Emerging Topics in Life Sciences (2019) **3** 423–432
https://doi.org/10.1042/ETLS20180177

Review Article

Data-driven challenges and opportunities in crystallography

Calina Glynn and Jose A. Rodriguez

Department of Chemistry and Biochemistry, UCLA-DOE Institute for Genomics and Proteomics, University of California, Los Angeles (UCLA), Los Angeles, CA 90095, U.S.A

Correspondence: Jose A. Rodriguez (jrodriguez@mbi.ucla.edu)

Structural biology is in the midst of a revolution fueled by faster and more powerful instruments capable of delivering orders of magnitude more data than their predecessors. This increased pace in data gathering introduces new experimental and computational challenges, frustrating real-time processing and interpretation of data and requiring long-term solutions for data archival and retrieval. This combination of challenges and opportunities is driving the exploration of new areas of structural biology, including studies of macromolecular dynamics and the investigation of molecular ensembles in search of a better understanding of conformational landscapes. The next generation of instruments promises to yield even greater data rates, requiring a concerted effort by institutions, centers and individuals to extract meaning from every bit and make data accessible to the community at large, facilitating data mining efforts by individuals or groups as analysis tools improve.

Introduction

The crystallographic analysis of macromolecules is increasingly dependent on data-driven approaches [1]. Growing data output is a central feature of the new tools that facilitate crystallographic analysis. These include ultra-bright ultra-fast X-ray sources, upgraded synchrotron facilities and accessible instruments for routine electron diffraction. Single experiments can now yield gigabytes or terabytes of data from one or more samples allowing unprecedented exploration of ultra-fast processes or from increasingly small crystallites [2–4]. Here, we discuss two areas of rapid growth in structural biology that are increasingly reliant on large-scale data gathering: serial X-ray crystallography [1,5,6] and electron crystallography of micro and nanocrystals [7–10]. While other structural biology approaches are without a doubt also reliant on large datasets [11], data demands in crystallography are rapidly growing. We focus on these two closely related crystallographic approaches and their evolution from conventional data gathering to large-scale data collection and analysis.

While the two crystallographic approaches we highlight are somewhat similar, some key differences are important to note. Most of the X-ray crystallography measurements we describe are performed in a small number of large, government-funded facilities with staff and computational infrastructure matched to the scale of the experiment. In contrast, electron diffraction efforts are generally driven by individual groups and are implemented across a large number of small-scale facilities. Nonetheless, data collected by each approach is largely sparse and must be reduced to a small set of measurements for structure determination. The focus of this review is, therefore, the challenge of data reduction and the extraction of information from large sets of individual diffraction patterns belonging to one or more crystals.

Conventional X-ray crystallography efforts set the stage for the analysis of this type of data by establishing methods for routine data collection and processing from macro-scale crystals. The success of these approaches has yielded a consistently high number of structures deposited in the protein data bank (PDB) per year, a number that remains unmatched by any other structural biology method (Figure 1). Likewise, efforts in electron cryo-microscopy (cryoEM) [11–19] have established analogous

Received: 1 May 2019
Revised: 13 June 2019
Accepted: 24 June 2019

Version of Record published:
26 July 2019

© 2019 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and the Royal Society of Biology and distributed under the Creative Commons Attribution License 4.0 (CC BY).

Emerging Topics in Life Sciences (2019) **3** 423–432
https://doi.org/10.1042/ETLS20180177

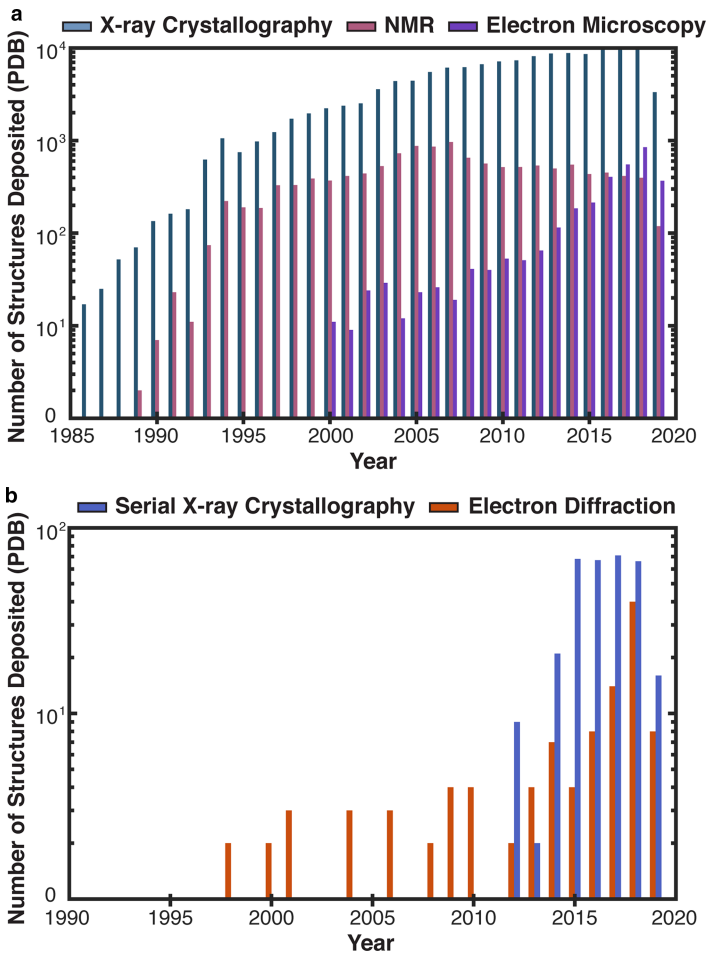

Figure 1. Number of structures deposited in the PDB per year [20].

(**a**) Structures deposited per year from 1985 to 2019 associated with all X-ray crystallographic methods, compared with nuclear magnetic resonance (NMR) and electron cryo-microscopy (cryoEM). Each counted structure is associated with X-ray crystallography (blue), NMR (pink) or electron microscopy (magenta). (**b**) A subset of structures deposited per year from 1990 to 2019 associated with serial X-ray crystallography or electron crystallography, including MicroED. Each counted structure is associated with data collection by serial X-ray crystallography methods, including those acquired at X-ray free electron laser facilities (purple), compared with those acquired by electron crystallography methods (orange).

pipelines for structure determination [11,19] that directly benefit electron crystallographic approaches [8], giving rise to a growing number of deposited structures per year (Figure 1) as catalogued by the protein data bank (PDB) [20].

The rise in structures determined by these new methods is accompanied by increasing challenges associated with storage, analysis and accessibility of data [20–23]. The Electron Microscopy Public Image Archive has been established to facilitate the public archival of electron microscopy and diffraction data [22]. Likewise, data repositories like the coherent X-ray imaging data bank are providing resources for depositing the terabytes of data collected at X-ray free electron laser facilities [23]. As the use of fast detectors grows routine in crystallography, the number of sites worldwide collecting large datasets will place additional demands on these and analogous resources.

