Derivation of the small-angle scattering profile of a target biomacromolecule from a profile deteriorated by aggregates. AUC–SAS

Ken Morishima, Rintaro Inoue and Masaaki Sugiyama*

Institute for Integrated Radiation and Nuclear Science, Kyoto University, 2-1010 Asashironishi, Kumatori, Sennan-gun, Osaka 590-0494, Japan. *Correspondence e-mail: sugiyama.masaaki.5n@kyoto-u.ac.jp

Aggregates cause a fatal problem in the structural analysis of a biomacromolecule in solution using small-angle X-ray or neutron scattering (SAS): they deteriorate the scattering profile of the target molecule and lead to an incorrect structure. Recently, an integrated method of analytical ultracentrifugation (AUC) and SAS, abbreviated AUC–SAS, was developed as a new approach to overcome this problem. However, the original version of AUC–SAS does not offer a correct scattering profile of the target molecule when the weight fraction of aggregates is higher than *ca* 10%. In this study, the obstacle point in the original AUC–SAS approach is identified. The improved AUC–SAS method is then applicable to a solution with a relatively larger weight fraction of aggregates (≤20%).

1. Introduction

Small-angle X-ray and neutron scattering (SAXS and SANS), collectively abbreviated as SAS, are increasingly being used to reveal structures of biomacromolecules in solution (Svergun & Koch, 2003; Bernadó *et al.*, 2018; Mahieu & Gabel, 2018). Modern computational analysis methods for SAS offer a detailed three-dimensional structural model (Grant, 2018; Bengtsen *et al.*, 2020; Gräwert & Svergun, 2020; Matsumoto *et al.*, 2020; Okuda *et al.*, 2021; Shimizu *et al.*, 2022; Yunoki *et al.*, 2022). To build a reliable structural model using these methods, it is crucial to obtain an experimental scattering profile that purely corresponds to the target molecule. However, even with a small content of aggregates (<10%), the scattering profile deteriorates from that of the target molecule and can result in an incorrect structural model. Moreover, there is another serious problem related to aggregates. Typically, an abnormal upturn of the scattering profile in the lowest scattering-angle region is recognized as experimental evidence of aggregate contamination. However, the scattering profile cannot show such clear evidence when the weight fraction of the aggregates is low. For example, the Guinier approximation holds for a sample with a small weight fraction of aggregates, and the scattering profile is expressed as a straight line in the Guinier plot, which gives the gyration radius of the sample biomacromolecule. However, when the gyration radius is larger than the expected radius, it is difficult to determine whether the solution includes aggregates or whether the target molecule itself is deformed from the expected structure. Accordingly, to solve the 'aggregation problems' of the identification and removal of aggregates, SAS coupled with other
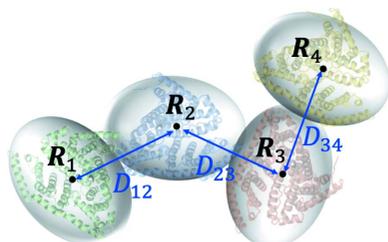
methods, such as size-exclusion chromatography (SEC–SAXS), has been explored (David & Pérez, 2009; Ryan *et al.*, 2018; Inoue *et al.*, 2019).

Recently, another integrated approach using analytical ultracentrifugation (AUC) and SAS, abbreviated AUC–SAS (Morishima *et al.*, 2020), has been developed to overcome aggregation problems. AUC–SAS derives a scattering profile of the target molecule in the solution including aggregates by utilizing the molecular distribution obtained with AUC. AUC–SAS reportedly offers precise scattering profiles of several biomacromolecules in solution (Hirano *et al.*, 2021; Okuda *et al.*, 2021). Because AUC–SAS does not require a large amount of sample or a very high intensity instrument, as needed by synchrotron-light SAXS, it has the potential to be applied to laboratory-based SAXS. AUC–SAS is also applicable to SANS, which faces the same aggregation problem.

Improvement of AUC–SAS will expand the scope of wider applications. For example, the first version of AUC–SAS ('first AUC–SAS') was constrained by the weight fraction of the aggregates (less than ~10%). In the present study, we have improved AUC–SAS, making it applicable to samples with relatively large weight fractions of aggregates (>10%). Furthermore, we provide software for the improved AUC–SAS, which is available to any SAS experimenter.

2. Experimental

2.1. Samples

Bovine serum albumin (BSA), apoferritin (AF), catalase (Cat), lysozyme (Lyz), ovalbumin (OVA) and ribonuclease A (RNaseA) were purchased from Merck, Sigma–Aldrich (Darmstadt, Germany). Human βB2-crystallin (βB2-cry) clone (consistent with NCBI sequence NM 000496) in a pET3a plasmid was obtained from Genscript (Piscataway, New Jersey, USA). The recombinant βB2-cry plasmid was then used to transform competent BL21(DE3)pLysS cells (Thermo Fisher Scientific, Waltham, Massachusetts, USA). Purification of βB2-cry was performed by following previous reports (Lampi *et al.*, 2006).

BSA, AF, Lyz, OVA and RNaseA were dissolved in 100 mM Tris–HCl buffer (pH 7.5) containing 100 mM NaCl. Cat and βB2-cry were dissolved in 50 mM Tris–HCl buffer (pH 8.0) containing 150 mM NaCl. The protein solutions were purified by SEC with a Superdex 200 Increase 10/300 GL column (for BSA, Cat, βB2-cry and OVA), Superose 6 Increase 10/300 GL column (for AF) and Superdex 75 Increase 10/300 GL column (for Lyz and RNaseA). The protein solutions were prepared by mixing the main component and its aggregate fractions while keeping the weight fraction of aggregates (r_a) $\leq$ 0.2. Sample codes are expressed as [protein + number] (*e.g.* BSA6), where the number corresponds to r_a. The mass concentrations for the AUC and SAXS measurements were 2.0 mg ml^{-1} for BSA6, BSA13, BSA20, AF5, AF15 and AF21; 2.1 mg ml^{-1} for Cat3 and Cat8; 1.1 mg ml^{-1} for Lyz6; 2.3 mg ml^{-1} for βB2-cry11; 2.2 mg ml^{-1} for OVA4; and 2.0 mg ml^{-1} for RNaseA8. BSA3 was subjected to AUC and SANS measurements after dialysis in D_2O buffer.

