

A beginner's guide to macromolecular crystallization

Fabrice Gorrec (MRC Laboratory of Molecular Biology, U.K.)

Obtaining diffraction-quality crystals is currently the rate-limiting step in macromolecular X-ray crystallography of proteins, DNA, RNA or their complexes, in the vast majority of cases. Since each sample has different and specific characteristics – which is the reason for wanting to study every single one of them in the first place – crystallization conditions cannot be predicted. Hence, researchers must enable crystal nucleation and growth through experimentation and screening. The size, shape and surface of the sample or complexes of interest are often altered through genetic and biochemical manipulation to facilitate crystallization, based on bioinformatics analyses and trial and error. Pure samples are trialled against a very broad range of crystallization conditions. The currently predominant method to achieve crystallization is sitting drop vapour diffusion with nanolitre-class robotic liquid handlers. Once initial screening yields crystals, further optimization experiments are usually required to obtain larger and diffraction-quality crystals.

Background

Macromolecular X-ray crystallography applied to crystals of biological molecules enables us to visualize structures of proteins, DNA, RNA and their complexes with near to full atomic resolution (~ 3.5 Å to less than 1 Å resolution). Crystallography's immense power lies in the fact that it makes visible complex atomic structures, sometimes with hundreds of thousands of atoms. It enables the understanding of biology through the lens of the laws of chemistry and physics. It is also employed to aid structure-based, rational drug design.

X-ray crystallography as a method has matured tremendously since its inception in the 1950s at the MRC Laboratory of Molecular Biology (LMB), the birthplace of protein crystallography. Through the use of much brighter X-ray beams at synchrotrons and sophisticated computer programmes, once high-quality crystals have been made available, structure determination has become a mere technicality. In contrast, obtaining crystals has been an increasingly severe bottleneck. Some projects take years before crystals can be obtained, if at all. Large proteins and complexes (>200 kDa) are particularly difficult to crystallize.

In order to predict how to crystallize a protein, what would help most would be details of its atomic structure, the very thing that the method tries to obtain. Since each protein has an unavoidably different structure, we are required to employ sophisticated screening procedures to empirically find the conditions under which the sample forms crystals. Furthermore, a sample to be trialled for crystallization needs to be as pure and homogeneous

as possible. It should reach a concentration of at least a few grams per litre, which is high for most samples. Samples that are not very stable should be trialled readily after their preparation; some even require low temperature handling or oxygen depletion during the entire crystallization process.

Summary box

- Successful structure determination is usually easier the stronger X-ray diffraction is from the crystals, and should be better than 3.5 Å resolution.
- Appearance of crystals can be misleading, and hence, their true usefulness is best established by X-ray diffraction.
- Often, many sample variations and preparations are required for solving novel or complicated crystal structures.
- Different types of crystallization screen formulations, integrating a plethora of suitable chemicals, are employed to alter the variables determining the crystallization experiments.
- When enough sample is available, a large initial screen increases the likelihood of obtaining useful crystals.
- Nucleation is often rate limiting for obtaining crystals.
- Crystal growth after nucleation is typically affected by defects that limit crystal size.
- Various seeding techniques can be applied to overcome the nucleation bottleneck.

The basis of crystallography is that a crystal is a nearly perfect periodic three-dimensional arrangement of molecules (the 'lattice'). A crystal amplifies the weak scattered X-ray signal from each molecule only in the direction where constructive interference between the diffracted beams occurs. Therefore, larger crystals often, but not always, diffract more strongly. The conditions for constructive interference are condensed in Bragg's law.

Diffraction patterns are recorded on two-dimensional detectors while the crystal is rotated. Once the intensities of all the beams have been determined, the three-dimensional electron density function/map can be built using a number of standard methods to solve the phase problem. The phase problem arises because only the intensities of the diffracted beams are recorded and not their relative phase angles, which are needed to solve the structure. However, for crystals that diffract beyond 3.5 Å resolution, standard methods exist to solve the phase problem, such as molecular replacement (MR; using a homologous structure), single anomalous diffraction using selenomethionine incorporation (SeMet SAD) or single/multiple isomorphous replacement (S/MIR, Max Perutz' original heavy metal-binding procedure).

