chiral centres (Å^3)	0.117 (0.200)	0.118 (0.200)	0.122 (0.200)
Planar groups (Å)	0.010 (0.021)	0.010 (0.021)	0.009 (0.021)
Average main-chain B (Å^2)	11.7	9.1	7.7
Average side chain B (Å^2)	15.1	10.8	10.5
Ramachandran plot			
Preferred regions (%)	95.3	95.4	94.5
Allowed regions (%)	3.4	3.8	5.1
Outliers (%)	1.3	0.8	0.4

† Values in parentheses are for the highest resolution shell. ‡ R_{merge} is defined as $100 \times \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I(hkl)$ is the intensity of the reflection. § $R_{cryst} = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$. ¶ R.m.s. deviations from ideal geometry (target values are given in parentheses).

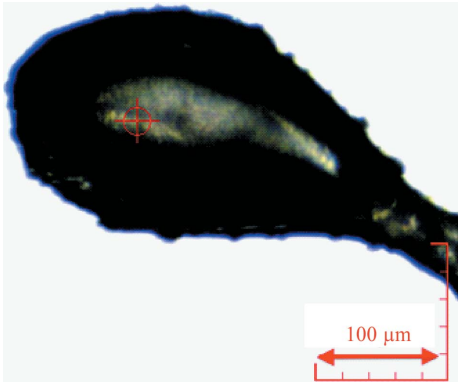

Figure 1
Crystal A1 in the loop at the Diamond Light Source. The red cross indicates the position from which the best diffraction was obtained.

(Fokine & Urzhumtsev, 2002), respectively, for medium-resolution structures. Asymmetric peak profile parameters were optimized for the experimental setup. The identification procedure further involves an approximate background subtraction of the experimental pattern.

In the verification step, unit-cell parameters were determined from a pattern-profile fit using *PROTPOW* (Hartmann *et al.*, 2010), in which a solvent-corrected pattern I_{calc} is calculated from the PDB coordinates and fitted to the experimental pattern I_{obs}. During the fit, the calculated relative intensities were fixed [individual peak fitting by, for example, the Le Bail (Le Bail *et al.*, 1988) and Pawley methods (Pawley, 1981) is not possible owing to the large overlap of peaks]. In addition to the unit-cell parameters, this procedure involves optimization of the k_{sol} parameter and the background. The fit is evaluated by the profile R factor given as

$$R_p = \frac{\sum |I_{obs} - I_{calc}|}{\sum |I_{obs}|}. \qquad (1)$$

2.3. Single-crystal diffraction of microcrystals from the production suspension

In addition to the microcrystalline material, the wild-type Savinase production suspension contained some larger needle-like crystals with dimensions of ~20 × 20 × 60 µm (still very small by normal standards). A single crystal (A1) was mounted in a diffraction loop and vitrified in Paratone oil without additional cryoprotectant, and data were collected on beamline I03 of the Diamond Light Source with a beam size of 20 × 20 µm. Initial experiments showed only weak diffraction with a high background. However, use of the grid-screening approach (Aishima *et al.*, 2010), *i.e.* stepping across the sample and illuminating only a small area of the loop at each grid point (Fig. 1), provided an impressive improvement in diffraction quality. Data were collected to a resolution of 1.4 Å and the structure was solved in space group $P2_12_12_1$, with unit-cell parameters a = 53.04, b = 57.55, c = 71.37 Å. For this and the two following structures, X-ray data were processed using programs from the *CCP*4 suite (Winn *et al.*, 2011). The images were processed using *xia*2 (Winter, 2010). The structure was solved by *MOLREP* (Vagin & Teplyakov, 2010), using the structure of native Savinase (PDB entry 1svn; Betzel *et al.*, 1992) as a search model. The structure was refined by *REFMAC* (Murshudov *et al.*, 2011) iterated with manual model building/correction in *Coot* (Emsley *et al.*, 2010), to give an R and R_{free} of 10.8 and 15.5%, respectively. Processing and refinement statistics for this and following structures are given in Table 1.

A second crystal (A2) was mounted analogously to A1 and data were collected on the Diamond Light Source beamline I04 to 1.57 Å resolution, again using a grid scan. This structure also belonged to space group $P2_12_12_1$, but with different unit-cell parameters a = 52.72, b = 57.13, c = 65.86 Å. The structure

was also solved starting from PDB entry 1svn, and was refined to an R and R_{free} of 14.1 and 18.0%, respectively.