2.2. AUC measurements

Sedimentation velocity AUC measurements were performed using ProteomeLab XL-I (Beckman Coulter, USA). The samples were loaded into cells equipped with 1.5 mm path length titanium center pieces (Nanolytics, Germany). All measurements were performed using Rayleigh interference optics at 298 K. The rotor speed was set at 45 000 r min^{-1} for BSA, AF, Cat, βB2-cry and OVA; and 60 000 r min^{-1} for Lyz and RNaseA. The time evolution of the sedimentation data was analyzed using the multi-component Lamm equation (Lebowitz *et al.*, 2002). The weight-concentration distribution $c(s_{20,w})$ as a function of the sedimentation coefficient and frictional ratio f/f_0 was computed using the *SEDFIT* software (version 15.01c) (Schuck, 2000). The sedimentation coefficient was normalized to be the value at 293 K in pure water, $s_{20,w}$. The weight fraction of the j-mer, r_j, was obtained from the corresponding peak area of $c(s_{20,w})$. The molecular weight, M_j, of the j-mer was calculated using the corresponding peak positions $s_{20,w,j}$ and f/f_0 (Brown & Schuck, 2006) as

$$M_j = \left[\frac{6\pi\eta N_A}{(1-\rho\bar{v})} \left(\frac{3\bar{v}}{4\pi N_A} \right)^{1/3} \left(\frac{f}{f_0} \right) \right]^{3/2} s_{20,w,j}^{3/2}, \qquad (1)$$

where η, ρ, N_A and $\bar{v}$ are the viscosity of water at 293 K, the density of water at 293 K, Avogadro's number and the partial specific volume of the protein, respectively.

2.3. SAXS measurements

SAXS measurements were performed using a laboratory-based instrument (NANOPIX, Rigaku, Japan) equipped with a high-brilliance point-focused generator of a Cu $K\alpha$ source (MicroMAX-007 HFMR, Rigaku, Japan) (wavelength = 1.54 Å). Scattered X-rays were measured using a HyPix-6000 hybrid photon counting detector (Rigaku, Japan) composed of 765 × 813 pixels with a spatial resolution of 100 μm. For all samples, the sample-to-detector distance (SDD) was set to 1330 mm, with which the covered q range was $0.01 \leq q \leq 0.20$ Å^{-1} (where q is the magnitude of the scattering vector). Two-dimensional scattering patterns were converted to one-dimensional scattering profiles using the *SAngler* software (Shimizu *et al.*, 2016). After correction by the transmittance and subtraction of buffer scattering, the absolute scattering intensity was obtained using the standard scattering intensity of water (1.632 × 10^{-2} cm^{-1}) (Orthaber *et al.*, 2000). All measurements were performed at 298 K.

2.4. SEC–SAXS measurements

SEC–SAXS measurements were conducted with a laboratory-based SEC–SAXS system (La-SSS) (Inoue *et al.*, 2019), which is made up of a NANOPIX combined with a Prominence high-performance liquid chromatography system (SHIMADZU, Japan). A Superdex 200 Increase 10/300 GL for BSA, Cat, βB2-cry and OVA, a Superose 6 Increase 10/300 GL for AF, and a Superdex 75 Increase 10/300 GL for Lyz and RNaseA were utilized as the SEC column. All measurements were performed at a flow rate of 0.02 ml min^{-1} at 298 K.

2.5. SANS measurements

SANS measurements were performed using the SANS-U instrument located at JRR-3 (Japan Atomic Energy Agency, JAEA). A neutron beam at a wavelength of 6.0 Å with 10% resolution was irradiated on the samples. Scattered neutrons were counted using a two-dimensional detector (Ordela, USA). The SDDs were set to 4000 and 1030 mm, which covered a q range of 0.010–0.35 Å^{-1}. Two-dimensional scattering patterns were converted to one-dimensional scattering profiles using the *Red2D* software (https://github.com/hurxl/Red2D). After correction by the transmittance and subtraction of buffer scattering, the absolute scattering intensity was obtained with the standard scattering intensity of H_2O (0.89 cm^{-1}) (Shibayama *et al.*, 2005). All measurements were performed at 298 K.

3. Methodology

In this section, we explain how to derive the scattering profile of a monomer from that of a solution that includes aggregates by the AUC–SAS method (see §1 of the supporting information for further details), and present the problems in applying the first AUC–SAS to a solution with a high weight fraction of aggregates.

3.1. Derivation of the scattering profile of protein monomer from an ensemble-averaged scattering profile

The scattering profile of the monomer and its aggregates, $I(q)$, is represented as

$$I(q) = \sum_{j=1}^{n} I_j(q) = \sum_{j=1}^{n} c_j i_j(q) = c \sum_{j=1}^{n} r_j i_j(q), \tag{2}$$

where j denotes the association number ($1 \leq j \leq n$); $I_j(q)$, c_j and $i_j(q)$ are the scattering profile, weight concentration and concentration-normalized scattering profile [$i_j(q) = I_j(q)/c_j$] for the j-mer, respectively; and c and r_j are the total concentration ($c = \sum_j c_j$) and weight fraction for the j-mer ($r_j = c_j/c$), respectively. Since a j-mer could have diverse configurations, $I_j(q)$ indicates the ensemble-average scattering profile of all j-mers. Here, c is low, as the scattering profile is free from the interparticle interference effect.

To solve equation (2) for $I_1(q)$, the weight fractions of all components, $\{r_j\}$ ($j \geq 1$) (#1), and the scattering profiles of aggregates, $i_j(q)$ ($j \geq 2$) (#2), are required. As a prerequisite, highly denatured proteins and high-order aggregates are removed from the sample solution through the purification for a general SAS measurement. Hence, it is reasonable to assume that the residual aggregates are 4-mer at most ($j \leq 4$) and that the total weight fraction of the aggregates, r_a ($\equiv 1 - r_1$), is <0.2. If this prerequisite is not satisfied (*i.e.* $j > 4$ and/or $r_a > 0.2$), the sample should be re-purified. Under these conditions, AUC offers information #1 ($\{r_j\}$) (§2 of the supporting information). Next, to obtain information #2 [$i_j(q)$ ($j \geq 2$)], we divided $i_j(q)$ into two q regions, $i_{jH}(q)$ and $i_{jL}(q)$, in the sufficiently high and lower q regions, respectively. Here, $i_{jH}(q)$ ($j \geq 2$) could be identical to $i_{1H}(q)$ [$i_{jH}(q) \simeq i_{1H}(q)$] because there is no difference in the inner local structure between the

monomer and the aggregates under the prerequisite conditions (no highly denatured aggregates in the sample). Therefore, $I_{1H}(q)$ is obtained using $I(q)$ and r_1 as follows (see §1 of the supporting information for further details):

$$I_{1H}(q) \simeq r_1 I(q), \tag{3}$$

where $I(q)$ and r_1 are experimentally offered by SAS and AUC, respectively.