Macromolecular crystals

The largest dimension of a typical small protein is about 10 nm. A crystal of the same protein, big enough to be visible under a microscope, is a lattice formed by billions of the protein molecules (Figure 1a). Crystal morphology is dictated by the lattice type and symmetry (<https://it.iucr.org/>).

The quality of the internal order of the crystals is often reflected in their appearance, where sharp edges can be a sign of crystals which will diffract well. However, even crystals that look promising have typically high solvent content (in the order of 50%–70%). Besides, they are mainly formed through weak interactions such as small hydrophobic surfaces, water-mediated, longer-range and flexible electrostatic interactions and sometimes even single hydrogen bonds. On the one hand, this is beneficial as the macromolecules are unlikely to be altered by the binding energies generated during the crystallization process. On the other hand, the weakness of the interactions means that protein crystals have a mosaic structure (Figure 1b). This so-called mosaicity broadens the diffracted beams and lowers their signal-to-noise ratio on the detector.

An even more severe crystal pathology, twinning, can occur when parts of the crystal grow in distinct, but pseudo symmetry-related orientations (Figure 1c). The crystal can, therefore, be packed in nearly identical ways to yield the same lattice structure, which can sometimes lead to an inability to solve the structure.

The sample is the main variable of crystallization

Samples of interest are generated using molecular biology methods. For example, recombinant proteins are produced with expression systems such as bacteria, insect cells or mammalian cells. For successful crystallization, many samples require the removal of

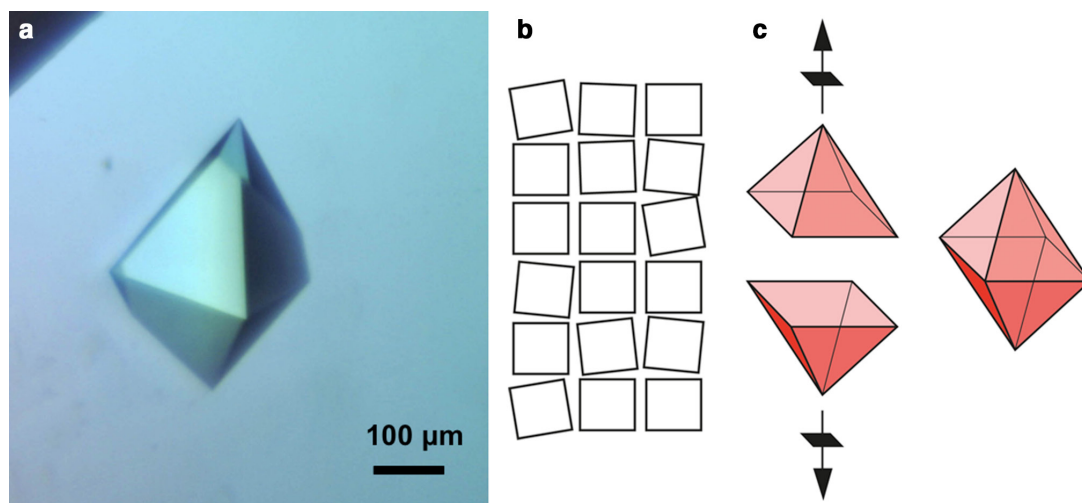


Figure 1. (a) A promising bipyramidal-shaped protein crystal ready to be tested for X-ray diffraction. (b) A highly exaggerated representation of crystal mosaicity (adapted from Phil Jeffrey's course on X-ray data collection, xray0.princeton.edu/~phil/Facility/). (c) Representation of a possible case of crystal twinning: crystal formed with two pyramid structures in opposite orientations (from Zbigniew Dauter's course on Twinning in protein crystal, <https://slideplayer.com/slide/4983758/>).

flexible and small disordered parts, problematic domains and/or subunits. Aggregation properties may need to be improved, as rapid, nonspecific aggregation competes with the crystallization process.

