REFLECTIONS ON A DECADE OF SYNCHROTRON X-RADIATION PROTEIN CRYSTALLOGraphy

R. HELLIWELL
Department of Physics, University of York, Heslington, York YO1 500, England and
SERC, Daresbury Laboratory, Warrington, Cheshire WA4 4AD, England

Introduction

Approximately 10 years have elapsed since the publication of the first experiences of using synchrotron X-radiation (SR, see Fig. 1) for protein crystal data collection [1,2] (Fig. 2). Since then this sphere of work within protein crystallography has matured with a vengeance. I record here some of the highlights in the development and exploitation of this versatile source of tunable, intense X-rays and demonstrate a combing of age involving important new structural results in ezymology and virology based on new technical capabilities.

The first diffraction experiments with SR were performed by investigators interested in fibre diffraction rather than of single crystals. [3]. This work was undertaken at the DESY synchrotron in Hambrug. It clearly revealed that SR was likely to be an important tool in studying all types of biological samples (i.e. crystals as well as solutions and fibres). In addition, it showed that particle accelerator facilities could also be accessible to the smaller research teams associated with biology, chemistry and materials science. However, the interests of the community involved in SR research areas were eventually to be best served by the running of separate, dedicated accelerators for SR use. This has culminated in dedicated sources of SR in the UK (SRS), Japan (Photon Factory) and USA (NSLS, Brookhaven).

Today the protein crystallographer enjoys routine access to tunable, monochromatic X-ray beams with intensities 10^3 times and white (polychromatic) beams 10^5 times more intense than the most intense laboratory source. It is now possible, therefore, to tackle not only very difficult samples where weak scattering arises, for example from small sample volume or large unit cells but also where radiation decay over long periods of time, even in weak X-ray beams, occurs. Also, most importantly, it is now possible to study kinetic events in crystals. The tunability is being used to attack the phase problem of X-ray crystallography when the source is used in conjunction with new, sophisticated electronic area detectors.

These capabilities are vital for tackling an array of problems where crystals may not be easy to obtain or heavy atom derivatives are not easy to make. Also, since the definition of an equilibrium structure based on coordinates and B factors will at some point be related to function by a rearrangement of some part of the structure within a short time-scale, then kinetic crystallography studies, at least on some proteins, will provide important insights into structure/function relationships.

The time required to collect a complete 3-D data set at high resolution has been reduced from about a week to 30 minutes real time with the strongest monochromatic beams available [4]. Hence, kinetic experiments at the SRS on the phosphorylase b enzyme substrate complex have been realized at temperatures above 0°C where the samples are routinely stable; the first example of its kind. [5]. Further gains in monochromatic intensities will become available by up to a factor of 10 on the SRS with the introduction of a new lattice or (extra) focusing magnets later this year. Further factors of up to two to three orders of magnitude increase in monochromatic intensity are routinely discussed now for protein crystallography [6,7,8] with the next generation of sources such as the ESRF (European Synchrotron Radiation Facility) or the USA 6GeV New Ring. White beam Laue methods provide another potentially very exciting approach to kinetic crystallography allowing sub-second to msec or so data collection times with present storage rings [9,6].

In this paper I will give a personal selection c highlights in the last decade drawn from work at SR sources around the world. In my introduction I have deliberately chosen to mention areas of ap-
application where SR was essential in allowing the work to be done. There is an obvious complementarity between laboratory X-ray source work and use of a centralized RSR facility, ie the SR source provides the technical capability beyond what can be made available in the n home laboratory.

Radiation Damage

One of the earliest worries about the utilization of intense SR sources was that radiation damage would be an insurmountable problem [10]. The same sort of worries have assailed the discussions on the potential of X-ray undulator sources on the ESRF [6,7]. In the mid 1970s the pessi-

![Fig. 1. Synchrotron X-radiation has the following properties: It is intense, tunable over a wide wavelength/energy range illustrated in (a) which depicts the universal spectral curves for several machines. It is also well collimated – see (b) – in the forward direction. In addition it is polarized with a pulsed time structure.](image)
mistic expectations were completely at variance with the initial experiences. Hence, it was clear at the earliest stage of the use of SR in protein crystallography that there was a surprise in store; to collect data quickly with an intense beam could be much better, with some samples, than using a long exposure time with a weaker laboratory source (even with the sample cooled in each case). This benefit of the use of SR manifested itself in the new ease with which very high-resolution data could be measured. The first demonstration of this surprising effect was at the Stanford Synchrotron Radiation Laboratory (SSRL)-see Fig. 3. The effect was confirmed in data collection at the DCI ring in Paris with crystals of the phosphorylase b enzyme [11]. The latter observation here was accurately qualified; approximately 5 x more data could be collected per sample at DCI which was approximately 50 x or so more intense than a laboratory source and moreover the data extended uniformly to a higher resolution. These benefits were utilized in providing the data necessary for a variety of high-resolution model refinements of this and other proteins.