On the other hand, extrapolation of equation (3) to the lower q region, $i_{jL}(q) \simeq i_{1L}(q)$, does not hold (open magenta circles in Fig. S1 of the supporting information.). Therefore, $I_{1L}(q)$ is considered as follows. First, the forward scattering intensity, $I_1(0)$, is obtained with $I(0)$, r_j and M_j, which are experimentally given by SAS and AUC, as follows (see §1 of the supporting information for further details):

$$I_1(0) = \frac{r_1 M_1}{\sum_{j=1}^{n} r_j M_j} I(0). \tag{4}$$

The remaining issue is a way to obtain $I_{1L}(q)$ ($q > 0$); namely, connecting between $I_{1H}(q)$ and $I_1(0)$. The first AUC–SAS (Morishima *et al.*, 2020) connects them with the Guinier formula:

$$I_{1L}(q) = I_1(0) \exp\left(-\frac{R_{g1}^2}{3} q^2\right), \tag{5}$$

where R_{g1} is the gyration radius of the target monomer. As R_{g1} is an adjustable parameter, a reasonable $I_{1L}(q)$ is found by a smooth joint with $I_{1H}(q)$ at joint point q_c. Finally, $I_1(q)$ is derived from $I_{1L}(q)$ ($q \leq q_c$) and $I_{1H}(q)$ ($q > q_c$), and the appropriate R_{g1} is also provided (see §1 of the supporting information for further details).

3.2. Problems in first AUC–SAS

Figs. 1(*a*)–1(*c*) show the concentration-normalized scattering profiles, $i_1(q)$ [$= I_1(q)/c_1$], that were derived with the first AUC–SAS. The samples were BSA solutions with different weight fractions of aggregates (*a*) $r_a = 0.06$, (*b*) 0.13 and (*c*) 0.20. The experimental AUC and SAXS data are shown in §2 of the supporting information and as open black circles in Figs. 1(*a*)–1(*c*), respectively. The black lines in Figs. 1(*a*)–1(*c*) represent the concentration-normalized scattering profiles, $i_1(q)_{Xtal}$, calculated from the crystal structure of the BSA monomer (PDB code 4f5s; Bujacz, 2012). Here, $i_1(q)_{Xtal}$ is identical to that obtained using SEC–SAXS for a BSA solution (Bucciarelli *et al.*, 2018). Fig. 1(*d*) shows the deviations between the scattering profile derived from the first AUC–SAS, $i_1(q)$, and that calculated from the crystal structure, $i_1(q)_{Xtal}$, *i.e.* $\Delta i_1(q)/\sigma(q)$. Here, $\Delta i_1(q) = i_1(q) - i_1(q)_{Xtal}$ and $\sigma(q)$ is the error of $i_1(q)$. The first AUC–SAS successfully offered reasonable $i_1(q)$ at $r_a = 0.06$ [$\Delta i_1(q)/\sigma(q) < 1$] but produced a large deviation in the middle q region ($0.5 \leq qR_{g1} \leq 3$) at $r_a = 0.13$ and 0.20 [$\Delta i_1(q)/\sigma(q) > 1$]. As a result, R_{g1} at $r_a = 0.06$ ($R_{g1} = 27.2 \pm 0.2$ Å) is consistent with that of the crystal structure ($R_{g1,Xtal} = 27.1$ Å), whereas the R_{g1}s at $r_a = 0.13$ and 0.20 ($R_{g1} = 27.5 \pm 0.2$ Å and $R_{g1} = 28.1 \pm 0.2$ Å,

respectively) are larger than $R_{\mathrm{g1,Xtal}}$ [R_{g1} and $i_1(0)$ are listed in Table S1 of the supporting information].

As shown in Figs. 1(e)–1(h), $i_{\mathrm{1H}}(q)$ [$= I_{\mathrm{1H}}(q)/c_1$], which is given by equation (3), deviated from $i_1(q)_{\mathrm{Xtal}}$ even more in the higher q region than in the Guinier region ($1.3 < qR_{\mathrm{g1}} < 3$) at $r_{\mathrm{a}} = 0.13$ and 0.20. The large deviations, $\Delta i_{\mathrm{1H}}(q)/\sigma(q)$, at $r_{\mathrm{a}} = 0.13$ and 0.20 in the middle q region make the connection points, q_{c}, shift to the out-of-Guinier region ($q_{\mathrm{c}}R_{\mathrm{g1}} \simeq 1.6$ and 1.9, respectively). Consequently, incorrect R_{g1} and scattering profiles were obtained. To solve this problem, the connection should be performed in the Guinier region, that is, $I_{\mathrm{1H}}(q)$ is

correctly extrapolated to the inside of the Guinier region. In this study, we have developed a method to correctly extrapolate $I_{\mathrm{1H}}(q)$ and offer a reasonable $I_1(q)$, even for relatively large r_{a}.

4. Results and discussion

4.1. Scattering profile of aggregates

The approximation of $i_{j\mathrm{H}}(q) \simeq i_{\mathrm{1H}}(q)$, which gives $I_{\mathrm{1H}}(q)$ [equation (3)], holds in the sufficiently high q region ($qR_{\mathrm{g1}} > 3$), as shown in Figs. 1(e)–1(h). To derive the appropriate
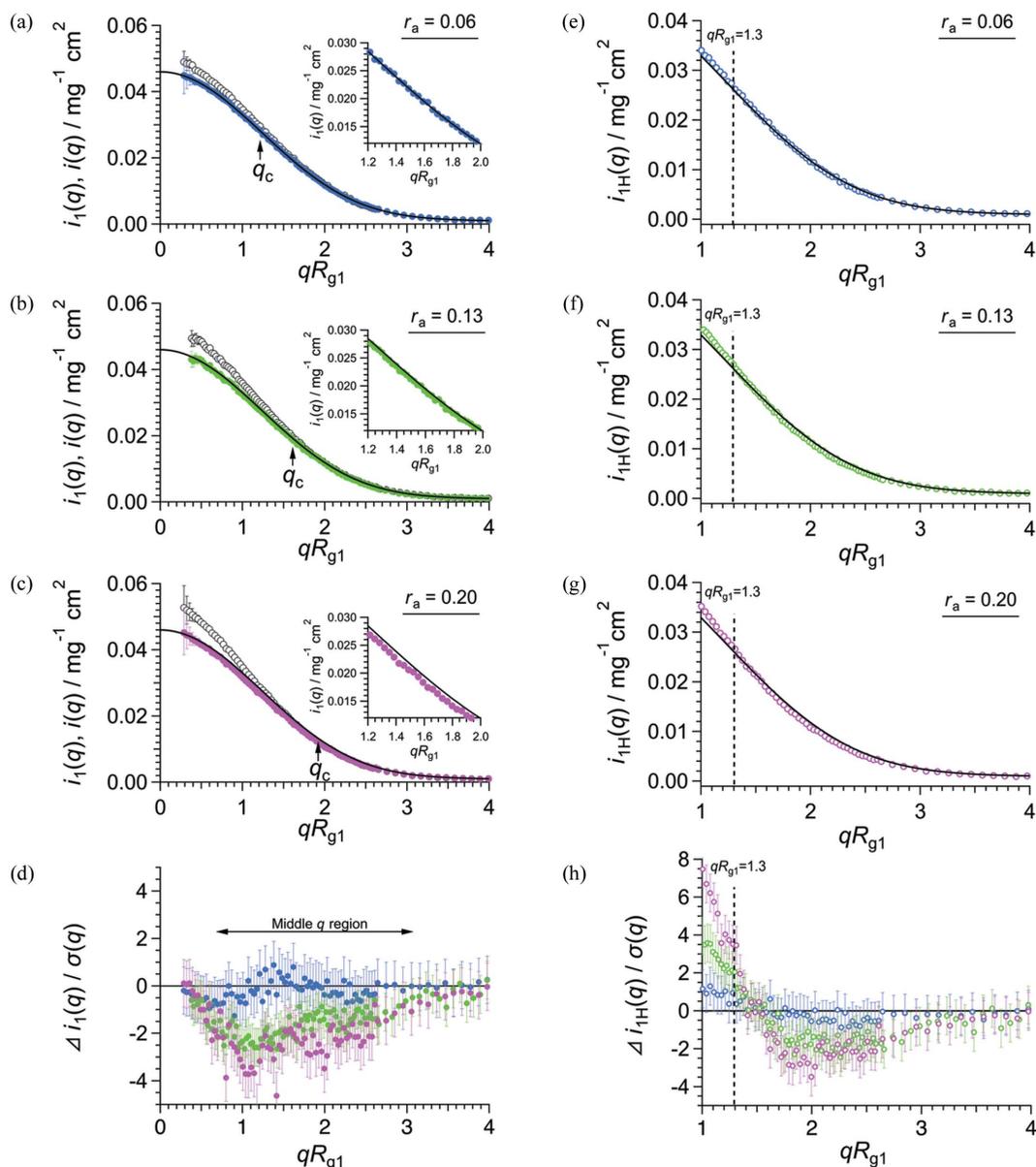

