J. Synchrotron Rad. (1997). **4**, 17–20

Protein Crystallography in the Soft X-ray Region: Crystal Lifetime and Diffraction Efficiency

I. Polikarpov

Laboratório Nacional de Luz Sincrotron/LNLS, CNPq, CP 6192, CEP 13083-970, Campinas, SP, Brazil

(*Received 2 April 1996; accepted 27 September 1996*)

Multiwavelength anomalous diffraction on the sulfur (5.01 Å) and phosphorus (5.77 Å) absorption *K*-edges seems to offer a solution to one of the most appealing problems of protein crystallography – the determination of the phases of structure amplitudes. However, a strong increase in absorption of X-rays in this region may impede the development of this method of structure solution. Analytical expressions for diffraction efficiency and normalized diffraction efficiency have been derived in the present paper. It is shown that the crystal lifetime in the soft X-ray region will be significantly shorter than the lifetime of a macromolecular crystal exposed to 1–1.5 Å X-rays during diffraction data collection. An optimum crystal size has been estimated.

Keywords: protein crystallography; soft X-rays; multiwavelength anomalous diffraction; crystal lifetime; diffraction efficiency.

1. Introduction

The basic problem in protein crystal structure determination is the problem of determining the phases of structure factors. Multiwavelength anomalous diffraction (MAD) is becoming increasingly popular in macromolecular structure solution due to the improvements in synchrotron radiation instrumentation, computing facilities and specialized software. In MAD experiments all the necessary data can be collected from one single crystal at several wavelengths around the absorption edge of a heavy metal bound to the macromolecule. The same crystal then serves as the derivative as well as the native at different wavelengths and the problem of non-isomorphism does not appear. However, apart from metal-containing proteins, useful heavy-metal derivatives or selenomethionyl-substituted have to be produced.

Soft X-rays are of use in protein crystallography due to the possibility of MAD experiments on the sulfur (5.01 Å) and phosphorus (5.77 Å) absorption *K*-edges. Two out of twenty naturally occurring amino acids (Cys and Met) contain sulfur. Ribonucleic acid (RNA) and deoxyribonucleic acid (DNA) contain phosphorus. It is also present in adenosine triphosphate (ATP) and guanosine triphosphate (GTP).

Hendrickson & Teeter (1981) successfully used sulfur as a principal anomalous scatterer in their crystallographic study of crambin with conventional Cu $K\alpha$ radiation even though the sulfur absorption edge at 5 Å was far removed from the 1.54 Å wavelength used.

The idea of using an anomalous signal from sulfur or phosphorus naturally occurring in proteins seems to be very attractive because in this way one can avoid the preparation of a derivative suitable for MAD. Recent experimental studies showed the feasibility of structure amplitude phase determination in soft X-ray diffraction studies (Stuhrmann & Lehmann, 1994; Stuhrmann, Huetsch, Trame & Stuhrmann, 1995). There are, however, questions that must be answered. What is the lifetime of the protein crystal exposed to soft X-rays? Can one collect a decent data set (or data sets) from one crystal? What is the optimum crystal size? In the present paper we will attempt to answer these questions.

The purpose of this work is to derive analytical expressions for diffraction efficiency and average normalized diffraction efficiency to estimate the optimum crystal size of a macromolecular crystal to be used in soft X-ray protein crystallography. It is shown that protein crystals will suffer less damage from soft X-rays if they have an optimum size, which is four to five times larger than the absorption length of soft X-rays in the crystal. However, protein crystals, even with an optimized size, will survive for a significantly shorter time under the soft X-ray beam than their larger compatriots irradiated with 1–1.5 Å radiation in the process of diffraction data collection.

