AIP Conference Proceedings

Software for serial data analysis measured by SEC-SAXS/UV-Vis spectroscopy

Cite as: AIP Conference Proceedings **2054**, 060082 (2019); https://doi.org/10.1063/1.5084713
Published Online: 16 January 2019

Kento Yonezawa, Masatsuyo Takahashi, Keiko Yatabe, Yasuko Nagatani, and Nobutaka Shimizu

View Online Export Citation

ARTICLES YOU MAY BE INTERESTED IN

BL-10C, the small-angle x-ray scattering beamline at the photon factory
AIP Conference Proceedings **2054**, 060041 (2019); https://doi.org/10.1063/1.5084672

New high-brilliance small angle x-ray scattering beamline, BL-15A2 at the photon factory
AIP Conference Proceedings **2054**, 060038 (2019); https://doi.org/10.1063/1.5084669

Further developments of the tender x-ray diffractometer at BL-15A2 of the photon factory
AIP Conference Proceedings **2054**, 060043 (2019); https://doi.org/10.1063/1.5084674

AIP Publishing

© 2019 Author(s).

Software for Serial Data Analysis Measured by SEC-SAXS/UV-Vis Spectroscopy

Kento Yonezawa[1, a)], Masatsuyo Takahashi[1], Keiko Yatabe[1], Yasuko Nagatani[1] and Nobutaka Shimizu[1]

[1] *Photon Factory, Institute of Materials Structure Science, High Energy Accelerator Research Organization (KEK) 1-1 Oho, Tsukuba, Ibaraki 305-0801, Japan*

[a)] Corresponding author: ykento@post.kek.jp

Abstract. Data processing software is commonly utilized for serial measurements in small-angle X-ray scattering with size exclusion chromatography (SEC-SAXS) to effectively process hundreds of scattering data files. However, software to automatically extrapolate SEC-SAXS data to an infinite dilution has yet to be developed. SEC-SAXS/UV-Vis measurements at the Photon Factory not only provide structural information but also give the concentration for each measurement at the sample cell position simultaneously. Since SEC-SAXS data contain various concentration data, the concentration dependence due to the interparticle interference effect can be accurately estimated. Herein we report a new program with a user-friendly interface for multiple automatic data processes to calculate the scattering intensity extrapolated to the infinite dilution. This software provides baseline corrections and a mapping optimization between the serial scattering intensity and the absorbance to obtain the most accurate scattering profile.

INTRODUCTION

Small-angle X-ray scattering (SAXS) is a powerful technique to obtain structure information at a low-resolution in biological molecular systems. Recently, SAXS experiments for biological macromolecules solution (BioSAXS) have employed size exclusion chromatography (SEC-SAXS) as the general standard because it can measure the monodispersed components isolated by a gel-filtration column. Continuously flowing samples in SEC-SAXS also allow data about the concentration variations to be measured without radiation damage to the sample.

Due to the prevalence of SEC-SAXS, analysis programs have been developed to process multiple forms of data. The *ATSAS* program package, which is widely used to analyze biological macromolecules, also includes a module for SEC-SAXS, *CHROMIXS* [1,2]. This module can select a suitable buffer and sample region from the chromatograms. Other software for SEC-SAXS has also been developed [3,4,5,6]. Some programs perform baseline corrections to adjust the baseline drift during the long-time measurements. *DERA* adopts the Guinier optimization to maintain a constant buffer scale using the Guinier approximation [5]. The *US-SOMO* HPLC-SAXS module employs the integral baseline method, which is based on capillary fouling [6]. Such software programs allow data obtained by SEC-SAXS to be easily analyzed. Although these programs can process hundreds of data files in the pipeline, they cannot appropriately process the influence of the interparticle interference effect, which appears due to various concentration dependencies. Interparticle interference can be reduced by adding NaCl, but it is difficult to remove the concentration dependence completely. To incorporate the concentration dependence into the analysis to extrapolate an infinite dilution, the concentration at the sample cell position must be measured.