Figure 1
First AUC–SAS for the BSA solutions with various weight fractions of aggregates. (a)–(c) Filled blue, green and purple circles show $i_1(q)$ [$= I_1(q)/c_1$] which is derived by the first AUC–SAS for the BSA solutions with $r_{\mathrm{a}} = 0.06$ (BSA6), 0.13 (BSA13) and 0.20 (BSA20), respectively. Open black circles show the experimental SAXS data $i(q)$ [$= I(q)/c$]. Black lines represent $i_1(q)_{\mathrm{Xtal}}$ calculated from the crystal structure of BSA monomer (PDB code 4f5s). Arrows indicate the connection points q_{c} between $i_{\mathrm{1L}}(q)$ and $i_{\mathrm{1H}}(q)$. Insets show the enlarged pictures in the range $1.2 \leq qR_{\mathrm{g1}} \leq 2.0$. (d) Filled blue, green and purple circles show the residuals $\Delta i_1(q)/\sigma(q)$ for BSA6, BSA13 and BSA20, respectively. Here, $\Delta i_1(q) = i_1(q) - i_1(q)_{\mathrm{Xtal}}$ and $\sigma(q)$ denotes the error of $i_1(q)$. (e)–(g) Open blue, green and purple circles show $i_{\mathrm{1H}}(q)$ [$= I_{\mathrm{1H}}(q)/c_1$] given by equation (3) for BSA6, BSA13 and BSA20, respectively. The black line in each panel represents $i_1(q)_{\mathrm{Xtal}}$ calculated from the crystal structure of the BSA monomer. (h) Open blue, green and purple circles denote the residuals $\Delta i_{\mathrm{1H}}(q)/\sigma(q)$ for BSA6, BSA13 and BSA20, respectively. Here, $\Delta i_{\mathrm{1H}}(q) = i_{\mathrm{1H}}(q) - i_1(q)_{\mathrm{Xtal}}$ and $\sigma(q)$ is the error of $i_{\mathrm{1H}}(q)$. The broken line denotes the upper limit of the Guinier approximation range ($qR_{\mathrm{g1}} = 1.3$).

$I_{1H}(q)$ that is correctly extrapolated to the inside of the Guinier region, we carefully reconsidered the scattering profile of an aggregate. First, the concentration-normalized scattering profile of the j-mer, $i_j(q)$, is represented as follows:

$$i_j(q) = \frac{i_1(0)}{j} \left\langle \sum_{k=1}^{j} \sum_{l=1}^{j} F_k(\mathbf{q}) F_l^*(\mathbf{q}) \exp[-i\mathbf{q} \cdot (\mathbf{R}_k - \mathbf{R}_l)] \right\rangle, \quad (6)$$

where $\mathbf{R}_{k,l}$ and $F_{k,l}(\mathbf{q})$ are the position vectors of the center of mass (COM) and the form factors of the k- or l-th subunit, respectively [Fig. 2(a)]. The form factor is normalized to be $\langle |F_{k,l}(0)|^2 \rangle = 1$, where $\langle \ldots \rangle$ denotes the orientational average. The asterisk (*) denotes the complex conjugate.

Next, we assumed that the subunits were randomly arranged in the aggregate. According to the 'decoupling approximation method' (Kotlarchyk & Chen, 1983), the form factor is independent of the position in the aggregate: $F_k(\mathbf{q})F_l^*(\mathbf{q})$ and $\exp[-i\mathbf{q} \cdot (\mathbf{R}_k - \mathbf{R}_l)]$ in equation (6) can be decoupled, as in equation (S3) of the supporting information. Therefore, $i_j(q)$ can be expressed as follows (also see §4 of the supporting information):

$$i_j(q) = \{1 + \beta(q)[T_j(q) - 1]\}i_1(q), \quad (7)$$

where

$$T_j(q) \equiv \frac{1}{j} \sum_{k=1}^{j} \sum_{l=1}^{j} \frac{\sin(qD_{kl})}{qD_{kl}} \quad (8)$$

and

$$\beta(q) \equiv \frac{|\langle F(\mathbf{q}) \rangle|^2}{\langle |F(\mathbf{q})|^2 \rangle}. \quad (9)$$

$T_j(q)$ is the inter-subunit structure factor defined by the Debye function [equation (8)] with the distance between the COMs of the kth and lth subunits, D_{kl}. Considering the random arrangement of the subunits, $T_j(q)$ is expressed with the random flight model as equation (10). This model was originally developed for a synthetic polymer chain (Burchard & Kajiwara, 1970) and has been subsequently applied to randomly associated proteins (Larsen et al., 2020).

$$T_j(q) = \frac{2}{1 - \sin(qD)/qD}$$
$$- \frac{2 - 2[\sin(qD)/qD]^j}{j[1 - \sin(qD)/qD]^2} [\sin(qD)/qD] - 1, \quad (10)$$

where D is the average distance between neighboring subunits ($= \langle D_{k,k+1} \rangle$). Assuming that the gyration radius of a subunit, R_{g1}, is the effective radius of the subunit, we defined $D \equiv 2R_{g1}$ (see §5 of the supporting information).