2. Theoretical considerations

2.1. *Integrated diffraction intensities*

An integrated intensity of X-rays diffracted by a mosaic crystal in the direction of diffraction is written as

$$P = \sum_{s=1,2} N^2\lambda^3 e^4 |F|^2 C_s^2 t S \exp(-\mu L)/[\sin(2\theta_b) 2m^2c^4] \quad (1)$$

where N is the number of atoms per unit volume, λ is the X-ray wavelength, F is a structure factor, θ_b is the Bragg angle, e and m are the charge and mass of an electron, respectively, c is the velocity of light, L is the distance travelled by X-rays in the crystal, μ is the

linear absorption coefficient, $C_1 = 1$ for radiation polarized perpendicular to the plane of diffraction (σ polarization) and $C_2 = \cos(2\theta_b)$ for X-rays polarized in the diffraction plane (π polarization), t is the crystal thickness, and S is its cross section. The linear absorption coefficient, μ, is, in a first approximation, also a function of λ^3:

$$\mu = a\lambda^3, \tag{2}$$

where $a \simeq 0.22\,\text{mm}^{-1}\,\text{Å}^{-3}$ (Arndt, 1984).

The integrated diffraction intensity as a function of wavelength and an average crystal size for a cubic shape crystal can be written as

$$P \propto \sum_{s=1,2} \lambda^2 t^3 |F|^2 C_s^2 \exp(-\mu t)/(1 - \lambda^2/4d^2)^{1/2}, \tag{3}$$

where d is the maximum resolution range of diffraction.

If we consider a plate-like crystal with a constant cross section, S, and a varying thickness, t, (3) can be rewritten as

$$P \propto \sum_{s=1,2} \lambda^2 t S |F|^2 C_s^2 \exp(-\mu L)/(1 - \lambda^2/4d^2). \tag{4}$$

Differentiating (3) by λ one can determine the optimum (in terms of integrated intensity of diffraction per incident photon) wavelength of the X-rays as a function of the size of the crystal (Rosenbaum & Holmes, 1980):

$$\lambda = [2/(3at)]^{1/3}. \tag{5}$$

A plot of an optimum synchrotron radiation wavelength as a function of crystal size is given in Fig. 1.

2.2. *Diffraction efficiency*

Part of the X-ray radiation passing through a protein crystal gets absorbed. The energy deposited in a cubic shape sample as a result of photoelectric absorption is

$$E_{\text{abs}} \propto I_0 t^2 (\hbar c/\lambda)[1 - \exp(-\mu t)]\,. \tag{6}$$

Here, I_0 is the incident intensity through the unit cross section of the sample, $\hbar$ is the Planck constant and c denotes the speed of light.

In protein crystallography one would like to maximize the integrated diffraction intensity of X-rays diffracted by the macromolecular crystal in order to obtain better diffraction data statistics. At the same time, energy absorbed in the crystal should be kept to the lowest possible limit because it is the primary source of radiation damage and, as a consequence, of degradation of diffraction data. In other words we are interested in the optimization of a ratio: integrated diffraction intensity from a protein crystal *versus* the energy of X-ray photons deposited in the crystal during the process of diffraction data collection. We will call this ratio the diffraction efficiency.

The energy of X-ray quanta deposited in the sample strongly increases at longer wavelengths because of the increase in absorption of X-rays by the crystal. For a particular wavelength the deposited energy grows non-linearly with the increase of the sample size. At first, while $\mu t \ll 1$, the integrated diffraction intensity as a function of the size of the crystal increases, and it then decreases when μt becomes larger than unity.