Before long, of course, a project would appear which needed intense SR to make it feasible at all.

Fig. 2. Schematic outline of the beam line optics and instrument at an SR source for protein or virus crystal data collection.

Fig. 3. Cone axis oscillation photographs of an azurin crystal recorded (a) at Stanford on SSRL λ=1.740 Å and (b) at CuKα on a sealed tube X-ray source. The SR based data extends to higher resolution. This is due to a time dependent damage effect whereby shorter exposures are optimal. (Ref. 1, see also Refs. 11 and 12)
Such a project is the structural study of the enzyme purine nucleoside phosphorylase (PNP) from human erythrocytes pursued in a collaborative effort between the University of Alabama and the SERC, Daresbury Laboratory at the SRS [12]. In this case, work on a conventional X-ray source was limited to a data resolution of \( \approx 4 \) Å at best. However, tests at the SRS, with a beam 200 x a laboratory source, revealed data extending to a resolution of \( \approx 2.8 \) Å. Also, 3-5 crystals yielded a complete data set: The improvement in resolution allowed the protein chain to be traced instead of simply defining the overall molecular envelope. Several heavy atom derivatives were prepared, rather easily in fact due to the large solvent content of the crystals (\( \approx 78\% \)), and data collected very rapidly on the SRS. The structure was very easily delineated from the electron density map (Fig. 4). The time taken to reach the structure (the rate constant for structure determination) was \( \approx 18 \) months to 2 years; and this on a previously considered intractable problem. The fact that this enzyme has an important medical effect as a destroyer of anti-cancer agents has brought home to the pharmaceutical industry that SR is a vital option as a tool to tackle those problems which it is interested in i.e. where well ordered crystals may not always be easily obtained.

The difficulties in predicting the behavior of protein crystals in intense SR beams has caused anguish in assessing the value of new sources of SR such as the ESRF or the USA 6 GeV ring. Fortunately a practical test of the theory [6,7] of the interaction of very intense monochromatic beams from the next generation of machine like the ESRF is available by using white beams from the correct sources. For example, the focused white beam from the SRS wiggler offers an intensity \( 10^5 \) times a laboratory source and \( 10^2 \) x the most intense monochromatic SR beams currently available [4]. Recent experience on the SRS with white beam Laue diffraction from protein crystals indicates that in a very rapid series of snap shots from a crystal such as insulin the data is unperturbed by radiation up to about 1 minute at which point radiation damage takes over (Helliwell, Moffat and Holden unpublished results).

Theoretical calculations have been made on the behavior of protein crystals in these beams in particular to predict the lifetime and the minimum

Fig. 4. The structure of purine nucleoside phosphorylase solved at 3 Å resolution using data collected at the SRS. (Ref 12).
sample size that may be considered tractable and the minimum exposure time which is considered feasible. Both these effects are related to radiation damage. The currently accepted view for kinetic crystallography is that sample heating and damage will become a severe problem as the beam intensity is increased but of course exposure time can always be reduced pro rata for a fixed sample volume [6,7]. In the small sample case the situation is less favorable. Basically in this instance the high intensity of the beam is traded off against sample volume, for a fixed exposure time. Ultimately, in the limit, a sample volume must be used where only one reflection could be derived per sample. This is hardly a practical basis for data collection. Instead by truncating the limit at the point where several thousand reflections may be derived per sample it is possible to show that 20 µm is perhaps a lower limit to the tractable sample size whatever intensities may be realized [6,7]. The physics of this limit being imposed by the absorption by protein crystals of X-rays divided by their X-ray scattering efficiency [6, 7, 8].

Small Sample Volume

In the above I have given some of the details and references dealing with the phenomenological treatment of radiation damage and its implications for the smallest sample volume that is tractable. Experimentally this is already a most promising area. Figure 5 shows that strong data can be achieved from a ≈20 µm³ sample volume (work done in a collaboration between Stanford and Daresbury). This whole area is at a very promising stage of development. The coupling of electronic area detectors with a high intensity beam will take the method to the stage of not only obviously seeing strong data but also making accurate measurements. Already, the somewhat less stringent case of samples of the order of 100 µm x 100 µm x 50 µm of the key enzyme β-lactamase has yielded the structure [13]. This is an important landmark.

Protein crystal growth was perhaps the least studied area scientifically and technically, until recently. Recently, of course, regular access to microgravity environments is stimulating considerable thought on the physics of protein crystal growth. One cannot but feel that this area will be turned from one of accepted recipes and guesses to a true science. Hence, since SR is easing the requirement for large, good quality crystals on the one hand and with these other studies as well there are grounds for considerable optimism about being able to determine any protein structure by
crystallographic methods.