2.4. Controlled growth of single crystals in the 1ndq-like crystal form

Following the identification of the 1ndq-like crystal form as the dominant form from XRPD experiments, it was decided to reproduce these crystals on a larger scale to confirm the ability of the present Savinase material to crystallize in this form, although this started from highly purified material free of any microcrystals as seen in the suspension. The highly purified Savinase was buffered in 50 m*M* H_3BO_3, 5 m*M* dimethylglutaric acid, 1 m*M* $CaCl_2$, 100 m*M* NaCl pH 6.0. The conditions reported by Betzel *et al.* (1988) were used. Crystals were grown in 15 µl hanging drops consisting of 20 mg ml^{-1} protein, 4% PEG 4000, 0.33 *M* NaCl, 1.5 m*M* $CaCl_2$, 18 m*M* citrate buffer pH 6.0. The drops were equilibrated against 1 ml reservoirs containing 10% PEG 4000, 1 *M* NaCl, 5 m*M* $CaCl_2$, 50 m*M* citrate buffer pH 6.0. The crystals grew to dimensions of 0.5 × 0.5 × 1.5 mm in 4–7 d.

Data were collected to 1.17 Å spacing on Diamond Light Source beamline I04; for these large crystals a grid scan was not required. The structure also belonged to space group $P2_12_12_1$, this time with unit-cell parameters $a = 47.23$, $b = 61.82$, $c = 75.25$ Å, similar to PDB entry 1ndq as expected. The structure was solved starting from PDB entry 1svn, as for the A1 and A2 crystals, and refined to an R and R_{free} of 9.2 and 11.8%, respectively.

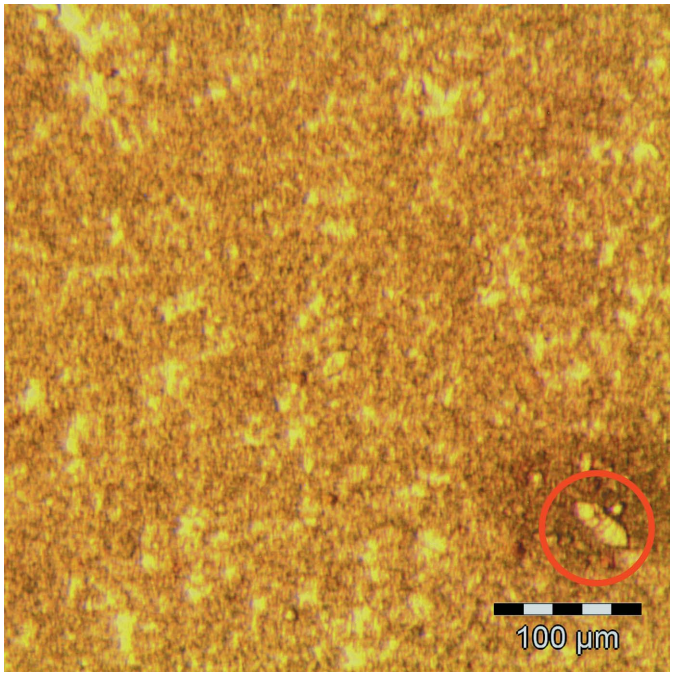

Figure 2
Microscope photograph of Savinase microcrystals in the production suspension. The microcrystals (dark areas) constitute the majority of the sample. A few larger single crystals (maximum dimension of about 60 µm) similar to the crystal used for single-crystal diffraction were observed (red circle).

2.5. Crystal contact analysis

The crystal interfaces were identified and analysed using the *PISA* server (Krissinel & Henrick, 2007). Before the analysis, the amino-acid sequences in the structures from the PDB were renumbered according to that used in the crystal structures reported in this work. Each crystal contact is characterized by the number of interfacing atoms (N_{at}) and residues (N_{res}) from the identity molecule i and its neighbour molecule n, respectively. Firstly, the interfaces are identified by rolling a water probe of diameter 1.4 Å over the surface of each monomer in the unit cell. Atoms/residues that have been in contact with the probe for both the identity molecule i and the neighbour molecule n are assigned to belong to the crystal interface between these molecules, and the accessible surface area (ASA_{in}) of the interface is calculated as

$$\begin{aligned} ASA_{in} &= \frac{ASA_i + ASA_n - ASA_{dimer}}{2} \\ &= \frac{2ASA_{monomer} - ASA_{dimer}}{2}, \end{aligned} \qquad (2)$$

where ASA_{dimer} is the accessible surface area of the dimer formed by molecules i and n. Since molecule i is equivalent to molecule n, this reduces to $2ASA_{monomer}$.

