Proceedings of the CAS–CERN Accelerator School: Free Electron Lasers and Energy Recovery Linacs, Hamburg, Germany, 31 May–10 June 2016, edited by R. Bailey, CERN Yellow Reports: School Proceedings, Vol. 1/2018, CERN-2018-001-SP (CERN, Geneva, 2018)

Crystallography and Molecular Imaging using X-ray Lasers

T.A. White
Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany

Abstract
A very successful application of X-ray free-electron lasers has been made in structural biology. Acquiring diffraction data using X-ray pulses with durations of a few tens of femtoseconds allows the conventional processes of radiation damage to be sidestepped, breaking limits that previously applied, while at the same time permitting experiments to probe chemical dynamics on short timescales. This contribution gives an introduction to applications of free-electron lasers in biochemistry, including potential future applications to single-molecule diffraction.

Keywords
Crystallography; biochemistry; free-electron laser.

1 The need for knowledge of the structure of biomolecules

The functioning of many biochemical systems relies primarily not on the chemical compositions of the molecules—the relative amounts of different elements making up the participating molecules—but rather on the structures of the molecules. Proteins, for example, are comprised primarily of carbon, hydrogen, nitrogen, and oxygen atoms in the ratios 40 : 62 : 10 : 12 respectively, with remarkably little variation. Proteins are built from amino acids, which can be consecutively bonded together to form a *polypeptide* chain. A protein consists of one or more polypeptide chains, perhaps in addition to a few extra chemical groups. Stabilized primarily by hydrogen bonds, the polypeptide chains form specific structures consisting of structural motifs, such as helices, sheets, and loops. The specific structure of a protein gives it its particular properties, for example, by creating an open region or 'pocket' where another molecule can fit during catalysis of a bond-breaking reaction. Many proteins can be considered as 'molecular machines'.

Discovering the structure of a protein is one of the first steps towards understanding how it performs its task. This understanding is directly applicable in areas such as medicine, one possible application being to block a binding pocket in a pathogenic protein and therefore hinder its activity. A good example of this is shown in Fig. 1, which illustrates a protein known as HIV-1 protease. At a certain stage in the life cycle of the human immunodeficiency virus, this protein cuts a long polypeptide chain into smaller segments, which have specific activities of their own. The active side of HIV-1 protease is the hole that can be seen in the picture, which is where the polypeptide chain fits while being cut. The action of the protein can be inhibited by blocking the active site with some other molecule. A class of drugs known as protease inhibitors work in exactly this way, by fitting into the hole and binding there more tightly than the polypeptide chain. By acquiring and applying knowledge of the structure of the protease, including seeing how the inhibitor binds into the hole, we may be able to make better protease inhibitors, which bind more strongly and more selectively to the target molecules. Stronger binding, in this case, would mean the drug is more effective, while more selective binding may mean that the drug has fewer side-effects.

2 Diffraction from molecules and crystals

In an ideal experiment, we could determine the structure of a single protein molecule simply by placing it in an X-ray beam and measuring the intensity of the scattered X-rays in different directions with an