Advances in X-ray diffraction

The advent of high brightness sources for crystallographic study of macromolecules took a quantum leap with the development and introduction of ultra-bright, pulsed X-ray lasers [6,24,25]. While initially developed with

© 2019 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and the Royal Society of Biology and distributed under the Creative Commons Attribution License 4.0 (CC BY).

the goal of determining structures of single molecules, acquiring a signal before inducing a coulombic explosion of each molecule [26,27], X-ray lasers have opened new doors for the crystallographic study of proteins and nucleic acids [1,6,28]. The first structure published by this approach was that of photosystem I, for which more than 3 million diffraction patterns were collected [2]. Among the first demonstrations of the feasibility of this approach was an effort to demonstrate the routine ability to determine structures of proteins from crystal slurries (Figure 2). This was demonstrated with the structure of lysozyme determined from a slurry of microcrystals [29], although many other sample delivery strategies have since been demonstrated [30] (Figure 2). This new kind of experiment, termed serial femtosecond crystallography (SFX) [2,6], allows the detection of high-resolution diffraction from individual microcrystals micrometers on edge. Lysozyme crystals were delivered to the 9.4 keV X-ray beam via a focused jet of liquid with a 10-micrometer focus, and out of 1.5 million individual diffraction frames collected, about 12 000 were indexed and integrated to yield a high-resolution structure [29]. Diffraction frames measured the interaction of crystals with individual 40 or 5 femtosecond pulses at a rate of 120 Hz, yielding a data stream of gigabytes per minute [29]. This approach has now been applied to a whole variety of samples, including entirely unknown structures, some of whose crystals are grown in cells [31–38]. A key feature of this data-driven crystallographic approach is the ability to probe dynamics, capturing the atomic scale motions within proteins [28,32,39,40].

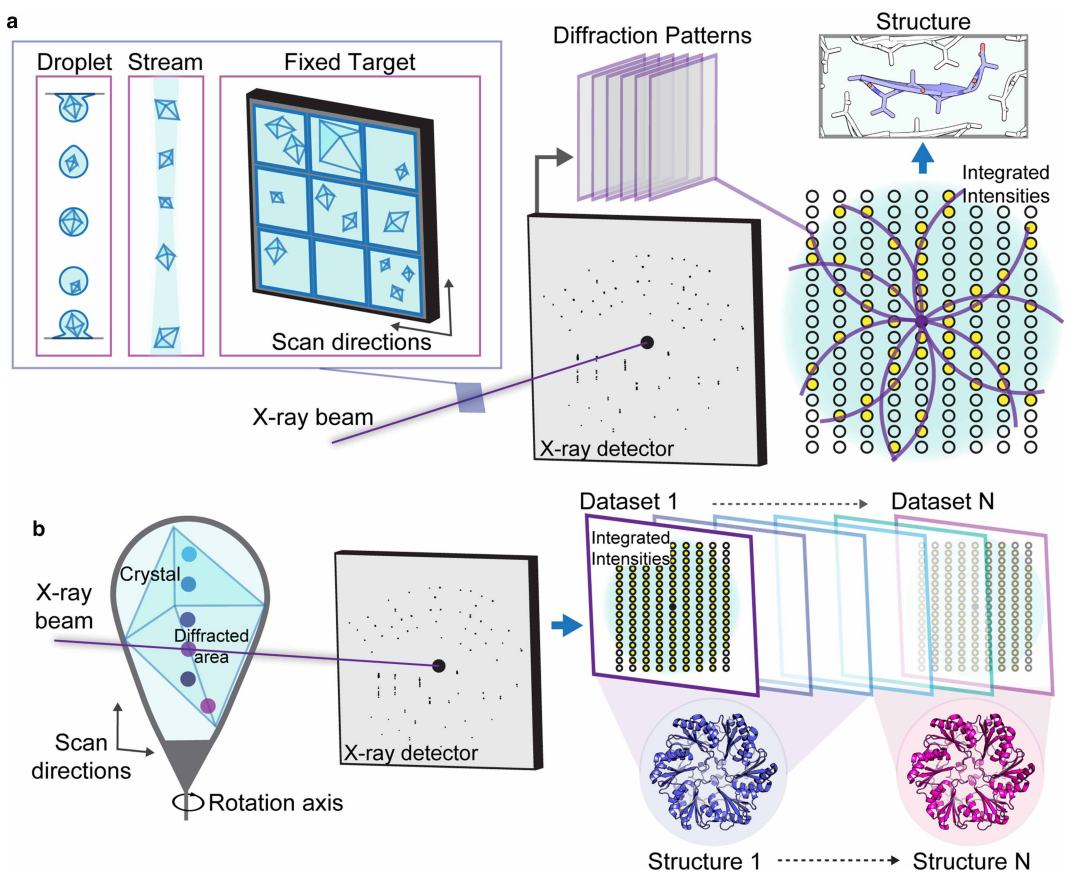

Figure 2. Examples of data gathering by serial X-ray crystallography.

(**a**) Ultra-fast pulses of high energy X-rays are directed at a sample interaction region to which samples can be delivered by varying means. Three methods of sample delivery are pictured: droplet-based delivery, stream or liquid-jet delivery, and fixed target delivery, typically by a windowed device. Data from a train of patterns recorded from different crystals are combined to yield one or more datasets that are used to determine a molecular structure. (**b**) Approaches that pursue multiple structures from a single crystal can train an X-ray beam on particular locations on a crystal, collecting multiple datasets that can then be combined or individually yield a structure. In this way, several structures can be obtained from a single crystal or a set of similar crystals. This example illustrates determination of the monoclinic structure of EutL, PDB ID 4TLH.

© 2019 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and the Royal Society of Biology and distributed under the Creative Commons Attribution License 4.0 (CC BY).

Emerging Topics in Life Sciences (2019) **3** 423–432
https://doi.org/10.1042/ETLS20180177

The latest generation of ultra-fast X-ray sources offers even greater throughput [41]. The megahertz X-ray free electron laser facility, part of the European XFEL project, delivers 10 pulse trains of megahertz X-ray sequences delivering up to 27 000 pulses per second [41]. However, efficiently utilizing a torrent of pulses presents a challenge for real-time processing and data storage. Diffraction produced by 50-pulse trains impinging on crystals in a liquid jet every 100 milliseconds was captured by a megapixel Adaptive Gain Integrating Pixel Detector [41,42]. A jet with flow rates of ∼50 m/s allows the fresh sample to be probed by the series of pulses in each train and the recording of individual diffraction patterns within a train, with pulses in a train separated by ∼900 ns [41].