$\beta(q)$ indicates the shape anisotropy of the subunit [equation (9)]. Because the form factor of a subunit, $F(\mathbf{q})$, is unknown prior to structural analysis of the monomer, we assumed that the subunit is an ellipsoid whose semi-axes are r and pr (p is the axial ratio), as shown in Fig. 2(b). Its form factor is then represented as follows:

$$F(q, \alpha) = \frac{3[\sin(qr_e) - (qr_e)\cos(qr_e)]}{(qr_e)^3}, \quad (11)$$

where

$$r_e = \left(\frac{5}{2 + p^2}\right)^{1/2} R_{g1}(\sin^2\alpha + p^2\cos^2\alpha)^{1/2}. \quad (12)$$

Then

$$|\langle F(\mathbf{q}) \rangle|^2 = \left[\int_0^{\pi/2} F(q, \alpha)\sin(\alpha)\,d\alpha\right]^2 \quad (13)$$

and

$$\left\langle |F(\mathbf{q})|^2 \right\rangle = \int_0^{\pi/2} F^2(q, \alpha)\sin(\alpha)\,d\alpha, \quad (14)$$

where α is the orientation angle between the axis of the ellipsoid and q [Fig. 2(b)]. $\beta(q)$ was obtained by substituting equations (11)–(14) into equation (9). The axial ratio, p, is estimated using the frictional ratio f/f_0, which is offered by the AUC measurement (Lebowitz et al., 2002) (see further details in §6 of the supporting information).

4.2. Improved AUC–SAS

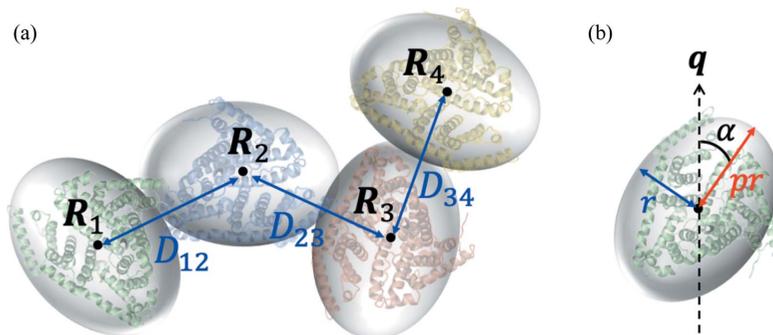By substituting equation (7) into equation (2), $I_1(q)$ is expressed as follows:

Figure 2
Schematic illustrations of an aggregate and a subunit. (a) A schematic illustration of an aggregate ($j = 4$) in which the subunits are randomly arranged. Black points and blue arrows represent the COMs of the subunits and the distances between the COMs of neighboring subunits, respectively. (b) A schematic illustration of the ellipsoidal approximation of a subunit. Blue and red arrows represent the semi-axes. The broken black arrow indicates the scattering vector.

$$I_1(q) = \left(\sum_{j=1}^{n} r_j \{ 1 + \beta(q)[T_j(q) - 1] \} \right)^{-1} r_1 I(q). \qquad (15)$$

For this improvement, $I_{1H}(q)$ was calculated using equation (15), instead of equation (3). To estimate R_{g1} in $T_j(q)$ and $\beta(q)$ [equations (10)–(14)], the first AUC–SAS was initially used.

The improved method was demonstrated for BSA and AF solutions with $r_a = 0.20$ (BSA20) and $r_a = 0.21$ (AF21), respectively. Their experimental AUC data are shown in §2 of the supporting information. Fig. 3(a) shows $i_1(q)$ [= $I_1(q)/c_1$] which was derived using the first AUC–SAS (purple circles) and improved AUC–SAS (cyan circles) for BSA20. As shown in Fig. 3(b), the deviations $\Delta i_1(q)/\sigma(q)$ for the improved
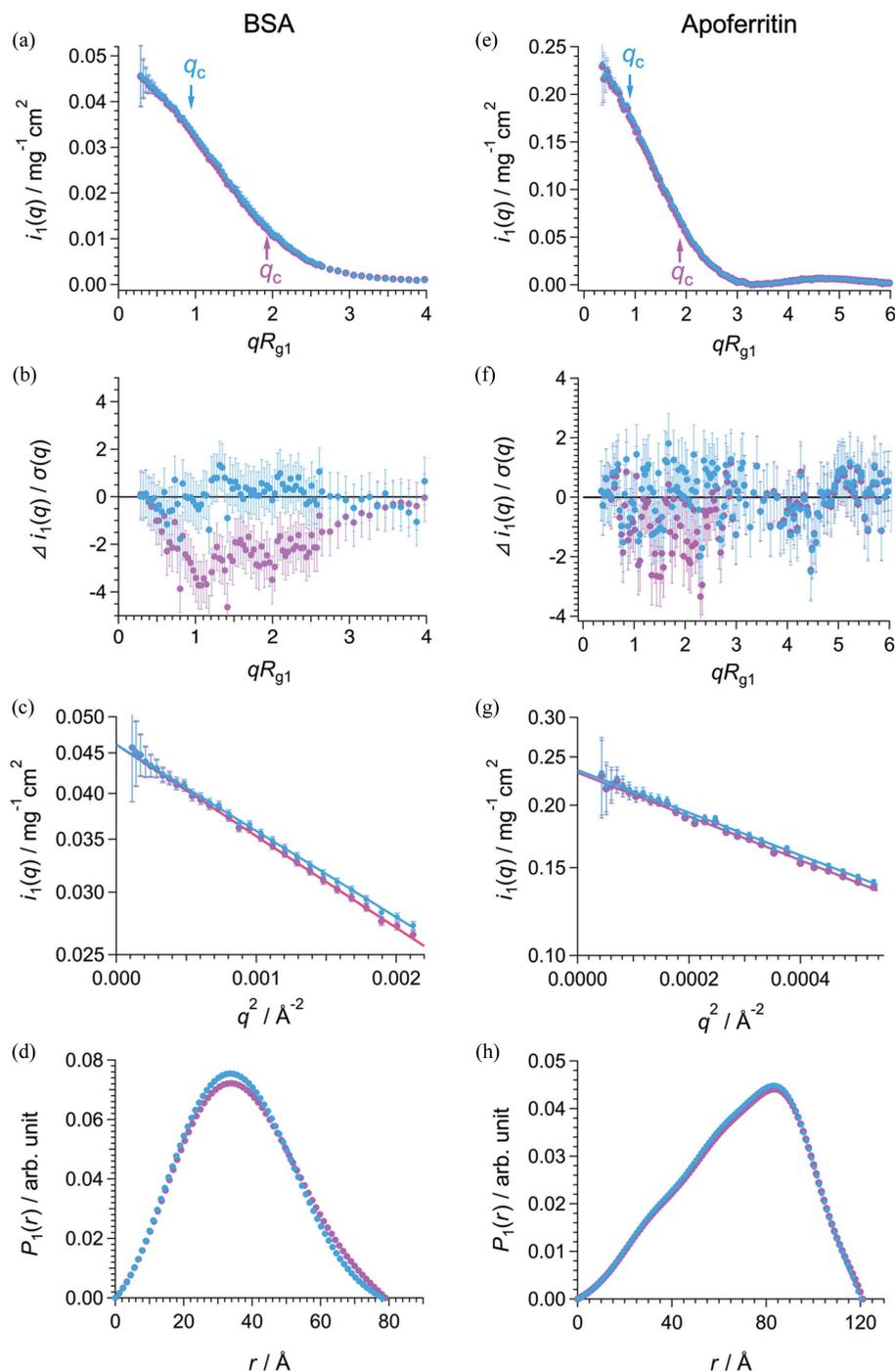

