International Tables for Crystallography (2006). Vol. F, Section 19.3.2, pp. 428–429.

19.3. Small-angle X-ray scattering

BY H. TSURUTA AND J. E. JOHNSON

19.3.1. Introduction

Mechanistic biology is frequently confronted with an experimental paradox. High-resolution structures are required to develop a chemical description of macromolecular interactions, but the processes themselves are dynamic and not amenable to high-resolution methods. Chemists and molecular biologists have been successful in generating homogeneous components and entire complex systems that are amenable to crystallization and high-resolution analysis. If the dynamics of the systems are small, as in the case of some enzymes and electron-transfer reactions, time-resolved studies can be performed in the context of the crystal lattice using Laue diffraction (Moffat, 1997; Genick *et al.*, 1997; Srajer *et al.*, 1996). If the dynamic features of the biological reactions are large, the motility must be studied in solution. During the last decade, small-angle X-ray scattering (SAXS) has emerged as an important method for studying large-scale dynamic processes, ranging from protein folding to virus particle polymorphism. The renaissance of this method has resulted from a variety of advances in molecular biology and X-ray instrumentation, and these have dramatically increased the information content of the derived results. Modern synchrotron X-ray sources and advanced detector systems have lead to higher-resolution data in both the spatial and time domains. Computational analyses of the data have improved dramatically, resulting, in some favourable cases, in *de novo* three-dimensional density functions from SAXS data that are comparable to a 20 Å-resolution electron-microscopy reconstruction. Model-based approaches to interpreting SAXS data have expanded their usefulness to the validation of mechanistic hypotheses involving movement or association of components independently determined at high resolution. Finally, SAXS studies have benefited from the same molecular-biology advances that provide large quantities of homogeneous material for crystallographic and NMR studies. The purpose of this chapter is to address practical aspects of SAXS as they relate to and complement macromolecular crystallography.

19.3.2. Small-angle single-crystal X-ray diffraction studies

We begin by briefly describing an obvious connection between small-angle solution X-ray scattering and single-crystal X-ray scattering, *i.e.* small-angle single-crystal diffraction (Tsuruta *et al.*, 1998; Miller *et al.*, 1999). The majority of macromolecular single-crystal studies ignore data at resolutions lower than 20 Å. Measuring these data accurately is generally difficult when collecting high-resolution data, because they are orders of magnitude stronger in intensity and frequently saturate the detector pixels or are intentionally blocked from reaching the detector by a large beam stop. It is technically challenging to record low-resolution data accurately, because they fall very close to the primary beam, where a variety of parasitic scattering effects occur. These data, however, contain information about the structure that may not be determined from only the high-resolution data. Low-resolution (∞–15 Å) data are sensitive to structures that are organized over large distances and that contribute primarily to the low-frequency terms in a Fourier series. In contrast, portions of the structure that contribute to high-frequency terms must be in precisely the same position in all the unit cells within the crystal. Two examples illustrate the importance of low-resolution terms.

Assume that an exposed loop on the surface of a protein is highly mobile. The scattering-factor curves for atoms in this loop will be highly attenuated by the large Debye–Waller factor. If the temperature factors are high enough, these atoms will contribute nothing to the high-resolution data. In contrast, these atoms will contribute strongly to the low-resolution data. If the low-resolution terms are included in the Fourier series, the atom positions will not appear as individual atoms; instead, an envelope of density describing the statistically distributed positions of these atoms will be visible, and this can aid significantly in the modelling of the loop.

A second example is the nucleic acid within a spherical virus. The nucleotides generally do not display the symmetry of the icosahedral capsid, although in some cases segments of RNA or DNA clearly interact with the protein and are therefore visible at high resolution. The RNA density is not usually visible if only higher-resolution data are used, and the virus particle appears empty. The reason for this is that the RNA is best described as a uniform sphere within the particle, and the Fourier transform of such a sphere falls off rapidly if the sphere is relatively large. A typical RNA virus will have an internal RNA core of about 100 Å in radius. The scattering contribution of this region is virtually zero beyond 20 Å resolution. If the low-resolution data are measured accurately, the region occupied by RNA and its general level of interaction with the coat protein can be clearly seen (Fig. 19.3.2.1).