https://doi.org/10.23730/CYRSP-2018-001.605

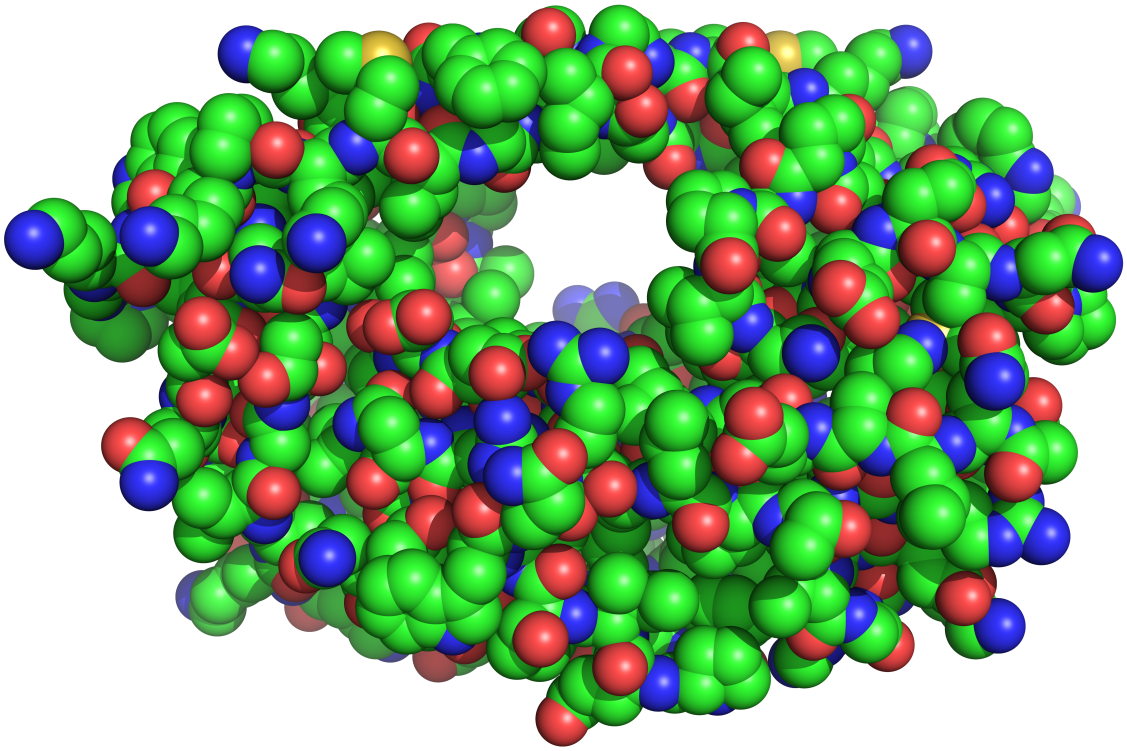

Fig. 1: Structure of HIV-1 protease, showing the hole where a long peptide chain fits while being cut into smaller segments. Atoms are represented by spheres. Carbon atoms are green; nitrogen, blue; oxygen, red; and sulphur, gold. Hydrogen atoms are not shown. Image generated from Protein Data Bank entry 3HVP [1].

area detector. Several features of a practical experiment have prevented us from doing this so far. First, the incident X-ray beam would have to be very intense to produce a measurable diffraction pattern. The required flux is of the order of 10^{13} photons per square micrometre, which corresponds to a radiation dose of the order of 10^{10} Gy [2]. The gray (Gy) is the unit of radiation dose, one gray being defined as one joule of radiation absorbed per kilogram of matter. These figures do not immediately say anything about the timescale over which the flux should be administered—it could be given using a very weak X-ray source over a long period of time—but 10^{10} Gy is a very large X-ray dose for a biological sample, several orders of magnitude greater than it would be able to withstand. The situation appears even more dire when we consider that the term ‘measurable diffraction pattern’, above, already takes into account that the detector should be able to measure individual scattered X-ray photons.

This problem can be solved by spreading the radiation dose among a large number of protein molecules. This can be done by using a protein crystal instead of a single protein molecule. A crystal consists of many copies of a translationally repeated unit cell, each of which contains one or more copies of the entire protein molecule. The regular arrangement of molecules within a crystal provides a very large increase in the scattered intensity, because the X-rays scattered by each unit cell of the crystal interfere constructively with those scattered by the others. This raises the signal to the point where a diffraction pattern can be measured from a protein crystal using an X-ray source of the type found in many laboratories. The increased signal, however, is compressed into sharp Bragg peaks, compared with the smoothly varying pattern that would be seen from a single molecule (albeit at a very low signal level). Figure 2 shows these two scenarios side by side. Reconstructing the structure in both cases would mean solving the well-known phase problem. The X-ray detector can measure only the intensities of