The ratio of integrated diffraction intensity to the energy of the X-ray photons deposited in a cubic shape crystal during the process of diffraction, *i.e.* diffraction efficiency, can be approximated as

$$(P/E_{\text{abs}}) \propto [\lambda^3 t \exp(-\mu t)]/[1 - \exp(-\mu t)]\,. \tag{7}$$

It follows from (7) that diffraction efficiency, (P/E_{abs}), is a function of two parameters: the size of the crystal and the wavelength of the X-rays. The wavelength dependence of the diffraction efficiency normalized to 100% at the short-wavelength limit for three different values of the parameter t is shown in Fig. 2. For all three crystal sizes the diffraction efficiency is equal to 100% at very short wavelengths. For a very small crystal (t = 15 µm) the diffraction efficiency decreases very slowly as the X-ray wavelength increases (see Fig. 2). At 5.77 Å it is still ~75% of the initial value. For the larger 25 µm crystal, the diffraction efficiency

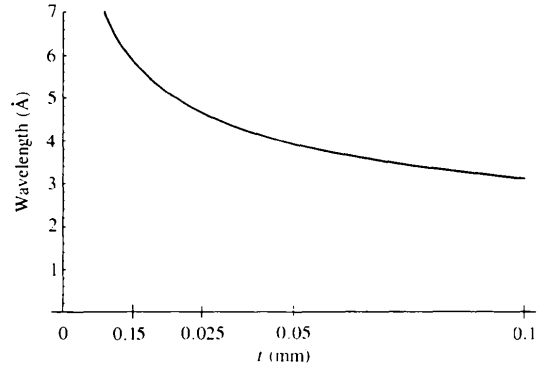

Figure 1
The long-wavelength limit as a function of the size of the crystal computed according to equation (5). The wavelengths plotted in this figure correspond to the ~75% level of the diffraction efficiency (see Fig. 2).

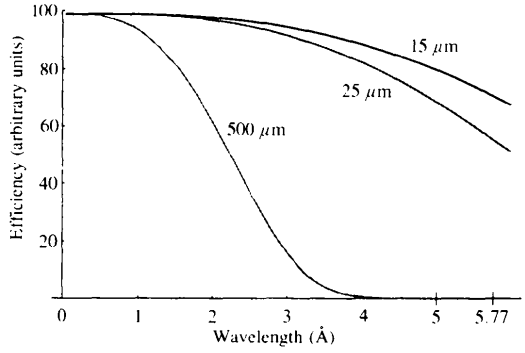

Figure 2
Wavelength dependence of the diffraction efficiency [equation (7)] normalized to 100% at the short-wavelength limit for three different thicknesses, t, of the crystal. Maximum optimum wavelengths 5 and 5.77 Å for the 25 and 15 µm crystals, respectively, are shown as ticks on the x axis.

declines more rapidly. It reaches the 75% level at ~5 Å and at 6 Å it is less than 60%. For a larger 500 μm crystal the diffraction efficiency decreases even faster. It reaches the 75% level at ~1.54 Å, and at 4 Å it practically vanishes.

This functional behaviour is not surprising. The integrated diffraction intensity for a thin crystal ($t \ll \mu^{-1}$) initially grows as λ^2 until the absorption of the X-ray quanta inside the crystal becomes less significant [see equations (1)–(4)]. A further increase in the wavelength of the X-rays will lead to diffraction in a thick crystal ($t \gg \mu^{-1}$), where most of the X-ray quanta will be absorbed before leaving the bulk of the protein crystal and the integrated diffraction intensities will drop exponentially. This is when the diffraction efficiency rapidly decreases.

For a plate-like crystal the term μt in (7) should be substituted by μL. In this situation diffraction efficiency is not only a function of crystal thickness and X-ray wavelength but also of the optical path length, L. The diffraction efficiency for some of the reflections in a diffraction data set will be lower than for others. This will be the case for reflections collected while the protein crystal is oriented with its long dimension along the incident or diffracted beam. If the crystal happened to have very different dimensions (*e.g.* a thin plate-like crystal), these reflections will have poor statistics. Absorption corrections therefore have to be applied and data rejection criteria have to be analyzed carefully.

The integrated diffraction intensity from a protein crystal *versus* the energy of X-ray photons deposited in a crystal during the process of diffraction data collection is almost the same for a 500 μm crystal at 1.54 Å X-ray wavelength as compared with a 25 μm crystal diffracting 5 Å radiation. There is one important difference: the number of macromolecules in a 500 μm crystal is significantly larger than the number of molecules contained in a 25 μm crystal. This means that the radiation load per molecule in the second case will be much higher.