In principle, increasing the rigidity of the sample, making it more soluble and stable in the chosen buffer, increases the chances of successful crystallization and also the resulting X-ray diffraction. However, rigidity and simplicity do not systematically produce the best outcome. Besides, secondary indicators of sample quality, such as stability, purity or amounts available, can be misleading. As a result, empirical approach is preferred in which the majority of the samples produced are directly trialled for crystallization. Fundamentally, perfect samples are those that give crystals and are meaningful in terms of the biological question to be answered.

Macromolecular crystallization

Crystallization essentially occurs through an increase in the saturation of the sample as produced by the crystallization solution (namely the 'condition') and the buffer used for preparing the sample. In order to crystallize, the solubility threshold of the sample must be crossed, so that it becomes energetically favourable for the sample to bind to itself repetitively, rather than to solvent molecules. Nonspecific aggregation fulfils the same condition, and hence, it is the principal obstacle to crystallization as it removes, normally irreversibly, the sample material from solution without yielding crystals.

Many different effects can cause the solubility threshold to be crossed. A 'precipitant salt' such as ammonium sulphate induces salting out of the sample: the

solvation of the protein decreases, causing its properties to change, such that hydrophobic interactions on the surface are accentuated. The effect on precipitation is proportional to the salt concentration and its Hofmeister ranking. Other types of precipitants, such as polyethylene glycols (PEGs), also compete with solvent molecules on the protein surface in different ways. Varying the buffer's pH will alter the density of charges on the surface of macromolecules, altering both their solubility and the possibilities of interactions. Other chemicals can be employed to affect solubility, but often have more specific effects that go beyond solubility and are termed 'crystallization additives'; they are cross-linkers, reducing agents, metals, sugars, nucleotides, drug-like molecules and many more. Finally, the solubility of many molecules is also temperature dependent.

Figure 2a shows the simplest form of a protein solubility phase diagram. Saturation levels are represented according to two example parameters that alter solubility: concentration of the sample and precipitant concentration. The solubility boundary curve divides the diagram into two areas (undersaturated and supersaturated states). The supersaturated area comprises three zones: metastable, nucleation and precipitation zones. As explained, in practice, many other parameters should be considered, and hence, phase diagram cannot be built to facilitate successful crystallization. For example, transition from aggregation to crystal growth is often observed for large and complex samples: They have many components and parts that can react very differently to the changes in parameters.

Several methods can be employed to experimentally increase saturation of the sample. Numerous setups exist; however, vapour diffusion with sitting drops is currently

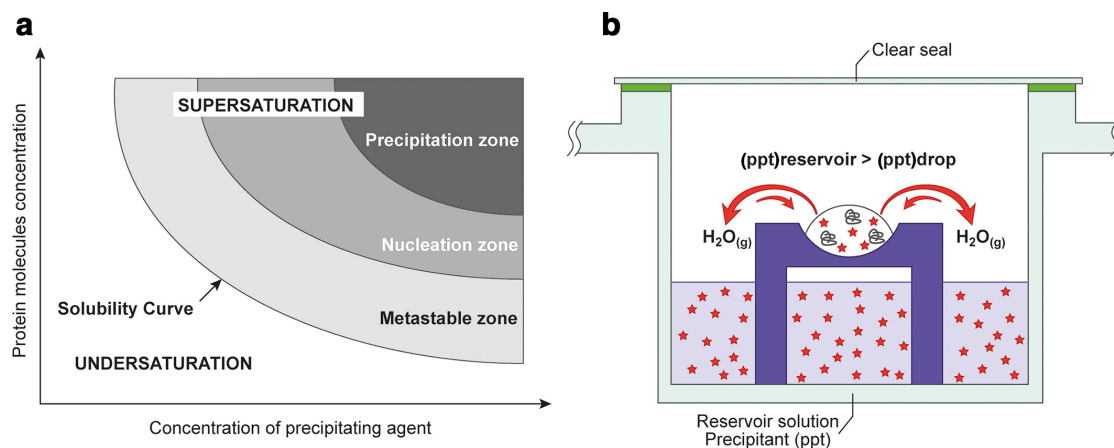


Figure 2. (a) Two-dimensional phase diagram depicting solubility of small globular proteins (University of Cambridge course on protein crystallography, <https://www-structmed.cimr.cam.ac.uk/course.html>). (b) Schematic representation of a vapour diffusion experiment with sitting drop. The precipitant in the crystallization condition and the drop is represented by red stars, and the macromolecular sample, which is part of the drop with the condition, by folded threads.

