

Monitoring protein precipitates by in-house X-ray powder diffraction

Kenny Ståhl,^{1,a)} Christian G. Frankær,¹ Jakob Petersen,¹ and Pernille Harris¹

¹ *Department of Chemistry, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark*

Key words: protein powder, fingerprinting, search-match, solvent correction, lysozyme, insulin

Powder diffraction from protein powders using in-house diffractometers is an effective tool for identification and monitoring of protein crystal forms and artifacts. As an alternative to conventional powder diffractometers a single crystal diffractometer equipped with an X-ray micro-source can be used to collect powder patterns from 1 μ l samples. Using a small-angle X-ray scattering (SAXS) camera it is possible to collect data within minutes. A streamlined program has been developed for the calculation of powder patterns from pdb-coordinates, and includes correction for bulk-solvent. A number of such calculated powder patterns from insulin and lysozyme have been included in the powder diffraction database and successfully used for search-match identification. However, the fit could be much improved if peak asymmetry and multiple bulk-solvent corrections were included. When including a large number of protein data sets in the database some problems can be foreseen due to the large number of overlapping peaks in the low-angle region, and small differences in unit cell parameters between pdb-data and powder data. It is suggested that protein entries are supplied with more searchable keywords as protein name, protein type, molecular weight, source organism etc. in order to limit possible hits.

^{a)} Author to whom correspondence should be addressed. Electronic mail: kenny@kemi.dtu.dk

I. INTRODUCTION

It has been demonstrated (Norrman *et al.*, 2006, Frericks Schmidt *et al.*, (2007) and Hartmann *et al.*, 2010) that fingerprints of micro-crystalline protein samples can be obtained by XRPD from in-house equipment. Powder patterns calculated from pdb-coordinates can be brought to very good agreement to experimental powder patterns when including bulk-solvent correction, unit cell and peak shape optimization (Hartmann *et al.*, 2010). The method has been found useful for polymorph recognition and crystallization monitoring. It is presently applied in industrial protein production both for quality assurance and problem solving. It is a fast and convenient way to distinguish between protein polymorphs, simple salt and amorphous precipitates. Although the angular resolution from an in-house powder diffractometer is limited, as compared to a synchrotron source (Margiolaki, 2008), minor changes as a cell axis doubling can be detected (Frankær *et al.*, 2012). Ultimately substrate inclusions should be possible to detect. In order to be generally applicable a streamlined computer program is needed, which will be described and discussed below.

Identification of an unknown protein crystal form is literally as like looking for a needle in a haystack. Presently the powder diffraction database (ICDD, the International Centre for Diffraction Data, www.icdd.com) contains a very limited amount of protein data. Below we demonstrate the possibility to use existing search-match procedures to identify protein powder patterns, and that powder patterns calculated from pdb-coordinates with proper care can be added to the database and included in the search-match procedure. Several problems can be foreseen when including large amount of protein data, which will be discussed below. Finally, it is demonstrated that alternative in-house data collection methods can be used for instance in order to minimize sample amounts or in combination with other techniques.

II. DATA TREATMENT STRATEGY AND DEVELOPMENTS

A streamlined Windows-based data treatment program has been developed. The data treatment can be divided into three major steps:

A) Direct calculation of powder diffraction patterns from coordinate files to create a library of patterns. The calculation automatically includes a flat bulk-solvent correction (Phillips, 1980; Fokine & Urzhumtsev, 2002). As demonstrated by Hartmann *et al.*, (2010; 2011) this is a crucial step also illustrated in Figure 1. Alternatively, powder patterns can be calculated from observed single crystal diffraction data, F_{obs} , by which the contribution from the bulk-solvent is already included. However, our experience is that the use of structure factors deposited in the Protein Data Bank (PDB, www.pdb.org) is problematic because low resolution data often are missing or poorly measured.