At the Photon Factory, a fiber spectrophotometer QEpro or QE65pro (Ocean Optics), which can be measured at the cell position, is attached to the sample cell holder to monitor the sample concentration of the SAXS data simultaneously [7]. SAXS analysis software *SAngler* has been developed to instantaneously convert hundreds of two-dimensional (2D) images into one-dimensional (1D) scattering profiles [8]. *SAngler* can also monitor the directory in which the azimuthal averaged *.dat* file resides and perform a background subtraction if the background is specified.

Proceedings of the 13th International Conference on Synchrotron Radiation Instrumentation – SRI2018
AIP Conf. Proc. 2054, 060082-1–060082-6; https://doi.org/10.1063/1.5084713
Published by AIP Publishing. 978-0-7354-1782-3/$30.00

After the background subtraction, the scattering intensities can be analyzed simultaneously using the absorbance at 280 nm.

In this study, we developed *Serial Analyzer*. This user-friendly software can supply a smart strategy to extrapolate the infinite dilution for serial data analysis such as SEC-SAXS with UV-Visible (UV-Vis) spectroscopy automatically [9]. To accurately calculate the extrapolation profiles, we propose a new concept for baseline corrections and mapping between SAXS and UV-Vis spectroscopy. The source code of *Serial Analyzer* is written by Python 3.6.5. The program can be installed into a 64-bit Windows PC. If *Serial Analyzer* and *ATSAS* are both installed, *AOTORG* and *ALMERGE,* which are command-line programs of *ATSAS* performing Guinier analysis and the extrapolation, respectively, are executed simultaneously [10,11].

EXPERIMENTAL PROCEDURE

Microcrystals of glucose isomerase (GI) were purchased from Hampton Research (HR7-102). GI was dissolved in 10 mM HEPES in a pH7.0 buffer and diluted about 2000 times by amicon ultra (MILLIPORE). The final GI concentration was 24 mg mL^{-1}. SEC-SAXS data collection at room temperature was performed at beamline BL-10C of the Photon Factory (Tsukuba, Japan) with the ACQUITY UPLC H-Class system (Waters). Superdex 200 Increase (GE Healthcare) was used as a gel-filtration column. The eluted solutions from the column were exposed to X-rays and UV-Vis light in a sample cell, which was a stainless-steel cell with a 1-mm light pass via a 0.02-mm-thick quartz glass window. The flow rate was set to 0.05 mL/min during the measurements. The scattering images were recorded on a PILATUS3 2M detector (Dectris), and the scattering intensities on the 2D images were azimuthally averaged and converted into 1D intensity profiles using *SAngler* [8]. The scattering intensities were also converted into an absolute scale using water as the standard. The background data was the average of 20 images of flowed buffer scatterings.

DATA ANALYSIS WITH THE *SERIAL ANALYZER*

This software automatically calculates the scattering profile using the loading data to extrapolate the infinite dilution. Figure 1 shows the flowchart of the software with an emphasis on serial data adjustments and automatic data processing.

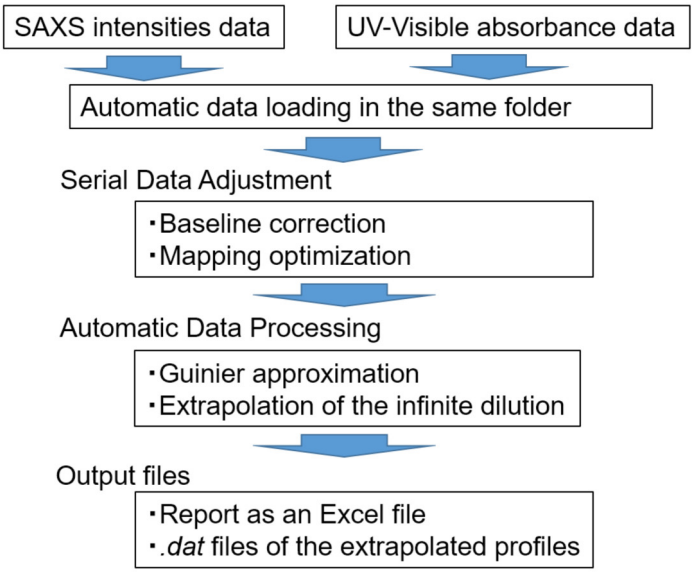

FIGURE 1. Flowchart of the data processing in this software.