Serial femtosecond crystallography inspired the adaptation of methods and technologies for serial data collection efforts at synchrotron beamlines [43–45]. There, millisecond timescale shutterless detection of the interaction between focused X-ray beams and a high viscosity crystal matrix or crystals held in fixed target mounts has facilitated routine room-temperature determination of structures from thousands of diffraction patterns within minutes to hours [43–46]. The large, fast detectors used in some of these experiments accelerate data collection [47]; 18-megapixel images collected at high frame rates (50 Hz) allow the collection of a single dataset in under 90 min [44]. Collection of room-temperature data at this rate in an accessible synchrotron beamline has enabled high-resolution structure determination [45,48], including determination of structures *de novo* and with bound ligands [44]. Acquisition frame rates of 50 Hz appear to improve the signal to noise, particularly at high resolution by allowing the collection of multiple diffraction patterns per crystal and the elimination of background from empty frames [44]. This was particularly advantageous for *de novo* structure determination through the improved measurement of the anomalous signal [44]. Importantly, information about ligand binding could be acquired from as few as 900 images (9 Gb) collected over a period of 5 min [44]. This opens the door to large-scale screening efforts for ligand binding at dedicated serial crystallography beamlines. This may be even further enabled by the development of compact X-ray sources that deliver high-repetition bright X-ray beams in a facility with the footprint of a conventional laboratory [49].

A different type of serial crystallography experiment uses bright X-ray sources to expand the amount of data collected from single crystals and extract more information out of each (Figure 2) [48,50,51]. These strategies train an X-ray beam on particular regions of a crystal or collections of crystals, collecting a partial or full dataset at each location (Figure 2). The data collected at each location can then be indexed and integrated, and the data between sampled locations correlated to yield multiple datasets per crystal [51]. This approach is especially advantageous when radiation damage is a challenge, and helps explore the possible conformational heterogeneity present within macro-scale crystals that may represent structural variations in their constituent proteins [52]. In the simplest case, regions within a crystal can be non-isomorphous and distinguished by varying unit cell parameters and described by a lattice distortion [51]. Combining data within isomorphous sets then allows for a single structure to be determined per set (Figure 2), a comparison of the structures reveals conformational differences that explain the lattice distortion [51]. With greater data sampling, the potential to mine differences in crystal structures in search for conformational heterogeneity becomes increasingly possible. These approaches, generally referred to as multiple structures from one crystal (MSOX), are not limited to spatial heterogeneity in crystals, but can also add a temporal component to assess changes induced by radiation [50,53] or changes due to environmental differences or molecular actions [54].

Advances in electron diffraction

Improvements to detector technology have played a major role in the interrogation of macromolecular structures from thin three-dimensional crystals by dramatically increasing the rate of accurate data gathering [9,10,55]. The field of electron crystallography of three-dimensional microcrystallites (MicroED) is a marriage of cryoEM sample manipulation and data collection with X-ray crystallography analysis for routine structure determination of macromolecular structures [8]. The MicroED technique, initially demonstrated on the well-known protein standard lysozyme [9], has now been successful in determining a variety of structures from small molecules to macromolecular assemblies (Figure 1) [56–58]. To obtain electron diffraction from 3D protein crystals, approaches like MicroED illuminate a sub-micron thick crystal with an extremely low-dose electron beam (Figure 3). The signal from the illuminated crystal can be isolated from that of its surroundings by means of a selected area aperture (Figure 3). Diffraction is then sampled from the selected area and recorded either as a movie from a continuously rotating crystal (Figure 3) [59] or as a series of precession photographs at discrete tilt angles [60]. Data is reduced by conventional crystallography approaches to yield integrated intensities that can cover part or all of the reciprocal lattice, depending on crystal symmetry and the fraction of

© 2019 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and the Royal Society of Biology and distributed under the Creative Commons Attribution License 4.0 (CC BY).

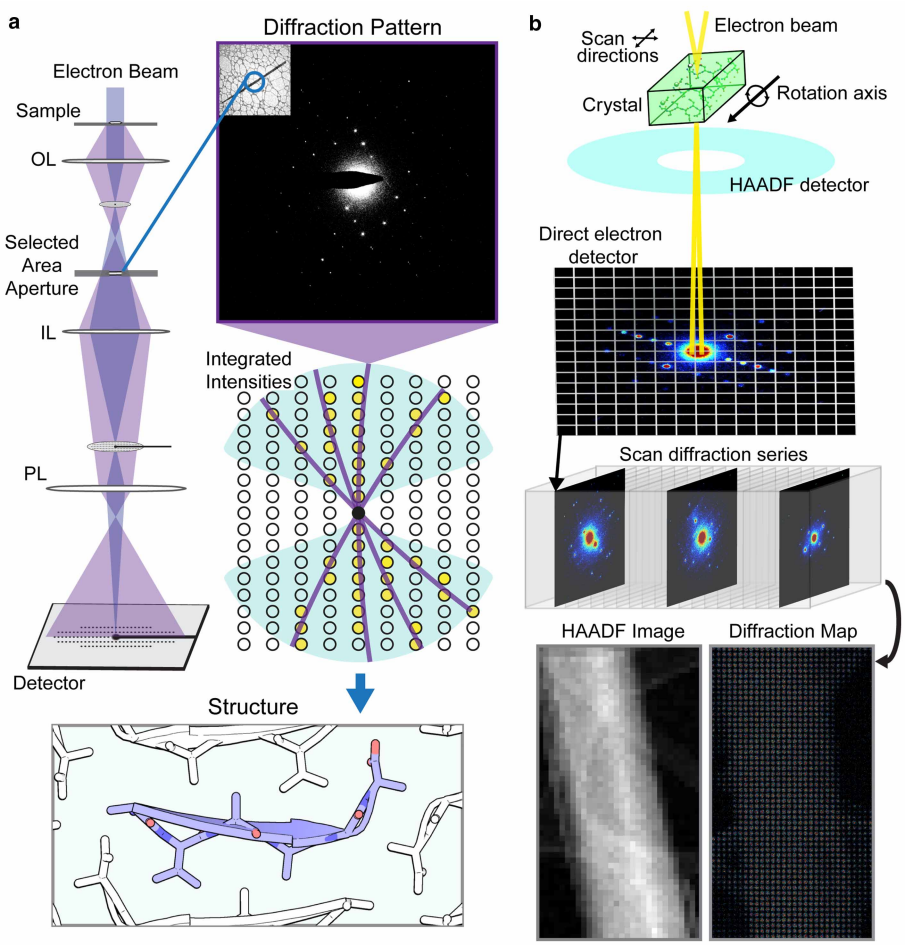

Figure 3. Examples of data gathering by selected area or scanning electron diffraction.
(**a**) In a conventional electron diffraction experiment, a crystal and its surroundings are illuminated with a low-dose electron beam. A selected area aperture samples diffraction from a portion or all of the illuminated crystal; diffraction patterns are sampled continuously as the crystal is unidirectionally rotated within the electron beam. Integrated intensities from these patterns populate a reciprocal lattice that is phased to determine a molecular structure. OL, objective lens; IL, intermediate lenses; PL, projection lenses. (**b**) In scanning nanobeam electron diffraction, a focused electron beam is scanned across a sample while diffraction patterns are recorded. Electrons scattered to high angle at each scan point are integrated by a dark field detector (HAADF) to yield a visual representation of the sample, while electrons scattered to a lower angle travel through a hole in that detector and are recorded by a fast readout, pixelated electron counting detector to yield a diffraction map of the scanned region. Both panels show data collected from nanocrystals of a hexapeptide.

