

Martha Brennich, Petra Pernot, and Adam Round

---

## Abstract

SAS is a powerful technique to investigate oligomeric state and domain organization of macromolecules, e.g. proteins and nucleic acids, under physiological, functional and even time resolved conditions. However, reconstructing three dimensional structures from SAS data is inherently ambiguous, as no information about orientation and phase is available. In addition experimental artifacts such as radiation damage, concentration effects and incorrect background subtraction can hinder the interpretation of even lead to wrong results. In this chapter, explanations on how to analyze data and how to assess and minimize the influence of experimental artifacts on the data. Furthermore, guidelines on how to present the resulting data and models to demonstrate the data supports the conclusion being made and that it is not biased by artifacts, will be given.

---

## Keywords

SAS • X-rays • Neutrons • Contrast • Solution scattering • SEC • BioSAS

---

## 4.1 Introduction

SAS is not a new technique, the first experiments date back to the 1930s (Guinier 1938), and the technique has been applied to biological

macromolecules early on (Hosemann 1939). In recent years the combination of advances in sample production, high flux (X-ray and neutron) sources with rapid access to automated systems and advanced modeling (taking advantage of

---

M. Brennich (✉)

European Molecular Biology Laboratory, Grenoble, 71  
avenue des Martyrs, 38042 Grenoble, France  
e-mail: [rejma@esrf.fr](mailto:rejma@esrf.fr)

P. Pernot

Structural Biology Group, European Synchrotron  
Radiation Facility, Grenoble, 71 avenue des Martyrs,  
38042 Grenoble, France

---

A. Round

SPB/SFX instrument, European XFEL GmbH,  
Holzkoppel 4, 22869 Schenefeld, Germany

Faculty of Natural Sciences, Keele University, Keele,  
Staffordshire ST5 5BG, UK

modern computing) has made BioSAS a valuable tool for structural biologists.

SAS experiments on biological macromolecules in solution (BioSAS) using both neutrons (SANS) or X-rays (SAXS) provide information on the size and shape of the scattering object. Using Guinier's law (Guinier 1938) the Radius of gyration ( $R_g$ ), a measure of the overall size, can be determined together with the forward scattering intensity ( $I_0$ ), which is proportional to the molecular mass and the macromolecular concentration. Additionally the hydrated volume of the scatterer can be determined using Porod's law (Porod 1982) and the maximum dimension ( $D_{max}$ ) within the scatterer can be estimated through the process of the inverse Fourier transformation (Glatter 1977; Svergun 1992).

As proteins in solution are mobile, all orientations are possible, and the SAS signal only contains orientation-averaged information. Combined with the intrinsic lack of phase information (only intensities can be measured), direct shape reconstruction by inverse Fourier transform is impossible and indirect shape reconstruction by model generation is by nature ambiguous.

Furthermore, the shape (form factor) of the particles of interest, scattering events between particles (structure factor) as well as scattering of the buffer, the sample holder and parasitic scattering of the instrument used contribute to the measured signal. Thus the first step required in data processing is data reduction from a raw 2D pattern to an idealized, artifact-free scattering curve representing the investigated particle only.

In order to aid those wishing to exploit BioSAS experiments this chapter covers the necessary data analysis steps for processing and interpretation of SAS data as well as instructions on how to present data for publication.

## 4.2 Data Reduction with Examples of Common Pitfalls

Data reduction is not a separate part of the experiment which starts once the data acquisition is complete, but an integral part of the data collection as preliminary results of data reduction and analysis provide valuable feedback on data quality. A typical BioSAXS experimental set-up is presented in Fig. 4.1 (Pernot et al. 2013). Many

**Fig. 4.1** Experimental setup at the ESRF BioSAXS beamline BM29. This experimental facility is dedicated to SAXS measurements of samples in solution offering both Static (batch) operation and online SEC measurements (both HPLC and FPLC). X-ray scattering images are acquired using a Pilatus 1 M detector 1 Air scattering is avoided by using an evacuated flight tube 2 A touch screen monitor 3 allows easy control of the dedicated sample changer 4 Inset photograph shows the sample changer from the top with sample storage opened



BioSAS instruments, have adopted automated approaches to data reduction as well as preliminary analysis (Brennich et al. 2016; Franke et al. 2012). These tools provide background corrected scattering curves and the useful invariants ( $R_g$ ,  $I_0$ , Porod Volume and  $D_{\max}$ ), which give valuable feedback regarding the sample behavior and data quality (Figs. 4.2, 4.3, and 4.4).

