

Structures from poorly diffracting crystals: a new start for macromolecular crystallography?

Janos Hajdu*

Laboratory of Molecular Biophysics, Department of Cell and Molecular Biology, Uppsala University, Husargatan 3 (Box 596), SE-751 24 Uppsala, Sweden, and The European Extreme Light Infrastructure, Institute of Physics, AS CR, Na Slovance 2, Prague 18221 8, Czech Republic. *Correspondence e-mail: janos@xray.bmc.uu.se

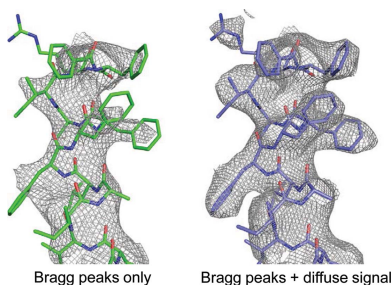
Keywords: continuous diffraction; disordered crystals; statistics; photosystem II; diffuse scattering.

In a ground-breaking study published in this issue of *Journal of Applied Crystallography*, Henry Chapman and colleagues (Chapman *et al.*, 2017) describe the analysis of the intensity statistics of continuous diffraction patterns from disordered macromolecular crystals, using micro- and nanocrystals of photosystem II as test objects. Photosystem II is a membrane protein responsible for splitting water and producing breathable oxygen during photosynthesis. The analysis method developed here uses previously published continuous diffraction data on photosystem II, measured at the Linac Coherent Light Source (LCLS) X-ray free-electron laser (FEL) (Ayyer *et al.*, 2016). The work has the potential to have a significant impact on macromolecular crystallography by solving the phase problem from the continuous molecular transform, accessible from translationally disordered crystals and nanocrystals, and by extending resolution beyond the resolution of the Bragg peaks.

All crystals of macromolecules are disordered to some degree; around half of the volume of the crystals is liquid of crystallization, and the macromolecules in the lattice are subject to static and dynamic disorder, including displacement from ideal lattice positions. Chapman *et al.* point out that the intensities of far-field diffraction patterns of aligned molecules obey Wilson statistics, whether those molecules are in isolation (giving rise to a continuous diffraction pattern) or arranged in a crystal, to produce Bragg peaks. Ensembles of molecules in several orientations, but uncorrelated in position, give rise to the incoherent sum of the diffraction from those objects, modifying the statistics in a similar way as crystal twinning modifies the distribution of Bragg intensities. Note also that single-molecule diffraction data add together in the same way when randomly oriented macromolecules are exposed one by one in random orientations (Maia *et al.*, 2009).

Chapman *et al.* have developed the analysis of the intensity statistics of continuous diffraction from disordered crystals to obtain parameters such as scaling, beam coherence and the number of contributing independent object orientations. The results show significant improvements (Fig. 1) in the structure of photosystem II when the diffuse background is included in the structure determination. Many of the statistical results will be useful in other contexts where there is a similar mixing of coherent and incoherent signal. Other systems may behave somewhat differently. Indeed, the literature on diffuse scattering suggests that there are various types of disorder with different levels of spatial correlation. Such systems will need to be investigated independently before it will be clear how widely this analysis applies.

The quest to measure the molecular transform directly in a protein crystal goes back to 1938 when Bernal *et al.* (1938) noticed that unit-cell dimensions of haemoglobin and chymotrypsin crystals changed and reflections moved as the crystals underwent drying or swelling. During this process, the Bragg peaks sampled different parts of the molecular transform, leading to alterations in the relative intensities of the Bragg reflections. The paper states that ‘Studies of these changes provide an opportunity of separating the effects of inter- and intra-molecular scattering. This may make possible the direct Fourier analysis of the molecular structure once complete sets of reflections are available in different states of hydration.’ Subsequent work by Shannon (1949), Sayre (1952), Fienup (1982) and others opened the way to derive phases directly from such continuous diffraction patterns by the technique of ‘over-sampling’.



As demonstrated here by Chapman and colleagues, the molecular transform is directly accessible in translationally disordered crystals, using coherent X-rays. Unlike strong Bragg peaks, the shadowy diffuse pattern is often hidden just out of sight. The remarkable thing about the diffuse ‘speckle patterns’ is that the phase problem is directly solvable. One way to think about that, explains Chapman, is to note that the autocorrelation of an object, formed by the Fourier transform of the diffraction intensities, is completely measured. This is quite different from the situation for a crystal. The autocorrelation in that case is called the Patterson map. As crystallographers know, it is a map of all the pair correlations (difference vectors between atoms) in the crystal. For crystals, this knowledge of all the differences between atoms is not usually enough to determine their actual positions in space. The problem is that one cannot distinguish if a particular difference relates to two atoms in the same molecule or to one in one molecule and another in the neighbouring molecule.

