Overview and new developments in softer X-ray (2Å < λ < 5Å) protein crystallography

John R. Helliwell

Department of Chemistry, University of Manchester, Manchester M13 9PL, UK. E-mail: john.helliwell@man.ac.uk

New methodologies with synchrotron radiation and X-ray Free Electron Lasers (XFELs) in structural biology are being developed. Recent trends in harnessing softer X-rays in protein crystallography for phase determination are described. These include reference to a data collection test at 2.6 Å wavelength with a lysozyme crystal on SRS station 7.2 (Helliwell 1983) and also use of softer X-rays (2Å wavelength) to optimise f " at the xenon L_1 absorption edge in the **S**ingle **I**somorphous **R**eplacement **O**ptimised **A**nomalous **S**cattering (**'SIROAS'**) structure determination of apocrustacyanin A1 with four, partially occupied, xenon atoms (Cianci et al 2001; Chayen et al 2000). The hand of the protein was determined using the f " enhanced sulphur anomalous signal from 6 disulphides in the protein dimer of 40kDa.. In a follow up study the single wavelength xenon L_1 edge f " optimised data set alone was used for phase determination and phase improvement by solvent flattening etc (CCP4 DM) (Olczak et al 2003). Auto-tracing of the protein was feasible but required additional diffraction data at higher resolution. This latter could be avoided in future by using improved tilted detector settings during use of softer X-rays ie towards back scattering recording (Helliwell 2002). The Olczak et al study has already led to optimisation of the new SRS beamline **MPW MAD 10** (see www.nwsgc.ac.uk) firstly involving the thinning of the beryllium windows as much as possible and planning for a MAR Research tilted detector 'desk top beamline' geometry. Thus the use of softer ie 2 to 3 Å wavelength range X-rays will allow optimisation of xenon and iodine L edge f " and enhancing of sulphur f " signals for higher throughput protein crystallography. Softer X-rays utilisation in protein crystallography includes work done on SRS bending magnet station 7.2 in the early 1980s by the author as station scientist (Helliwell (1984)). In the future development of XFELs these softer X-ray wavelengths could also be harnessed and relax the demands to some extent, on the complexity and cost, of an XFEL. Thus, by use of say 4Å XFEL radiation and use of a back scattering geometry area detector the single molecule molecular transform could be sampled to a spatial resolution of 2 Å sufficient, in principle, for protein model refinement (Miao et al 1999). Meanwhile Miao et al (2003) report the first experimental recording of the diffraction pattern from intact Escherichia coli bacteria using coherent x-rays, with a wavelength of 2 A, at a resolution of 30nm and a real space image constructed. The new single particle X-ray diffraction-imaging era has commenced.

1. Introduction

The pace of gene sequencing, from which protein amino acid sequences are derived, is phenomenal. The pace of three-dimensional protein structure determination is also accelerating quickly, offering experimental capabilites in structure and function definition on a large numbers scale. Today the Protein Data Bank holds some 20000+ protein structures of which some 90% are derived from protein crystallography. The remainder are mainly derived from NMR solution and also electron diffraction structure determination as well as neutron protein crystallography and molecular modelling. The number derived from using synchrotron radiation is growing rapidly (Helliwell, 1992; Cassetta et al 1999). In the future the prediction of protein fold from amino acid sequences may become possible, which will further accelerate the pace of experimental protein structure and function determination eg via new forms of molecular replacement.

2. A 'phylogenetic tree' of SRS PX instrumentation as an example

The development of synchrotron-radiation protein-crystallography beamline instrumentation initially, some 20 years ago, encompassed two separate approaches (high-intensity optics versus rapidly tunable optics; e.g. see Helliwell, 1979). A major challenge was how to harness the typical first- and second-generation synchrotron radiation source emittance (source size and divergence) available at that time so as to match the available protein crystal sample acceptance (crystal size and mosaicity). The emittances of synchrotron radiation sources have improved considerably over the years. On a third-generation high-brilliance synchrotron radiation source, rapid tunability, needed for measuring more than one wavelength around an elemental absorption edge to vary the anomalous dispersion signal, can be provided whilst simultaneously having a high-intensity X-ray beam at the sample. The reflection intensities can thereby be measured quickly, precisely and accurately.