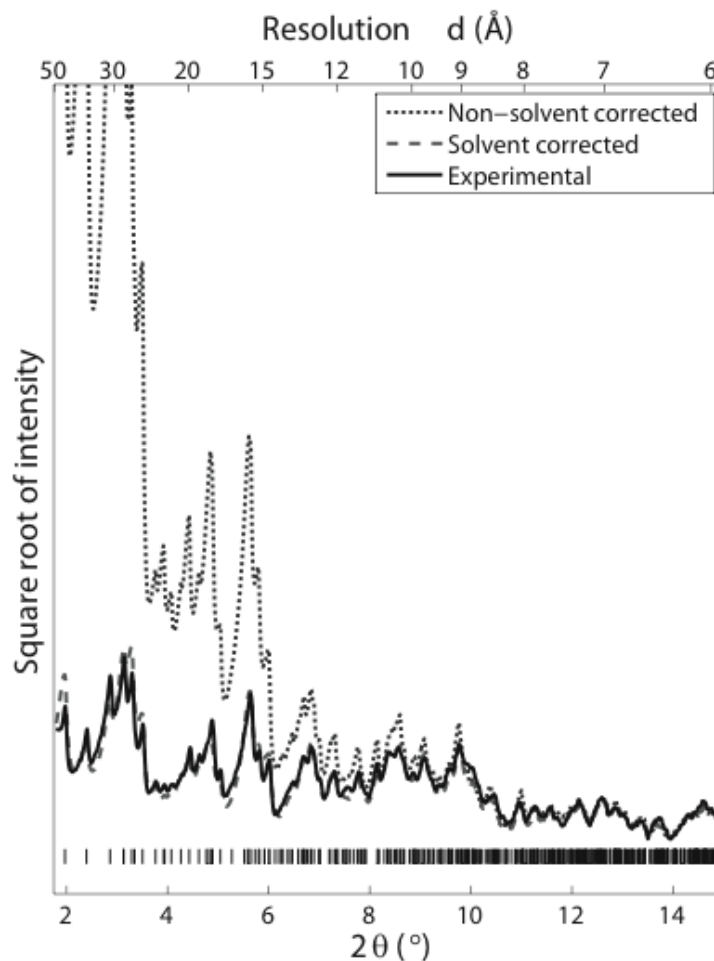


Figure 1: Calculated powder diffraction patterns for orthorhombic lysozyme before (dotted line) and after (dashed line) bulk-solvent correction. An experimental powder pattern (solid line) is shown for comparison. N.B. The effect from the bulk-solvent is reduced by showing the square root of the intensity.

B) Identification. A crystal form can be identified by performing a search-match of the experimental powder pattern against a database containing experimental patterns as well as patterns calculated from pdb-files.

C) Verification. When a good match is found the fit can be further improved by recalculation of the matched powder pattern after refining the unit cell and peak profile parameters. A flow chart describing the data treatment procedure by the new software using a pdb-file and the experimental data as inputs is shown in Figure 2.