the most widely employed because of its simplicity and amenability to automation. The macromolecular sample solution and the condition are mixed together to form a drop that sits next to a 'reservoir', which contains a much larger volume of the same crystallization condition (Figure 2b). The experiment is swiftly sealed airtight, allowing equilibration of the osmolarity to take place between the sample-containing crystallization drop and the reservoir. Since the concentration of solutes in the reservoir is larger than in the sample drop (which was diluted by mixing), water will diffuse through the vapour phase from the protein drop to the reservoir (in an attempt to equilibrate the solute concentrations/osmolarity). This means the size of the sample drop will reduce gradually, increasing the concentration of all components and allowing the sample to be crystallized.

If done correctly, the vapour diffusion will move the experiment through the phase diagram such that crystal growth is achieved. Since many parameters are changed at once during this process, systematic explorations of the influence of single parameters is not feasible (and possibly, also not desirable). However, it makes it possible to test empirically many different precipitants, concentration variations, pHs and other parameters with a very simple setup, which gives enough control to make simple deductions of what works and what does not.

Automation of crystallization experiments

Liquid handling devices and robotic handling of specialized crystallization microwell plates, which typically provide preformed wells for 96 experiments, have enabled a steady increase in the number of experiments that can be set up rapidly, with little effort, and most importantly with a limited amount of sample.

Crystallization droplets contain between 100 and 200 nl of sample. Screens dispensed in modern,

commercially available plates for vapour diffusion experiments consume 10–20 μl of sample for 96 experiments. Many samples can, therefore, be initially trialled for crystallization against a very broad range of conditions, at different temperatures and concentrations. In fact, the chances of producing useful crystals increase almost linearly with the number of individual trials.

More specialized crystallization setups exist; for example, automated handling of liquids also facilitates crystallization of specific classes of macromolecules such as membrane proteins, notably through lipidic cubic phase (LCP) setups. Automation also facilitates subsequent optimization that is normally needed to grow better-quality and larger crystals.

Crystallization screens

The crystallization screens are prepared by combining different precipitants at various concentrations with buffers and additives. The aim is to alter many variables associated with the main parameters of crystallization, often at once, in order to increase the accessible parameter space while keeping the number of conditions relatively low.

Because of the very large number of reagents that can promote crystallization, their systematic permutations to formulate an extensive screen would quickly run to millions of conditions. Systematic 'grid' screens are hence mostly employed during later optimization, once a small number of crystallization reagents have been selected and their relative concentrations need optimizing. Other, non-systematic approaches to screen formulations are widely employed during initial trials. Important historic examples are Carter and Carter's (1979) 'incomplete factorial', where reagents/combinations were omitted randomly, and Jancarik and Kim's (1991) 'sparse matrix' formulation, where conditions were selected empirically based on previous results with other samples. Figure 3 shows these three main types of screen formulations