## 4.2.1 Azimuthal Integration

At most modern BioSAS facilities, the scattering signal is detected with area detectors in order to detect as many of the scattered photons or neutrons as possible, resulting in 2D scattering images. As in general the scattering of randomly oriented particles in solution is isotropic, these images can be reduced to 1D curves without any loss of information by azimuthal integration (Fig. 4.2a–c). The following information is necessary:

- Type of detector used (its pixel size and geometry)
- Sample-to-detector distance
- Photon or neutron energy
- Direct beam coordinates on the detector
- “integration mask” which lists all pixels to ignore in the integration process (e.g. those hidden by the beamstop, etc.)

A variety of azimuthal integration software packages exist for different operating systems and data formats (Ashiotis et al. 2015; Benecke et al. 2014; Rodriguez-Navarro 2006; Hammersley 1997). The reader can find a listing of commonly used software at <http://smallangle.org/content/Software#Reduction-Visualisation>. Integrators suited for SANS data also take the distortion of the 2D data due to gravity into account. At most neutron and X-ray facilities suitable (sometimes even automated) software and guidance on how to use it are available.

The results of azimuthal integration are (in some cases already normalized) intensity and its standard deviation versus the scattering vector. Conventions on the scattering vector

differ, and it is important to note its units ( $\text{nm}^{-1}$  or  $\text{\AA}^{-1}$ ) and whether the scattering vector is equal to  $4\pi \sin \theta/\lambda$  or  $2 \sin \theta/\lambda$ . When comparing data from different instruments, it might be necessary to convert them according to conventions used.

Artifacts appearing after azimuthal integration can be caused by many factors, such as incorrect masking (typically concerns pixels close to the direct beam), integrating anisotropic patterns, ‘crazy’ pixels (Fig. 4.2d) etc. All such factors, which affect the data, must be identified and corrected before further data processing.