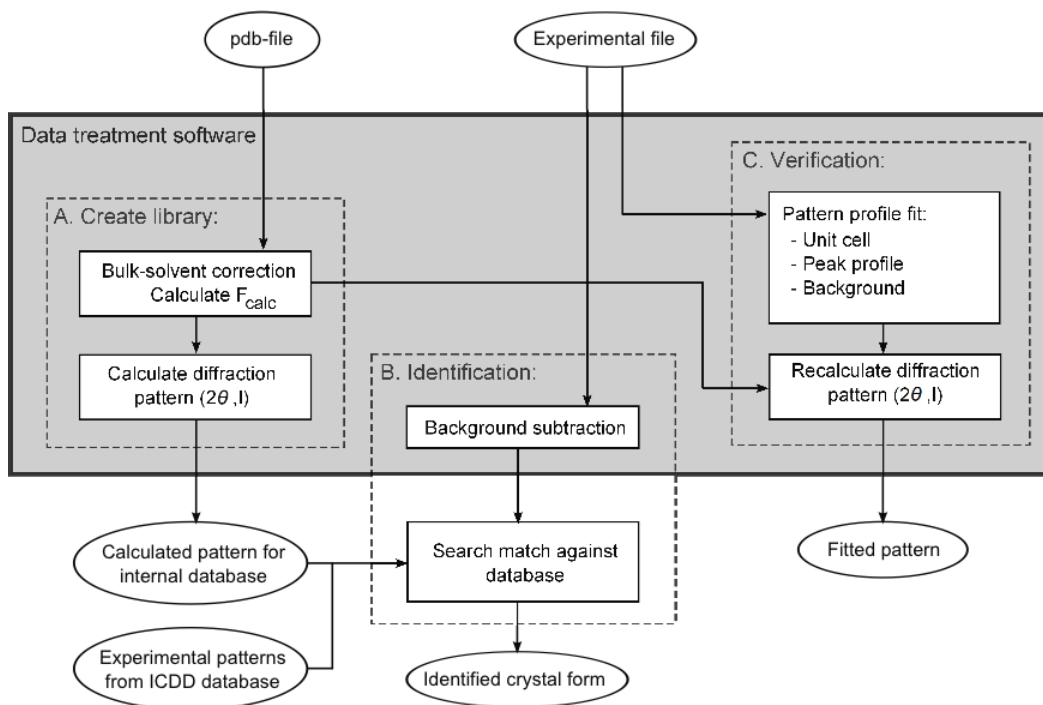


Figure 2: Flow chart describing the three major steps of the data treatment. A) Calculation of powder diffraction patterns from pdb-files for a database. B) Identification of a crystal form by searching calculated powder patterns in the database. C) Verification of a crystal form where the fit between calculated and experimental powder patterns is optimized. The program takes a pdb-file and experimental data as input.

III. IDENTIFICATION FROM ESTABLISHED DATABASES

The ICDD database contains powder diffraction data so far from eleven different crystal forms of insulin (Norrman *et al.*, 2006). In analogy to inorganic and small molecule compounds, identification by search-match procedures can be carried out for proteins as well. In spite of a severe peak overlap the powder patterns serve as a unique fingerprints and the different crystal forms can easily be distinguished, also from data recorded on in-house XRPD equipment. An existing search-match procedure (Crystallographica Search-Match, 1999) has been tested on

three different crystal forms of insulin. Rhombohedral T₆- and R₆-insulin micro-crystals were prepared as described by Frankær *et al.* (2012), and cubic insulin as described by Hartmann *et al.* (2010). Powder diffraction data were collected for 4 h for all samples on an in-house Huber G670 diffractometer using Cu K α ₁ radiation using the method described by Frankær *et al.* (2011). Each powder pattern was subject to a rough background subtraction and then identified by search-match against the ICDD database using powder patterns in the low angle 2 θ -range (2–10°). The result from R₆-insulin is shown in Figure 3.