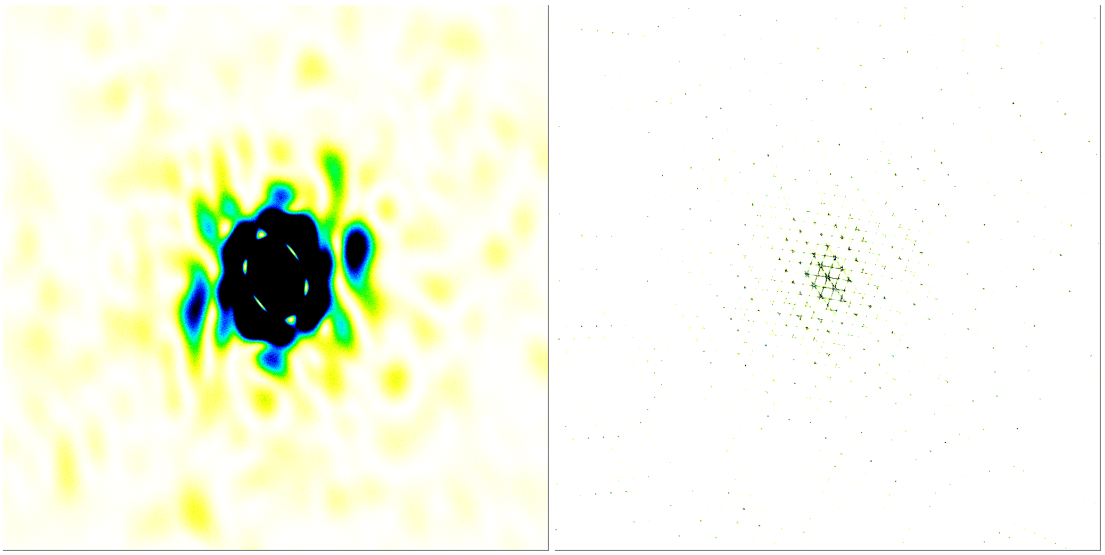

Fig. 2: Comparison of diffraction patterns from a single molecule (left) and a crystal (right). The colour scale, for low to high intensity, goes from white through yellow, green, cyan, blue, and, finally, black.

the diffraction signal, when in reality it is a complex-valued function and therefore has phase values as well. No atomic-resolution lens exists for X-rays, so we cannot build an imaging system to turn the amplitude and phase information directly into an image. Instead, we must reconstruct the phase information computationally.

It might appear that the crystal diffraction pattern, although it has the big advantage of being easily measurable, contains less information, and this is indeed the case. The crystal diffraction pattern in fact contains exactly half of the information that would theoretically be needed to reconstruct the structure [3]. The single-molecule pattern contains sufficient information to reconstruct the structure using a constraint satisfaction procedure [4], at least for the ideal experiment without background scattering and detector artefacts.

For the crystal case, a variety of structure solution techniques have been developed for making up the information shortfall. The most widely used of these is known as *molecular replacement*, and essentially involves comparing the measured intensities with those that would be produced by a molecule that we hypothesize is similar in structure to the protein under investigation. Other techniques exploit the variation of scattering power of heavy atoms with X-ray wavelength (the so-called anomalous diffraction techniques), or the changes to the intensities that arise when heavy atoms are embedded into the structure (the isomorphous replacement method). Notice that all of these techniques involve making further measurements or introducing external information. Direct methods, which are 'pure' solutions to the phase problem without doing either of these things, also exist. However, direct methods are usually not applicable to diffraction data from protein crystals because they require information to higher resolution than is usually permitted by their degree of order [5].

