High-throughput Characterization of Expressed Proteins by Wide-angle Solution X-ray Scattering

R.F. Fischetti, D.J. Rodi, L. Makowski

Argonne National Laboratory, Argonne, IL, U.S.A.

Introduction

Small-angle x-ray scattering (SAXS) from biological macromolecules in solution is a technique that yields information on the overall shape of the molecule, and it can monitor conformational changes, including molecular associations in solution. Although wide-angle x-ray scattering (WAXS) has the potential to provide higher-resolution structural information, practical use of the data has been limited by the difficulty of measuring the weak protein scattering that is superimposed on a much stronger background of scatter from the solvent and sample container. Recently, it has been demonstrated not only that WAXS patterns obtained on insertion device beamlines at third-generation synchrotron sources are sensitive to protein conformation states but also that they can be quantitatively compared to data calculated from detailed structural models [1, 2]. These data provide a rich source of structural information that has not yet been exploited. A combination of SAXS and WAXS analysis has the potential to generate information on the size, shape, and structural class (i.e., fold) of the large fraction of proteins that may not form crystals. It is applicable to all classes of proteins, including membrane proteins, large protein complexes, and proteins with disordered regions. An unambiguous determination of fold cannot be obtained directly from solution scattering data [3], but a comparison of solution scattering from proteins of unknown structure with data from proteins of known structure has the potential to reduce the number of possible folds for a protein to a short list, if not a unique designation.

A potential obstacle to obtaining structural information by WAXS is radiation-induced damage to the protein structure from the intense x-ray beam on undulator source beamlines. This has been shown to be a significant problem in protein x-ray crystallography. Radiation exposure can induce chemical and structural damage to protein crystals, such as cleavage of disulfide bonds, decarboxylation of acidic residues, and an increase in both atomic B-factors and unit cell volume. However, previous studies could not determine if the increased rate of damage was in proportion to the total dose or the dose rate. The use of crystals at liquid nitrogen temperatures has greatly alleviated the problem of radiation damage in crystallography, and a similar approach may be adaptable for WAXS. However, it is anticipated that this approach would lead to serious problems in background scaling and subtraction because of the increased scattering from the ice crystals, thus limiting the use of WAXS for dynamic studies. We approach the problem by using a sample flow cell that minimizes the radiation dose to individual proteins, and we show that this approach eliminates observable changes in scattering due to radiation damage.

Here we demonstrate the capability of collecting x-ray scattering data from proteins in solution to spacings of 2.2 Å (or q = 2.8 Å-1), show that these data are consistent with expectations based on known crystallographic coordinates, show that these data are sensitive indicators of conformational changes of a protein in solution, demonstrate the use of WAXS as a monitor of the effect of chemical denaturing conditions on proteins, and show that the effect of radiation damage can be observed and evaluated by using this method.

Methods and Materials

Equine heart cytochrome C and equine skeletal muscle myoglobin were obtained from Sigma, and bovine erythrocyte hemoglobin was obtained from CalBiochem. Protein samples were centrifuged through a Nanosep® centrifugal device for 10 minutes prior to beam exposure to remove high-molecular-weight protein aggregates from solution. Protein solutions were denatured by adding crystalline guanidine hydrochloride.

WAXS data were collected at the Bio-CAT undulator beamline (18-ID) of the APS. The experimental layout was arranged as follows. A nitrogen-gas-filled ion chamber was used to record the x-ray beam intensity. In-vacuum guard slits were set to remove low-angle scatter from the x-ray optics and upstream windows. Another set of guard slits was positioned about 4 mm in front of the sample to remove scatter from the vacuum exit window (mica) and small air gap. The sample cell consisted of a thin-walled quartz capillary (1 mm i.d.) attached to a programmable pump that could be adjusted to deliver either continuous or discontinuous flow through the capillary. A helium-gas-filled path with a mica entrance window and a 0.5-mil Mylar® exit window was positioned between the capillary and the detector. Scattering from the Mylar exit window was blocked by positioning the beam stop to touch the Mylar window. The x-ray scattering pattern was detected with a charge- coupled device (CCD) detector