Bragg reflections obstructed by the supporting grid and holder (Figure 3). Structures are determined from this data either by direct methods [61–63], when the resolution of the data is atomic (~1.2 Å) or by molecular replacement when a suitable probe exists [8]. Over one hundred macromolecular structures have now been deposited in the PDB, determined by electron diffraction since 1990 (Figure 1). Nearly half of these are determined from 3D protein crystals, and add to a growing number of small molecule structures [64–67].

While the number of structures deposited into the PDB by MicroED is miniscule in comparison with X-ray crystallography efforts (Figure 1), growth in the number of sites collecting MicroED data and the implementation of automation will pose a challenge for efforts promoting archival and access to raw data. This challenge is most prominent for diffraction data collected on fast electron detectors [3,47], especially where multiple datasets must be combined to yield information about a single structure. The simplest of these experiments is one where data from several tilt series is merged to create a single high completeness dataset [8,68], but even this effort can be a challenge since the specific combination of datasets that yields the best merge is not known

© 2019 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and the Royal Society of Biology and distributed under the Creative Commons Attribution License 4.0 (CC BY).

Emerging Topics in Life Sciences (2019) **3** 423–432
https://doi.org/10.1042/ETLS20180177

a priori. In cases where polymorphism is high, and diffraction quality is highly variable between crystals, obtaining a quality merge may require a collection of data from dozens or more individual crystals. Widespread adoption of this approach necessitates methods for automation in data collection and processing [69,70]. Partial automation is already facilitating more complex experiments, including scanning nanobeam diffraction where thousands of patterns are collected at a given orientation from a single crystallite (Figure 3) [3,71], and serial crystallography experiments that combine diffraction from hundreds or thousands of crystallites using SerialED software [69].

Scanning diffraction experiments are particularly data rich [3]. In a scanning electron diffraction experiment, a focused electron beam is positioned at specific locations across a sample in a defined pattern (Figure 3), analogous to the experiments performed when scanning an X-ray beam across a much larger crystal (Figure 2) [51]. While the beam is positioned at each location for only a small fraction of a second, a diffraction pattern is collected on a fast detector [72]. Data volume in scanning diffraction scales varies with the number of individual scan steps. Depending on the size of a scan area and the density of locations sampled, this procedure can generate tens of thousands of diffraction patterns per scan, multiplied by the number of crystallite tilt angles sampled (Figure 3). Data in each scan is plentiful, but information is sparse; each diffraction pattern in a scan has few Bragg reflections, and each reflection may only be partially sampled [3]. This makes averaging over similar patterns advantageous. Ultimately, all patterns containing a usable signal could be combined into a single set of integrated intensities representing a subset or the whole of reciprocal space. By spatially linking reciprocal space information to specific locations on a crystal, scanning diffraction offers a rich source of information unavailable in a conventional crystallography experiment [3,51]. While a simple analysis of total diffracted intensity can reveal regions of strong or weak diffraction within a crystal, more subtle changes in the diffracted signal can be extracted by algorithms that compare deviations across patterns. Comparison of patterns across a scan can ultimately reveal changes in lattice orientation [3], and potentially changes in unit cell parameters or perhaps changes in the structure of molecules in a crystal.

The role of detectors in big data gathering

The use of advanced detectors for photon and electron counting is expediting data collection in both X-ray and electron crystallography, facilitating the interrogation of crystals with short exposure times using high-flux sources. An outstanding challenge is posed by the rapid readout Adaptive Gain Integrating Pixel Detector [42], a system capable of delivering bursts of 350 images with a 6.5-mHz frame rate [41,73]. This and other fast detectors are now key to enabling new efforts in serial crystallography [43], where their high dynamic range and sensitivity allow rapid readout from single X-ray pulses or short electron beam exposures [43,47,74].

Electron microscopy as a whole has largely benefitted from the development of advanced detectors, first by the introduction of slow readout charged couple device (CCD) sensors [75], then by faster active pixel sensors including the complementary metal-oxide semiconductors (CMOS) including direct electron detectors [76,77]. Each improvement in detector technology has grown data output, facilitating not only improved speed and accuracy of data collection but also more efficient data archival and processing [77]. While the readout rates of several electron detectors are now in the kilohertz range [65,74,76], most experiments are not performed in this regime. Fast readout is instead used to preserve resolution in electron microscopy images by limiting the influence of motion induced artifacts [78]. In continuous rotation electron diffraction, readout rates are adjusted to limit overexposure due to brightly diffracted beams, limit readout noise and ensure adequate sampling of the reciprocal lattice (Figure 3) [8,68,74].

Performing electron diffraction of single crystals at kilohertz frame rates could allow the collection of a 140-degree wedge of reciprocal space while sampling a quarter of a degree per frame in approximately half a second. However, the use of direct electron detectors for measuring diffraction from macromolecular crystals not only facilitates rapid, low-noise, low-dose data collection but can also increase the data rate per experiment by over an order of magnitude. This startling pace of acquisition would have a dramatic effect on our ability to rapidly screen samples, probe heterogeneity in crystallites and ultimately sample more experimental parameters. However, the collection of 14-megapixel, 8-bit images at 400 frames per second yields a torrential 5.6 gigabytes of raw data per second. In line conversion of this data from its archival format to one suitable for crystallographic analysis is the first of many bottlenecks. Since crystallographic data is highly sparse, the greatest improvement to processing of kilohertz diffraction will come from algorithms for rapid and automated data reduction [68,69].

© 2019 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and the Royal Society of Biology and distributed under the Creative Commons Attribution License 4.0 (CC BY).

Toward real-time processing

Algorithms for accurate data reduction from sparse diffraction data are key enabling technologies for emerging crystallographic experiments. Rapid hit-finding and indexing algorithms are routinely used in X-ray free electron laser facilities during serial crystallography experiments [79–85], where the rate of image collection far outpaces the amount of data that can be parsed manually in real time. Algorithms used to parse 120 Hz data generated by the Cornell-SLAC Pixel Array Detector can operate concurrently with data collection, giving the crystallographer a notion of how many frames have usable diffraction and how many reflections are present in a frame [81,83–86]. Many of these same packages can create and refine experimental parameter models, index reflections and integrate reflection intensities. Programs such as KAMO [87] and Blend [88] facilitate the identification of consistent unit cells and evaluate the capacity of data subsets to be merged. Their algorithms rely on the relationships between cell parameters or intensity correlations to select suitable merge sets. Whilst a similar feat of automated data processing has only recently been demonstrated in electron diffraction [69]. Although no fundamental hurdles limit the full applicability of these methods to continuous rotation, serial or nanobeam electron diffraction, sparse single electron diffraction patterns may be especially difficult to index without sufficient prior constraints on cell dimension and experimental geometry. The need for a centralized access to a comprehensive set of tools and protocols for analysis of large crystallographic datasets will continue to be underscored by the continued and rapid development of sources, instruments, methods, programs and algorithms [89].