Graphical User Interface of *Serial Analyzer*

We designed a user-friendly interface, which uses a simple operation for almost automatic analysis. Thus, even novice SEC-SAXS users can use *Series Analyzer.* Once the directory is selected, all of the *.dat* files and a UV-Vis text

file in that directory are loaded automatically (Fig. 2(a)). After clicking the 'analyze all' button, three chromatograms (UV-Vis, SAXS, Superimposed) are shown subsequently (Fig. 2(b)). The default points to illustrate the 1D scanned chromatogram for the SAXS intensities and the UV-Vis absorbance are set at 0.02 Å^{-1} and 280 nm, respectively. The detailed baseline correction and the automatic mapping method of two superimposed chromatograms are described in the following section. Users can set the data extrapolation range to an infinite dilution. After adjusting the baseline and data range by clicking 'OK' in Fig. 2(b), additional options can be utilized (Fig. 2(c)). The number of averaged scattering curves can be changed to improve the signal to noise (S/N) ratio. Additionally, the averaged files are automatically saved. The concentration value from the UV-Vis absorbance can also be calculated if the extinction values of target proteins are inputted (Fig. 2(c)). The calculations for the Guinier analysis and the extrapolation to the infinite dilution begin upon clicking the 'Run' button in Fig. 2(c). Although the calculations depend on the number of data files, a calculation for hundreds of scattering profiles takes 3–10 minutes. This software outputs the extrapolated scattering *.dat* file of the ascending and the descending sides of the peak and an analytical report as an excel file, which contains the results of the Guinier analysis and the extrapolations.