specially designed for imaging measurements requiring high sensitivity (a dynamic range of 10,000 to 1) and high spatial resolution (48 × 48 μm, 1798 × 1028 pixels). The specimen-to-detector distance was 148 mm. The x-ray beam was focused to 40 × 180 μm (vertical × horizontal, full width at half-maximum [FWHM]) at the detector. Because of the long depth of focus, the beam was only slightly larger at the specimen. The beamline is capable of delivering approximately 2×10^{13} photons/s per 100 mA of beam current. However, previous experience on the Bio-CAT beamline demonstrated that proteins under a variety of physical conditions are damaged after exposure times of a few tenths of a second to a few seconds at these intensity levels. In these experiments, thin aluminum foils were used as x-ray beam attenuators to control the incident beam flux. The 2-D scattering patterns were integrated radially to 1-D scattering intensity profiles by using the program Fit2D, version 9.129 [4-6]. The origin of the diffraction pattern was determined by calculating the center of powder diffraction rings from silver-behenate powder. The incident x-ray beam intensity was recorded and used to normalize the individual exposures.

In order to assess the effect of radiation dose on proteins, three scattering data collection modes were employed. In the first mode, data were collected as a series of 0.7-s exposures from protein samples sitting stationary within the sample cell in the beam path (ST or stationary mode). In the second mode, data were collected as a series of 8.3-s exposures from protein samples that were oscillated within the sample capillary at a rate of 10.3 oscillations/min during beam exposure (FR or fry mode). In the third mode, data were collected as a series of 8.3-s exposures from protein samples that were kept flowing unidirectionally through the beam during exposure, so that no one part of the solution was exposed more than once in the direct beam (RD or reduced exposure mode). The flow rate was adjusted so that no single protein spent more than 100 ms in the direct beam (2.4 μL/s). Data were also collected from the empty sample capillary and from the capillary containing buffer. Exposures from the sample and buffer were alternated to minimize the possible effects of drift in any experimental parameter.

Scattering from samples should be separable into four individual components as a result of scattering from the (1) protein, (2) bulk solvent, (3) solvent of hydration, and (4) capillary. Scattering from the solvent of hydration (boundary layer), although potentially important at small angles of scattering, is generally at least two orders of magnitude weaker than any other contribution in the range of angles studied here [1]. Scattering from protein was estimated according to:

$$I_{prot} = I_{obs} - I_{cap} - (1 - vol~\%)I_{solvent} , \qquad (1)$$

where I_{obs} is the measured scattering from the protein sample, I_{cap} is the measured scattering from the empty capillary, vol % is the estimated proportion of the solution taken up by the protein (and thereby excluding solvent), and $I_{solvent}$ was estimated by:

$$I_{solvent} = I_{bkgd} - I_{cap} , \qquad (2)$$

where I_{bkgd} is the measured scattering from the capillary containing buffer. The cytochrome C sample was approximately 2.5 vol % (30 mg mL^{-1}), and, on the basis of that estimate, the intensity of diffraction from the protein was calculated. Scaling was based on the total dose estimated as described above. Small ambiguities (less than one-tenth of 1%) in the relative scaling of the scattering from solvent, capillary, and protein solution lead to some uncertainty in the scaling of higher-angle features relative to the features in the 0.1-A^{-1} range. These were resolved on the basis of self-consistency of features in patterns from homologous proteins.

WAXS patterns were calculated from crystallographic coordinates by using the CRYSOL program [7], using 50 spherical harmonics and default parameters for calculating the solvation shell and particle envelope. A maximum allowed number of Fibonacci grid points of 18 was used for all calculations. No attempt was made to directly fit the CRYSOL-calculated pattern to the experimental data.

Results

Background-subtracted scattering profiles for hemoglobin in all three beam exposure modes (i.e., RD-, FR-, and ST-mode-generated profiles) were calculated and contrasted. The patterns for hemoglobin demonstrated significant deviations between the RD and FR modes (virtually identical) and the ST mode. Figure 1 contrasts the average from 10 ST shots (10×0.5 s = 5 s of total exposure, shown in black) with the average of four RD shots (8.1 s $\times$ 4 = 32.4 s of total exposure, shown in red). The protein sample in the ST mode shows clear signs of degradation across the entire pattern, indicating a breakdown of features within the size range for both secondary and tertiary structure. (Note the partial loss of peaks at spacings of roughly 0.08 and 0.095 Å, and note the filling in of minima.) The different noise levels apparent in the two plots are the result of different total exposure times, not radiation damage.