Summary

- New crystallographic approaches are facing new challenges and opportunities delivered by data-rich experiments facilitated by the development of new ultra-fast detectors.
- Electron crystallography experiments are following a similar trend, where faster detectors are allowing the rapid collection of information-rich data from single crystals or collections of crystals.
- New algorithms for automated data collection and analysis are leveraging the rise in data production to yield faster structure solutions, and new insights into the macromolecular structure at the time and storage space limits.

Abbreviations

CCD, charged couple device; CMOS, complementary metal-oxide semiconductors; MSOX, multiple structures from one crystal; MicroED, microcrystallites; SFX, serial femtosecond crystallography.

Funding

This work is supported by DOE Grant DE-FC02-02ER63421, NSF Grant DMR-1548924 and the NIH-NIGMS Grant R35 GM128867. J.A.R. is supported as a Searle Scholar a Pew Scholar and a Beckman Young Investigator.

Acknowledgements

We thank Marcus Gallagher-Jones (UCLA) for his help in preparing figures and contributions to the writing of this manuscript, and Johan Hattne (UCLA) for critical discussion.

Competing Interests

The Authors declare that there are no competing interests associated with the manuscript.

References

1 Martin-Garcia, J.M., Conrad, C.E., Coe, J., Roy-Chowdhury, S. and Fromme, P. (2016) Serial femtosecond crystallography: a revolution in structural biology. *Arch. Biochem. Biophys.* **602**, 32–47 https://doi.org/10.1016/j.abb.2016.03.036

© 2019 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and the Royal Society of Biology and distributed under the Creative Commons Attribution License 4.0 (CC BY).

Emerging Topics in Life Sciences (2019) **3** 423–432
https://doi.org/10.1042/ETLS20180177

2 Chapman, H.N., Fromme, P., Barty, A., White, T.A., Kirian, R.A., Aquila, A. et al. (2011) Femtosecond X-ray protein nanocrystallography. *Nature* **470**, 73–77 https://doi.org/10.1038/nature09750
3 Gallagher-Jones, M., Ophus, C., Bustillo, K.C., Boyer, D.R., Panova, O., Glynn, C. et al. (2019) Nanoscale mosaicity revealed in peptide microcrystals by scanning electron nanodiffraction. *Commun. Biol.* **2**, 26 https://doi.org/10.1038/s42003-018-0263-8
4 Gallagher-Jones, M., Rodriguez, J.A. and Miao, J. (2016) Frontier methods in coherent X-ray diffraction for high-resolution structure determination. *Q. Rev. Biophys.* **49**, e20 https://doi.org/10.1017/S0033583516000147
5 Johansson, L.C., Stauch, B., Ishchenko, A. and Cherezov, V. (2017) A bright future for serial femtosecond crystallography with XFELs. *Trends Biochem. Sci.* **42**, 749–762 https://doi.org/10.1016/j.tibs.2017.06.007
6 Schlichting, I. (2015) Serial femtosecond crystallography: the first five years. *IUCrJ* **2**, 246–255 https://doi.org/10.1107/S205225251402702X
7 Clabbers, M.T.B., van Genderen, E., Wan, W., Wiegers, E.L., Gruene, T. and Abrahams, J.P. (2017) Protein structure determination by electron diffraction using a single three-dimensional nanocrystal. *Acta Crystallogr. D Struct. Biol.* **73**, 738–748 https://doi.org/10.1107/S2059798317010348
8 Shi, D., Nannenga, B.L., de la Cruz, M.J., Liu, J., Sawtelle, S., Calero, G. et al. (2016) The collection of MicroED data for macromolecular crystallography. *Nat. Protoc.* **11**, 895–904 https://doi.org/10.1038/nprot.2016.046
9 Shi, D., Nannenga, B.L., Iadanza, M.G. and Gonen, T. (2013) Three-dimensional electron crystallography of protein microcrystals. *eLife* **2**, e01345 https://doi.org/10.7554/eLife.01345
10 Yonekura, K., Kato, K., Ogasawara, M., Tomita, M. and Toyoshima, C. (2015) Electron crystallography of ultrathin 3D protein crystals: atomic model with charges. *Proc. Natl Acad. Sci. U.S.A.* **112**, 3368–3373 https://doi.org/10.1073/pnas.1500724112
11 Baldwin, P.R., Tan, Y.Z., Eng, E.T., Rice, W.J., Noble, A.J., Negro, C.J. et al. (2018) Big data in cryoEM: automated collection, processing and accessibility of EM data. *Curr. Opin. Microbiol.* **43**, 1–8 https://doi.org/10.1016/j.mib.2017.10.005
12 Glaeser, R.M. (2008) Cryo-electron microscopy of biological nanostructures. *Phys. Today* **61**, 48 https://doi.org/10.1063/1.2835153
13 Taylor, K.A. and Glaeser, R.M. (1974) Electron diffraction of frozen, hydrated protein crystals. *Science* **186**, 1036–1037 https://doi.org/10.1126/science.186.4168.1036