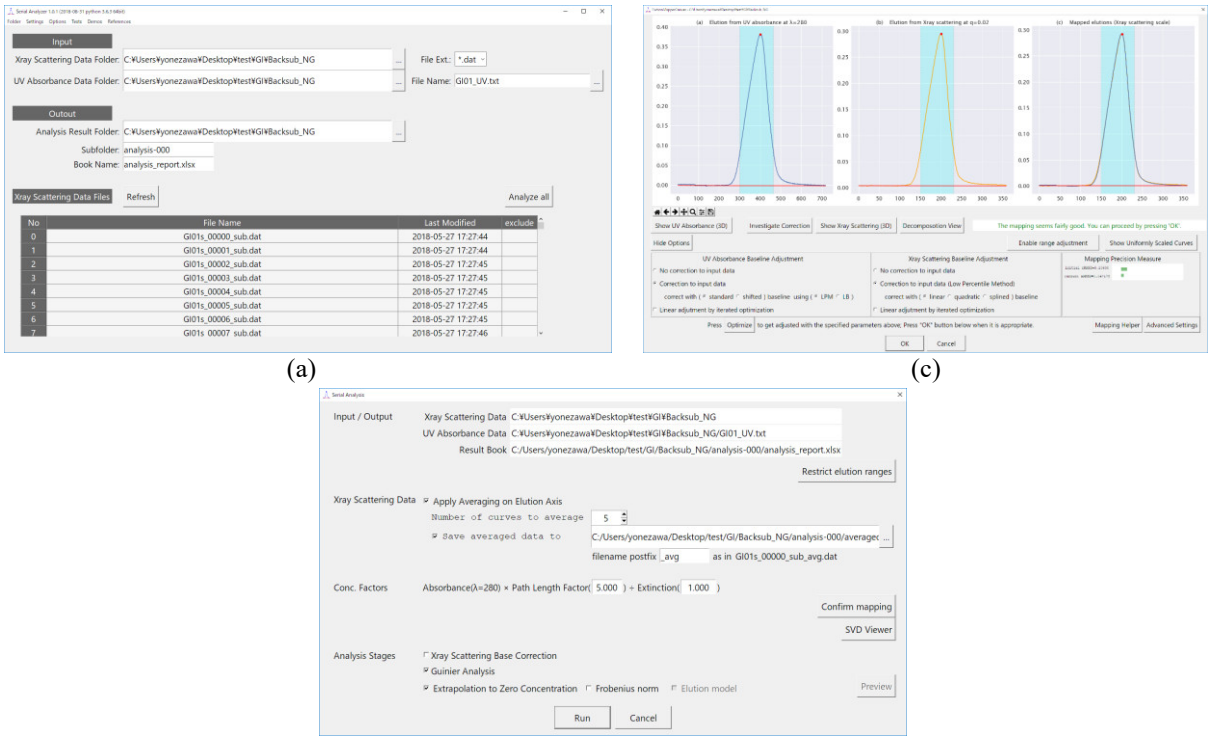

FIGURE 2. (a) Graphic user interface for data loading. (b) Chromatogram of the absorbance at 280 nm and the SAXS intensities at Q=0.02 Å^{-1}. (right) Scaled absorbance data are superimposed onto the SAXS data. (c) Optional settings for the Guinier analysis and the extrapolation to the infinite dilution.

Serial Data Adjustment between UV-Vis Absorbance and SAXS Intensities

To obtain the actual concentration, abnormal baseline drifts, which occur during a measurement, must be corrected in both the absorbance and the SAXS intensities. Our baseline correction method (Low Percentile Methods, LPM) is based on a linear regression for low percentile points of chromatograms on each Q value. Figure 3(a) shows an example of a chromatogram of the intensities at 0.0075 Å^{-1}. Initially, low percentile points (below 25%) against the total frame points (orange dots) are regressed to decide the slope of the corrected baseline. Secondly, the regressed line is parallel-shifted to all points except for the peaks. These points are determined by the low percentile value based on a Monte Carlo simulation, which depends on the width ratio of the peak frame numbers per the total frame numbers ($\sigma_{peak}/\sigma_{Total}$) and the S/N ratio. These calculations are repeated until they converge.

To verify whether the LPM method works, we prepared background-subtracted data, which are calculated incorrectly. Figure 3(b) shows the scattering profile of GI on the top of the peak. The solid black line and the blue dotted line are the calculated data of $[I_{sample}(Q)-0.95\times I_{buffer}(Q)]$ and that of $[I_{sample}(Q)-I_{buffer}(Q)]$, respectively. The high Q value in the scattering profile of GI shows a difference due to the buffer scaling. Then an LPM baseline correction is performed using this erroneous subtraction file (black line). The corrected scattering (solid red line) is consistent with $[I_{sample}(Q)-I_{buffer}(Q)]$, indicating that LPM can improve an abnormal baseline.

After determining the corrected baselines, the vertical and horizontal scales of the chromatogram from the UV-Vis measurement automatically adjust the chromatogram obtained from SAXS (Fig. 2(b)). The root mean square deviation (r.m.s.d.) value between the scaled UV-Vis and the SAXS chromatogram are evaluated by superimposing the UV-Vis absorbance onto the SAXS intensities. Mapping optimization between the UV-Vis data and the SAXS data is performed by minimizing r.m.s.d. The final concentration can be estimated by comparing the peak heights based on these corrections. Additionally, different baseline slopes can be aligned. These dual probe measurements can help identify when unusual phenomena, including baseline drift, occur and how to correct the data.