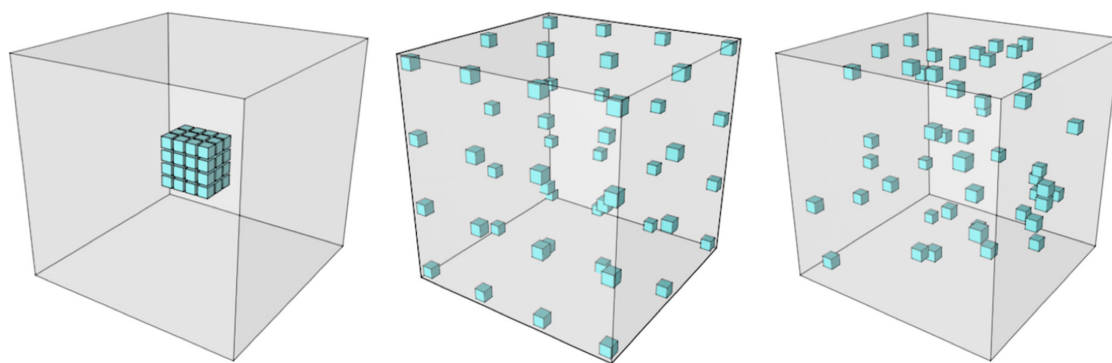


Figure 3. The three main types of crystallization screen formulations, represented within an idealized conceptual three-dimensional space. From left to right: grid screen, incomplete factorial screen and sparse matrix screen. The small cubes represent 48 combinations of three components (typically, a precipitant concentration, a pH value and a type of additive).

schematically. An orthogonal approach is to formulate a screen with mixes of additives and a limited number of precipitants and buffers, which is the basis of McPherson and Cudney's (2006) 'silver bullets' screen.

As the amount of sample is often limiting, various initiatives aim at optimizing the use of the large number of commercially available screens. For example, the 'C6 webtool' (CSIRO, Australia) allows assessment of the similarity between commercially available screens (<https://c6.csiro.au>).

Crystal nucleation and growth

Crystal growth must be preceded by nucleation. This process is most often rate limiting because it requires the first macromolecules to come together in the right orientation, and is therefore characterized by higher-order kinetics. Observations using time-resolved cryo-transmission electron microscopy suggest that nucleation can progress through specific interactions between only a few macromolecules at a very early stage of nucleation (Figure 4). Depending on the condition

used, some interactions are more or less favoured and different building blocks are formed initially. This mechanism explains why sometimes different crystals are observed during protein crystallization experiments with the same or very similar conditions, such as with the protein glucose isomerase.

Once nucleation has occurred, further addition of macromolecules is normally made easier by the fact that the crystal nucleus provides ready-made binding surfaces for the incoming molecules. This explains why crystal growth typically requires a lower level of saturation than nucleation ('metastable zone', Figure 2).

One mechanism of crystal growth is sequential layer addition, where multiple nuclei on the surface of the crystal become two-dimensional patches. The two-dimensional patches grow tangentially and merge to form a layer. The constraints imposed by internal symmetry and the continuous creation of step edges on the growing surfaces can lead to spiral structures ('spiral dislocation', Figure 5) that in turn lead to crystal imperfections as the gaps fill.