Is there any hope of measuring the diffraction from a single molecule? This would be possible if we could somehow suspend the usual rules of radiation damage and deliver an extremely large X-ray dose without the molecule being damaged. It turns out that this can be achieved by delivering the X-rays in a pulse with a very short duration indeed, and this is, of course, exactly the type of pulse provided by an X-ray free-electron laser. Before the first X-ray free-electron laser had been built, it was theorized, based on computer simulations, that radiation damage could be 'sidestepped' by delivering the entire X-ray dose in a single pulse with a duration of a few femtoseconds [6]. The molecule would be completely destroyed shortly afterwards, the electrons having been stripped from the atoms by the intense electric field, but the destruction would happen on a timescale longer than the pulse. Since the diffraction pattern

is recorded during the pulse itself and not the time afterwards, the destruction of the molecule should not significantly affect the diffraction signal.

A few years after it was first proposed, this 'diffraction before destruction' principle was demonstrated in an experiment at FLASH, a soft X-ray free-electron laser facility in Hamburg, Germany [7]. Using a single radiation pulse, a diffraction pattern was recorded from a two-dimensional pattern etched into a silicon nitride membrane. The pattern was destroyed in the process, as could be seen from a pattern recorded using a subsequent pulse on the same sample.

3 Crystallographic data collection and processing

Interpretable diffraction signals from single protein molecules have not been achieved to date. In the meantime, the ability of X-ray free-electron laser pulses to sidestep the usual radiation damage processes has been put to good use on crystalline samples. In biomolecular crystallography, radiation damage is still a serious problem, although the dose is spread among a large number of molecules. Conventional crystallography, as has been practised for several decades, is based on the rotation method, where a series of diffraction patterns is recorded while the crystal is continuously rotated. This produces a three-dimensional dataset. However, it requires the crystal to be exposed to the X-ray beam for a relatively long period of time. The maximum tolerable dose must be apportioned over the entire rotation series.

Using X-ray pulses from a free-electron laser sidesteps the radiation damage limit, but unfortunately means that only one diffraction pattern can be recorded from a single crystal. Many different views of the crystal, in different orientations, are needed to reconstruct the three-dimensional structure. If a very large crystal is available, subsequent frames can be recorded from a different position on it, away from the region affected by damage from the first shot. Applications of this method have been described [8]. However, if this is not possible then many crystals will be required to form a complete dataset. The situation where only one diffraction 'snapshot' is recorded from each crystal has been dubbed 'serial crystallography'. There are many ways to achieve this, but one of the most popular is to inject a 'jet' of crystal-laden liquid into the path of the X-ray beam [9]. Another popular method is to spread the crystals over a solid support, and then to raster the X-ray beam across it (although in practice the sample is moved, not the X-ray beam). When serial crystallography is performed using femtosecond X-ray pulses, it is known as *serial femtosecond crystallography*.

Data processing in serial crystallography is broadly similar to data processing for conventional rotation data. It is an active field of research in its own right, and has been described extensively in the literature [10]. The process is briefly described here. First, the Bragg peaks are found in each detector frame, and a decision made about whether the frame actually contains a usable diffraction pattern or not. Depending on the sample delivery method and the density of well-ordered crystals, typically only a small fraction of detector frames actually contain usable patterns. Once the 'hits' have been identified, the locations of the Bragg peaks are used to determine the orientation of the crystal, and hence to calculate the locations in the image where Bragg peaks should appear. This way, measurements can be made of all the peaks, even if they were not all found by the peak search, and even if some of them are very weak. Of course, the information that a particular Bragg peak is very weak, or even completely absent, is just as important as if it were very strong. Once the intensities have been measured from all the patterns, the measurements are combined, and the merged intensities used with a variety of algorithms to solve the structure. For serial crystallography, several data processing packages are now available, the most popular of which is CrystFEL [11, 12]. The subsequent steps—to solve the phase problem, determine the structure, and refine the structural model—are essentially the same for the merged data from free-electron laser experiments as in conventional rotation crystallography.