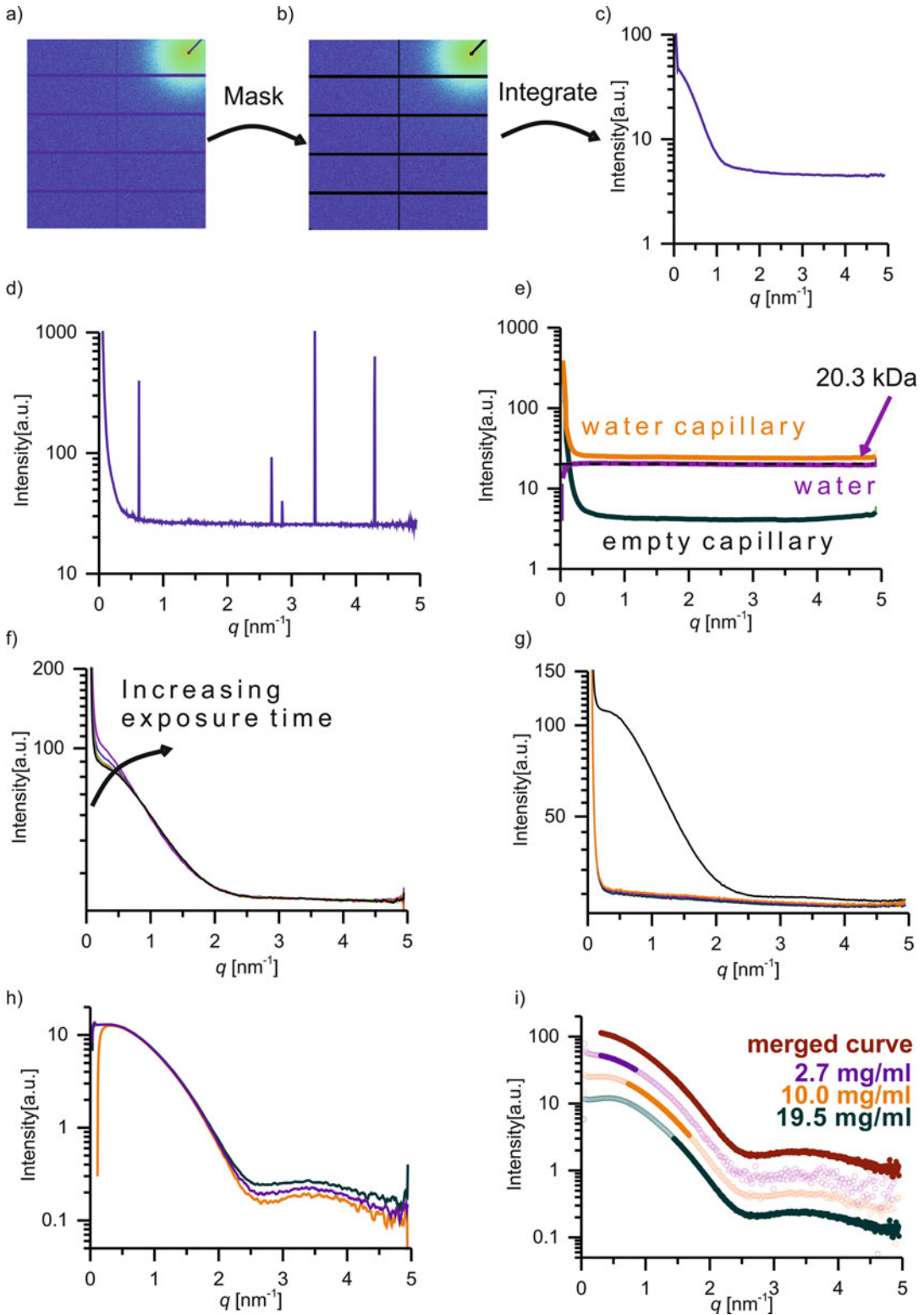
## 4.2.2 Normalization

The normalization of intensities from set-up dependent arbitrary units ( $I_{arb}$ ) to absolute units ( $I_{abs}$ ) by multiplication with a calibration factor is necessary in order to calculate the sample mass from the forward scattering and to correctly compare between different setups. This calibration factor can be determined by measuring the SAXS signal  $I_{st}(q)$  of a calibration standard, such as glassy carbon or water, and the conversion to absolute units is given by

$$I_{abs}(q) = \left( \frac{\partial \Sigma}{\partial \Omega} \right) (q) \frac{T_{st} d_{st} I_{arb}(q)}{T_s d_s I_{st}(q)}$$

where  $\left( \frac{\partial \Sigma}{\partial \Omega} \right) (q)$  is the known scattering intensity of the standard,  $T_{st}$  and  $T_s$  are the X-ray transmissions and  $d_{st}$  and  $d_s$  the thickness (X-ray path length) of the standard and the sample, respectively (Brian Richard 2013).

Generally, the  $q$ -dependence of the calibration factor is negligibly small. Therefore it can also be determined from the forward scattering of well-behaved proteins, such as  $\beta$ -amylase from sweet potato. In case of water the flat scattering intensity of  $0.01632 \text{ cm}^{-1}$  at  $20^\circ \text{C}$  and atmospheric pressure, can be measured at higher angles (Orthaber et al. 2000) as shown in Fig. 4.2e. For protein samples, it can be useful to provide the scattering intensity in units of kDa instead of  $\text{cm}^{-1}$ . The constant scattering intensity of water at  $20^\circ \text{C}$  when scaled for a standard protein



**Fig. 4.2** Data reduction from images to an idealized curve. (a) A raw image from a 2D detector. (b) The blacked out regions, corresponding to gaps between

detector modules, the beamstop, strong parasitic scattering, hot pixels, etc. are neglected (“masked”) in further processing. (c) Azimuthal integration around the beam

in buffer corresponds to 20.3 kDa (Mylonas and Svergun 2007). Lipids, nucleic acids and protein complexes due to the difference in contrast require modification to this scaling factor, which is dependent on the ratio of protein, lipid and nucleic acid in the investigated particle (see “Guinier approximation” Sect. 3.1 for more details). In practice, different facilities use different calibration methods and one should verify the method and units of the intensity normalization.