Background diffraction was subtracted from each protein solution scattering pattern by using Eq. (1), then compared with the diffraction pattern calculated from atomic coordinates (PDB files: 1CRC [using only chain A] for cytochrome C, 1WLA for myoglobin, and 1GZX for hemoglobin) by using the most recent version of CRYSOL [1]. The correspondence of calculated and
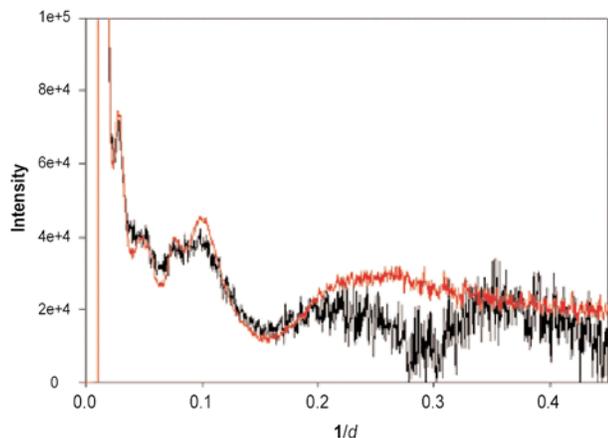
FIG. 1. Comparison of the average solution scattering profiles obtained from hemoglobin by using the RD data-collection mode (red) and ST data-collection mode (black). Note that the ST mode shows clear signs of degradation. The abscissa is in arbitrary units for relative intensity, and the ordinate is in units of $Å^{-1}$. © 2003 by The International Union of Crystallography, http://journals.iucr.org/.
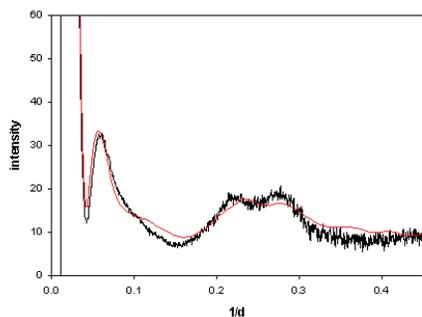

FIG. 2. WAXS curves for cytochrome C. The black curve is the theoretical solution scattering curve calculated by using CRYSOL with 50 spherical harmonics, as described in Results. The red curve is the measured scattering solution curve obtained with data-collection mode RD, as described in Methods. The abscissa is in arbitrary units for relative intensity, and the ordinate is in units of $Å^{-1}$. © 2003 by The International Union of Crystallography, http://journals.iucr.org/.

observed diffraction for cytochrome C, shown in Fig. 2, indicated that the measured scattering from the protein, although low compared with background scatter, was consistent with expectations based on the atomic coordinates of cytochrome C (R factor = 0.125). There is good agreement in both peak position and relative heights for both plots. A somewhat lower degree of correspondence was obtained for both myoglobin and hemoglobin, with respective R factors of 0.21 and 0.23 (data not shown).
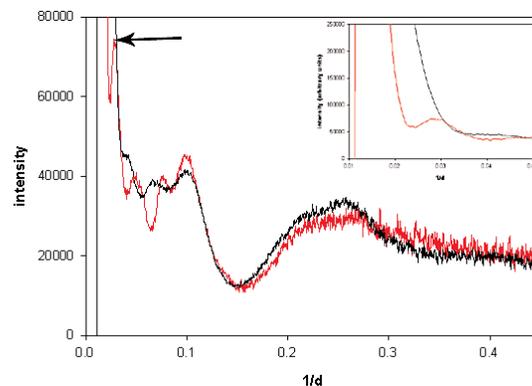

FIG. 3. Comparison of measured solution scattering curves for the related proteins myoglobin and hemoglobin. The black curve is myoglobin, and the red curve is hemoglobin. Note that the quaternary structure of hemoglobin as compared with the monomeric myoglobin leads to additional high-frequency fluctuations in the 1/d range of 0.021 ± 0.09 $Å^{-1}$. The abscissa is in arbitrary units for relative intensity, and the ordinate is in units of $Å^{-1}$. Details on the peak at 0.03 $Å^{-1}$ in the scattering from hemoglobin, contrasted with the myoglobin scattering curve devoid of that peak, are shown in the inset. © 2003 by The International Union of Crystallography, http://journals.iucr.org/.