**Large Unit Cells**

Data collection from very large unit cells benefits from the natural narrow beam collimation of SR to resolve adjacent spots [14] (Fig. 6). This was clearly spelt out as an important application area in the early work from DESY [2]. In addition, especially with mammalian viruses, the exposure time in the home laboratory can be so long that radiation damage effects take hold. The shortening of exposure times in the SR data collection from crystals of the rhinovirus was crucial in the remarkable study by Rossmann and coworkers [15] of the first intact picarno virus (Fig. 7). This study took the remarkably short time of one year or so to complete.

The processing of the virus crystal SR data by Rossmann et al. (1986 to be published) took advantage of a series of papers based on work at the SRS [16] which stressed the importance of a well characterized instrument to minimize the factors which smear the angular rocking width of reflections and provided the correct formulation to treat the inevitable slight deviations from the ideal geometry. Intriguingly in parallel with the SR theory a similar theory was being derived for the exactly analogous white neutron source case [17].

Further reductions in exposure times and hence radiation damage in virus crystallography may accrue from use of white beam Laue methods. In the case of say a unit cell of 200 Å with a polychromatic wavelength range of $0.2 \ < \lambda \ < \ 3$ Å a total of $\approx$200,000 reflections can be stimu-

---

**Fig. 7.** The intact rhinovirus solved using SR data from the Cornell SR source (CHESS). The particle consists of an icosahedral protein shell surrounding an RNA core. The notation refers to the unique proteins in the virus shell. The immunogenic determining proteins lie on the surface of the shell and these were revealed by the study.

**Fig. 8.** Laue diffraction pattern from a large unit cell protein crystal (E. coli purine nucleoside phosphorylase $a = 107$ Å, $c = 240$ Å P6$_{3}$22 or P6$_{5}$22), exposure time 4 seconds, SRS Wiggler (Helliwell and Moffat unpublished results).
lated at one crystal setting. With cubic symmetries this would represent almost a complete sampling of the independent data. The practical problems that have to be addressed in the recording of such patterns are primarily associated with fighting signal to noise. In very recent work at the SRS (Helliwell and Moffat unpublished). Laue patterns were recorded from crystals of the enzyme *E. coli* purine nucleoside phosphorylase (space group P6$_1$22 or P6$_5$22, $a=107$ Å $c=240$ Å). The structure is being determined by more conventional monochromatic SR methods in a collaboration between Ealick and Helliwell. However, the system also provides a test bed for attempting large unit cell Laue photography. Figure 8 shows one of the patterns obtained recently; the tactics adopted to maximize signal to noise included a minimum air path and use of an extruded polymer capillary (Ealick and Helliwell unpublished) rather than a traditional glass capillary.

**Time Resolved Crystallography**

This is an area which has been stimulated by methodologies developed for work on non-crystalline biological systems especially muscle diffraction [18]. A break-through in this area came at DORIS [19] in work on carbonmonox myoglobin crystals where structural changes were induced by the photolysis of the CO ligand by a laser on the msec time scale (Fig. 9). The difficulty associated with fast monochromatic data collection via fast refresh electronic area detectors is concerned with intensity changes that may be derived from movement of reciprocal lattice points off the Ewald sphere. Studies involving "slow" kinetics (eg where enzyme-substrate turnover rates involve time constants of many minutes), are not vulnerable to this since the data collection is safely less than the "perturbation time". This latter approach has been used to good effect in the study of the phosphorylase b enzyme substrate interactions. The first experiments performed by this group involved the substrates heptenitol and phosphate, where a large incubation time (50 h) gave rise, in the crystal itself, to a product, heptulose-2-phosphate, bound at the active site [5] (Fig. 10a). In the most recent experiments the group have been able to trap the actual enzyme-substrate complex [5]. Data collection times for =100,000 reflections to 2.8Å resolution of approximately 25 minutes have allowed a variety of reactions to be explored with the crystals held at temperatures about 0°C where they are routinely stable [5]. It seems somewhat paradoxical that the first kinetic crystallographic study should be on the largest single polypeptide chain protein (843 amino acids) being studied crystallographically. However, the presence of allosteric control sites on the enzyme allows the turnover rate of the enzyme to be suitably adjusted [5].