### 4.2.3 Averaging of Multiple Frames from Each Sample

Data acquisition is usually split up into several sub-exposures in order to detect (possible) sample degradation corrupting the signal. Especially when using X-rays, radiation often induces sample degradation (radiation damage). To minimize the effect of radiation damage, the sample is typically moved during data acquisition, with the aim that every sub-exposure can be taken on fresh sample. If the sample suffers radiation damage, contains air bubbles or is inhomogeneous, these sub-exposures will not give identical scattering patterns. Therefore, it is necessary to control for outliers. In many cases, this can be done by a qualitative visual control, but in more subtle cases statistical tests such as CORMAP from the ATSAS package need to be used (Franke et al. 2015; Petoukhov et al. 2007). Figure 4.2f shows a case of obvious radiation damage, with at least two frames showing clear radiation induced

aggregation. More careful analysis with CORMAP revealed that first systematic changes already occurred as early as in the fifth frame. Only artifact free sub-exposures should be averaged for a better signal to noise ratio and used for subsequent data processing.

### 4.2.4 Background Subtraction for Individual Concentrations

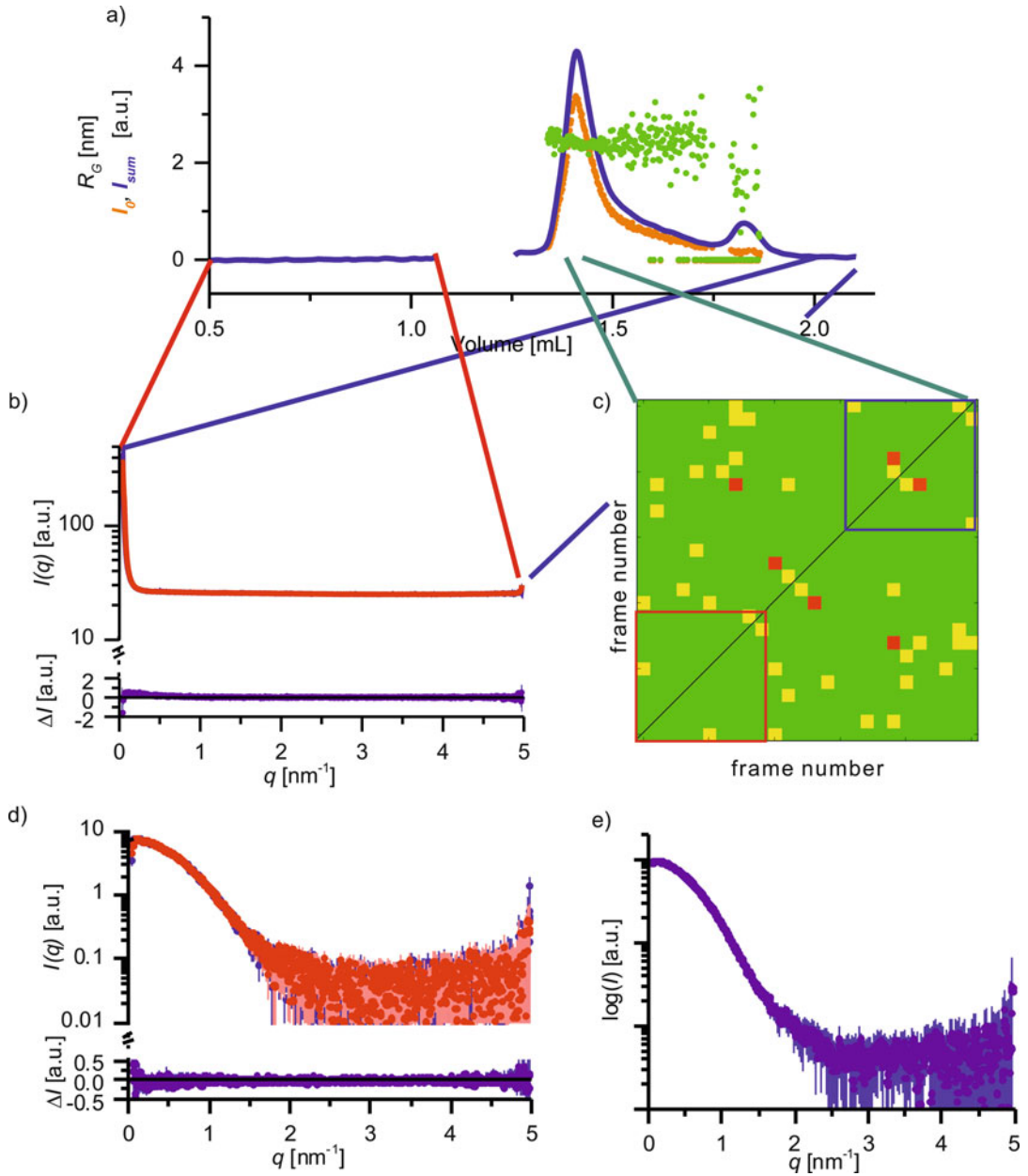
A SAS curve free of contributions of the buffer and set-up can be obtained by simply subtraction of a corresponding buffer measurement from the averaged sample curve. Ideally, the buffer should have been measured in the same sample environment and exposure parameters before and after the sample measurement. If the scattering from these two measurements is not identical (as tested for by e.g. CORMAP (Franke et al. 2015)), the reason has to be investigated (for example inadequate post sample cleaning of the exposure cell, inhomogeneities in the buffer, etc.). In the ideal case both buffer measurements will match and can be averaged (to improve signal to noise ratio) and subtracted from the scattering curve of the sample. In addition, the resulting subtracted curve is often normalized (divided) to the sample concentration (if known).