14 Taylor, K.A. and Glaeser, R.M. (1976) Electron microscopy of frozen hydrated biological specimens. *J. Ultrastruct. Res.* **55**, 448–456 https://doi.org/10.1016/S0022-5320(76)80099-8
15 Henderson, R., Baldwin, J.M., Ceska, T.A., Zemlin, F., Beckmann, E. and Downing, K.H. (1990) Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J. Mol. Biol.* **213**, 899–929 https://doi.org/10.1016/S0022-2836(05)80271-2
16 Adrian, M., Dubochet, J., Lepault, J. and McDowall, A.W. (1984) Cryo-electron microscopy of viruses. *Nature* **308**, 32–36 https://doi.org/10.1038/308032a0
17 Dubochet, J., Adrian, M., Chang, J.-J., Homo, J.-C., Lepault, J., McDowall, A.W. et al. (1988) Cryo-electron microscopy of vitrified specimens. *Q. Rev. Biophys.* **21**, 129–228 https://doi.org/10.1017/S0033583500004297
18 Cheng, Y., Grigorieff, N., Penczek, P.A. and Walz, T. (2015) A primer to single-particle cryo-electron microscopy. *Cell* **161**, 438–449 https://doi.org/10.1016/j.cell.2015.03.050
19 Baker, M.L., Zhang, J., Ludtke, S.J. and Chiu, W. (2010) Cryo-EM of macromolecular assemblies at near-atomic resolution. *Nat. Protoc.* **5**, 1697–1708 https://doi.org/10.1038/nprot.2010.126
20 Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H. et al. (2000) The protein data bank. *Nucleic Acids Res.* **28**, 235–242 https://doi.org/10.1093/nar/28.1.235
21 Lawson, C.L., Baker, M.L., Best, C., Bi, C., Dougherty, M., Feng, P. et al. (2011) EMDatabank.org: unified data resource for CryoEM. *Nucleic Acids Res.* **39**, D456–D464 https://doi.org/10.1093/nar/gkq880
22 Iudin, A., Korir, P.K., Salavert-Torres, J., Kleywegt, G.J. and Patwardhan, A. (2016) EMPIAR: a public archive for raw electron microscopy image data. *Nat. Methods* **13**, 387–388 https://doi.org/10.1038/nmeth.3806
23 Maia, F.R.N.C. (2012) The coherent X-ray imaging data bank. *Nat. Methods* **9**, 854–855 https://doi.org/10.1038/nmeth.2110
24 Pellegrini, C. (2016) X-ray free-electron lasers: from dreams to reality. *Phys. Scr. T* **169**, 014004 https://doi.org/10.1088/1402-4896/aa5281
25 Schlichting, I. and Miao, J. (2012) Emerging opportunities in structural biology with X-ray free-electron lasers. *Curr. Opin. Struct. Biol.* **22**, 613–626 https://doi.org/10.1016/j.sbi.2012.07.015
26 Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. and Hajdu, J. (2000) Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **406**, 752–757 https://doi.org/10.1038/35021099
27 Barty, A., Caleman, C., Aquila, A., Timneanu, N., Lomb, L., White, T.A. et al. (2012) Self-terminating diffraction gates femtosecond X-ray nanocrystallography measurements. *Nat. Photon.* **6**, 35–40 https://doi.org/10.1038/nphoton.2011.297
28 Spence, J.C.H., Weierstall, U. and Chapman, H.N. (2012) X-ray lasers for structural and dynamic biology. *Rep. Prog. Phys.* **75**, 102601 https://doi.org/10.1088/0034-4885/75/10/102601
29 Boutet, S., Lomb, L., Williams, G.J., Barends, T.R., Aquila, A., Doak, R.B. et al. (2012) High-resolution protein structure determination by serial femtosecond crystallography. *Science* **337**, 362–364 https://doi.org/10.1126/science.1217737
30 Martiel, I., Müller-Werkmeister, H.M. and Cohen, A.E. (2019) Strategies for sample delivery for femtosecond crystallography. *Acta Crystallogr. D Struct. Biol.* **75**, 160–177 https://doi.org/10.1107/S2059798318017953
31 Johansson, L.C., Arnlund, D., Katona, G., White, T.A., Barty, A., DePonte, D.P. et al. (2013) Structure of a photosynthetic reaction centre determined by serial femtosecond crystallography. *Nat. Commun.* **4**, 2911 https://doi.org/10.1038/ncomms3911
32 Kupitz, C., Basu, S., Grotjohann, I., Fromme, R., Zatsepin, N.A., Rendek, K.N. et al. (2014) Serial time-resolved crystallography of photosystem II using a femtosecond X-ray laser. *Nature* **513**, 261–265 https://doi.org/10.1038/nature13453
33 Liu, W., Wacker, D., Gati, C., Han, G.W., James, D., Wang, D. et al. (2013) Serial femtosecond crystallography of G protein-coupled receptors. *Science* **342**, 1521–1524 https://doi.org/10.1126/science.1244142
34 Zhang, H., Unal, H., Gati, C., Han, G.W., Liu, W., Zatsepin, N.A. et al. (2015) Structure of the angiotensin receptor revealed by serial femtosecond crystallography. *Cell* **161**, 833–844 https://doi.org/10.1016/j.cell.2015.04.011
35 Sawaya, M.R., Cascio, D., Gingery, M., Rodriguez, J., Goldschmidt, L., Colletier, J.-P. et al. (2014) Protein crystal structure obtained at 2.9 Å resolution from injecting bacterial cells into an X-ray free-electron laser beam. *Proc. Natl Acad. Sci. U.S.A.* **111**, 12769–12774 https://doi.org/10.1073/pnas.1413456111