The Daresbury SRS was the first dedicated synchrotron radiation X-ray source and as such, being non-parasitic, is a second-generation synchrotron radiation source. It came on-line in 1981. The first SRS protein crystallography instrument was station 7.2 on the very first X-ray beamline at SRS. The station optics (Helliwell et al., 1982) comprises a vertically focusing mirror in 1:1 focusing mode and an oblique-cut focusing monochromator: a single crystal of Ge(111) cut at 10° to the surface. The SRS bending-magnet beamline 7 source sizes of 0.4 × 14 mm2 were thus focused to 0.4 × 1.4 mm2 at the sample position, quite a reasonable match to typical sample sizes (at that time) of 0.5 mm cross section. Station 7.2 has served a national and international user community. The SRS had a superconducting wiggler magnet inserted in 1983. This allowed the development of a second PX station but with an order of magnitude higher intensity at wavelengths around 0.9 Å, the critical wavelength of emission of the wiggler (compared with 4 Å for the equivalent parameter on the bending-magnet beamline 7). Station 9.6 came on-line in 1984. The beamline optics again were tailored to the rather large SRS wiggler source size in the horizontal of ~14 mm but with a fine vertical source size of again ~0.5 mm. Thus 1:1 focusing in the vertical direction via a focusing curved mirror and a 10:1 oblique-cut demagnifying Si(111) monochromator was used (Helliwell et al., 1986). Improved brilliance (or brightness) of the SRS came in 1985, whereby the horizontal source size challenges referred to above were greatly alleviated. A new beamline optic became possible whereby the 1:1 focusing of a (toroid) mirror alone was sufficient for many protein crystal samples and the monochromator did not need then to have a focusing role. Rapid wavelength tuning (XAFS style) became possible with reasonable intensity *on the same station*. Rapidly tunable MAD experiments were thus going to be feasible. A rapidly tunable wiggler station 9.5 design based on a toroid mirror optic and a double-crystal monochromator was made (Brammer et al., 1988). The use of station 9.5 has allowed the development of rapidly tunable MAD experiments at SRS.

More recently SRS brought on line two new stations on a new multipole wiggler beamline (Duke et al., 1998). These are of the slow tunable design, like SRS 7.2/9.6. The need for rapid tuning of the wavelength has perhaps eased in that cryocooling of the sample allows whole data sets to be collected with little or no radiation

 © 2004 International Union of Crystallography • Printed in Great Britain – all rights reserved

damage. Therefore, time-dependent variations in the measurements are now due to beam-intensity fluctuations alone. There has also been a standardization towards the most popular element and absorption edge being selenium (Hendrickson et al., 1990). It has been practical, with this more homogeneous set of conditions, to work with the `slow tunable' optic design of 7.2/9.6 for Se MAD data collection.

Within Europe ESRF BM14 has become a very successful MAD station (Cassetta et al 1999). It is an order of magnitude higher than the intensity of the monochromatic SRS 9.5 intensity, thus matching the SRS 9.6 intensity `slow tunable' station design. The beamline BM14 optics scheme was based on the SRS 9.5 approach but with the addition of a collimating pre-mirror. The new SRS MPW MAD 10 beamline (see www.nwsgc.ac.uk) effectively has to match and surpass the BM 14 benchmark of success.

3. Overview of phasing strategies

Multiple (i.e. three or more) wavelength 'MAD' (e.g. see review by Hendrickson and Ogata (1997)) techniques have dominated synchrotron radiation based phasing of protein crystal structures in the last decade. Two-wavelength 'TW' phasing (Okaya and Pepinsky (1956)) at SR sources (Helliwell 1979, Peterson et al., (1996), Gonzalez (2003)) is now growing in popularity as a more beam time efficient method. Seleno-methionine incorporation into recombinant proteins (Horton et al., 1989) is the most favoured method of preparing samples suitable for MAD (Multiple-wavelength anomalous dispersion) phasing.