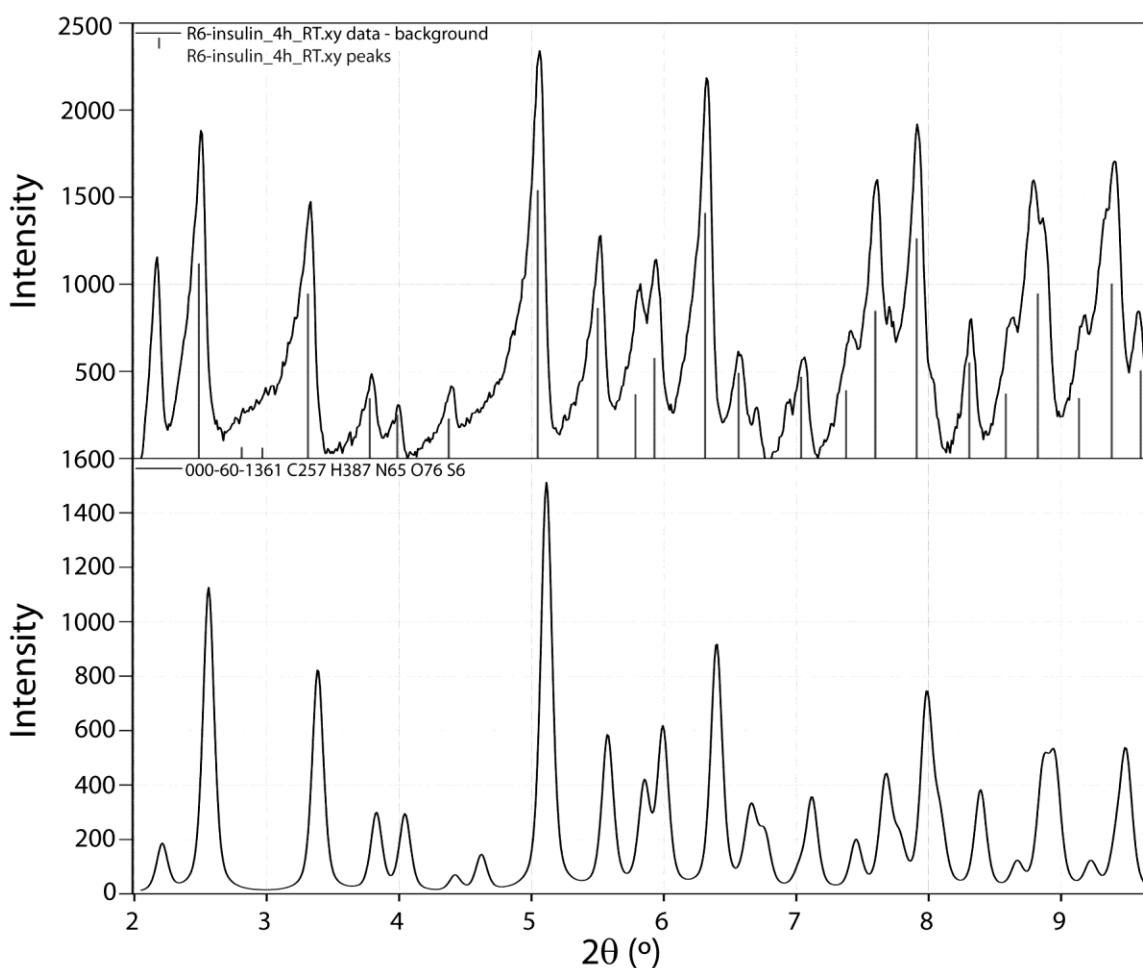


Figure 3: Identifications of rhombohedral R₆-insulin using the search-match procedure for data deposited in the ICDD database.

The correct identifications come up as first choice for all three insulin polymorphs, with good figure-of-merit. It is important to note that not all the peaks occurring at 2 θ > 5° are found by the search-match program, only the most intense peaks in the profile are found. This means that only

the most intense peaks are necessary for identification even though these peaks may be constituted by an overlap of several reflections. Slightly shifted peak positions are observed due to small variations in unit cell parameters. Furthermore, peak asymmetry is not taken into account, which impairs the visual impression.

IV. IDENTIFICATION BASED ON CALCULATED POWDER PATTERNS

Presently it is possible to include your own powder data in a custom database, which we have used for testing whether or not a protein powder can be identified from a calculated powder diffraction pattern.

As an ultimate goal XRPD could be used for identifying unknown proteins. Being able to recognise similarities in folding and packing from protein structures deposited in the Protein Data Bank may provide low resolution information, which can have future perspectives for phasing new structures. This will require a large number of powder patterns generated from coordinates. It may be worth considering to calculate powder patterns for all structures deposited in the PDB and to include them in the ICDD database.

Two different crystal forms of hen egg white lysozyme (HEWL) were used for search-match identification from calculated powder patterns. Micro-crystals of tetragonal lysozyme crystals were grown from a solution containing 25 mg/ml lysozyme, 4% (w/w) sodium chloride, and 0.1 M sodium acetate, pH 4.6 at 4 °C (Hampton Research). Orthorhombic lysozyme crystals were grown by the method described by Sato *et al.* (2001). Samples were prepared as described by Frankær *et al.* (2011), and powder diffraction data were collected on a Huber G670 diffractometer for 2 h using Cu $K\alpha_1$ -radiation.

Powder diffraction patterns from six different crystal forms of HEWL were calculated from pdb-files deposited in the PDB. The crystal forms included one monoclinic (pdb-entry 1LYS) three different orthorhombic (pdb-entries 1WTM, 1F0W, 1F10), one hexagonal (pdb-entry 2FBB), and one tetragonal (pdb-entry 1LZ8). Solvent correction was applied using average values of the solvent parameters found by Fokine & Urzhumtsev, (2002), and the unit cell parameters used for the calculations were taken from the pdb-file. The calculated powder patterns were included in a custom database, where each located peak was registered as the peak position and intensity. The search-match results from the identification of the two lysozyme forms are shown in Figure 4.