A comparison of diffraction from the structurally related proteins myoglobin and hemoglobin (Fig. 3) provides additional insight into the kind of information that can be obtained from solution scattering data. Hemoglobin is a tetramer of four polypeptides, each of which exhibits a fold very similar to that of myoglobin. This quaternary structure of hemoglobin results in a higher-frequency modulation of data out to a spacing of about 0.15 A^{-1}. Note that there is an additional peak in the hemoglobin scattering at about 0.031 $Å^{-1}$ (black arrows in Fig. 3). At spacings higher than 0.15 $Å^{-1}$, the data from the two proteins are very similar, reflecting their very similar secondary structures. Although the secondary structure of a protein also contributes to the scattering at intermediate spacings (0.05 to 0.15 A^{-1}), this is modulated by the effect of both the tertiary and quaternary structure.

To explore the use of WAXS for the study of protein unfolding, hemoglobin samples were intentionally denatured by the addition of crystalline guanidine hydrochloride to 2 molar and then 4 molar final concentrations, then exposed to the beam in RD mode (Fig. 4). Upon the addition of guanidine-HCl to 2 M, there is a significant loss of features at both high spacings (over 0.3 $Å^{-1}$) and lower spacings (red curve in Fig. 4). Note that there is a complete loss of the tetrameric form-associated peak at 0.029 to 0.030 $Å^{-1}$ spacing. Further addition of guanidine-HCl to 4 M
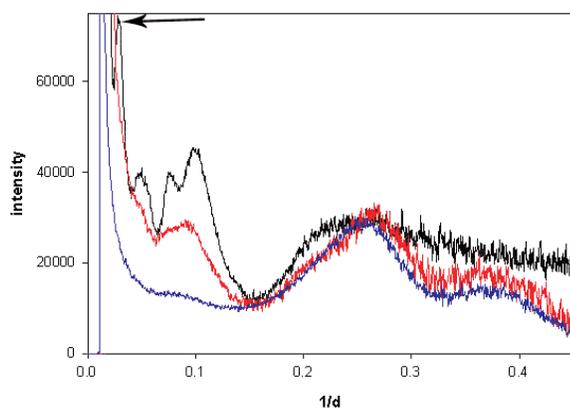
FIG. 4. *Effect of increasing concentration of guanidine hydrochloride on the solution scattering profile from hemoglobin. The black curve is 0 M guanidine hydrochloride, the red curve is 2 M guanidine hydrochloride, and the blue curve is 4 M guanidine hydrochloride. Note that the addition of guanidine hydrochloride to 2 M completely obliterates the peak at 1/d of around 0.029 Å^{-1}, a spacing that is consistent with the tetrameric form of hemoglobin. The abscissa is in arbitrary units for relative intensity, and the ordinate is in units of Å^{-1}. © 2003 by The International Union of Crystallography,* http://journals.iucr.org/.

shows a catastrophic loss of structure over the entire pattern (blue curve in Fig. 4). Both treatment of hemoglobin with 2 M guanidine-HCl and ST mode data collection result in the loss of the tertiary structure-related doublet at spacings of about 0.1 Å^{-1} and the peak at 0.031Å^{-1} associated with the presence of the tetrameric structure. These data infer that in spite of the different mechanisms of the actions, both chemical denaturants and radiation result in the dissociation of the hemoglobin tetramer. Guinier plots of the SAXS data from hemoglobin in 0, 2, and 4 M guanidine-HCl indicate radii of gyration consistent with tetrameric, dimeric, and monomeric structures, respectively. At 0 M guanidine-HCl, the observed radius of gyration was 25.7 Å, which is within 3% of the 26.4 Å calculated for a tetramer from crystallographic coordinates. For 2 M guanidine-HCl, the observed radius of gyration was 23.0 Å, or within 13% of that calculated for an αβ-dimer (20.3 Å). For 4 M guanidine-HCl, the observed radius of gyration was approximately 50 Å, corresponding to unfolded monomers.