4 Human membrane proteins

All organisms must keep some volumes separated from others. For example, a cell contains different compartments, which allow many specialized processes to occur separately, and these compartments may contain very different chemical environments, such as varying levels of acidity. A completely sealed cell compartment would not be of much use, and some means of transmitting signals or controlling the movement of substances in and out of compartments is required. These movements and transmissions are controlled by proteins embedded in the membranes enclosing the compartments. Since so many biological processes involve this type of protein, they are the target of most pharmaceutical substances and consequently the proteins of which we would most like to know the structures. Unfortunately, extracting such proteins from the membrane is difficult, because often pressure from the membrane helps to maintain their structure and they become unstable once extracted. The proteins in this category, known as integral membrane proteins, are therefore some of the most difficult ones to study crystallographically.

This category of protein has proven to be a success story for serial femtosecond crystallography, with several examples now published. Crystals can be grown inside a *lipidic cubic phase*, in which the proteins remain embedded in a membrane. Rather than subsequently extracting the crystals from the lipidic cubic phase matrix, the whole thing can be injected into the X-ray beam using an injection device designed for viscous fluids [13], which greatly simplifies sample handling. In addition, using a viscous medium has the great advantage that the sample can flow very slowly, meaning that more of it is probed by the X-ray beam rather than flowing past the interaction point between X-ray pulses and being wasted.

Proteins that have been studied include the serotonin receptor $5\text{-}HT_{2B}$ [14], δ-opioid receptor [15], AT_1 receptor [16], A_{2A} receptor [17], and *smoothened* receptor [13]. All of these are human proteins, involved with the regulation of such things as pain and blood pressure. In all these cases, the protein was crystallized while containing a drug molecule or analogue of one. From a biochemical point of view, this is useful for precisely the reasons outlined in Section 1. From the point of view of technique development, it is also useful because it provides a test of quality of the data. If the structure is solved correctly and the intensity measurements are sufficiently accurate and precise, the drug molecule should be visible in the resulting electron density map. Indeed, this was the case for all these examples.

A recent success for serial femtosecond crystallography was the structure of a light-sensitive protein found in the retina, rhodopsin, bound to another protein called arrestin. The action of arrestin reverses the structural changes that take place in rhodopsin when light interacts with it, preparing it for a new cycle of light detection. The combination of both molecules is very delicate, making it very challenging to grow large, well-ordered crystals, and data could be acquired only to a resolution of 7.7 Å using conventional rotation crystallography at a synchrotron. In contrast, the structure was solved to a resolution of about 3.5 Å using serial femtosecond crystallography [18].

5 Time-resolved crystallography

Although the use of free-electron lasers for determining static structures has been the target of interest in structural biology, their potential for determining dynamic structures, or so-called 'molecular movies', is much greater. The short duration of the X-ray pulses means that the time resolution can be very high. Time-resolved serial femtosecond crystallography has now been applied to several systems where a protein responds to illumination in the visible region of the electromagnetic spectrum. In these cases, the reaction was triggered using a short pulse of light from a laser, a small fraction of a second before the arrival of the X-ray free-electron laser pulse at the sample. This is the so-called 'pump-probe' scheme, and is used in conjunction with many techniques other than X-ray crystallography.

The process of photosynthesis is a very important and interesting light-activated chemical reaction. It has therefore been (and continues to be) a high-profile target for time-resolved femtosecond crystallography [19, 20], even more so because of the large sizes of the protein complexes involved and the consequent difficulties with making large crystals.

Recently, the high time resolution achievable with an X-ray free-electron laser was demonstrated for two proteins, myoglobin [21] and photoactive yellow protein [22], in both cases achieving a time resolution close to 100 fs. Many other ways to start a chemical reaction have been proposed, such as to mix a protein with a substance it can act on [23], and acquire diffraction snapshots after a controlled amount of time for mixing and diffusion to occur.