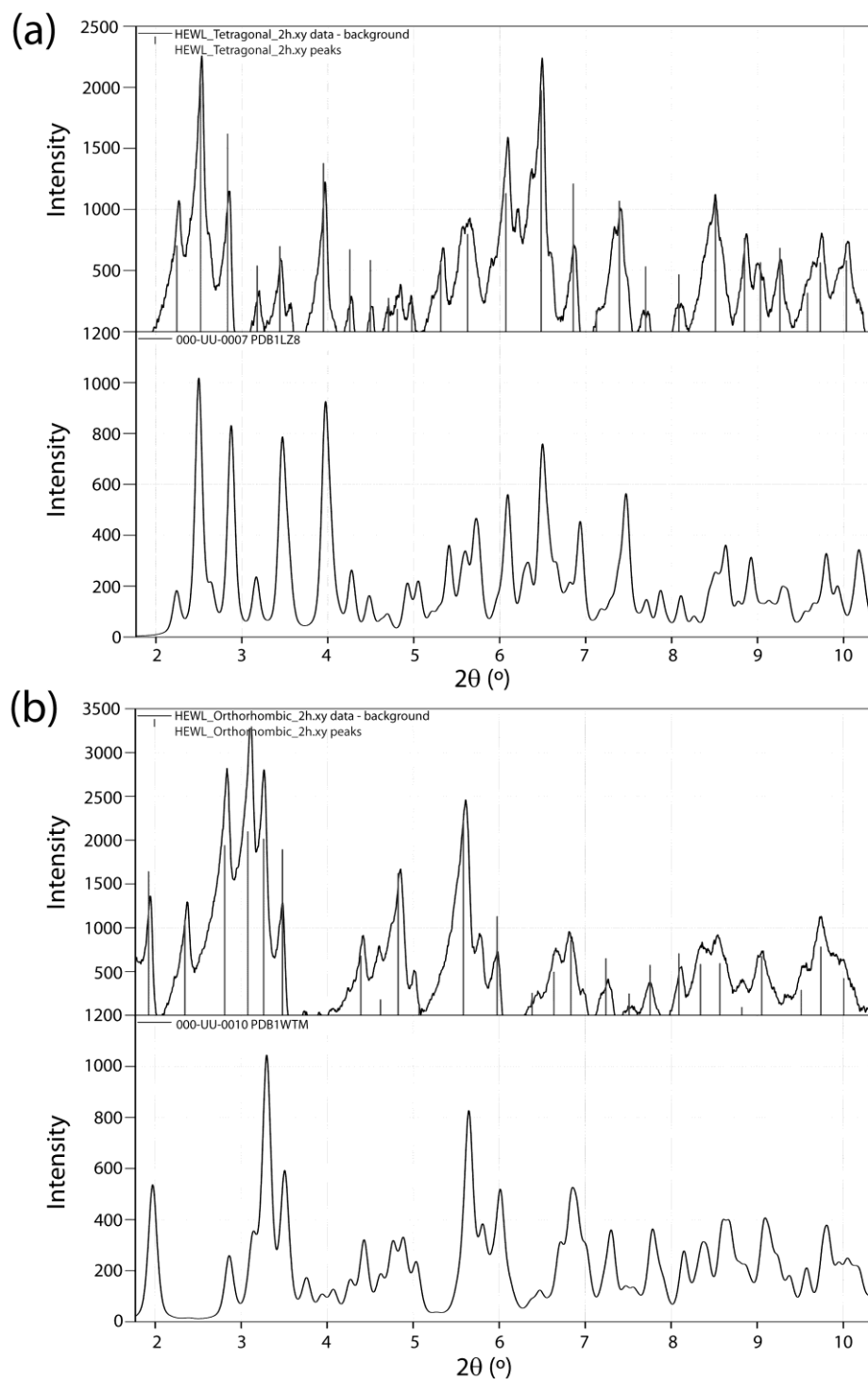


Figure 4: Identifications of (a) tetragonal and (b) orthorhombic lysozyme using the search-match procedure.