There is mounting evidence that one-wavelength anomalous scattering (OAS, otherwise known as SAD) may be sufficient to solve protein structures (starting with Wang (1985); and more recently see for example Liu et al (2000), Hao (2000) and Dauter (2002)). Very recently Olczak et al 2003 address an alternative strategy involving use of softer X-rays at a single wavelength (which they call Softer-SWAT; SWAT i.e. single wavelength anomalous technique) in optimizing Xe f ". It is also relevant in enhancing the use of sulphur f ". 2 Å wavelength was termed *softer X-rays* (Chayen et al., 2000), since the tradition for diffraction involves CuKα radiation (1.54 Å) or harder X-rays delivered by synchrotron sources (typically 0.9 Å), and soft is usually taken to mean 5 Å or longer wavelength (Carpentier et al., 2000; Behrens et al., 1998). The feasibility of using softer X-rays to collect PX diffraction data had already been determined with experimental work using a lysozyme crystal up to 2.6 Å wavelength on station 7.2 (Helliwell, J.R., 1983), figure 1, in turn based on the development and utilization of 2 Å, 1.89 Å, 1.74 Å and 1.488 Å wavelengths with a variety of protein crystals and anomalous dispersion utilisation (Helliwell, 1984). The routine use of softer Xrays has been emphasized by Weiss et al 2001 using Elettra data. Evans et al (2003) successfully used 2 Å softer X-rays for iodine L edge f " based phasing tests with elastase using Elettra data. Liu, Ogata and Hendrickson (2001) conducted test 4-wavelength MAD experiments at the M-IV edge of uranium (3.326 Å), giving very large f ' = -70e and f " = 80e signals, with crystals of porcine elastase derivatized with uranyl nitrate, and phase information of good accuracy to 3.2 Å resolution. Even longer wavelengths have been explored by Stuhrmann and coworkers (Stuhrmann et al 1997) in protein crystallography namely around the sulphur K edge of 5 Å for trypsin as test and at the phosphorous K edge (5.78 Å) for the 30S ribosomal subunit of Thermus thermophilus (Stuhrmann et al 1995).

New protein structural results harnessing X-ray wavelengths longer than the more conventional CuKα (1.54 Å) or SR ~0.9 to 1.1 Å instrument settings, include the following:- an anomalous dispersion study at the Mn K edge using a wavelength of 1.86 Å at SRS 7.2, which allowed the Mn ion sites to be distinguished from the Ca ion sites in pea lectin (Einspahr et al 1985); enhanced sulphur anomalous scattering at a wavelength of 1.74 Å at APS SERCAT allowed the de novo structure of obelin to be determined (Liu et al 2000); similarly to obelin was use of 1.77 Å wavelength data to determine the de novo structure of apocrustacyanin C1 (Gordon et al 2001). Also Micossi et al 2002 undertook de novo phasing of the structures of two crystal forms of tryparedoxin II from Crithidia fasciculata using single-wavelength anomalous diffraction techniques exploiting only the small anomalous signal from the S atoms intrinsic to the native protein with data collected at 1.77Å wavelength.

Recently the first de novo protein structure has been solved with X-rays >2 Å wavelength namely of apocrustacyanin A_1 (Cianci et al., 2001). This has been solved by means of the single isomorphous replacement (Xe) with optimized wavelength (2 Å) anomalous scattering method (SIROAS). The hand was determined using the sulphur anomalous signal. As a result of a large xenon f " at this wavelength (11.5e, 4e more than the value at CuKα (7.2e), and 8e more than at 0.9 Å (~ 3e), more than a thousand reflections with $\Delta F_{ano} > 3\sigma$ (ΔF_{ano}) to 2.3 Å resolution were collected. Olczak et al 2003 used the apocrustacyanin A_1 case to assay viability of a single wavelength anomalous technique (SWAT) being applied to phase the structure, even with a low occupancy of xenon sites (4 sites in all with occupancy ~0.4, ~0.4, ~0.3, ~0.1 respectively).

It is also worth recalling that Blow (1958) discussed the use of chromium Kα radiation (2.2 Å) for protein crystallography. Also MSC Rigaku now have a chromium Kα rotating anode with large area IP device for protein crystallography sustained by extensive phasing tests involving use of sulphur or other light elements based phasing (Ferrara et al 2002).

In Cassetta et al 1999 we reviewed the development of MAD methods and instrumentation. Thus this turning of our attention to single wavelength, but tuned, phasing is an important change of our thinking. The spirit of the original proposition (Helliwell 1979) worrying about the optimal use of SR beamtime is still highly relevant ie for realising highest throughputs possible from a beamline. The overall efficiency of the global complement of PX beamlines has a major bearing on the rate of depositions into the PDB. Softer X-rays target important anomalous scattering elements in protein crystallography namely xenon and iodine; their L edges deliver large f " signals. There are wider long term implications in the XFEL arena too, as now described in the last section below.