In the analysis reported here, the ASA for each patch is given in Å^2 and as percentage of the ASA of the entire protein ($ASA_{monomer}$). Each interface is further characterized by a solvation free-energy gain upon formation of the interface (ΔG_{solv}), which is calculated as described by Krissinel & Henrick (2007) and includes the areas and solvation energies for each atom type involved in the interface, the number of hydrogen bonds, salt bridges and disulfide bonds, as well as the entropy change ΔS upon complex dissociation. The number of potential hydrogen bonds across the interface (N_{HB}) is limited within a donor–acceptor distance of 3.89 Å. Salt bridges are defined within 4.0 Å between non-H atoms. Furthermore, the presence and number of bridging sulfate ions μ-SO_4 between the molecules are reported.

3. Results

3.1. Inspection of the samples by microscopy

A microscope photograph of the production suspension is shown in Fig. 2, in which the dark areas correspond to the microcrystalline powder of less than 1 µm, which constitutes the bulk of the sample used for the powder diffraction experiment. Among the microcrystals there are a few larger crystals with dimensions of ∼20 × 20 × 60 µm, two of which were fished out and mounted for the single-crystal diffraction experiments. Only about one in ten photographs taken at random locations of the sample contained visible larger fragments similar to that highlighted in Fig. 2.

3.2. Structures of the single crystals from the production suspension

The structure of Savinase in the A1 and A2 crystals is very similar to that published for the wild-type enzyme, with

r.m.s.d.s of 0.24 and 0.42 Å, respectively, for all C^{α} atoms compared with PDB entry 1svn. Cation-binding sites have been identified in *Bacillus* subtilisins: a high-affinity site which binds calcium and a lower affinity pair of sites which can bind calcium or a monovalent cation such as sodium in a mutually exclusive manner (Pantoliano *et al.*, 1988). Both the A1 and A2 structures clearly contain a calcium coordinated to seven O atoms with distances around 2.4 Å at the strong binding site, and a sodium coordinated to five O atoms with distances around 2.3 Å at the second metal site. The metal identities are strongly supported by the height of the peaks in the Fourier syntheses and the ligand-coordination distances and geometry. In the A2 crystal structure there are six sulfate ions bound, all located within the first solvation shell around the Savinase molecule and involved in contacts between adjacent molecules in the lattice. No sulfate ions were observed in the structure of A1.