© 2019 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and the Royal Society of Biology and distributed under the Creative Commons Attribution License 4.0 (CC BY).

36 Duszenko, M., Redecke, L., Mudogo, C.N., Sommer, B.P., Mogk, S., Oberthuer, D. et al. (2015) *In vivo* protein crystallization in combination with highly brilliant radiation sources offers novel opportunities for the structural analysis of post-translationally modified eukaryotic proteins. *Acta Crystallogr. F Struct. Biol. Commun.* **71**, 929–937 https://doi.org/10.1107/S2053230X15011450
37 Redecke, L., Nass, K., DePonte, D.P., White, T.A., Rehders, D., Barty, A. et al. (2013) Natively inhibited *Trypanosoma brucei* cathepsin B structure determined by using an X-ray laser. *Science* **339**, 227–230 https://doi.org/10.1126/science.1229663
38 Colletier, J.-P., Sawaya, M.R., Gingery, M., Rodriguez, J.A., Cascio, D., Brewster, A.S. et al. (2016) *De novo* phasing with X-ray laser reveals mosquito larvicide BinAB structure. *Nature* **539**, 43–47 https://doi.org/10.1038/nature19825
39 Tenboer, J., Basu, S., Zatsepin, N., Pande, K., Milathianaki, D., Frank, M. et al. (2014) Time-resolved serial crystallography captures high-resolution intermediates of photoactive yellow protein. *Science* **346**, 1242–1246 https://doi.org/10.1126/science.1259357
40 Coquelle, N., Sliwa, M., Woodhouse, J., Schirò, G., Adam, V., Aquila, A. et al. (2018) Chromophore twisting in the excited state of a photoswitchable fluorescent protein captured by time-resolved serial femtosecond crystallography. *Nat. Chem.* **10**, 31–37 https://doi.org/10.1038/nchem.2853
41 Grünbein, M.L., Bielecki, J., Gorel, A., Stricker, M., Bean, R., Cammarata, M. et al. (2018) Megahertz data collection from protein microcrystals at an X-ray free-electron laser. *Nat. Commun.* **9**, 3487 https://doi.org/10.1038/s41467-018-05953-4
42 Henrich, B., Becker, J., Dinapoli, R., Goettlicher, P., Graafsma, H., Hirsemann, H. et al. (2011) The adaptive gain integrating pixel detector AGIPD a detector for the European XFEL. *Nucl. Instrum. Methods Phys. Res. A* **633**, S11–S14 https://doi.org/10.1016/j.nima.2010.06.107
43 Martin-Garcia, J.M., Conrad, C.E., Nelson, G., Stander, N., Zatsepin, N.A., Zook, J. et al. (2017) Serial millisecond crystallography of membrane and soluble protein microcrystals using synchrotron radiation. *IUCrJ* **4**, 439–454 https://doi.org/10.1107/S205225251700570X
44 Weinert, T., Olieric, N., Cheng, R., Brünle, S., James, D., Ozerov, D. et al. (2017) Serial millisecond crystallography for routine room-temperature structure determination at synchrotrons. *Nat. Commun.* **8**, 542 https://doi.org/10.1038/s41467-017-00630-4
45 Botha, S., Nass, K., Barends, T.R.M., Kabsch, W., Latz, B., Dworkowski, F. et al. (2015) Room-temperature serial crystallography at synchrotron X-ray sources using slowly flowing free-standing high-viscosity microstreams. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 387–397 https://doi.org/10.1107/S1399004714026327
46 Ebrahim, A., Appleby, M.V., Axford, D., Beale, J., Moreno-Chicano, T., Sherrell, D.A. et al. (2019) Resolving polymorphs and radiation-driven effects in microcrystals using fixed-target serial synchrotron crystallography. *Acta Crystallogr. D Struct. Biol.* **75**, 151–159 https://doi.org/10.1107/S2059798318010240
47 Casanas, A., Warshamanage, R., Finke, A.D., Panepucci, E., Olieric, V., Nöll, A. et al. (2016) EIGER detector: application in macromolecular crystallography. *Acta Crystallogr. D Struct. Biol.* **72**, 1036–1048 https://doi.org/10.1107/S2059798316012304
48 Coquelle, N., Brewster, A.S., Kapp, U., Shilova, A., Weinhausen, B., Burghammer, M. et al. (2015) Raster-scanning serial protein crystallography using micro- and nano-focused synchrotron beams. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 1184–1196 https://doi.org/10.1107/S1399004715004514
49 Graves, W.S., Bessuille, J., Brown, P., Carbajo, S., Dolgashev, V., Hong, K.-H. et al. (2014) Compact x-ray source based on burst-mode inverse Compton scattering at 100 kHz. *Phys. Rev. ST Accel. Beams* **17**, 120701 https://doi.org/10.1103/PhysRevSTAB.17.120701
50 Horrell, S., Antonyuk, S.V., Eady, R.R., Hasnain, S.S., Hough, M.A. and Strange, R.W. (2016) Serial crystallography captures enzyme catalysis in copper nitrite reductase at atomic resolution from one crystal. *IUCrJ* **3**, 271–281 https://doi.org/10.1107/S205225251600823X
51 Thompson, M.C., Cascio, D. and Yeates, T.O. (2018) Microfocus diffraction from different regions of a protein crystal: structural variations and unit-cell polymorphism. *Acta Crystallogr. D Struct. Biol.* **74**, 411–421 https://doi.org/10.1107/S2059798318003479
52 van den Bedem, H. and Fraser, J.S. (2015) Integrative, dynamic structural biology at atomic resolution–it's about time. *Nat. Methods* **12**, 307–318 https://doi.org/10.1038/nmeth.3324
53 Horrell, S., Kekilli, D., Sen, K., Owen, R.L., Dworkowski, F.S.N., Antonyuk, S.V. et al. (2018) Enzyme catalysis captured using multiple structures from one crystal at varying temperatures. *IUCrJ* **5**, 283–292 https://doi.org/10.1107/S205225251800386X
54 Ren, Z., Chan, P.W.Y., Moffat, K., Pai, E.F., Royer, W.E., Šrajer, V. et al. (2013) Resolution of structural heterogeneity in dynamic crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 946–959 https://doi.org/10.1107/S0907444913003454
55 Nederlof, I., van Genderen, E., Li, Y.-W. and Abrahams, J.P. (2013) A Medipix quantum area detector allows rotation electron diffraction data collection from submicrometre three-dimensional protein crystals. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1223–1230 https://doi.org/10.1107/S0907444913009700
56 de la Cruz, M.J., Hattne, J., Shi, D., Seidler, P., Rodriguez, J., Reyes, F.E. et al. (2017) Atomic-resolution structures from fragmented protein crystals with the cryoEM method MicroED. *Nat. Methods* **14**, 399–402 https://doi.org/10.1038/nmeth.4178
57 Rodriguez, J.A., Eisenberg, D.S. and Gonen, T. (2017) Taking the measure of MicroED. *Curr. Opin. Struct. Biol.* **46**, 79–86 https://doi.org/10.1016/j.sbi.2017.06.004
58 Clabbers, M.T.B. and Abrahams, J.P. (2018) Electron diffraction and three-dimensional crystallography for structural biology. *Crystallogr. Rev.* **24**, 176–204 https://doi.org/10.1080/0889311X.2018.1446427
59 Nannenga, B.L., Shi, D., Leslie, A.G.W. and Gonen, T. (2014) High-resolution structure determination by continuous-rotation data collection in MicroED. *Nat. Methods* **11**, 927–930 https://doi.org/10.1038/nmeth.3043
60 Midgley, P.A. and Eggeman, A.S. (2015) Precession electron diffraction—a topical review. *IUCrJ* **2**, 126–136 https://doi.org/10.1107/S2052252514022283
61 Hauptman, H. (1986) The direct methods of X-ray crystallography. *Science* **233**, 178–183 https://doi.org/10.1126/science.233.4760.178
62 Usón, I. and Sheldrick, G.M. (1999) Advances in direct methods for protein crystallography. *Curr. Opin. Struct. Biol.* **9**, 643–648 https://doi.org/10.1016/S0959-440X(99)00020-2
63 Sawaya, M.R., Rodriguez, J., Cascio, D., Collazo, M.J., Shi, D., Reyes, F.E. et al. (2016) Ab initio structure determination from prion nanocrystals at atomic resolution by MicroED. *Proc. Natl Acad. Sci. U.S.A.* **113**, 11232–11236 https://doi.org/10.1073/pnas.1606287113
64 Jones, C.G., Martynowycz, M.W., Hattne, J., Fulton, T.J., Stoltz, B.M., Rodriguez, J.A. et al. (2018) The CryoEM method MicroED as a powerful tool for small molecule structure determination. *ACS Cent. Sci.* **4**, 1587–1592 https://doi.org/10.1021/acscentsci.8b00760
65 Gruene, T., Wennmacher, J.T.C., Zaubitzer, C., Holstein, J.J., Heidler, J., Fecteau-Lefebvre, A. et al. (2018) Rapid structure determination of microcrystalline molecular compounds using electron diffraction. *Angew. Chem. Int. Ed.* **57**, 16313–16317 https://doi.org/10.1002/anie.201811318
66 van Genderen, E., Clabbers, M.T.B., Das, P.P., Stewart, A., Nederlof, I., Barentsen, K.C. et al. (2016) *Ab initio* structure determination of nanocrystals of organic pharmaceutical compounds by electron diffraction at room temperature using a Timepix quantum area direct electron detector. *Acta Crystallogr. A Found. Adv.* **72**, 236–242 https://doi.org/10.1107/S2053273315022500

© 2019 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and the Royal Society of Biology and distributed under the Creative Commons Attribution License 4.0 (CC BY).