4. Extrapolations of the softer (2Å < λ < 5Å) X-ray single crystal data collection approach to the XFEL era

The basic ideas of using softer X-rays and extending the approach to backscattering data collection has potential in the XFEL arena that is upcoming in the next decade. Single molecule protein data collection via FEL X-ray flash methods, using multiple orientations (eg upto 100,000 shots) would allow important proteins that will not crystallise to have their structures determined. Feasibility of the algorithms has been successfully reported (Hodgson et al 2001). The viability of data collection in the face of radiation damage effects has been reported by Neutze et al (1999); a principal challenge is to record single protein molecule diffraction data before the onset of a protein molecule structure breaking up. There are also technical complexities associated with the XFEL itself not least the cost of producing 1 Å X-rays as in the TESLA design or 1.5A X-rays in the LCLS design. In protein crystallography even longer wavelengths have been explored by Stuhrmann and coworkers (Stuhrmann et al 1995, 1997) namely around the sulphur and phosphorus K edges of 5 and 5.78 Å, as referred to above. The first de novo protein crystal

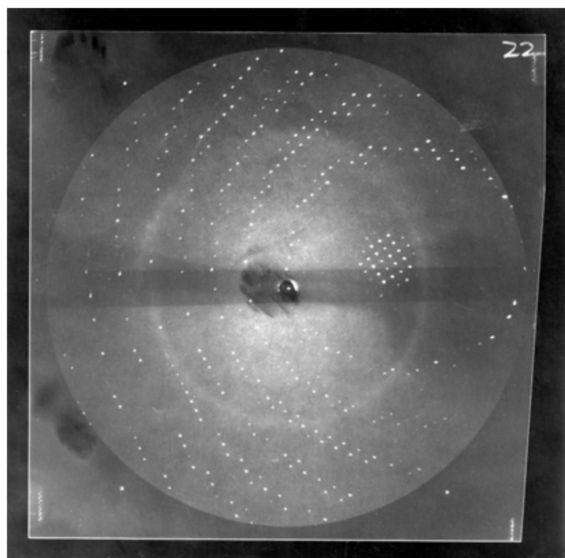
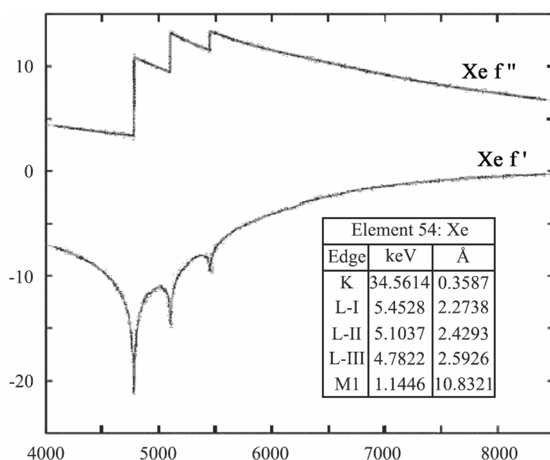


Figure 1

PX data collection at 2.6Å wavelength has been shown to be viable using SRS 7.2; example a lysozyme single crystal wrapped in a ‘roll your own’ mylar capillary (Helliwell 1983).



Element 54: Xe		
Edge	keV	Å
K	34.5614	0.3587
L-I	5.4528	2.2738
L-II	5.1037	2.4293
L-III	4.7822	2.5926
M1	1.1446	10.8321

Figure 2

Anomalous dispersion curves for the xenon L edges.

structure has now been solved at atomic resolution using X-rays >2 Å wavelength (Cianci et al 2001). At full back scattering the use of 5 Å X-rays would restrict the data resolution to 2.5 Å. A better working wavelength than 5 Å would be 4 Å, which would allow 2 Å resolution diffraction data to be collected. There may also remain interest in the use of XFEL wavelengths at the xenon L edges eg to control the value of xenon f ' values for protein structure determination where differences at two wavelengths in single molecule diffraction data could reveal the location of xenon atoms. A similar two wavelength, delta f ', approach would work for iodine or sulphur to locate their positions in a protein single molecule imaging experiment as well. Such stepwise structural approaches may prove valuable in challenging single molecule protein structure determinations via the XFEL route.

Meanwhile Miao et al (2003) report the first experimental recording of the diffraction pattern from intact Escherichia coli bacteria using coherent X-rays, with a wavelength of 2 Å, at a resolution of 30nm and a real space image constructed. The new single particle X-ray diffraction-imaging era has commenced.