6 Future outlook

It would be the dream of many structural biologists to determine the structure of biomolecules, at near atomic resolution, without crystallization, and progress is being made in that direction. Even though the problem of radiation damage is avoided, there are several other difficulties to be overcome. Since the signal level is still very low, only a few photons at high resolution, background scattering must be reduced to almost nothing, in turn, affecting how the protein molecules are delivered to the X-ray beam. The X-ray beam must be carefully characterized, and the detector must be well-calibrated and produce low noise. Alongside this, an interesting discovery was made recently; that certain types of disorder in a protein crystal can reveal the molecular scattering [24]. In this case, the signal is superimposed on the Bragg scattering of the crystal and extends to a higher resolution than the Bragg peaks.

It is clear that structural biology will continue to be a very important and productive application for current and future X-ray free-electron laser facilities.

Acknowledgements

I thank Henry Chapman, Dominik Oberthür, and Carolin Seuring for feedback on the manuscript, as well as funding from the Helmholtz Association through project oriented funds.

References

[1] A. Wlodawer *et al.*, *Science* **245** (1989) 616. https://doi.org/10.1126/science.2548279
[2] Q. Shen *et al.*, *J. Synchrotron Radiat.* **11** (2004) 432. https://doi.org/10.1107/S0909049504016772
[3] D. Sayre, *Acta Crystallog.* **5** (1952) 843. https://doi.org/10.1107/S0365110X52002276
[4] S. Marchesini, *Rev. Sci. Instrum.* **78** (2007) 011301. https://doi.org/10.1063/1.2403783
[5] C.J. Gilmore, *Acta Crystallogr.* **D56** (2000) 1205. https://doi.org/10.1107/S0907444900009355
[6] R. Neutze *et al.*, *Nature* **406** (2000) 752. https://doi.org/10.1038/35021099
[7] H.N. Chapman *et al.*, *Nat. Phys.* **2** (2006) 839. https://doi.org/10.1038/nphys461
[8] K. Hirata *et al.*, *Nat. Methods* **11** (2014) 734. https://doi.org/10.1038/nmeth.2962
[9] H.N. Chapman *et al.*, *Nature* **470** (2011) 73. https://doi.org/10.1038/nature09750
[10] T.A. White *et al.*, *Acta Crystallogr.* **D69** (2013) 1231. https://doi.org/10.1107/S0907444913013620
[11] T.A. White *et al.*, *J. Appl. Crystallogr.* **45** (2012) 335. https://doi.org/10.1107/S0021889812002312
[12] T.A. White *et al.*, *J. Appl. Crystallogr.* **49** (2016) 680. https://doi.org/10.1107/S1600576716004751
[13] U. Weierstall *et al.*, *Nat. Commun.* **5** (2014) 3309. https://doi.org/10.1038/ncomms4309
[14] W. Liu *et al.*, *Science* **342** (2013) 1521. https://doi.org/10.1126/science.1244142
[15] G. Fenalti *et al.*, *Nat. Struct. Mol. Biol.* **22** (2015) 265. https://doi.org/10.1038/nsmb.2965
[16] H. Zhang *et al.*, *Cell* **161** (2015) 833. https://doi.org/10.1016/j.cell.2015.04.011
[17] A. Batyuk *et al.*, *Sci. Adv.* **2** (2016) e1600292. https://doi.org/10.1126/sciadv.1600292
[18] Y. Kang *et al.*, *Nature* **523** (2015) 561. https://doi.org/10.1038/nature14656
[19] J. Kern *et al.*, *Science* **340** (2013) 491. https://doi.org/10.1126/science.1234273
[20] C. Kupitz *et al.*, *Nature* **513** (2014) 261. https://doi.org/10.1038/nature13453
[21] T.R.M. Barends *et al.*, *Science* **350** (2015) 445. http://dx.doi.org/10.1126/science.aac5492

[22] K. Pande *et al.*, *Science* **352** (2016) 725. https://doi.org/10.1126/science.aad5081

[23] M. Schmidt, *Adv. Condens. Matter Phys.* **2013** (2013) 167276. https://doi.org/10.1155/2013/167276

[24] K. Ayyer *et al.*, *Nature* **530** (2016) 202. https://doi.org/10.1038/nature16949