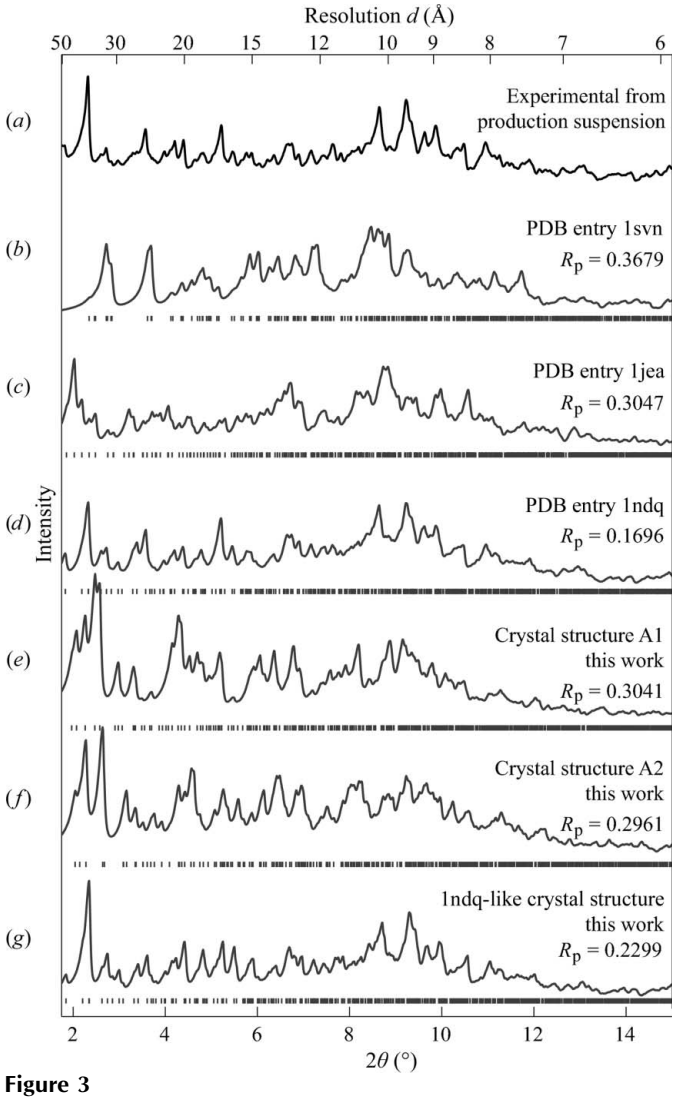

Figure 3
(*a*) Experimental powder diffraction pattern of the production suspension. (*b–g*) Patterns for the known crystal forms of *B. lentus* subtilisin calculated from (*b–d*) native Savinase structures deposited in the Protein Data Bank or (*e*), (*f*) and (*g*) the crystal structures A1, A2 and the 1ndq-like form reported here, respectively. Average bulk-solvent parameters (k_{sol} = 0.35 e Å^{-3} and B_{sol} = 46 Å^2) were used in the solvent correction of all calculated patterns except for the high-resolution structure (*g*), where k_{sol} = 0.37 e Å^{-3} was used. Theoretical positions of the Bragg peaks and profile *R* values for the fit with the experimental pattern are included for all calculated patterns. The crystal form present in the bulk of the sample corresponds to that computed for PDB entry 1ndq (Pan *et al.*, 2003).

3.3. Structure of the 1ndq-like crystal

This crystal form was first reported by Betzel *et al.* (1988) and a three-dimensional structure (PDB entry 1ndq) was solved at 1.8 Å resolution by Pan *et al.* (2003). The present structure of Savinase in the 1ndq-like crystal form is, as expected, very similar to others published for the wild-type enzyme, with r.m.s.d.s of 0.25 Å with A1, 0.44 Å with A2 and, not surprisingly, the smallest, 0.19 Å with PDB entry 1ndq. The metal content and coordination is the same as for the A1 and A2 forms, with a calcium in the strong binding site and a sodium at the second site, closer to the substrate-binding region.

3.4. Powder diffraction

The powder diffraction pattern of the production sample is shown in Fig. 3(*a*), together with patterns calculated from native *B. lentus* subtilisin crystal structures deposited in the PDB (Figs. 3*b*, 3*c* and 3*d*), as well as the single-crystal A1 and A2 structures solved from the crystals taken from exactly the same production sample as that used for powder diffraction (Figs. 3*e* and 3*f*). The calculated patterns are clearly distinguishable from one another and serve as unique fingerprints of each crystal form. The dominant crystal form in the powder sample (Fig. 3*a*) can be clearly identified as that in PDB entry 1ndq (R_p = 0.1696; Pan *et al.*, 2003) and reported previously by Betzel *et al.* (1988), although the presence of minor amounts of other forms lies below the sensitivity of this analysis. Another powder pattern was calculated from the 1ndq-like crystal structure solved in this work (Fig. 3*g*). This pattern is highly similar to that calculated for 1ndq (Fig. 3*d*) and therefore also matches the experimental pattern (R_p = 0.2299). However, it was calculated using a k_{sol} of 0.37 e Å^{-3} to compensate for the large number of ordered water molecules in the high-resolution structure reported here (362 in total) compared with 1ndq, which only contain 90 waters.

Crystal forms A1 and A2 are different from all the *B. lentus* subtilisin structures in the PDB, and the experimental powder pattern does not match the patterns calculated from either of these two forms (Figs. 3*e* and 3*f*). In particular, the experimental pattern (Fig. 3*a*) lacks the intense double peak located at 2θ = 2.48/2.57° characteristic of crystal form A1 (Fig. 3*e*), as well as the intense peak at 2θ = 2.63° characteristic of crystal form A2 (Fig. 3*f*), demonstrating that the proportion of these crystal forms is too low to be detected by powder diffraction. We estimate that XRPD would not be able to detect abundances of less than 5–10% of the total sample.

As the MX (110 K) and the XRPD (293 K) experiments were run at different temperatures, a reference XRPD pattern was collected at 100 K (see Supporting Information[1]). The

[1] Supporting information has been deposited in the IUCr electronic archive (Reference: TZ5050).

Table 2
Comparison of unit-cell parameters, solvent content, packing and crystal contact parameters.

	1svn	1jea†	1ndq	A1	A2
Space group	$P2_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters					
a (Å)	40.5	75.2	47.6	53.0	52.7
b (Å)	64.2	61.4	62.4	57.6	57.1
c (Å)	42.9	53.4	75.8	71.4	65.9
β (°)	118.8	—	—	—	—
Solvent content (%)	32.8	46.7	41.7	39.7	33.8
No. of neighbouring molecules	12	8	10	12	12
No. of unique contacts	6	4	5	6	6
Total No. of interfacing atoms	606	452	564	454	682
Total No. of interfacing residues	226	162	206	160	238
Total ASA involved in contacts (Å^2)	2675.8	2018.4	2515.4	2031.8	3007.8
Total ASA involved in contacts (%)	28.8	21.6	26.8	21.8	31.9
Total solvation energy, ΔG_{solv}‡ (kcal mol^{-1})	−0.4	−6.0	−13.2	−5.8	−9.0
Total No. of potential hydrogen bonds	26	24	32	34	40

† 1jea has been reindexed in order to orient the Savinase molecule analogously with the other orthorhombic structures. This means that a and c are interchanged, while b is inverted. ‡ Error estimates of ΔG_{solv} are up to 5 kcal mol^{-1}.

resulting XRPD patterns at 100 K showed analogous fingerprints to those collected at ambient temperature, indicating that no crystal rearrangement had occurred during cooling. The peaks were slightly shifted towards higher 2θ angles as a result of a contraction of around 0.5–1.0 Å along each unit-cell parameter.