Figure 3
Demonstration of the first and improved AUC–SAS for BSA20 and AF21. In all panels, purple and cyan circles represent the results of the first AUC–SAS and improved AUC–SAS, respectively. Concentration-normalized scattering profiles, $i_1(q)$, for (a) BSA monomer and (e) AF 24-mer. Residuals $\Delta i_1(q)/\sigma(q)$ for (b) BSA20 and (f) AF21. Guinier plots of $i_1(q)$ for (c) BSA20 and (g) AF21. Solid purple and cyan lines express the least-squares fitting lines with the Guinier formula. Pair distance distribution functions, $P_1(r)$, for (d) BSA20 and (h) AF21.

AUC–SAS were sufficiently small $[\Delta i_1(q)/\sigma(q) < 1]$ in the entire q region. As shown in Figs. 3(c) and 3(d) and Table 1, the improved AUC–SAS yielded more reasonable structural parameters $[R_{g1}, i_1(0), P_1(r)$ (pair distance distribution function) and D_{max}] than the first AUC–SAXS. For the larger protein, AF solution (AF21), the improved AUC–SAS successfully gave reasonable $i_1(q)$ and structural parameters [Figs. 3(e)–3(h) and Table 1]. Thus, the improved AUC–SAS was applicable to a solution with a relatively large r_a (≤ 0.2), which is the general condition for most SAS measurements.

Furthermore, we demonstrated the improved AUC–SAS for various proteins with different shapes and sizes (AUC results of the samples are shown in §2 of the supporting information). As shown in Fig. 4 and Table 2, the scattering profiles $i_1(q)$ and structural parameters $[R_{g1}$ and $i_1(0)]$ offered by the improved AUC–SAS are consistent with those of SEC–SAXS for these proteins at various r_a (≤ 0.2).

AUC–SAS is applicable to SANS, which faces the same aggregation problem, as well as SAXS. We examined the AUC–SANS for a BSA solution (BSA3) using the improved AUC–SAS (§7 of the supporting information). For the SANS data of BSA3, the improved AUC–SAS successfully offered a reasonable scattering profile and gyration radius ($R_{g1} = 26.5 \pm 0.2$ Å) that were consistent with those of the crystal structure ($R_{g1,Xtal} = 26.7$ Å). For neutron facilities without a SEC–SANS system (Jordan *et al.*, 2016; Johansen *et al.*, 2018; Sato *et al.*,
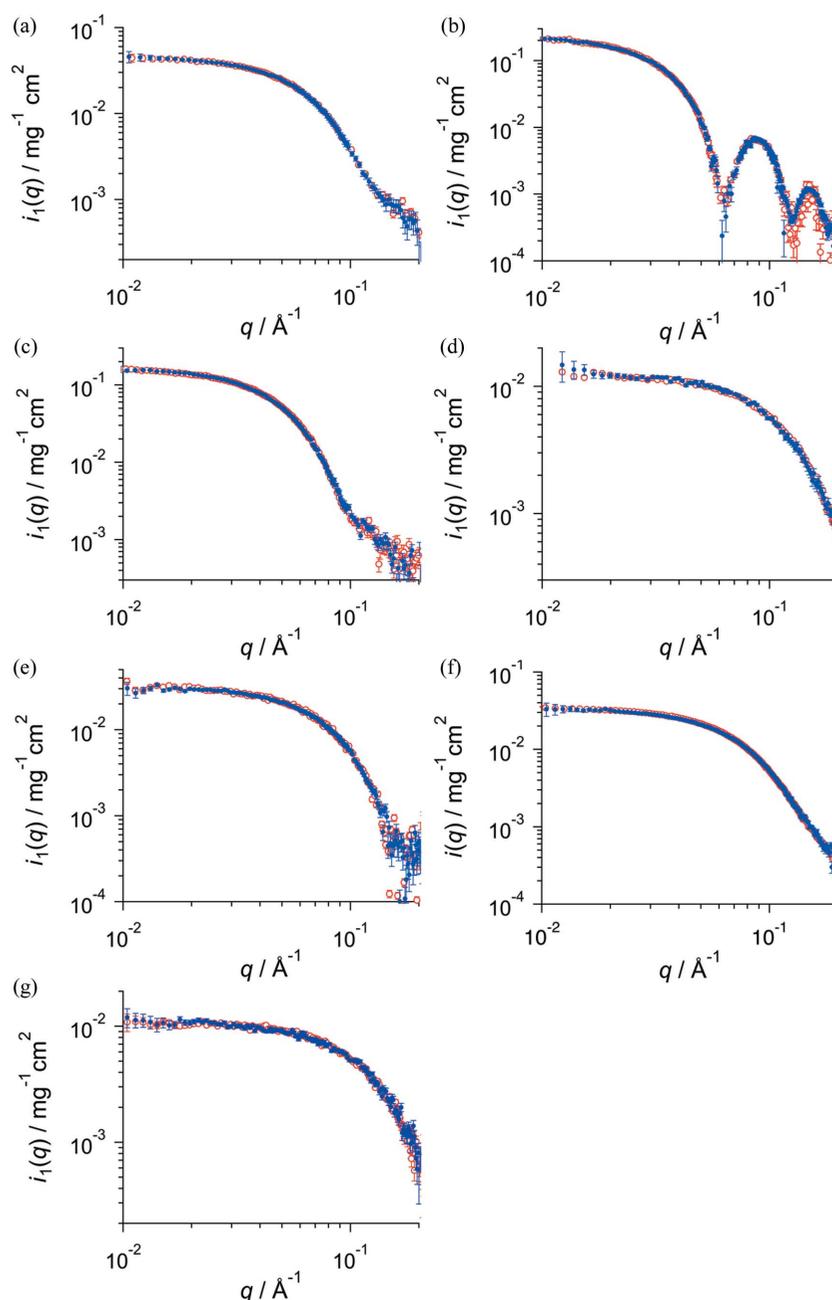

Figure 4
Open red and filled blue circles show the scattering profile $i_1(q)$ given by SEC–SAXS and improved AUC–SAS, respectively, for (a) BSA20, (b) AF21, (c) Cat8, (d) Lyz6, (e) βB2-cry11, (f) OVA4 and (g) RNaseA8.