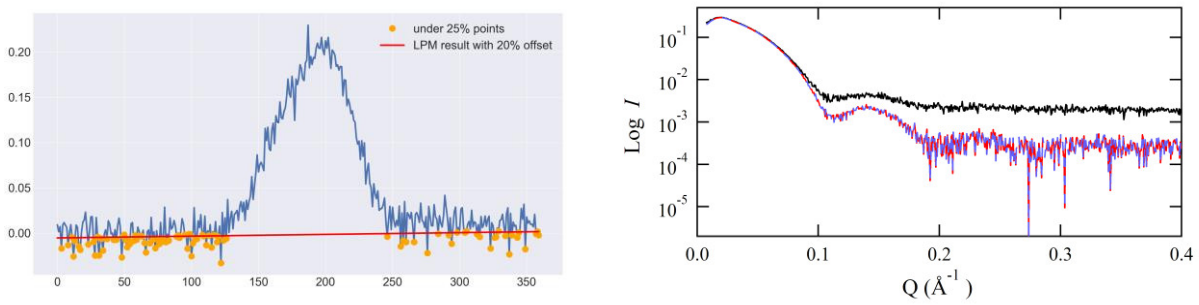

FIGURE 3. (a) Chromatogram of the SAXS intensities (Q=0.0075 Å^{-1}). Orange dots are low percentile points collected below 25%. Fitted line below 20% is shown. (b) Scattering profiles of GI on the top of the peak. Solid black line and blue dotted line are the calculated data of $[I_{sample}(Q)-0.95\times I_{buffer}(Q)]$ and that of $[I_{sample}(Q)-I_{buffer}(Q)]$, respectively. Solid red line superimposed onto $[I_{sample}(Q)-I_{buffer}(Q)]$ is the LPM baseline corrected scattering profile from $[I_{sample}(Q)-0.95\times I_{buffer}(Q)]$.

Serial Automatic Data Process of the Guinier Approximation

Analytical reports are outputted as an Excel file. The results of the Guinier analysis and the extrapolation to an infinite dilution are separated into several individual sheets within the file. Figure 4 shows the resulting figures of the Guinier analysis. The newly developed algorithm for the Guinier approximation analyses and *AUTORG* of *ATSAS* [10] are executed simultaneously. The Q ranges are determined under the condition of $Q\times R_g$ <1.3. Low Q regions with abnormal scatterings, including aggregations and/or interparticle interference, are avoided by recognizing the curvatures, maintaining the accuracy of the calculation.

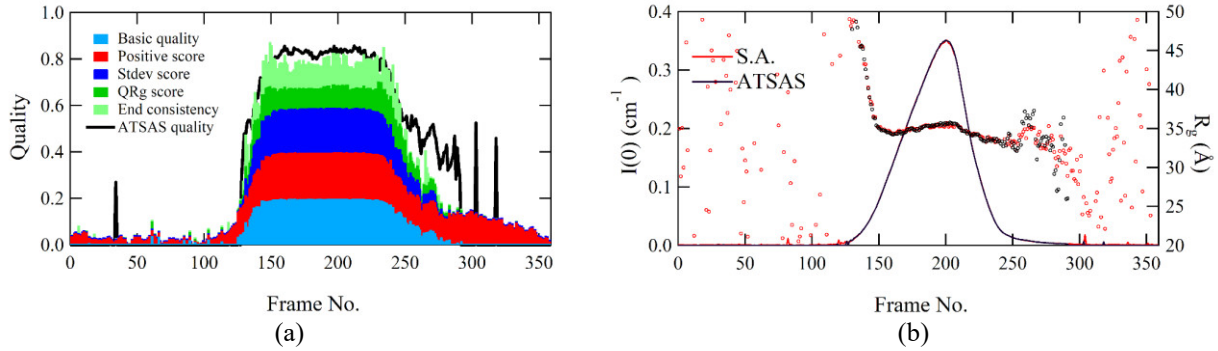

FIGURE 4. Result of the Guinier analysis for each scattering profile, which is outputted into an Excel file. (a) Quality scores. (b) $I(0)$ and R_g values of each scattering.