At this stage, it is necessary to check whether the buffer measurement matches the sample (Jacques and Trewhella 2010). Differences in the chemical composition of buffer and sample affect the transmission of X-rays and thereby the



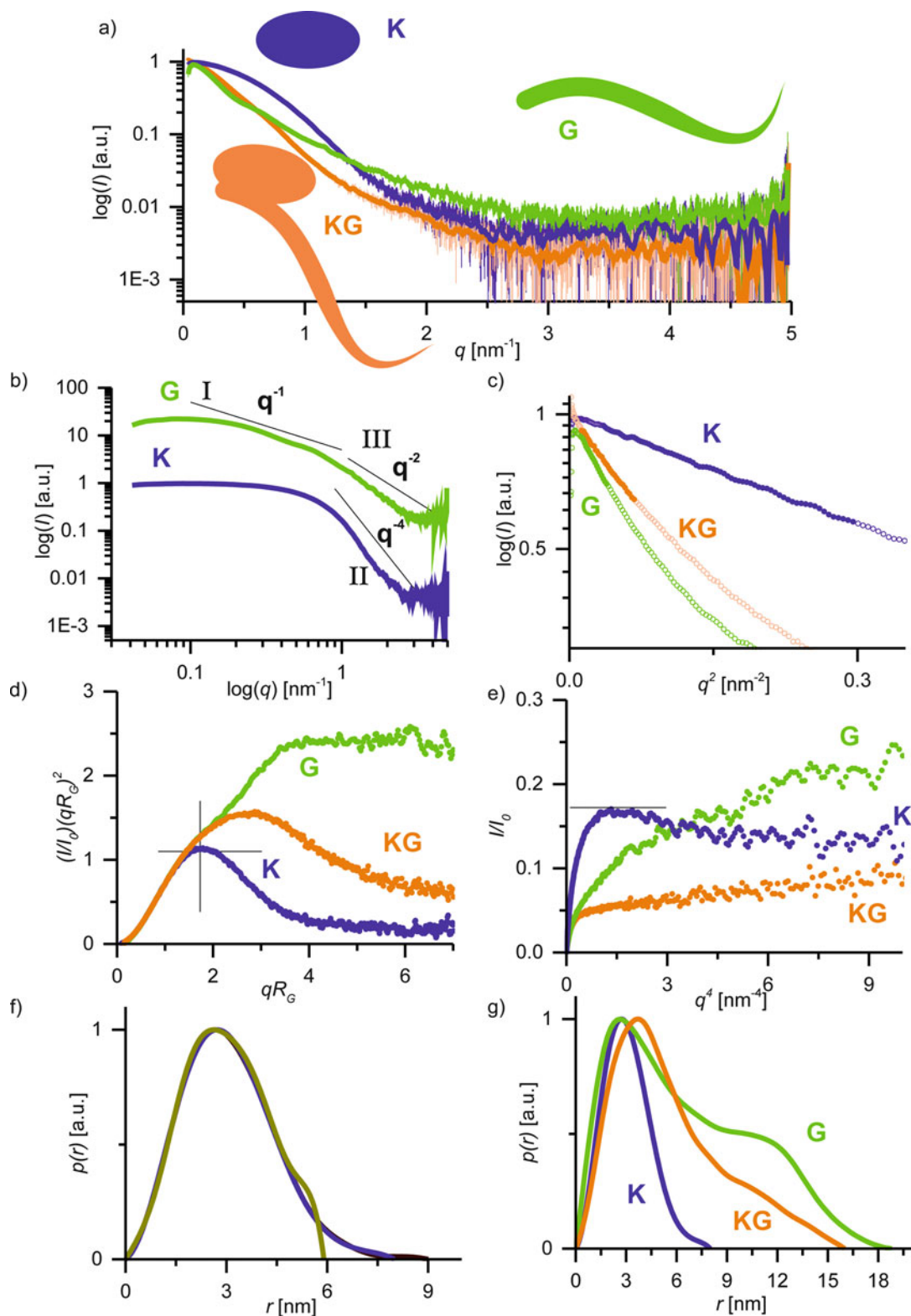
**Fig. 4.2** (continued) center provides the 1D SAXS curve. (d) Failure to mask hot pixels results in characteristic “spikes” in the 1D curve. (e) Calibration to absolute units is performed by subtraction of a measurement of an empty capillary (*green*) from that of a water filled capillary (*orange*). The resulting constant signal (*violet*) corresponds to 20.3 kDa or  $0.0163 \text{ cm}^{-1}$  (when correctly scaled). (f) Radiation damage usually results in continuous, systematic changes in signal. Sub-exposures need to be controlled for its onset and affected frames are discarded. (g, h) Subtracting a buffer containing too