The crystal form was further verified from a pattern profile fit in the 2θ range 1.7–15.0° (Fig. 4) using the coordinates deposited in PDB entry 1ndq (Pan *et al.*, 2003). The unit-cell parameters were determined[2] to be a = 47.65 (5), b = 62.43 (5), c = 75.74 (5) Å, which are close to those deposited for this crystal form (a = 47.6, b = 62.4, c = 75.8 Å; PDB entry 1ndq; Pan *et al.*, 2003), and the electron-density level in the bulk-solvent regions k_{sol} was determined to be 0.33 e Å^{-3}. The optimization of unit-cell parameters, k_{sol} and background improved the profile R factor, R_p, from 0.1696 to 0.0868.

3.5. Crystal contact analysis

The MX and XRPD results confirm the presence of three different crystal forms (PDB entry 1ndq, A1 and A2) in the production suspension. The crystal packing in these three structures were compared with two other crystal forms of native *B. lentus* subtilisin (PDB entries 1svn and 1jea; Graycar *et al.*, 1999) by analysing the contact interfaces between the molecules. Before the crystal contact analysis, re-indexing of 1jea was carried out in order to orient the subtilisin molecule analogously with respect to the crystallographic axes in all the orthorhombic structures.

For each structure the different crystal contacts have been identified, and information about the unique contacts can be found in the Supporting Information. The crystal packing for all five structures is shown in Fig. 5. Contacts present in more than one structure are coloured alike. To ease the comparison, the identity molecule is oriented the same way in all structures, and the part of the surface involved in the crystal contacts is shown without and with the neighbouring molecules.

The crystal contact analysis (Fig. 5) reveals that none of the contacts in the monoclinic crystal form (PDB entry 1svn) coincide with those found in the four orthorhombic crystal forms, and thus reflect a fundamentally different packing. In all of the orthorhombic structures, the b-directed screw contact 1 (magenta) is the same, and another b-directed screw contact 2 (green) is common to PDB entry 1ndq, A1 and A2. For the A1 and A2 single-crystal structures presented here, the residues involved in the contacts are the same and the orientation of neighbouring molecules in the two structures are alike. A1 and A2 thereby constitute a new overall crystal variant of Savinase which includes two different crystal packings.

The number of interfacing atoms and residues, the surface areas of the contacts, the solvation free energies and the number of potential hydrogen bonds have been summed for each structure. These parameters are listed in Table 2 together with unit-cell parameters and solvent content.

The 1ndq-like form is estimated to have the lowest solvation energy (Table 2) and may be the best stabilized form. However, the error estimates of the energy calculations are up to 5 kcal mol^{-1}, meaning that any conclusions on stabilization of the crystal form must be drawn with care. Furthermore, the salt present in the surrounding mother liquor will influence the stability of the protein fold itself as well as the lattice interactions.

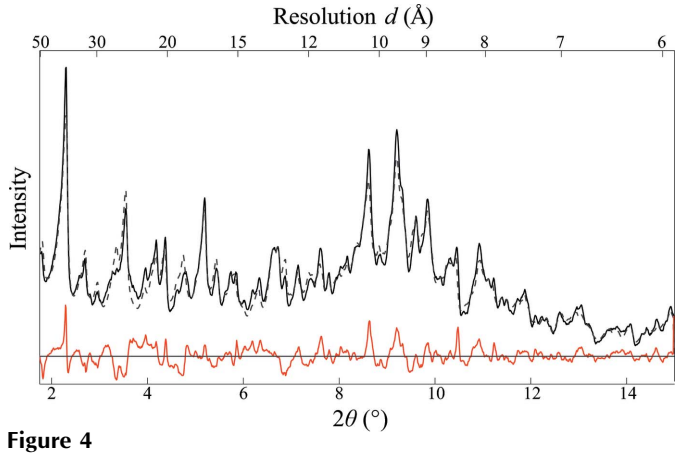

Figure 4
Optimized pattern profile fit for crystal form verification showing the experimental pattern (continuous line) and calculated pattern (dashed line). The latter was calculated from PDB entry 1ndq (Pan *et al.*, 2003) using optimized unit-cell parameters and k_{sol}, as well as an optimized background. Profile R factor R_p = 0.0868. The difference is shown in red.