As seen from Figure 4, the two crystal forms were identified from calculated powder patterns. The correct identification comes up among the first few choices when reasonable estimations of 2θ -zero shift and FWHM are used. However there are some disagreements. Firstly, the

parameters used for the bulk-solvent are average values and could be further optimized for each system. This explains the difference in intensities for the match in the low angle region ($2\theta = 2-5^\circ$), e.g. the peak at 3.5° in the powder pattern for the tetragonal crystal form. It may be worthwhile to include calculated powder patterns with several solvent correction levels in the database. Secondly, the asymmetry of particularly the peaks in the low angle region has not been taken into account. In addition to the visual impression the mistreated asymmetry affects the location of the peak maxima and thereby the 2θ -zero shift. Thirdly, there may be small variations in unit cell parameters, as the majority of the structures deposited in PDB are solved from low temperature data (typically 100 K), while our powder data were recorded at room temperature. Using an average isotropic expansion coefficient, may compensate for some of this disagreement.

In our test case the number of candidates is relatively small, and thus we relatively easily can obtain a match. If the database had contained larger amount of candidates we would have obtained numerous matches, and we can foresee a lot of challenges. It will most likely be necessary to include protein specific information as protein name, protein class, molar weight, source organism, etc. in the database in order to narrow the searches in analogy with the search procedures in the PDB. When including experimental protein powder data sets a quality marker based on instrumental resolution would be useful: Well resolved synchrotron data is clearly preferred as they in some cases can be indexed and polymorph mixtures can be detected (Collings *et al.*, 2010; Karavassili *et al.*, 2012).

V. ALTERNATIVE EXPERIMENTAL SETUPS

Depending on the specific application and availability of the protein, other powder diffraction setups may have advantages for different types of sample.

Using focussed X-ray micro-sources it is possible to collect powder diffraction data with a much reduced sample amount. Figure 5 shows the 2-dimensional powder pattern from approximately 1 μl of tetragonal lysozyme powder suspension, which was mounted in a commercial loop (Hampton Research) with a diameter of 1.0 mm. The loop was mounted on a diffractometer with micro-source beam: GeniX, Cu $K\alpha$ -radiation ($\lambda = 1.5418 \text{ \AA}$) source.

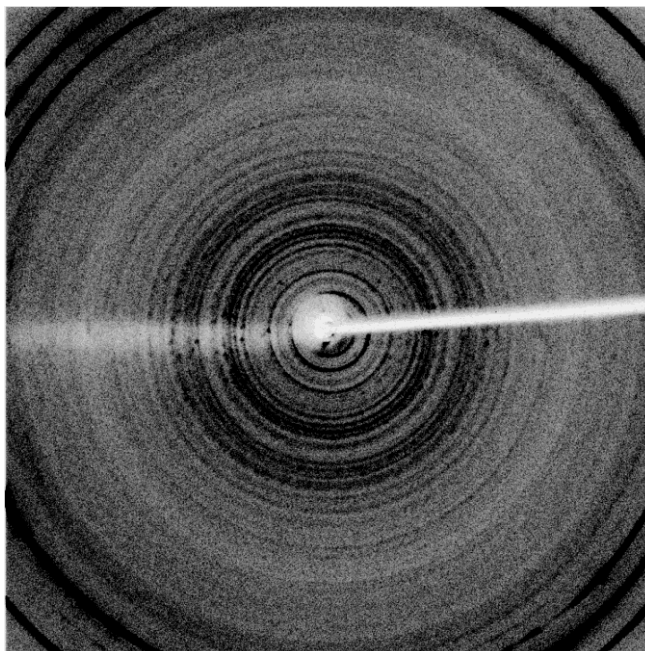


Figure 5: X-ray powder diffraction rings for tetragonal lysozyme powder. Data were collected in 45 minutes on 1 μl sample mounted in a cryo-loop on a micro-source single crystal diffractometer.

The beam is focussed on the sample with a spot size of 230 μm . The sample was cooled to 130 K to prevent it from drying out. Cryo-protectant was not used. Data were accumulated for 45 minutes on a Rigaku R-Axis IV++ 2D-detector in a single frame covering $\Delta\phi$ of 90°.