2.3. *Averaged normalized diffraction efficiency*

An averaged normalized diffraction efficiency, D, which is the integrated diffraction intensity per average energy deposited in a crystal per molecule, is

$$D \propto [\lambda^3 t V \exp(-\mu t)]/[1 - \exp(-\mu t)], \qquad (8)$$

where V is the volume of the crystal exposed to the X-rays. For a cubic-shape crystal, $V = t^3$.

Some representative cases of functional behaviour of D for a cubic-shape crystal are presented at Fig. 3. If the crystal is very small, *e.g.* 25 μm, then D does not vary too much with the wavelength (see Fig. 3), and at 5 Å it is still ~75% of what it was at short wavelengths. However, the absolute value of D is very low. If the crystal is larger, *e.g.* 50 μm, then in the short-wavelength range it has a two times higher absolute value of D as compared with the value at 5 Å. It decreases more rapidly with wavelength than in the case of a 25 μm crystal, but at 5 Å it still has a higher absolute level. The same is true for a 70 μm crystal (see Fig. 3).

For a 100 μm crystal the value of the integrated diffraction intensity per average energy deposited in a crystal per molecule is much higher in a short-wavelength range, but although it drops down very rapidly with increasing wavelength it is still higher than in all previous cases. The ratio of D at 1 Å to the same parameter at 5 Å is equal to five, which means that the molecules of the 100 μm crystal will be exposed to a five-times larger dose of X-rays during diffraction data collection at 5 Å as compared with data collection at 1 Å. This is due to the increase in absorption of radiation, provided that the data set has the same quality. At the same time the absolute value of D for a 100 μm crystal is almost 12 times higher than for a 25 μm crystal. In other words, although the absorption of 5 Å X-rays in a 100 μm crystal is much stronger, the energy absorbed is redistributed among a significantly larger number of protein molecules and the radiation load for each molecule will be smaller. For crystals of 200 μm or more, D declines much more rapidly and at 5 Å it reaches smaller absolute values than in the case of a 100 μm crystal.

This means that there is an optimum crystal size for crystallographic data collection at soft wavelengths in terms of averaged normalized diffraction efficiency and this optimum size is not optimum from the point of view of maximum integrated diffraction intensity (*i.e.* close to the absorption length). For the sulfur K-edge, crystal dimensions should be between 90 and 130 μm, *i.e.* approximately four to five times larger than the size predicted by (5). For the 5.77 Å edge of phosphorus the value is correspondingly between 70 and 100 μm.

The absolute value of D is, however, rather low. A 100 μm crystal will scatter approximately 100 times less 5 Å photons than a 300 μm crystal irradiated by 1 Å photons at an equal average accumulated dose of radiation. As a consequence, in the process of data collection with 5 Å X-rays, a two orders of magnitude larger dose of X-rays will be deposited per molecule of a 100 μm crystal to provide a diffraction data set of equal quality.

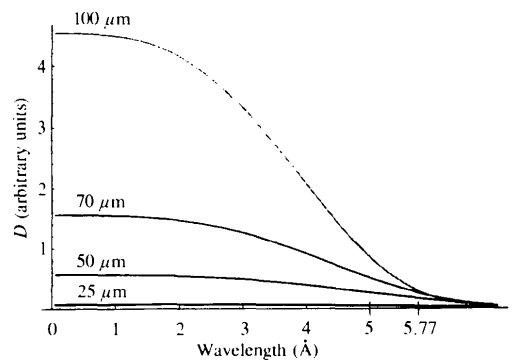

Figure 3
The averaged normalized diffraction intensity, D, as a function of the wavelength of X-rays for four different crystal sizes. Although D declines most rapidly for t = 100 μm at 5 Å it still has a larger absolute value than for 25, 50 and 70 μm crystal sizes.