The original quality score for each scattering profile is also provided to identify the ranges for further analyses (shown in Fig. 4(a)). The *Basic quality* is a preliminary evaluation based on the monotonicity in the small angle region. The *positive score* shows the rate of positive values contained in the scattering profile. As the negative intensity values in the scattering profile increase, this score decreases. $Q\times R_g$ *score* evaluates the width of the Guinier region described by $Q\times R_g$. As the width of the Guinier region (e.g., $0.35 < Q\times R_g < 1.30$, the value is 0.95) becomes wider, this score increases. R_g *Stdev score* evaluates the standard error of R_g in the Guinier region. *End consistency* evaluates the linearity of the Guinier plot at lower Q and a higher Q in the Guinier region. The total quality is an equally weighted sum of these five scores. These scores and the regions of data frames for the extrapolation can help users identify issues contained in the data.

In addition to these scores, $I(0)$ and R_g from the Guinier approximations are also calculated. A gradual increase or decrease in these values near the peak position suggests the existence of interparticle interference (Fig. 4(b)).

Extrapolation to Zero Concentration

The extrapolated scattering profiles using the ascending and descending sides of the peak are obtained as *.dat* files. In addition to executing *ALMERGE* of *ATSAS* [11], a new algorithm to extrapolate to an infinite dilution is also executed after the Guinier approximation. On each Q_i point, the scattering component of the form factor $A(Q_i)$ and that of structure factor $B(Q_i)$ are decomposed using the following formula.

$$I_n(Q_i)/c_n = A(Q_i) + c_n \times B(Q_i) \quad (1)$$

where $I_n(Q)$ is the scattering profile observed on the n-th scattering curve and c_n is the concentration values on the n-th scattering curve. The scattering profile without interparticle interference is obtained from $A(Q_i)$.

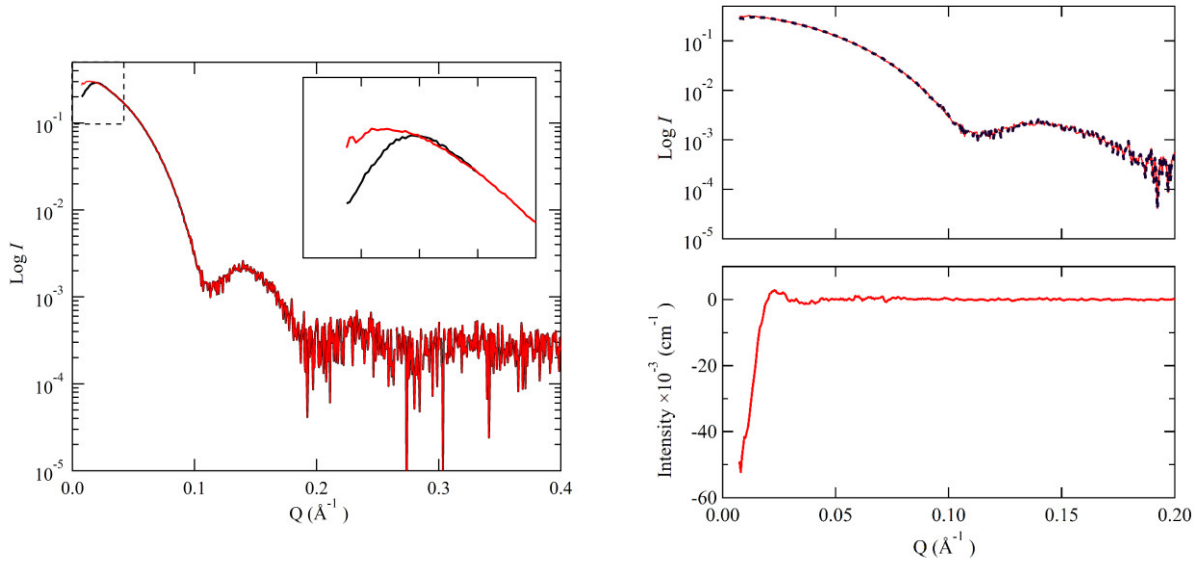

FIGURE 5. Extrapolation results to the infinite dilution. (a) Highest concentration scattering (black line) and extrapolated scattering profile (red line). Inset is an expansion on the smaller Q angle. (b) (upper) Scattering profiles extrapolated to the infinite dilution. Black dotted line and solid red line correspond to the scattering profiles calculated by *ALMERGE* or this program. (lower) Scattering of interparticle interference decomposed from formula (1).