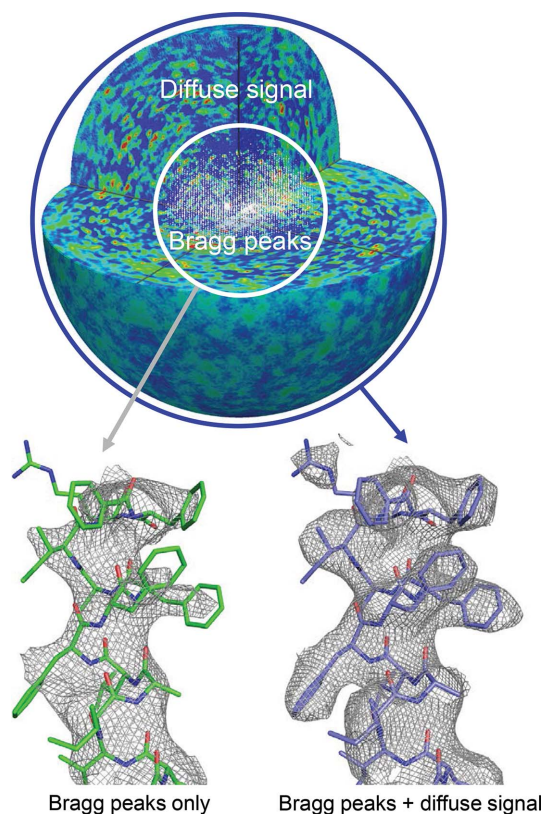


Figure 1

Lattice disorder reveals the continuous molecular transform of photo-system II and aids direct structure determination. The top image shows the entire three-dimensional diffraction volume of crystalline photo-system II, including Bragg peaks (in white at the centre) and diffuse scattering (blue–green), originating from molecules that do not follow strict lattice symmetry (about 2 Å r.m.s. deviation). This diffuse signal permits direct structure determination, using iterative phase retrieval methods developed for coherent diffractive imaging. As can be seen, the diffuse scatter extends well beyond the resolution of the Bragg peaks, because the signal does not depend on crystal quality in the same way as the Bragg peaks do. The electron density map shows substantial improvements (bottom right) relative to the map derived from the Bragg peaks (shown on the bottom left). Image components courtesy of Henry Chapman and Oleksandr Yefanov, CFEL, Hamburg.

The Patterson map is a truncated and aliased form of the autocorrelation of the unit cell. But when there is no crystal, only a single object, one can never be in doubt. If the object is 100 Å wide, it is clear that a pair correlation of 100 Å has to have those points located at the two extremes of the object. Such fully measured autocorrelation functions can be solved by a class of iterative algorithms – the simplest of which are akin to solvent flattening. Sayre managed to finally demonstrate this in a landmark paper headed by John Miao (Miao *et al.*, 1999).

With that background, and a flourishing of the field of coherent diffraction imaging with X-ray FELs, Chapman and colleagues were attuned to what properties to look for in diffraction of single objects. FELs brought such objects into the field of crystallography with the first experiments on nano- and microcrystals of photosystem II at the LCLS (Chapman *et al.*, 2011). These first experiments were actually designed as a way to test if ‘diffraction before destruction’ (Neutze *et al.*, 2000) holds at the atomic scale. The Bragg peaks of crystals would give a sensitive indication of whether the structure was intact, since in an explosion, correlation would be lost, killing Bragg peaks. Nanocrystallography worked, and it became a useful technique, but even in follow-up FEL experiments, photosystem II crystals never gave useful Bragg signals beyond about 4 or 5 Å resolution in the hands of Chapman and colleagues. Umena *et al.* (2011) showed, however, that photosystem II crystals could be coaxied into giving 1.9 Å resolution, suggesting the loss of diffraction was related to the way crystals were formed. If so, where was the diffraction going? The scattering of X-rays from atoms does not diminish just because those atoms are not in an ordered array, so if the X-rays were not making Bragg peaks they should still be detected at high resolutions in the same numbers.

After summing together thousands of diffraction patterns, oriented into a common frame of reference of the lattice with the help of the Bragg peaks, Yefanov *et al.* (2014) were able to obtain a strong speckle pattern, which extended out to high resolution beyond the Bragg peaks. The nature of the speckle pattern immediately suggested continuous diffraction from single objects, and constructing the autocorrelation function (by a Fourier transform of the three-dimensional intensity distribution) showed that these objects had the size and shape of the photosystem II dimer. Such diffraction would arise if there were random displacements of photosystem II molecules with about 2 Å r.m.s. The nature of the continuous diffraction was further confirmed by the fact that it could indeed be phased by the methods of coherent diffractive imaging to give an electron density map (Fig. 1), just starting with an envelope function taken from the low-resolution crystal structure. In the language of crystallographers, they generated an omit map where the entire molecule was omitted.