[2] Determination of the unit-cell parameters strongly depends on the actual set of peak profile parameters used. The unit-cell parameters and following standard deviations determined here were a = 47.65 (1), b = 62.43 (1), c = 75.74 (2) Å using profile parameters optimized to the experimental setup. However, a fit with similar quality can be obtained from other sets of profile parameters. Using different sets of peak profile parameters within a physically meaningful range, the unit-cell parameters varied within ±0.05 Å, which may be considered as a more realistic standard deviation.

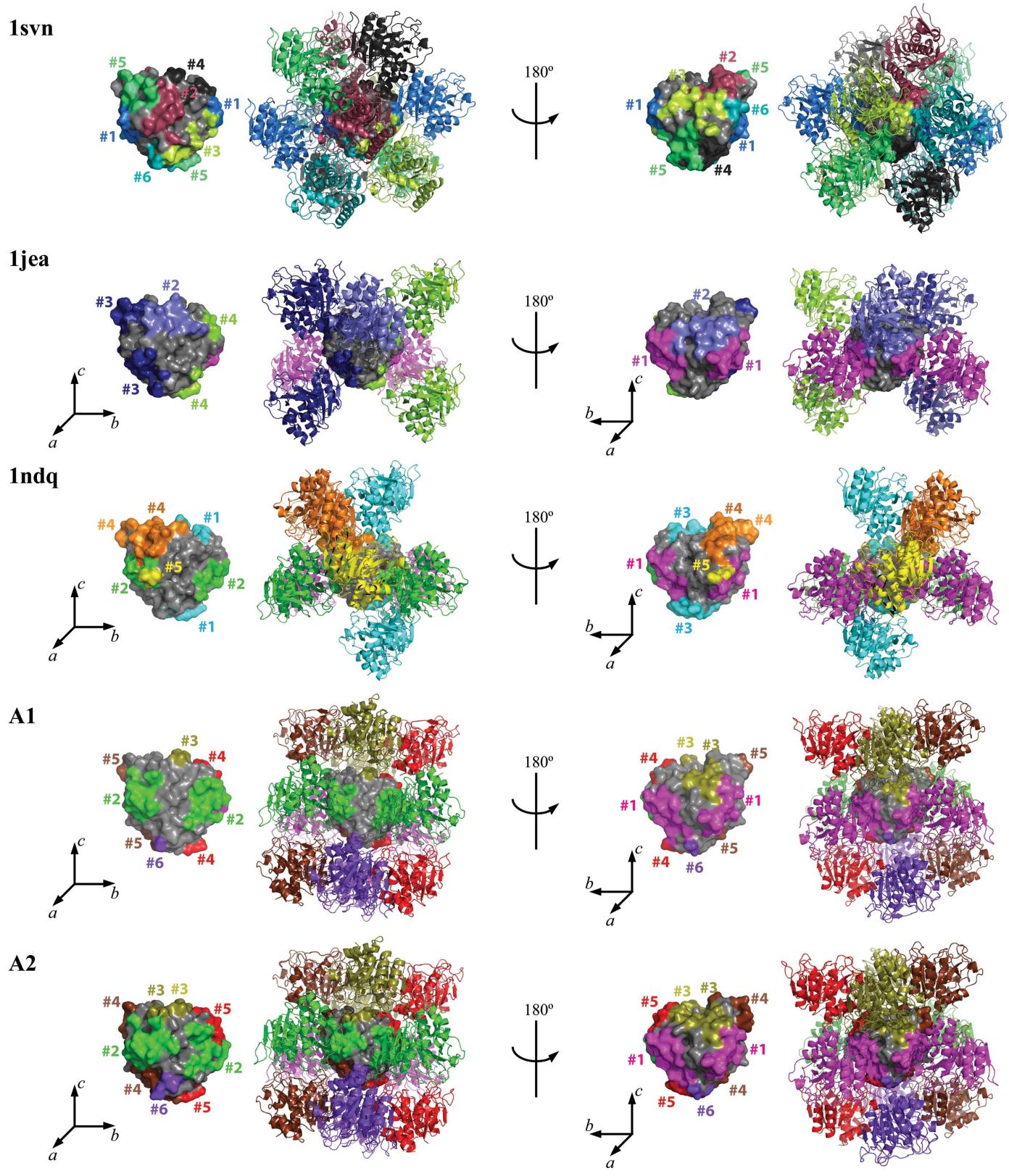

Figure 5
Crystal packing in five *B. lentus* subtilisin crystal forms. The crystal contacts were identified by *PISA* (Krissinel & Henrick, 2007) and are differently coloured and numbered. To facilitate the comparison, the identity molecule is shown in the same orientation in all structures from two different views (front and back). For each view the part of the surface involved in the crystal contacts is shown without and with the neighbouring molecules. Contacts present in more than one structure are coloured alike. Crystal structure 1jea has been re-indexed in order to orient the subtilisin molecule analogously with respect to the crystallographic axes as the other orthorhombic structures.