References

Behrens, W., Otto, H., Sturmann, H.B. and Heyn, M.P. (1998) *Biophys. J.* 75, 255-263.
Blow D M (1958) Proc Roy Soc A 302-336.
Carpentier, P., Berthet-Colonines, C., Capitan, M., Chesne, M.L., Fonclon, E., Leguien, S., Sturhmann, H., Thiadiere, D., Vicart, J., Zelinski, P. and Kahn, R. (2000) *Cell Mol. Biol.*, ***46,*** 915-935.
A. Cassetta, A. M. Deacon, S. E. Ealick, J. R. Helliwell and A. W. Thompson *J. Synchrotron Rad.* (1999). 6, 822-833
Cianci, M., Rizkallah, P. J., Olczak, A., Raftery, J., Chayen, N. E., Zagalsky, P. F. & Helliwell, J. R. (2001Acta Cryst. D57, 1219-1229.
Chayen, N. E., Cianci, M., Olczak, A., Raftery, J., Rizkallah, P. J., Zagalsky, P. F. & Helliwell, J. R. (2000). Acta Cryst. D56, 1064-1066.
Djinovic-Carugo, K., Everitt, P., Tucker, P.A. (1998) *J.Appl.Cryst*. **31**, 812-814.
Dauter, Z., Dauter, M. and Dodson, E. (2002) *Acta Cryst*. D**58**, 494-506.
Einspahr, H, Suguna, K, Suddath, F L, Ellis, G, Helliwell, J R & Papiz, M Z (1985) Acta Cryst B41, 336-341.
Evans G, Polentarutti, M, Djinovic Carugo, K and Bricogne G (2003) D59, 1429-1434.
Ferrara J D, Yang C, Pflugrath J and Courville D A (2003) Acta Cryst A58 (Supplement) C300.
Fourme, R. Shepard, W., Schiltz, M., Prangé, T., Ramin, M., Kahn, R., de la Fortelle, E. and Bricogne, G. (1999) *J.Synchrotron Rad*., **6**, 834-844.
Gonzalez, A. (2003) *Acta Cryst.* D**59**, 315-322
Gordon, E J , Leonard, G A, McSweeney, S and Zagalsky P F (2001) Acta Cryst D57, 1230-1237.
Hao, Q., Gu, Y. X., Zheng, C. D. & Fan, H. F. (2000). *J. Appl. Cryst.* **33**, 980-981.
Helliwell, J.R., Greenhough, T.J., Carr, P.D., Rule, S.H., Moore, P.R., Thompson, A.W., and Worgan, J.S. (1982) *J.Phys.* E**15**(12), 1563-1372.
Helliwell, J.R., (1979) *Daresbury Study Weekend Proceedings* DL/SCI **R13**, 1-6.
Helliwell J R (1983) unpublished
Helliwell J R (1984) Reports on Progress in Physics 47, 1403-1497.
Helliwell J R (2002). J. Synchrotron Rad. **9**, 1-8
Hendrickson W.A. & Ogata C.M. (1997) *Methods in Enzymology* 276, 494-523.
Hendrickson W.A., Horton J.R., LeMaster D.M. (1990) *EMBO J* **9**,1665-72.
Liu Y, Ogata CM & Hendrickson WA (2001) PNAS USA 98 (19): 10648-10653.
Liu, Z-J, Vysotski, E S, Chen, C-J, Rose, J P, Lee J and Wang B C (2000) Protein Science 9 2085-2093.
Miao JW, Hodgson KO and Sayre D PNAS
98 (12): 6641-6645 2001
Miao JW, Hodgson KO, Ishikawa T, Larabell CA, LeGros MA and Nishino Y. PNAS 100 (1): 110-112 2003
Micossi E, Hunter W N and Leonard G A (2002) Acta Cryst D 58, 21-28.
Neutze R, Wouts R, van der Spoel D, Weckert E, Hajdu J (2000)
Nature 406 752-757
Okaya and Pepinsky (1956) Phys. Rev. 103, 1645-1647.
Peterson M Peterson, M. R., Harrop, S. J., McSweeney, S. M., Leonard, G. A., Thompson, A. W., Hunter, W. N. & Helliwell, J. R. (1996) *J. Synchrotron Rad.* **3**, 24-34.
S. Stuhrmann, M. Hütsch, C. Trame, J. Thomas and H. B. Stuhrmann *J. Synchrotron Rad.* (1995). **2**, 83-86
S. Stuhrmann, K. S. Bartels, W. Braunwarth, R. Doose, F. Dauvergne, A. Gabriel, A. Knöchel, M. Marmotti, H. B. Stuhrmann, C. Trame and M. S. Lehmann*J. Synchrotron Rad.* (1997). **4**, 298-310
Wang, B.C. (1985) *Methods in Enzymol*, 115, 90-112.
M. S. Weiss, T. Sicker, K. Djinovic-Carugo and R. Hilgenfeld
Acta Cryst. (2001). D**57**, 689-695.

Received 30 April 2003 • Accepted 15 October 2003