Table 1
Gyration radii, forward scattering intensities, molecular weights calculated from forward scattering intensities, and maximum pair distances for BSA20 and AF21.

R_g and $i(0)$: gyration radius and concentration-normalized forward scattering intensity for non-treated SAXS, respectively. R_{g1} and $i_1(0)$: gyration radius and concentration-normalized forward scattering intensity of the monomer, respectively, which were derived using AUC–SAS. M: molecular weight calculated from the forward scattering intensity. D_{max}: maximum pair distance from $P_1(r)$. The error of the gyration radius is the standard deviation. The errors of the concentration-normalized forward scattering intensity and molecular weight were calculated from the standard deviations of the forward scattering intensity and concentration.

	R_g, R_{g1} (Å)	$i(0)$, $i_1(0)$ (mg^{-1}cm^2)	M (kDa)	D_{max}
BSA20				
Non-treated SAXS	30.9 ± 0.3	0.0565 ± 0.0005	84.2 ± 0.7	98.5
First AUC–SAS	28.1 ± 0.2	0.0461 ± 0.0004	68.7 ± 0.6	78.8
Improved AUC–SAS	27.3 ± 0.2	0.0468 ± 0.0004	69.6 ± 0.6	77.9
Crystal structure†	27.1	0.0465	69.2	77.0
AF21				
Non-treated SAXS	61.5 ± 0.8	0.280 ± 0.010	604 ± 20	175
First AUC–SAS	55.1 ± 0.5	0.235 ± 0.009	507 ± 19	122
Improved AUC–SAS	54.1 ± 0.5	0.239 ± 0.009	516 ± 19	120
Crystal structure‡	54.0	0.221	477	119

† PDB code 4f5s.   ‡ PDB code 4v1w; Russo & Passmore (2014).

Table 2
Gyration radii and forward scattering intensities given by AUC–SAXS (improved AUC–SAXS) and SEC–SAXS for various proteins.

R_{g1} and $i_1(0)$: gyration radius and concentration-normalized forward scattering intensity, respectively. The error of the gyration radius is the standard deviation. The error of the concentration-normalized forward scattering intensity was calculated from the standard deviations of the forward scattering intensity and concentration.

Protein	Method	Sample code	R_{g1} (Å)	$i_1(0)$ (mg^{-1}cm^2)
BSA	Improved AUC–SAXS	BSA6 ($r_a = 0.06$)	27.2 ± 0.2	0.0461 ± 0.0006
		BSA13 ($r_a = 0.13$)	27.3 ± 0.2	0.0460 ± 0.0005
		BSA20 ($r_a = 0.20$)	27.3 ± 0.2	0.0468 ± 0.0004
	SEC–SAXS	–	27.2 ± 0.3	0.0455 ± 0.0019
AF	Improved AUC–SAXS	AF5 ($r_a = 0.05$)	54.4 ± 0.6	0.249 ± 0.008
		AF15 ($r_a = 0.15$)	54.3 ± 0.5	0.243 ± 0.009
		AF21 ($r_a = 0.21$)	54.1 ± 0.5	0.239 ± 0.009
	SEC–SAXS	–	53.9 ± 0.6	0.241 ± 0.015
Catalase	Improved AUC–SAXS	Cat3 ($r_a = 0.03$)	37.2 ± 0.2	0.155 ± 0.006
		Cat8 ($r_a = 0.08$)	37.3 ± 0.2	0.153 ± 0.005
	SEC–SAXS	Cat8 ($r_a = 0.08$)	37.1 ± 0.2	0.160 ± 0.008
Lyz	Improved AUC–SAXS	Lyz6 ($r_a = 0.06$)	15.0 ± 0.2	0.0125 ± 0.0008
	SEC–SAXS	–	15.1 ± 0.1	0.0124 ± 0.0012
βB2-cry	Improved AUC–SAXS	βB2-cry11 ($r_a = 0.11$)	22.6 ± 0.2	0.0319 ± 0.0007
	SEC–SAXS	–	22.7 ± 0.3	0.0325 ± 0.0013
OVA	Improved AUC–SAXS	OVA4 ($r_a = 0.04$)	23.9 ± 0.2	0.0346 ± 0.0005
	SEC–SAXS	–	23.9 ± 0.4	0.0340 ± 0.0009
RNaseA	Improved AUC–SAXS	RNaseA8 ($r_a = 0.08$)	15.0 ± 0.2	0.0110 ± 0.0004
	SEC–SAXS	–	14.8 ± 0.2	0.0116 ± 0.0003

2021), AUC–SANS is the most promising method for obtaining the aggregation-free scattering profile.

In §8 and §9 of the supporting information, we evaluate the maximum errors originated by the random flight model and ellipsoidal approximation. The error in $I_1(q)$ is several per cent at most, even though the extreme cases are assumed.

It is often worthwhile analyzing the structure of the aggregate (Kovalchuk *et al.*, 2019). Programs such as *SASREFMX* and *OLIGOMER* in the *ATSAS* package

(Petoukhov *et al.*, 2012; Manalastas-Cantos *et al.*, 2021) are well known for modeling of aggregates. However, these programs require the structure of the monomer. Hence, the complementary use of AUC–SAS and these programs is a promising strategy.

Implementing the improved AUC–SAS, *Igor Pro*-based software (Kline, 2006) has been developed for the utilization of AUC–SAS by SAS experimenters. The required information is the data set of molecular weights (or association number), weight fractions and the frictional ratio, which are given by AUC. The scattering profile of the target monomer is obtained just by inputting the AUC information and SAS profile for the solution. The software is available at https://www.rri.kyoto-u.ac.jp/NSBNG/activity.html. Its usage is described in §10 of the supporting information.

5. Related literature

The following additional references are only cited in the supporting information for this article: Perkins (2001), Perrin (1934), Pierce *et al.* (2014).

of the Institute for Solid State Physics (ISSP), The University of Tokyo, at the Japan Atomic Energy Agency (JAEA) under proposal Nos. 21531, 21537, 22569 and 222914.