Figure 5 shows an example of GI. The black line in Fig. 5(a) is the highest concentration profile. The inset shows an expanded view of the small Q region. The interparticle interference effect appears in the small Q region. To remove this, the infinite dilution is extrapolated and then the scattering profile is corrected (red line in Fig. 5(a)). Figure 5(b) shows the decomposed scatterings to $A(Q)$ and $B(Q)$. As a default, the wider angle can be replaced with the highest concentration curve to improve the S/N ratio. These replaced Q points can be regarded as the region where the scattering curve of $B(Q)$ converges to zero. The R_g values before and after the extrapolation are 35.7±0.4 Å and

34.1±0.5 Å. Although the reference R_g value (32.5±0.7 Å) is smaller than the corrected one due to different experimental conditions, this correction is reasonable because it approaches the reference R_g value [12]. Almost all sample buffers contain NaCl to reduce interparticle interference effects. However, we need to confirm whether this effect is completely removed in the scattering profile. The SEC-SAXS with UV-Vis spectroscopy system and this software can improve these problems and relax the restrictions on the buffer conditions (e.g., NaCl must be removed or a high concentration of glycerol must be added).

SUMMARY

Herein new software for SEC-SAXS is developed that can automatically process the data obtained from SEC-SAXS with UV-Vis spectroscopy. This program can perform a baseline correction using the Low Percentile Method and the re-subtracted linear baseline drift on each Q. The Guinier approximation and the extrapolation to an infinite dilution are shown simultaneously and immediately. Hence, this software can analyze the most appropriate data for shape predictions.

ACKNOWLEDGMENTS

This research is partially supported by the Platform Project for Supporting Drug Discovery and Life Science Research (Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)) from the Japan Agency for Medical Research and Development (AMED).

REFERENCES

1. D. Franke, M. V. Petoukhov, P. V. Konarev, A. Panjkovich, A. Tuukkanen, H.D.T. Mertens, A.G. Kikhney, N.R. Hajizadeh, J.M. Franklin, C.M. Jeffries, and D.I. Svergun, *J. Appl. Cryst.* **50**, 1212-1225 (2017).
2. A. Panjkovich and D.I. Svergun, *Bioinformatics*. **34**, 1944-1946 (2017).
3. http://www.bioisis.net/tutorial
4. A. V. Shkumatov and S. V. Strelkov, *Acta Crystallogr. D* **71**, 1347-1350 (2015).
5. A.W. Malaby, S. Chakravarthy, T.C. Irving, S. V. Kathuria, O. Bilsel, and D.G. Lambright, *J. Appl. Cryst.* **48**, 1102-1113 (2015).
6. E. Brookes, P. Vachette, M. Rocco, and J. Pérez, *J. Appl. Cryst*. **49**, 1827-1841 (2016).
7. P. Bernadó, N. Shimizu, G. Zaccai, H. Kamikubo and M. Sugiyama, *BBA GEN. SUB.* **1862**, 253-274 (2018).
8. N. Shimizu, K. Yatabe, Y. Nagatani, S. Saijyo, T. Kosuge, and N. Igarashi, *AIP Conf. Proc*. **1741**, 50017 (2016).
9. http://pfwww.kek.jp/saxs/SerialAnalyzer.html
10. M. V. Petoukhov, P. V. Konarev, A.G. Kikhney, and D.I. Svergun, *J. Appl. Cryst*. **40**, s223-s228 (2007).
11. M. V. Petoukhov, D. Franke, A. V. Shkumatov, G. Tria, A.G. Kikhney, M. Gajda, C. Gorba, H.D.T. Mertens, P. V. Konarev, and D.I. Svergun, *J. Appl. Cryst.* **45**, 342-350 (2012).
12. E. Mylonas and D.I. Svergun, *J. Appl. Cryst*. **40**, s245-s249 (2007).