In their present paper, Chapman and colleagues make a detailed investigation of continuous diffraction of crystals with translational disorder and present strong statistical tests that indicate the nature of the speckle pattern. It is perhaps no surprise that these tests rely upon the statistics derived by Wilson, which have provided so much guidance over the years

to crystallographers. Speckle patterns follow the same negative exponential distributions of intensities as Bragg intensities, since both are coherent diffraction patterns of collections of atoms. For a translationally disordered protein crystal, the continuous diffraction can be thought of as the uncorrelated sum of the patterns of the rigid bodies in each of their orientations (incoherent addition). For the $P2_12_12_1$ symmetry of photosystem II crystals, there are four such orientations. The incoherent sum of their patterns is just like the incoherent sum of crystals in symmetry-related orientations, which occurs when crystals are twinned. The statistics of the continuous diffraction are modified as a result of this incoherent sum, and can be used to count the number of independent objects that contribute to the diffraction. This was found to be four in the case of photosystem II crystals, equal to the number of orientations, and quite distinct from the statistics of the Bragg peaks.

The solvent that carries and permeates the crystals gives a strong background that must be accurately subtracted for the phase retrieval method to converge, but it is not obvious how to subtract a slowly varying background from the continuous diffraction pattern. The authors extended the statistical description to include this component. They derived a fairly simple method to estimate the background, based only on the mean and variance of regions of the measured pattern, and showed a remarkable improvement of the extracted continuous diffraction as seen with a high degree of correlation between this and simulated diffraction.

The work is yet to be fully digested. At the moment there are some misunderstandings in the literature. For instance, in a recent review, Meisburger *et al.* (2017) seem to miss the point that, for two independent molecular orientations, one would get $|a|^2 + |b|^2$ for the continuous diffraction, whereas the Bragg peaks see the whole unit cell with $|a + b|^2$. These are not equal.

Ab initio phasing from continuous diffraction in crystallography requires further developments: extending to higher resolution, seeing whether this approach works with other proteins, how sensitive phasing is to different types of correlations that may occur in the displacements of molecules or other types of disorder, or whether one can in fact use the

extra information to give knowledge about conformational variability.

One interesting finding from the quantitative analysis is just how strong the continuous diffraction is, compared to the Bragg peaks. There were about four times as many X-ray photons contributing to the continuous diffraction than the Bragg peaks. The diffraction looks weak because it is spread over all pixels in the detector, not just into a few hundred Bragg peaks, but it is by no means insignificant. This extremely valuable information, which was hidden from view for so long, must be brought into the spotlight as a means to improve direct structure determination and reduce our reliance on *a priori* models.

Acknowledgements

I am grateful to Henry Chapman for discussions and to Oleksandr Yefanov for components of Fig. 1.

References

- Ayyer, K. *et al.* (2016). *Nature*, **530**, 202–206.
 Bernal, J. D., Fankuchen, I. & Perutz, M. F. (1938). *Nature*, **141**, 523–524.
 Chapman, H. N. *et al.* (2011). *Nature*, **470**, 73–77.
 Chapman, H. N., Yefanov, O. M., Ayyer, K., White, T. A., Barty, A., Morgan, A., Mariani, V., Oberthuer, D. & Pande, K. (2017). *J. Appl. Cryst.* **50**, 1084–1103.
 Fienup, J. R. (1982). *Appl. Opt.* **21**, 2758–2769.
 Maia, F. R. N. C., Ekeberg, T., Timneanu, N., van der Spoel, D. & Hajdu, J. (2009). *Phys. Rev. E*, **80**, 031905.
 Meisburger, S. P., Thomas, W. C., Watkins, M. B. & Ando, N. (2017). *Chem. Rev.* **117**, 7615–7672.
 Miao, J., Charalambous, P., Kirz, J. & Sayre, D. (1999). *Nature*, **400**, 342–344.
 Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. (2000). *Nature*, **406**, 752–757.
 Sayre, D. (1952). *Acta Cryst.* **5**, 843.
 Shannon, C. E. (1949). *Proc. Inst. Radio Eng.* **37**, 10–21.
 Umena, Y., Kawakami, K., Shen, J. R. & Kamiya, N. (2011). *Nature*, **473**, 55–60.
 Yefanov, O., Gati, C., Bourenkov, G., Kirian, R. A., White, T. A., Spence, J. C. H., Chapman, H. N. & Barty, A. (2014). *Philos. Trans. R. Soc. London Ser. B*, **369**, 20130333.