3.6. Crystal packing and solvent content

The tightest packing is observed in the monoclinic form (PDB entry 1svn; Betzel *et al.*, 1992), in which each molecule is surrounded by 12 neighbouring molecules and which has only 32.8% solvent content (Table 2). The most loosely packed structure is the orthorhombic 1jea form (Graycar *et al.*, 1999), with only eight symmetry-related neighbouring molecules and 46.7% solvent content. For the A1 structure the Matthews coefficient of 2.04 $\text{Å}^3\ \text{Da}^{-1}$ corresponds to a solvent content of 39.7%; it is slightly more tightly packed than the orthorhombic 1ndq form (Pan *et al.*, 2003). The A2 structure has a Matthews coefficient of 1.6 $\text{Å}^3\ \text{Da}^{-1}$, corresponding to a solvent content of 33.8%; it is thus more tightly packed than A1, as seen from the large contact areas in Fig. 5, and almost as tightly packed as the monoclinic form 1svn. The presence of the bridging sulfate ions helps to glue the molecules together, thereby providing a tighter packing, particularly along the crystallographic *c* axis. In both the A1 and A2 structures the low solvent content is shown by the packing of 12 symmetry-related monomers around the central molecule (Fig. 5).

4. Discussion

4.1. XRPD for identification of protein microcrystalline suspensions

Our results demonstrate that MX and XRPD can provide complementary information in the analysis of a microcrystalline protein suspension produced during large-scale production. Most importantly, the XRPD experiments on Savinase revealed a dominant crystal form in the bulk sample, which by comparison with the deposited PDB entries could be identified as crystal form 1ndq (Pan *et al.*, 2003), a form originally reported by Betzel *et al.* (1988) as form 2 in their crystallization of wild-type Savinase. However, we emphasize that this experiment can only reveal the major component(s) and does not exclude the presence of other polymorphic forms as small fractions of the suspension. The qualitative analysis allowing the identity of the crystal form to be determined is furthermore dependent on comparison with a set of known forms determined by conventional MX experiments: the *ab initio* determination of protein structures from XRPD data poses considerable challenges.

4.2. The microcrystals and polymorphism

What surprised us was that the two larger single crystals, while also in space group $P2_12_12_1$, were quite different from the dominant species identified by XRPD in terms of unit cell and lattice interactions, confirming that different polymorphs of Savinase are present in the same production sample, albeit in very different amounts. It is evident from the photomicrograph in Fig. 2 that the larger crystals are present at a very low level, well below that required to detect them by XRPD. One potential source of difference between the different crystal polymorphs in the XRPD and the MX experiments is that using the in-house X-ray source for XRPD the samples are analysed at ambient temperature, whereas the MX data are recorded at 110 K. This issue was addressed by recording an XRPD pattern at 100 K, which differed from that at ambient temperature only in a small contraction of the unit cell (see Supporting Information).

XRPD thus emerges as a powerful tool for identifying the dominating component(s) of such a microcrystalline suspension, with the crystal form easy to identify provided that a single-crystal model is available. In contrast, single-crystal MX may yield results that are not representative of the bulk production sample, particularly in cases such as Savinase where it turns out that there are two or more crystal forms present. Multiple crystal forms have previously been reported in conventional crystallization experiments for single-crystal diffraction, where highly purified proteins are crystallized under the same tightly controlled conditions (*i.e.* protein purity, precipitant and protein concentrations, temperature and pH). One example is the crystallization of lysozyme in the presence of nitrite, which resulted in a mixture of triclinic and monoclinic crystals (Sieker, 1988; Walsh *et al.*, 1998). Perhaps more relevant here is earlier work on Savinase itself: Betzel *et al.* (1988) reported three crystal forms grown under close to identical conditions, one of these being the 1ndq form.

4.3. Crystal contact analysis

A total of four different orthorhombic crystal forms of wild-type Savinase are now known: the three presented here, one of which matches 1ndq, and in addition PDB entry 1jea. The analysis of crystal contacts revealed common characteristics within the four orthorhombic crystal forms, as a *b*-directed screw contact was common to all of them. This goes some way to rationalizing the co-existence of the three different crystal forms in the production suspension, and suggests that conversion between them may be possible. The similarity of crystal forms A1 and A2 is more distinct than the other forms as the identity molecule is surrounded by 12 neighbouring molecules with analogous orientations. The difference is a closer packing in the A2 form caused by the bridging effect of the interface sulfate ions.