![Fig. 9. Very fast (msec) time resolved protein crystallography. Feasibility experiment to evaluate data collection from CO myoglobin following laser photolysis of the ligand. (Ref. 19).](image)
Another approach to these experiments is also being followed involving the full white beam from the SRS Wiggler. Laue patterns can be recorded in approximately 250 msec (Fig. 10b) which contains ≈30% of the unique data in one snapshot. The geometry of the Laue method, involving a stationary crystal, is ideally suited to work with a flow cell. These patterns have already

Fig. 10. Kinetic crystallography on the phosphorylase b enzyme-substrate interactions. In (a) the substrate heptulose-2-phosphate, formed in the crystal, is shown at the active site. Very fast monochromatic data collection of up to 200 reflection intensities/sec peak rate are measured with SR; a whole data set to 2.8 Å can be collected in 25 minutes. To improve the time resolution into the sub-second region full white beam (0.5 Å < λ < 2.5 Å) Laue data collection is being used. In (b) is shown a prediction of the pattern collected in a single snapshot comprising about 30% of the unique data to 2.8 Å which can be recorded in 250 msec (Ref 5).
yielded an interesting result—diffusion of phosphate into the crystals results in a 60 second or so transient, uniaxial increase in mosaic spread which is clearly revealed via the continuous mapping of reciprocal space onto the film that is "instantly" available with the Laue geometry.

**Statistical Quality of SR Data**

It has been a common observation that data collected at an SR source extends to higher resolution, as mentioned earlier and concommitantly the data quality has improved significantly. Hence not only can the data be recorded very rapidly but with better agreement statistics. Of the options available at the SR source the optimum choice will involve short wavelengths, for example, since here the sample absorption surface can be completely flattened, an improvement over even the best attempts to fit an empirical surface.

The quality of white beam Laue data can be judged from the statistics derived for pea lectin [20]. This work shows that the data quality of Laue patterns is adequate for many time dependent studies.

However, film is not good enough for multiple wavelength phasing methods that demand the accuracy brought with an electronic area detector (see below). The combination of an area detector with an SR source and its routine use for multi-\(\lambda\) measurements has so far eluded users of these facilities.

**MAD (multiple wavelength anomalous dispersion) Phasing Methods**

The use of the variation in atomic scattering factors of heavy atoms especially close to their X-ray absorption edges was identified from the start as a key application area. It has taken the longest to develop because of the need for instrumentation to make the most accurate measurements of the structure amplitudes as a function or \(\lambda\). The differences, due to \(f'\) and \(f''\), that one is attempting to measure, are smaller than those derived from isomorphous substitution. To achieve the maximum signal to noise, careful selection and stability of \(\lambda\) is paramount and optimum data statistics are required. These requirements are attainable but both instrument and source must perform optimally.

On the experimental side there is still discussion as to whether a narrow \(\delta\lambda/\lambda\) (10\(^{-4}\)) is important to access absorption edge fine structure [21], and in particular its polarization dependence [22] or whether to use a somewhat less \(\delta\lambda/\lambda\) (10\(^{-3}\)) to smooth out these variations [23] or indeed to induce spatially dispersed wavelength dependent profiles [24]. The ubiquity of X-ray absorption edges for atoms commonly binding to proteins allows widespread utilization of these effects unlike the neutron multi-\(\lambda\) case where few nuclei are suitable. Other trends are emerging. Lanthanides are attractive from the point of view that large white line effects (XANES) occur [21] but the chemistry of lanthanide derivatisation is more difficult. Chemically the high atomic number elements easily form derivatives but the XANES is less marked [24]. Finally, in the wavelength range 0.5 Å to 3 Å the technical specification and needs for the detector change, ie the absorption efficiency, window losses, pixel size and collection aperture requirements change [25].

All these factors have slowed down the routine utilization of these effects. The area is, however, still one of great promise for routine use in many problems. Not only will it allow those crystal structures to be tackled which cannot easily be solved with isomorphous replacement it may also...
provide the means for very rapid solving of a structure. This is especially true if combined with new theoretical approaches to phasing that are being adapted, such as maximum entropy [26]. Work at Stanford (SSRL) on lamprey haemoglobin [27] and at Paris (DCI) on Tb-parvalbumin [23] as test systems has now resulted in good quality electron density images being produced. This is an important landmark. Figure 11 shows the dispersion curves for Tb used in the DCI work (Ref. 23).

Conclusion

A remarkable amount has been achieved in the last 10 years. Structural results have emerged which could not have been achieved in any other way. Details of the enzyme-substrate complex in phosphorylase b have been obtained via very fast monochromatic data collection methods. The observation of reduced radiation damage with short exposure times and a consequent increase in resolution has been utilized to solve the structure of purine nucleoside phosphorylase and the rhino virus. The field has come of age. There is more to come with the routine use of small sample volumes, faster kinetics and multiple wavelength phasing methods.

Acknowledgements

I wish to thank the University of York and SERC, Daresbury Laboratory for support and the SERC, EMBO, NATO and the Royal Society for Research Support. In particular I wish to record my thanks to Dorothy Hodgkin who drew my attention, as a research student, to the existence of synchrotron radiation in 1975.

References