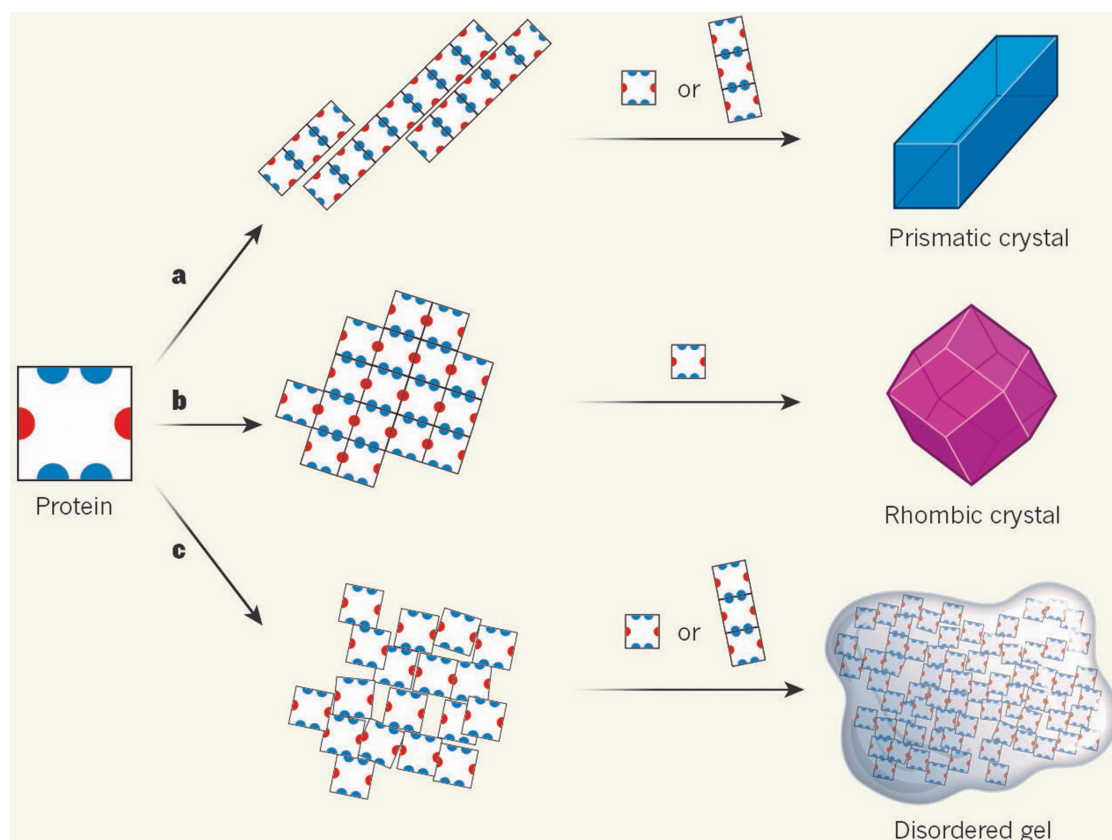


Figure 4. The pathways to crystal formation for glucose isomerase (including two different crystal forms and a gel-like structure) are determined by specific nucleation building blocks. Coloured semicircles represent sites of possible interactions between glucose isomerase, represented as a box-shaped protein. Adapted from Alberstein and Akif Tezcan (2018) *Nature* (doi: 10.1038/d41586-018-03801-5).

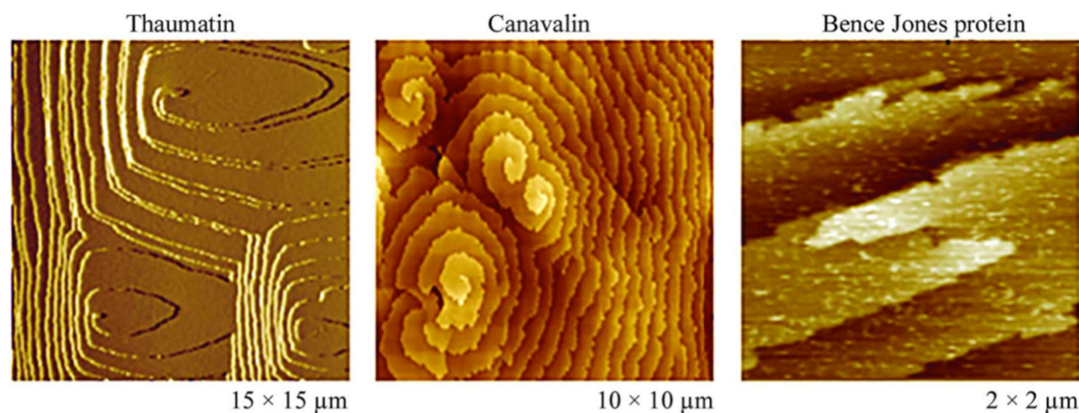


Figure 5. Crystal growth of three different protein crystals by spiral dislocation as observed with atomic force microscopy (AFM). Adapted from McPherson and Kuznetsov (2014) *Acta Cryst F* (doi: 10.1107/S2053230X14004816).

Because of the relative plasticity of a typical protein crystal lattice, impurities such as partially denatured macromolecules, proteolysed proteins or other contaminants can easily be integrated during growth. This can create severe crystal defects that, in unfavourable cases, cause termination of crystal growth. This effect is called ‘poisoning’ of growth and is the most common reason why many crystals refuse to grow beyond a certain size.