much (*orange*) or too little (*green*) salt mostly affects the high  $q$  and very low  $q$  regions, as seen in comparison to the ideal subtraction (*violet*). Artefacts like that of the orange curve at very small  $q$  are clear warnings. (i) To eliminate concentration artefacts while maintaining a good signal to noise ratio, data from different concentrations can be combined. The ideal curve (*red*) was constructed using the filled symbols of the three shown concentrations, while the data corresponding to the open symbols was only used for consistency checks



**Fig. 4.3** Data reduction for online SEC-SAXS. (a) SEC-SAXS chromatogram presents the total scattering intensity (blue) for all frames of the peak and forward scattering (orange) and radius of gyration (green) for all sample frames. (b) Comparison of the buffer collected before (red, upper part) and after (blue, upper part) the sample peak. Their difference (violet, lower part) confirms that the buffer signal stays constant. (c) Results of the DATCMP tests for frames in the region of interest. (d) Comparison of frames collected at the beginning of the sample peak (red, upper part) and the end of the sample peak (blue, upper part). Their difference (violet, lower part) confirms that the signal stays constant throughout the peak. (e) Final, averaged SAXS curve

Green squares stand for matching frames ( $p \geq 0.9$ ), red squares correspond to non-matching frames ( $p \leq 0.1$ ), and yellow squares represents all other cases. The blue and red lines limit the regions compared in Fig. 4.3d). (d) Comparison of frames collected at the beginning of the sample peak (red, upper part) and the end of the sample peak (blue, upper part). Their difference (violet, lower part) confirms that the signal stays constant throughout the peak. (e) Final, averaged SAXS curve



**Fig. 4.4** Primary data analysis of a complex. (a) Background corrected SAXS curves of the proteins K, G and their complex KG in semi-logarithmic representation.

The curve corresponding to K is flatter at small angles than the curves of G and KG, indicating its smaller size. The curve corresponding to G decays more slowly than K

scaling of the scattering curve. In many cases, the sample contributes only little to the scattering at high angles (scattering vector above  $4 \text{ nm}^{-1}$ ), so mismatches can often be identified in this region. Indicators for buffer mismatch are non-matching scattering curves at higher angles, systematically negative regions in the subtracted curve, deviations from Porod's law or differences at high angles between different sample concentrations. However, some of these indicators can also occur for small scatterers at high concentrations and non-globular samples such as intrinsically disordered proteins (IDPs) (Bernado and Svergun 2012).

Figure 4.2g, h show how subtraction of the wrong buffer affects the scattering curve. For both over-subtraction (orange) and under-subtraction (green) the scattering at high angles deviates from the matching case (magenta). In the case of the buffer over-subtraction, the difference in the relative contribution of the capillary scattering additionally results in a sharp downturn of the curve at small angles, which can even affect the determination of the radius of gyration.