5. Conclusions

In conclusion, XRPD is a robust tool for analysing the dominant crystal form(s) in submicrometre crystal suspensions and hence has potential as a powerful tool for quality control in such commercial processes. It can clearly distinguish with confidence cells with significantly different unit-cell parameters. Fundamentally different diffraction patterns arise when unit-cell parameters vary by more than $\sim$2 Å, but the detection limit of such differences needs to be established. An important remaining question is the minimum percentage of a component that can be detected in such mixtures. Very minor components such as A1 and A2 reported here for Savinase, which are well below 1% as estimated from the microscope photographs, are clearly well below the sensitivity. The detection limit must form part of a future study.

We gratefully acknowledge the The Danish National Advanced Technology Foundation for funding of this project. We thank Diamond Light Source for access through our BAG allocation to beamlines I03 and I04 that contributed to the results presented here.

References

Aishima, J., Owen, R. L., Axford, D., Shepherd, E., Winter, G., Levik, K., Gibbons, P., Ashton, A. & Evans, G. (2010). *Acta Cryst.* D**66**, 1032–1035.

Betzel, C., Dauter, Z., Dauter, M., Ingelman, M., Papendorf, G., Wilson, K. S. & Branner, S. (1988). *J. Mol. Biol.* **204**, 803–804.

Betzel, C., Klupsch, S., Papendorf, G., Hastrup, S., Branner, S. & Wilson, K. S. (1992). *J. Mol. Biol.* **223**, 427–445.

Chapman, H. N. *et al.* (2011). *Nature (London)*, **470**, 73–77.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Fokine, A. & Urzhumtsev, A. (2002). *Acta Cryst.* D**58**, 1387–1392.

Frankaer, C. G., Harris, P. & Ståhl, K. (2011). *J. Appl. Cryst.* **44**, 1288–1290.

Graycar, T., Knapp, M., Ganshaw, G., Dauberman, J. & Bott, R. (1999). *J. Mol. Biol.* **292**, 97–109.

Hartmann, C. G., Harris, P. & Ståhl, K. (2011). *Z. Kristallogr. Proc.* **1**, 163–168.

Hartmann, C. G., Nielsen, O. F., Ståhl, K. & Harris, P. (2010). *J. Appl. Cryst.* **43**, 876–882.

Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.

Le Bail, A., Duroy, H. & Fourquet, J. L. (1988). *Mater. Res. Bull.* **23**, 447–452.

Margiolaki, I., Giannopoulou, A. E., Wright, J. P., Knight, L., Norrman, M., Schluckebier, G., Fitch, A. N. & Von Dreele, R. B. (2013). *Acta Cryst.* D**69**, 978–990.

Margiolaki, I. & Wright, J. P. (2008). *Acta Cryst.* A**64**, 169–180.

Moukhametzianov, R., Burghammer, M., Edwards, P. C., Petitdemange, S., Popov, D., Fransen, M., McMullan, G., Schertler, G. F. X. & Riekel, C. (2008). *Acta Cryst.* D**64**, 158–166.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Pan, X., Bott, R. & Glatz, C. E. (2003). *J. Cryst. Growth*, **254**, 492–502.

Pantoliano, M. W., Whitlow, M., Wood, J. F., Rollence, M. L., Finzel, B. C., Gilliland, G. L., Poulos, T. L. & Bryan, P. N. (1988). *Biochemistry*, **27**, 8311–8317.

Pawley, G. S. (1981). *J. Appl. Cryst.* **14**, 357–361.

Sieker, L. C. (1988). *J. Cryst. Growth*, **90**, 31–38.

Ståhl, K., Frankaer, C. G., Petersen, J. & Harris, P. (2013). *Powder Diffr.* **28**, s458–S469.

Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.

Von Dreele, R. B. (1999). *J. Appl. Cryst.* **32**, 1084–1089.

Von Dreele, R. B., Stephens, P. W., Smith, G. D. & Blessing, R. H. (2000). *Acta Cryst.* D**56**, 1549–1553.

Wagner, A., Duman, R., Stevens, B. & Ward, A. (2013). *Acta Cryst.* D**69**, 1297–1302.

Walsh, M. A., Schneider, T. R., Sieker, L. C., Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1998). *Acta Cryst.* D**54**, 522–546.

Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.

Winter, G. (2010). *J. Appl. Cryst.* **43**, 186–190.