Seeding

The fact that growth preferentially occurs at lower saturation than nucleation means that a technique like vapour diffusion can be limiting because nucleation competes with growth (while saturation keeps increasing). A way to decouple nucleation and growth that can much improve crystals is seeding. Microcrystals (the seeds) are taken from previous trials that yielded poor crystals or crystals that are too small. They are then used to initiate crystal growth at a lower level of saturation that does not allow spontaneous nucleation.

A modern extension of this principle is macro-seed matrix screening. Here, seeds are produced by crushing previously obtained crystals, including fully grown ones. The resulting crystal fragments are used as seeds in crystallization experiments against a crystallization screen, integrating a broad variety of conditions such as a common sparse matrix screen. The method works because sometimes crystal growth can be jump-started, even if the seeds do not have the correct lattice for the new condition – it is still lowering the barrier of molecule addition enough to allow growth of crystals, essentially overcoming the nucleation bottleneck.

Other approaches to seeding use nucleants that have a different nature than the sample itself: epitaxial nucleation is being used that makes use of

mineral or polymer substrates as nucleation seeds for macromolecular crystallization. In cross-seeding, seeds are prepared with different proteins that are structurally related to the targeted sample.

Perspectives

The limits of macromolecular crystallography get tested less frequently nowadays because of the attractiveness of cryo-electron microscopy (cryo-EM) that enables the structure determination of very large and heterogeneous complexes without crystallization (see the corresponding beginner's guide here <https://doi.org/10.1042/BIO04102046>). However, according to the number of depositions in the RCSB Protein Data Bank (PDB, Table 1), X-ray crystallography is currently still the predominant technique in terms of numbers of structures solved, especially at high resolution and for small and medium-sized samples. An application that utilizes crystallography's greatest strengths – high-throughput experimental protocols and ultra-fast structure determination – is fragment screening where hundreds to thousands of crystals soaked with small molecules reveal binding pockets and binding modes. It is anticipated that cryo-EM will not be able to compete with the throughput and precision of this application of crystallography for a while, if ever.

In addition, X-ray crystallography is not limited to the determination of static structures; dynamic aspects of protein function can be elucidated in some instances. Femtosecond time-resolved crystallography with X-ray free electron lasers (XFEL) enables visualization of conformational dynamics and functionally important motions throughout the catalytic cycle of enzymes, as long as the crystal lattice does not interfere with the mechanism or the conformational changes required by it.

Table 1. Structures deposited in the PDB over the last 3 years, listed by the technique employed (X-ray crystallography vs cryo-EM) and the resolution achieved. The average molecular weight is calculated from data with resolutions better than 3.5 Å. Entries were selected according to their release dates from 1 January 2018 to 18 November 2020. More statistics are available at <https://www.rcsb.org/stats>

Year	Technique	Number of structures			Av. MW (kDa)
		All	Resolution <3.5 Å	Resolution <2.0 Å	
2020	X-ray	9917	9694	5159	88
2020	Cryo-EM	2079	1060	17	699
2019	X-ray	9673	9452	4620	93
2019	cryo-EM	1421	554	3	784
2018	X-ray	9889	9690	4866	92
2018	Cryo-EM	848	222	2	860

The weakness of macromolecular crystallography is the rate-limiting step of successful crystallization. Besides better control of nucleation and seeding, a practical aspect that urgently needs attention is the further miniaturization of crystallization experiments. For this, technology based on acoustic liquid handling is interesting. Crystallization droplets of 10 nl are feasible and practical, instead of the 100 nl currently used with standard pipetting technology. Further miniaturization will enable much larger screens and/or the reduction of sample needed – both extremely important parameters for crystallography to remain relevant for difficult and

challenging samples. It seems reasonable to suggest that for the time being, macromolecular samples should initially be tried both with cryo-EM and crystallography, as long as this is feasible in terms of sample and time consumption.