67 Palatinus, L., Brázda, P., Boullay, P., Perez, O., Klementová, M., Petit, S. et al. (2017) Hydrogen positions in single nanocrystals revealed by electron diffraction. *Science* **355**, 166–169 https://doi.org/10.1126/science.aak9652
68 Hattne, J., Reyes, F.E., Nannenga, B.L., Shi, D., de la Cruz, M.J., Leslie, A.G.W. et al. (2015) MicroED data collection and processing. *Acta Crystallogr. A Found. Adv.* **71**, 353–360 https://doi.org/10.1107/S2053273315010669
69 Smeets, S., Zou, X. and Wan, W. (2018) Serial electron crystallography for structure determination and phase analysis of nanocrystalline materials. *J. Appl. Crystallogr.* **51**, 1262–1273 https://doi.org/10.1107/S1600576718009500
70 de la Cruz, M.J., Martynowycz, M.W., Hattne, J. and Gonen, T. (2019) MicroED data collection with SerialEM. *Ultramicroscopy* **201**, 77–80 https://doi.org/10.1016/j.ultramic.2019.03.009
71 Lanza, A., Margheritis, E., Mugnaioli, E., Cappello, V., Garau, G. and Gemmi, M. (2019) Nanobeam precession-assisted 3D electron diffraction reveals a new polymorph of hen egg-white lysozyme. *IUCrJ* **6**, 178–188 https://doi.org/10.1107/S2052252518017657
72 Bustillo, K.C., Panova, O., Chen, X.C., Takacs, C.J., Ciston, J., Ophus, C. et al. (2017) Nanobeam scanning diffraction for orientation mapping of polymers. *Microsc. Microanal.* **23**, 1782–1783 https://doi.org/10.1017/S1431927617009576
73 Allahgholi, A., Becker, J., Delfs, A., Dinapoli, R., Goettlicher, P., Greiffenberg, D. et al. (2019) The adaptive gain integrating pixel detector at the European XFEL. *J. Synchrotron Radiat.* **26**, 74–82 https://doi.org/10.1107/S1600577518016077
74 Tinti, G., Fröjdh, E., van Genderen, E., Gruene, T., Schmitt, B., de Winter, D.A.M. et al. (2018) Electron crystallography with the EIGER detector. *IUCrJ* **5**, 190–199 https://doi.org/10.1107/S2052252518000945
75 Fan, G.Y. and Ellisman, M.H. (2000) Digital imaging in transmission electron microscopy. *J. Microsc.* **200**, 1–13 https://doi.org/10.1046/j.1365-2818.2000.00737.x
76 Contarato, D., Denes, P., Doering, D., Joseph, J. and Krieger, B. (2012) High speed, radiation hard CMOS pixel sensors for transmission electron microscopy. *Phys. Procedia* **37**, 1504–1510 https://doi.org/10.1016/j.phpro.2012.04.103
77 Ruskin, R.S., Yu, Z. and Grigorieff, N. (2013) Quantitative characterization of electron detectors for transmission electron microscopy. *J. Struct. Biol.* **184**, 385–393 https://doi.org/10.1016/j.jsb.2013.10.016
78 Li, X., Mooney, P., Zheng, S., Booth, C.R., Braunfeld, M.B., Gubbens, S. et al. (2013) Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10**, 584–590 https://doi.org/10.1038/nmeth.2472
79 Brewster, A.S., Sawaya, M.R., Rodriguez, J., Hattne, J., Echols, N., McFarlane, H.T. et al. (2015) Indexing amyloid peptide diffraction from serial femtosecond crystallography: new algorithms for sparse patterns. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 357–366 https://doi.org/10.1107/S1399004714026145
80 Sauter, N.K., Hattne, J., Brewster, A.S., Echols, N., Zwart, P.H. and Adams, P.D. (2014) Improved crystal orientation and physical properties from single-shot XFEL stills. *Acta Crystallogr. D Biol. Crystallogr.* **70**, 3299–3309 https://doi.org/10.1107/S1399004714024134
81 White, T.A. (2019) Processing serial crystallography data with *CrystFEL*: a step-by-step guide. *Acta Crystallogr. D Struct. Biol.* **75**, 219–233 https://doi.org/10.1107/S205979831801238X
82 White, T.A., Mariani, V., Brehm, W., Yefanov, O., Barty, A., Beyerlein, K.R. et al. (2016) Recent developments in *CrystFEL*. *J. Appl. Crystallogr.* **49**, 680–689 https://doi.org/10.1107/S1600576716004751
83 Sauter, N.K., Hattne, J., Grosse-Kunstleve, R.W. and Echols, N. (2013) New Python-based methods for data processing. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1274–1282 https://doi.org/10.1107/S0907444913000863
84 Nakane, T., Joti, Y., Tono, K., Yabashi, M., Nango, E., Iwata, S. et al. (2016) Data processing pipeline for serial femtosecond crystallography at SACLA. *J. Appl. Crystallogr.* **49**, 1035–1041 https://doi.org/10.1107/S1600576716005720
85 Ginn, H.M., Evans, G., Sauter, N.K. and Stuart, D.I. (2016) On the release ofcppxfelfor processing X-ray free-electron laser images. *J. Appl. Crystallogr.* **49**, 1065–1072 https://doi.org/10.1107/S1600576716006981
86 Lyubimov, A.Y., Uervirojnangkoorn, M., Zeldin, O.B., Brewster, A.S., Murray, T.D., Sauter, N.K. et al. (2016) *IOTA*: integration optimization, triage and analysis tool for the processing of XFEL diffraction images. *J. Appl. Crystallogr.* **49**, 1057–1064 https://doi.org/10.1107/S1600576716006683
87 Yamashita, K., Hirata, K. and Yamamoto, M. (2018) *KAMO*: towards automated data processing for microcrystals. *Acta Crystallogr. D Struct. Biol.* **74**, 441–449 https://doi.org/10.1107/S2059798318004576
88 Mylona, A., Carr, S., Aller, P., Moraes, I., Treisman, R., Evans, G. et al. (2017) A novel approach to data collection for difficult structures: data management for large numbers of crystals with the BLEND software. *Crystals* **7**, 242 https://doi.org/10.3390/cryst7080242
89 Grimes, J.M., Hall, D.R., Ashton, A.W., Evans, G., Owen, R.L., Wagner, A. et al. (2018) Where is crystallography going? *Acta Crystallogr. D Struct. Biol.* **74**, 152–166 https://doi.org/10.1107/S2059798317016709

© 2019 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and the Royal Society of Biology and distributed under the Creative Commons Attribution License 4.0 (CC BY).