3. Conclusions

There are two major problems in protein crystallographic structure solution: how to grow large, well diffracting crystals and how to obtain phases of structure amplitudes. MAD of naturally occurring sulfur and phosphorus *K*-edges can potentially give a solution for one-and-a-half of these two problems. One does not need to screen for derivatives – native crystals can be used as such. Also, one does not need to have large crystals – crystals between 25 and 100 μm would be sufficient. There are, however, some intrinsic problems connected with the drastic increase in absorption in the soft X-ray region.

The theoretical considerations summarized above show that good-quality diffraction data can be collected from very small (25–100 μm) protein crystals in the soft X-ray region. However, the dose of X-ray radiation deposited per molecule of these small crystals might be significantly higher than that deposited on larger crystals in the process of traditional protein crystallography data collection at shorter (1–1.5 Å) radiation wavelengths.

Another intrinsic problem of long-wavelength data collection is the correction for absorption of the diffraction data. This may be made either by empirical methods (Blundell & Johnson, 1976) or, once the orientation matrix of the crystal has been determined, by direct computation of the X-ray beam paths through the crystal, mother liquor and capillary, provided that adequate information about the crystal shape is available. In practice, local scaling against data taken at short wavelengths is probably the easiest way.

Even after flash-freezing at 100 K, protein crystals will suffer from radiation damage. Using experience from electron microscopy, Henderson (1990) has predicted a limit of $\sim 1.3 \times 10^{17}$ keV mm^{-3} absorbed energy in the specimen before significant radiation damage occurs at cryotemperatures. Gonzalez, Thompson & Nave (1992) showed that during Laue diffraction data collection radiation damage can be observed after an absorbed dose of $\sim 4 \times 10^{17}$ keV mm^{-3}. Gonzalez & Nave (1994) confirmed the estimation of Henderson (1990) and calculated that a 0.3 mm crystal with a 100 Å unit cell will produce a good-quality data set at 0.9 Å, absorbing $\sim 10^{16}$ keV mm^{-3} of radiation. Assuming a dose of 8×10^{17} keV mm^{-3} as the maximum the sample can absorb while still giving processable and useful data, 80 data sets can be collected from such a crystal (Gonzalez & Nave, 1994). It follows from our estimations that on irradiation with 5 Å X-rays, a 100 μm crystal will survive approximately 100 times less allowing for, for example, one data set to be collected.

One can circumvent the problem of radiation damage to some extent by choosing needle-like crystals and translating them along the spindle axis during data collection. This is probably the best crystal shape for soft MAD data collection.

The author would like to thank Ingo Koelln and Hans Bartunik for fruitful discussions, and FAPESP and CNPq for financial support.

References

Arndt, U. W. (1984). *J. Appl. Cryst.* **17**, 118–119.
Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography*. New York: Academic Press.
Gonzalez, A. & Nave, C. (1994). *Acta Cryst.* D**50**, 874–877.
Gonzalez, A., Thompson, A. W. & Nave, C. (1992). *Rev. Sci. Instrum.* **63**(1), 1177–1180.
Henderson, R. (1990). *Proc. R. Soc. London Ser. B*, **241**, 6–8.
Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
Rosenbaum, G. & Holmes, K. C. (1980). *Synchrotron Radiation Research*, edited by H. Winick & S. Doniach, ch. 16. New York: Plenum.
Stuhrmann, H. B. & Lehmann, M. S. (1994). *Resonant Anomalous X-ray Scattering: Theory and Application*, edited by G. Materlik, C. J. Sparks & K. Fischer, pp. 175–194. Amsterdam: Elsevier.
Stuhrmann, S., Huetsch, M., Trame, C. & Stuhrmann, H. B. (1995). *J. Synchrotron Rad.* **2**, 83–86.