The relation between single-crystal and solution X-ray scattering is clearly illustrated in Fig. 19.3.2.2, where the scattering of a single crystal and the solution scattering of a 320 Å-diameter RNA virus are compared. The figure shows the scattering expected for a uniform sphere 160 Å in radius, the observed solution X-ray scattering from this virus and the single-crystal diffraction from

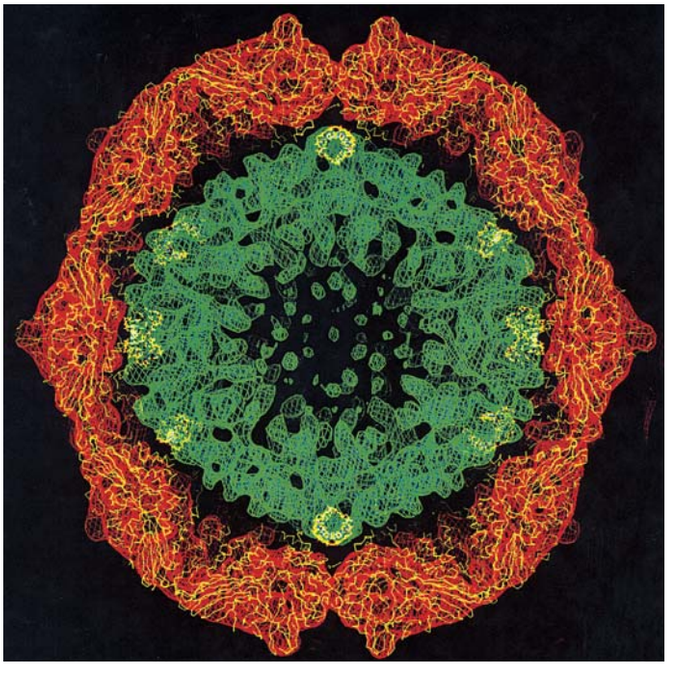

Fig. 19.3.2.1. A cross section of the electron-density map of sFHV at 14 Å resolution. The density shown in red corresponds to the protein capsid, while density in green corresponds to the bulk RNA or dynamic protein segments. The Cα backbone of the atomic model of the protein capsid seen in the high-resolution structures of sFHV is shown as yellow traces, and the model of the ordered pieces of duplex RNA is shown as a stick model. The high-resolution protein model fits the electron-density map very well. The ordered RNA model is buried in the RNA density in green.

Copyright © 2006 International Union of Crystallography

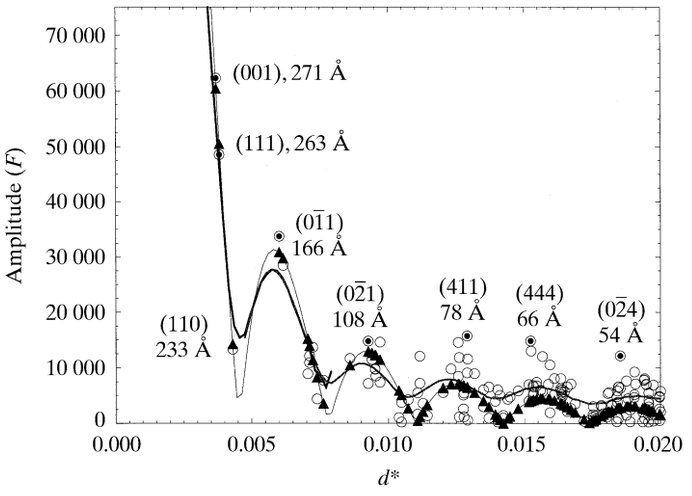

Fig. 19.3.2.2. Comparison of the absolute values of single-crystal reflection amplitudes (circles), the solution scattering intensity (thick curve) and the calculated scattering intensity from a uniform sphere (thin curve).

the virus. All three diffraction patterns are in close agreement at resolutions below 50 Å, with the only difference being the continuous sampling of the transform in the computed and solution scattering curve and the discrete sampled transform of the crystalline virus. The RNA density in Flock house virus (FHV) shown in Fig. 19.3.2.1 illustrates the information content of the low-resolution data. DNA density functions in polyoma virus were mapped by employing similar methods (Griffith *et al.*, 1992). Low-resolution single-crystal data are also critical for applying the methods of *de novo* phase determination. Geometric solids, such as a sphere, or a low-resolution electron cryo-microscopy (cryoEM) reconstruction may serve as initial phasing models for such a strategy, and the procedure of refining and extending phases is much more robust if the low-resolution data are measured accurately (Tsuruta *et al.*, 1998).