Acknowledgements

I would like to thank Jan Löwe (LMB) for very helpful discussions, Jo Westmoreland for the preparation of figures and Rachel Kramer Green (RCSB PDB) for the data in Table 1. ■

Further reading and viewing

- Bragg's Law and X-ray Diffraction Animation: <https://www.youtube.com/watch?v=wtvs1t3YZPw>
- Cracking the Phase Problem: <https://www.nobelprize.org/prizes/chemistry/1962/perspectives/>
- Wlodawer, A., Dauter, Z., and Jaskolski, M. (2017) *Protein Crystallography*, Springer protocols, Humana Press, New York, DOI: 10.1007/978-1-4939-7000-1
- Derewenda, Z.S. (2004) The use of recombinant methods and molecular engineering in protein crystallization, *Methods*, **34**, 354–363. DOI: 10.1016/j.ymeth.2004.03.024
- Dessau, M.A. and Modis, Y. (2011) Protein crystallization for X-ray crystallography, *J. Vis. Exp.* **47** 2285. DOI: 10.3791/2285
- Stock, D., Perisic, O. and Löwe, J. (2005) Robotic nanoliter protein crystallisation at the MRC Laboratory of Molecular Biology, *Prog. Biophys. Mol. Biol.* **88**, 311–327. DOI: 10.1016/j.pbiomolbio.2004.07.009
- Li, D. and Caffrey, M. (2020) Structure and functional characterization of membrane integral proteins in the lipid cubic phase, *J. Mol. Biol.* **432**, 5104–5123. DOI: 10.1016/j.jmb.2020.02.024
- Van Driessche, A.E.S., Van Gerven, N., Bomans, P.H.H. et al. (2018) Molecular nucleation mechanisms and control strategies for crystal polymorph selection. *Nature*, **556**, 384–403. DOI: 10.1038/nature25971

Continued

- McPherson, A. and Shlichta, P. (1988) Heterogeneous and epitaxial nucleation of protein crystals on mineral surfaces. *Science*, **239**, 385–387. DOI: 10.1126/science.239.4838.385
- D'Arcy, A., Bergfors, T., Cowan-Jacob, S.W. and Marsh, M. (2014) Microseed matrix screening for optimization in protein crystallization: what have we learned? *Acta Cryst. F*, **70**, 1117–1126. DOI: 10.1107/S2053230X14015507
- To know more about seeding (and other) protocols see the tutorials of Terese Bergfors: <https://xray.teresebergfors.com>
- Dasgupta, M., Budday, D., de Oliveira, S.H.P. et al. (2019) Mix-and-inject XFEL crystallography reveals gated conformational dynamics during enzyme catalysis. *Proc. Natl Acad. Sci. U.S.A.* **116**, 25634–25640. DOI: 10.1073/pnas.1901864116
- Gorrec, F. (2014) Progress in macromolecular crystallography depends on further miniaturization of crystallization experiments. *Drug Discov. Today*, **19**, 1505. DOI: 10.1016/j.drudis.2014.07.002
- Baker, L.M. Aimon, A., Murray, J.B. et al. (2020) Rapid optimisation of fragments and hits to lead compounds from screening of crude reaction mixtures. *Commun. Chem.* **3**, 122. DOI: 10.1038/s42004-020-00367-0
- Šrajer, V. and Schmidt, M. (2017) Watching proteins function with time-resolved X-ray crystallography. *J. Phys. D Appl. Phys.* **50**, 373001. DOI: 10.1088/1361-6463/aa7d32
- Nakane, T., Kotecha, A., Sente, A. et al. (2020) Single-particle cryo-EM at atomic resolution. *Nature*, **587**, 152–156, DOI: 10.1038/s41586-020-2829-0



For almost two decades, Fabrice Gorrec has been developing technologies for protein crystallization, including robotic systems, software, screens and microplates. Since 2007, he has been responsible for the crystallization facility at the MRC Laboratory of Molecular Biology (LMB, Cambridge, UK) where support is being given to dozens of users. Email: fgorrec@mrc-lmb.cam.ac.uk. Website: <https://www3.mrc-lmb.cam.ac.uk/sites/protein-crystallisation/>