#### 4.2.5 Merging of Different Concentrations, Extrapolation to Zero-Concentration

BioSAS experiments are almost exclusively carried out with the aim to determine the form factor of scattering particles. However, only the combination of form factor and structure factor is experimentally accessible. Conveniently, the structure factor depends, in contrast to the form factor, on the particle concentration and becomes negligible at sufficiently low concentrations where inter-particle distances are sufficiently large to prevent interactions (Bonneté et al. 1999). Therefore, the effect of the structure factor in BioSAS measurements can be minimized by measuring samples in dilute conditions and crosschecking at multiple concentrations. In addition, other concentration dependent artifacts such as aggregation also diminish at lower concentration.

A concentration which can be assumed to be free of interparticle effects needs to fulfill the following criteria:

**Fig. 4.4** (continued) and KG at high angles, suggesting a higher degree of disorder present. **(b)** SAXS curves of the proteins K and G in double-logarithmic representation. G curve follows a  $q^{-1}$  trend at small angles (region I) which is a signature of an elongated form. At high angles, G curves follows a  $q^{-2}$  trend (region III), indication of disordered regions. In contrast K curve follows a  $q^{-4}$  trend (region II) as expected in the case of a globular protein. **(c)** Guinier plots of proteins K and G and KG complex, with the Guinier region indicated by closed symbols. The small protein K has the longest Guinier region. Increased radius of gyration and anisotropy decrease the length of the Guinier region as observed for protein G. **(d)** Normalized Kratky plot of proteins K and G and KG complex. The symmetric peak of protein K is found at the position predicted for globular proteins. In the case of the KG complex the peak is shifted upwards and right indicating an anisotropic shape. The decay to zero is flatter than for K, suggesting presence of disordered regions. Due to the fact that the area under the peak is not clearly defined, the Porod volume is not reliable. The curve corresponding to protein G has no peak at all

but only a plateau, a signature of a high degree of flexibility. The strong up-right shift additionally indicates anisotropy. **(e)** Porod plot of proteins K and G and KG complex. The curve corresponding to globular protein K displays a clear plateau, whereas the signal for the more flexible complex KG and protein G continuously increases. **(f)** Pair distribution  $p(r)$  curve of protein K determined using different values of  $D_{max}$ . A too small  $D_{max}$  value results in an abrupt turn of the curve to 0 (*dark yellow curve*). A too large  $D_{max}$  (*red curve*) value used for  $p(r)$  calculation, results in a long trailing tail close to 0. The curve calculated with  $D_{max}$  chosen for further analysis (*dashed blue curve*) approaches 0 gently. **(g)**  $p(r)$  curves for proteins K and G and complex KG. The  $p(r)$  the small globular K is nearly symmetric. In the case of the more anisotropic KG complex and protein G, the position of the main peak barely moves but the curves becomes more asymmetric and the  $D_{max}$  increases. The additional peak at about 12 nm in the curve for G indicates the presence of spatially separated sub-domains



- The signal (subtracted curve) is identical for lower concentrations
- The data at low angles fulfil the Guinier approximation well
- The radii of gyration determined *via* the Guinier approximation and *via* the pair distribution function respectively match each other.

If the signal to noise ratio at this concentration is sufficiently good even at higher angles, it can be used for all further analysis. Otherwise, it is necessary to include data from higher concentrations. For this, one first identifies the point from which on differences between the concentrations are only due to noise, then scales the concentrations to each other in an overlap region and takes the lower angle data from the lower concentration and the higher angle data from the higher concentration. Figure 4.2i presents the creation of an idealized curve based on three concentrations (2.7 mg/ml, 10 mg/ml and 19.5 mg/ml). The regions of each curve used for building the idealized curve are represented by the filled symbols, whereas data represented by open symbols was not included. In the experiment even lower concentration data was collected, however no significant difference to the data at concentration equal to 2.7 mg/ml was found. All initial data points that do not follow Guinier's law are removed for the creation of the idealized curve. At 10 mg/ml and 19.5 mg/ml one can observe strong contributions from inter-particle scattering, which have to be removed before further analysis. The affected regions can be identified by comparison to the next lower concentration. Points at low angles that show significant differences to the data collected at lower concentrations are ignored. The higher  $q$  end of the regions that contribute to the idealized curve follows directly from the lower  $q$  end of the next higher region and extends just a few data points beyond it. Alternatively, many SAS packages provide routines which allow automatic extrapolation of all measured concentrations to an idealized “concentration zero” curve (Franke et al. 2012).

The result of these approaches is an idealized curve, which should be free from any inter-

particle artifacts. This idealized curve is used for further analysis and modeling but the degree of variation observed in the different concentrations should be kept in mind with regards to the confidence in the interpretation.