Even though the micro-crystals were not cryo-protected, it was possible to obtain good signal from the sample. Cryo-protection of the micro-crystals is possible (Jenner *et al.*, 2007), but requires thorough optimization as the cryoprotectant in many cases slightly etches the surface of the crystals. This is not noticeable when treating large single crystals, but for a powder sample the crystallite size is significantly reduced or even dissolved. Working with in-house sources, radiation damage is moderate, and cryo-protection may in many cases complicate the experiments unnecessarily.

Yet another alternative is to collect powder patterns on a Small Angle X-ray Scattering (SAXS) camera. Figure 6 shows a powder pattern collected on a Bruker Nanostar equipped with a rotating anode Cu $K\alpha$ -source. Data were collected on approximately 15 μl sample for 5 minutes. As seen from Figure 6, the angular resolution is quite low due to large slit size, 1.5 by 1.5 mm. A SAXS camera has the advantage of being fast and optimized for the lowest angles, which can be

difficult on standard powder diffractometers due to slit and beam stop adjustments. Furthermore, samples containing non-crystalline precipitates from aggregation and fibrillation can be further characterized from this method. Applying the method to high-throughput synchrotron SAXS beamlines could provide fast polymorph screening of micro-crystalline protein samples, possibly using principal component analysis (PCA) as a first fast sorting step (Norrman *et al.*, 2006).

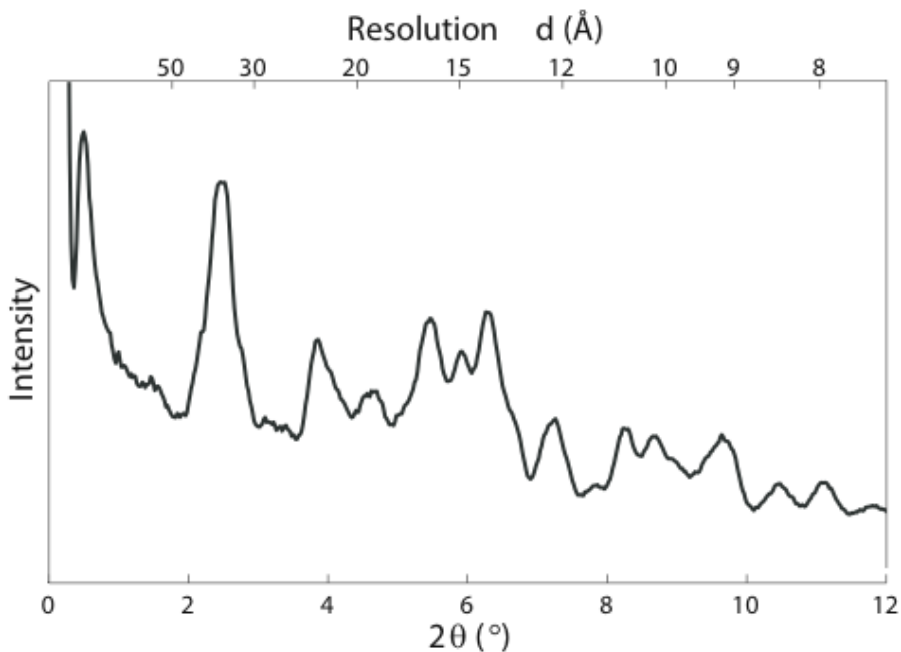


Figure 6: X-ray powder diffraction pattern for tetragonal lysozyme collected on a SAXS setup for 5 minutes.

VI. CONCLUSION

- It is possible to calculate protein powder diffraction patterns from pdb-coordinates. Including the bulk-solvent correction very good agreements can be obtained between calculated and observed powder patterns. All necessary calculations can be performed in one streamlined program.
- Existing search-match procedures can identify protein powders using a *limited set* of either measured or calculated powder patterns included in the powder database. Deviations in the matches can be related to missing peak asymmetry, variation in bulk-solvent level, and peak shifts due to (minor) unit cell variations.
- In the case of an *extended set* of protein data in the powder database we can foresee problems with too many hits due to the large number of peaks in a limited angular range.

Including additional searchable keywords for protein entries will be necessary for narrowing the searches.