Discussion

It is becoming increasingly apparent that 3-D structural information will be critical in making a comprehensive functional analysis of many, if not most, proteins. For the majority of proteins that cannot be readily crystallized, new methods of structural characterization are needed. For those proteins that can

be crystallized, methods for characterization of functional processes that are accompanied by large structural changes will be required. X-ray scattering from proteins in solution provides direct structural information about the secondary, tertiary, and quaternary organization of a protein. With optimized hardware and software, these data can be collected in a high-throughput fashion to provide information about 3-D structures and structural changes that occur in solution.

A major concern for x-ray scattering of protein solutions is radiation-induced degradation by the high intensities available at the third-generation synchrotron radiation source utilized in these studies. Earlier work performed at the JASRI at Spring8 [2] involving 60-s exposure times to collect data at spacings of ~ 0.003 to ~0.4 Å^{-1} did not specifically address the issue of protein degradation at third-generation sources. This may be a function of its use of a bending magnet beamline and a large flat cell, both of which would be expected to lower radiation-induced damage to the protein molecules. The data presented here indicate that while x-ray damage is observable when proteins are intentionally overexposed to x-rays by using an insertion device beamline, experimental protocols can be designed to minimize that damage while maximizing the signal intensity obtained.

These solution scattering experiments on cytochrome C, myoglobin, and hemoglobin indicate that accurate solution scattering data can be collected to spacings approaching 2.2 Å, adequately predicted from the atomic coordinates of crystallized proteins; used for comparative structural analyses; and used to monitor structural changes that occur in the sample. Unlike circular dichroism (CD) spectroscopy, which provides extensive short-range information on the percentage content of α-helices, β-sheets, etc., the data shown here demonstrate a sensitivity to tertiary and quaternary structural influences that are not apparent in CD spectra. This suggests that WAXS may ultimately prove to be a valuable tool for the rapid confirmation or rejection of structural hypotheses derived from amino acid sequence data via bioinformatic analysis. The impact of WAXS data would grow substantially if an extensive database of solution scattering from proteins of known structure was constructed. This database could provide the basis for making predictions about the domain and fold structure of proteins of unknown structure and make detailed structural analyses of dynamic functional processes possible.

Acknowledgments

Health under RR-08630. Details on this report can be found in the *Journal of Synchrotron Radiation* [8].

References

[1] D. Svergun, C. Barferato, and M.H.J. Koch, "CRYSOL — A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates," J. Appl. Cryst. **28**, 768-773 (1995).

[2] M. Hirai, H. Iwase, T. Hayakawa, K. Miura, and K. Inoue, "Structural hierarchy of several proteins observed by wide-angle solution scattering," J. Synchrotron Rad. **9**, 202-205 (2002).

[3] D.I. Svergun, M.V. Petoukhov, and M.H. Koch, "Determination of domain structure of proteins from x-ray solution scattering," Biophys. J. **80**(6), 2946-2953 (2001).

[4] A.P. Hammersley, FIT2D: *An Introduction and Overview*, ESRF Internal Report ESRF97HA02T (1997).

[5] A.P. Hammersley, *FIT2D V9.129 Reference Manual V3.1*, ESRF Internal Report ESRF98HA01T (1998).

[6] A.P. Hammersley, S.O. Svensson, M. Hanfland, A.N. Fitch, and D. Häusermann, "Two-dimensional detector software: From real detector to idealised image or two-theta scan," High Pressure Res. **14**, 235-248 (1996).

[7] International Union of Crystallography, Crysol Manual, vol. 2.3; http://www.embl-hamburg.de/External Info/Research/Sax/manual_crysol.html.

[8] R.F. Fischetti, D.J. Rodi, A. Mirza, T.C. Irving, E. Kondrashkina, and L. Makowski, "High resolution wide angle x-ray scattering of protein integrity," J. Synchrotron Radiat. **10**, 398-404 (2003).