19.3.3. Solution X-ray scattering studies

Protein crystals have high solvent content: typically 50% and as much as 80% in some cases (*e.g.* Johnson & Hollingshead, 1981; Wikoff *et al.*, 1998). Some solvent is well ordered and visible in the crystal structure. Some, referred to as bulk solvent, is dynamic, not specifically associated with parts of the repetitive lattice and can be readily exchanged within the crystal using a flow cell. The high solvent content of macromolecular crystals makes it probable that most of the features of a protein molecule that are important for its function in solution are preserved in the crystallographic structure. Cases are known, however, in which lattice forces or crystallization conditions have masked functional features of the macromolecule or induced functionally irrelevant oligomerization. A straightforward application of solution scattering is the verification of the relevance of the oligomerization state of a macromolecule observed in the crystal. The methods are complementary, because the crystal structure almost invariably leads to the correct secondary and tertiary structures for the monomer, and the SAXS experiment allows the determination of the quaternary structure of the macromolecule in solution.

The solution scattering pattern of the macromolecule can be calculated directly from a set of atomic coordinates. Scattering computed from models of different oligomerization states or from models of the monomer can be readily compared with the observed scattering pattern. In many cases, not only can the oligomer state be determined, but also the point-group symmetry of the oligomer; hence more information is available than through a solution molecular-mass determination by other physical methods. Using the same approach, effects of various ligands on the state of a protein can be studied in solution. Any large-scale structure alteration of a protein, including changes in quaternary structure or folding and unfolding, can be readily studied in solution, either as a static experiment or as a time-resolved investigation.

Recently, the search for crystallization conditions of macromolecules has been refined by the use of dynamic light scattering. Solution X-ray scattering can be even more effective because it is sensitive to molecular aggregates that are smaller than those detected by light scattering, and this broadens the radius of convergence for conditions that will produce good-quality crystals. For example, the interparticle interaction potential of proteins with a mass of less than 100 kDa can be readily measured in the SAXS experiment, and this has been shown to be an important factor in their crystallization (Bonneté *et al.*, 1997). In a particularly successful experiment, a set of small-angle solution X-ray scattering results was used to improve crystallization conditions, and this lead to single crystals that diffracted beyond 1.9 Å (Mourey *et al.*, 1997). A similar approach was used to distinguish between virus coat protein polymerization and crystallization (Petitpas *et al.*, 1998). Studies in which a number of solution conditions must be tested can be done most effectively using synchrotron small-angle scattering facilities, though it is possible to conduct such studies with a conventional X-ray source.

19.3.3.1. *Information content of solution scattering*

Solution scattering deals with samples that are randomly oriented in a solvent. Three-dimensional information is compressed to one dimension, and a one-dimensional self-correlation function (a rotationally averaged Patterson function) of the protein is obtained when the scattering function is properly Fourier transformed. Excellent books describing the information content of small-angle scattering include those by Guinier & Fournet (1955) and Glatter & Kratky (1982), and for modern perspectives see Feigin & Svergun (1987). Concise reviews with a strong focus on structural biology have also appeared, including those by Trewhella (1998) and Koch (1991). These reviews describe the theoretical background, experimental considerations and ways of extracting structural information. None of these take our perspective, which is addressed to the sceptical crystallographer.

Assume a monodisperse protein solution. X-ray scattering from the solution originates from the excess electron density of protein particles over that of the solvent, $\Delta\rho(\mathbf{r}) = \rho(\mathbf{r})_{\rm protein} - \rho_{\rm solvent}$ (Fig. 19.3.3.1). Note that $\rho(\mathbf{r})_{\rm protein}$ is typically only about 30% higher than $\rho_{\rm solvent}$, and that $\Delta\rho(\mathbf{r})$, the electron-density 'contrast', depends on the solvent composition. The scattering intensity observed for such a system is given by

$$\begin{aligned} I(Q) &= F(Q)F^*(Q) \\ &= \int_{\mathbf{r}_1\mathbf{r}_2} \Delta\rho(\mathbf{r}_1)\Delta\rho(\mathbf{r}_2)\exp[-iQ(\mathbf{r}_1-\mathbf{r}_2)]\,{\rm d}V_1\,{\rm d}V_2, \quad (19.3.3.1)\end{aligned}$$

where $Q = 4\pi\sin\theta/\lambda$ (2θ is the scattering angle and λ is the wavelength of the X-radiation used). Owing to the rotationally averaged nature of solution scattering, the exponential term may be substituted with the Debye function, $(\sin Q\mathbf{r})/Q\mathbf{r}$, giving

$$I(Q) = \int_{\mathbf{r}} 4\pi\mathbf{r}^2\Delta\rho(\mathbf{r})^2(\sin Q\mathbf{r})/Q\mathbf{r}\,{\rm d}\mathbf{r}. \quad (19.3.3.2)$$

Guinier & Fournet (1955) showed that equation (19.3.3.2) can be approximated by a Gaussian function at relatively small scattering angles resulting in