- Depending on sample type and amount of sample other powder diffraction setups may be advantageous: An X-ray micro-source single-crystal diffractometer allows for sample amounts of 1 μl . Furthermore, sample cooling will give unit cell parameters in closer correspondence to single-crystal data. A SAXS camera allows fast measurements of the lowest angles and characterization of non-crystalline precipitates and protein aggregations.

ACKNOWLEDGEMENTS

We gratefully acknowledge Professor Anette Henriksen and Carlsberg Laboratories for use of their micro-source diffractometer and Professor Jan Skov Pedersen, Aarhus University, for use of the SAXS camera.

Collings, I., Watier, Y., Giffard, M., Dagogo, S., Kahn, R., Bonneté, F., Wright, J. P., Fitch, A. N.

and Margiolaki, I. (2010). "Polymorphism of microcrystalline urate oxidase from *Aspergillus flavus*," Acta Crystallogr., Sect. D: Biol. Crystallogr. **66**, 539–548

Oxford Cryosystems (1999). "Crystallographica Search-Match," J. Appl. Crystallogr. **32**, 379–380.

Fokine, A. and Urzhumtsev, A. (2002). "Flat bulk-solvent model: obtaining optimal parameters," Acta Crystallogr., Sect. D: Biol. Crystallogr. **58**, 1387–1392.

Frankær, C. G., Harris, P. and Ståhl, K. (2011). "A sample holder for in-house X-ray powder diffraction studies of protein powders," J. Appl. Crystallogr. **44**, 1288–1290.

Frankær, C. G., Knudsen, M. V., Norén, K., Nazarenko, E., Ståhl, K. and Harris, P. (2012). "The structures of T₆, T₃R₃ and R₆ bovine insulin: combining X-ray diffraction and absorption spectroscopy," Acta Crystallogr., Sect. D: Biol. Crystallogr. **68**, 1259–1271.

Frericks Schmidt, H. L., Sperling, L. J., Gao, Y. G., Wylie, B. J., Boettcher, J. M., Wilson, S. R. and Rienstra, C.M. (2007). "Crystal Polymorphism of Protein GB1 Examined by Solid-State NMR Spectroscopy and X-ray Diffraction," J. Phys. Chem. B. **111**, 14362–14369.

Hartmann, C. G., Nielsen, O. F., Ståhl, K., and Harris, P. (2010). "In-house characterization of protein powder," J. Appl. Crystallogr. **43**, 876–882.

Hartmann, C. G., Harris, P., and Ståhl, K. (2011). "In-house characterization of protein powder," Z. Kristallogr. Proc. **1**, 163–168.

- Jenner, M. J., Wright, J. P., Margiolaki, I., and Fitch, A. N. (2007). "Successful protein cryocooling for powder diffraction," *J. Appl. Crystallogr.* **40**, 121–124.
- Karavassili, F., Giannopoulou, A. E., Kotsiliti, E., Knight, L., Norrman, M., Schluckebier, G., Drube, L., Fitch, A. N., Wright J. P. and Margiolaki, I. (2012). "Structural studies of human insulin cocrystallized with phenol or resorcinol via powder diffraction," *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **68**, 1632-1641.
- Margiolaki, I., and Wright, J. P. (2008). "Powder crystallography on macromolecules," *Acta Crystallogr., Sect. A: Found. Crystallogr.* **64**, 169–180.
- Norrman, M., Ståhl, K., Schluckebier, G. and Al-Karadaghi, S. (2006). "Characterization of insulin microcrystals using powder diffraction and multivariate data analysis," *J. Appl. Crystallogr.* **39**, 391–400.
- Phillips, S. E. V. (1980). "Structure and Refinement of Oxymyoglobin at 1.6 Å Resolution," *J. Mol. Biol.*, **142**, 531–554.
- Sato, T., Yamada, Y., Saijo, S., Hori, T., Hirose, R., Tanaka, N., Sazaki, G., Nakajima, K., Igarashi, N., Tanaka, M. and Matsuura, Y. (2001). "Improvement in diffraction maxima in orthorhombic HEWL crystal grown under higher magnetic field," *J. Crystal Growth.* **232**, 229–236.