References

Bengtsen, T., Holm, V. L., Kjølbye, L. R., Midtgaard, S. R., Johansen, N. T., Tesei, G., Bottaro, S., Schiøtt, B., Arleth, L. & Lindorff-Larsen, K. (2020). *Elife*, **9**, e56518.

Bernadó, P., Shimizu, N., Zaccai, G., Kamikubo, H. & Sugiyama, M. (2018). *Biochim. Biophys. Acta*, **1862**, 253–274.

Brown, P. H. & Schuck, P. (2006). *Biophys. J.* **90**, 4651–4661.

Bucciarelli, S., Midtgaard, S. R., Nors Pedersen, M., Skou, S., Arleth, L. & Vestergaard, B. (2018). *J. Appl. Cryst.* **51**, 1623–1632.

Bujacz, A. (2012). *Acta Cryst.* D**68**, 1278–1289.

Burchard, W. & Kajiwara, K. (1970). *Proc. R. Soc. Lond. Ser. A*, **316**, 185–199.

David, G. & Pérez, J. (2009). *J. Appl. Cryst.* **42**, 892–900.

Grant, T. D. (2018). *Nat. Methods*, **15**, 191–193.

Gräwert, T. W. & Svergun, D. I. (2020). *J. Mol. Biol.* **432**, 3078–3092.

Hirano, R., Arimura, Y., Kujirai, T., Shibata, M., Okuda, A., Morishima, K., Inoue, R., Sugiyama, M. & Kurumizaka, H. (2021). *Commun. Biol.* **4**, 191.

Inoue, R., Nakagawa, T., Morishima, K., Sato, N., Okuda, A., Urade, R., Yogo, R., Yanaka, S., Yagi-Utsumi, M., Kato, K., Omoto, K., Ito, K. & Sugiyama, M. (2019). *Sci. Rep.* **9**, 12610.

Johansen, N. T., Pedersen, M. C., Porcar, L., Martel, A. & Arleth, L. (2018). *Acta Cryst.* D**74**, 1178–1191.

Jordan, A., Jacques, M., Merrick, C., Devos, J., Forsyth, V. T., Porcar, L. & Martel, A. (2016). *J. Appl. Cryst.* **49**, 2015–2020.

Kline, S. R. (2006). *J. Appl. Cryst.* **39**, 895–900.

Kotlarchyk, M. & Chen, S. H. (1983). *J. Chem. Phys.* **79**, 2461–2469.

Kovalchuk, M. V., Boikova, A. S., Dyakova, Y. A., Ilina, K. B., Konarev, P. V., Kryukova, A. E., Marchenkova, M. A., Pisarevsky, Y. V. & Timofeev, V. I. (2019). *J. Biomol. Struct. Dyn.* **37**, 3058–3064.

Lampi, K. J., Amyx, K. K., Ahmann, P. & Steel, E. A. (2006). *Biochemistry*, **45**, 3146–3153.

Larsen, A. H., Pedersen, J. S. & Arleth, L. (2020). *J. Appl. Cryst.* **53**, 991–1005.

Lebowitz, J., Lewis, M. S. & Schuck, P. (2002). *Protein Sci.* **11**, 2067–2079.

Mahieu, E. & Gabel, F. (2018). *Acta Cryst.* D**74**, 715–726.

Manalastas-Cantos, K., Konarev, P. V., Hajizadeh, N. R., Kikhney, A. G., Petoukhov, M. V., Molodenskiy, D. S., Panjkovich, A., Mertens, H. D. T., Gruzinov, A., Borges, C., Jeffries, C. M., Svergun, D. I. & Franke, D. (2021). *J. Appl. Cryst.* **54**, 343–355.

Matsumoto, A., Sugiyama, M., Li, Z., Martel, A., Porcar, L., Inoue, R., Kato, D., Osakabe, A., Kurumizaka, H. & Kono, H. (2020). *Biophys. J.* **118**, 2209–2219.

Morishima, K., Okuda, A., Inoue, R., Sato, N., Miyamoto, Y., Urade, R., Yagi-Utsumi, M., Kato, K., Hirano, R., Kujirai, T., Kurumizaka, H. & Sugiyama, M. (2020). *Commun. Biol.* **3**, 294.

Okuda, A., Shimizu, M., Morishima, K., Inoue, R., Sato, N., Urade, R. & Sugiyama, M. (2021). *Sci. Rep.* **11**, 5655.

Orthaber, D., Bergmann, A. & Glatter, O. (2000). *J. Appl. Cryst.* **33**, 218–225.

Perkins, S. J. (2001). *Biophys. Chem.* **93**, 129–139.

Perrin, F. (1934). *J. Phys. Radium*, **5**, 497.

Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D. T., Konarev, P. V. & Svergun, D. I. (2012). *J. Appl. Cryst.* **45**, 342–350.

Pierce, B. G., Wiehe, K., Hwang, H., Kim, B.-H., Vreven, T. & Weng, Z. (2014). *Bioinformatics*, **30**, 1771–1773.

Russo, C. J. & Passmore, L. A. (2014). *Science*, **346**, 1377–1380.

Ryan, T. M., Trewhella, J., Murphy, J. M., Keown, J. R., Casey, L., Pearce, F. G., Goldstone, D. C., Chen, K., Luo, Z., Kobe, B., McDevitt, C. A., Watkin, S. A., Hawley, A. M., Mudie, S. T., Samardzic Boban, V. & Kirby, N. (2018). *J. Appl. Cryst.* **51**, 97–111.

Sato, N., Yogo, R., Yanaka, S., Martel, A., Porcar, L., Morishima, K., Inoue, R., Tominaga, T., Arimori, T., Takagi, J., Sugiyama, M. & Kato, K. (2021). *J. Biochem.* **169**, 701–708.

Schuck, P. (2000). *Biophys. J.* **78**, 1606–1619.

Shibayama, M., Nagao, M., Okabe, S. & Karino, T. (2005). *J. Phys. Soc. Jpn*, **74**, 2728–2736.

Shimizu, M., Okuda, A., Morishima, K., Inoue, R., Sato, N., Yunoki, Y., Urade, R. & Sugiyama, M. (2022). *Sci. Rep.* **12**, 9970.

Shimizu, N., Yatabe, K., Nagatani, Y., Saijyo, S., Kosuge, T. & Igarashi, N. (2016). *AIP Conf. Proc.* **1741**, 050017.

Svergun, D. I. & Koch, M. H. (2003). *Rep. Prog. Phys.* **66**, 1735–1782.

Yunoki, Y., Matsumoto, A., Morishima, K., Martel, A., Porcar, L., Sato, N., Yogo, R., Tominaga, T., Inoue, R., Yagi-Utsumi, M., Okuda, A., Shimizu, M., Urade, R., Terauchi, K., Kono, H., Yagi, H., Kato, K. & Sugiyama, M. (2022). *Commun. Biol.* **5**, 184.