#### 4.2.6 Background Subtraction and Averaging for SEC-SAXS Experiments

A SEC-SAXS experiment typically consists of several hundreds to thousand acquired frames either continuously throughout the elution process or only in regions of interest, e.g. buffer before the elution of the sample from the column, the sample itself and buffer again after the elution of the sample (Watanabe and Inoko 2009; Round et al. 2013; Mathew et al. 2004; Lambright et al. 2013; Grant et al. 2011; Graewert et al. 2015; David and Pérez 2009).

Before further processing, the stability of the buffer baseline needs to be confirmed by comparison of individual measurements. The following effects can negatively affect the baseline:

- Column not completely equilibrated
- Slow drift in the experimental setup
- Sample eluting in column void volume
- Mismatch between running buffer and sample buffer
- Spoiling of the sample environment by additives (or contaminants) in the buffer (typically radiation damage of the buffer and its deposition on walls of the sample exposure cell)
- Spoiling of the sample environment by the sample

Figure 4.3a shows a typical SEC-SAXS chromatogram, providing the total SAXS intensity for each frame as well as the radius of gyration and forward scattering from later processing steps. Between 1.1 and 1.25 mL the shutter was closed to avoid spoiling from the aggregate peak. Between 1.75 and 1.9 mL the total scattering increases due to excess salt injected with the sample (salt peak). When choosing a suitable

buffer, one therefore needs to avoid these two regions. Comparing the average of the buffer frames acquired before the aggregate peak (Fig. 4.3b) shows that there is a slight mismatch at small angles, indicative of mild capillary spoiling.

If the buffer signals are matching, measurements can be averaged and subtracted from the individual sample measurements. In some cases, small changes can be interpolated to provide a suitable buffer subtraction for each sample measurement (Brookes et al. 2013). For each individual subtraction, one can then calculate the radius of gyration and forward scattering. If the sample concentration has been measured simultaneously or with a known delay, this information can be combined to estimate the molecular weight. In our example, due to the slight spoiling, we subtracted the buffer recorded before the aggregate peak and obtained the forward scattering and radius of gyration shown in Fig. 4.3a.

Although the aim of performing online size exclusion chromatography (SEC) on mixtures is to separate the different species before collecting SAS data, sometimes the peaks will elute too close to one another. This can lead to overlapping peaks and in these regions the data measured will represent the mixed scattering from the overlapping species, with the proportions contributed to the total observed scattering by each species changing with time. It may still be possible to find regions with only one species, corresponding measurements can be merged together to give the scattering for that species. To identify such regions, one first finds sufficiently large (in the range of one injection volume) regions of stable  $R_g$ . When such a region is identified, one verifies that the individual (suitably scaled) SAS curves match each other. In our example, a region of 25 frames was identified as potentially stable. To confirm this hypothesis, the subtracted SAXS curves were scaled and a pairwise comparison using CORMAP was performed. The results of the test are visualized in Fig. 4.3c, displaying clearly matching frames ( $p \geq 0.9$ ) as green squares, non-matching frames ( $p \leq 0.1$ )

as red squares and all others as yellow squares. Out of the 300 individual tests in this case, only two gave  $p$ -values smaller than 0.1. In the case of low concentration or high noise, it is further advisable, to compare the averages of different sub-regions. If all these tests confirm a stable signal, corresponding measurements can be averaged and used for further processing. A (rare) special case of this scenario appears when protein concentrations are high enough for inter-particle effects to cause a decrease of scattering at low angles. An approach similar to the one described in part 2.5 can be used to combine data from different parts of the chromatogram.

If no stable signal can be found, direct merging of the data is not valid, as the underlying hypothesis of homogeneity and purity does not hold for a mixture of species. Deconvolution using the assumption of overlapping Gaussian peaks can in some cases recover the scattering from the individual species (Brookes et al. 2013). However, it is recommended where possible to re-measure the sample using a better resolving column to separate the peaks experimentally.

---

## 4.3 1D Curve Analysis

### 4.3.1 Calculation of Model Independent Parameters

#### 4.3.1.1 Initial Assumptions

Interpretation of data from a SAS experiment gives average parameters of the scattering particles. Each model independent parameter provides a single number which is less informative if the sample is not monodisperse, i.e. a single oligomeric species in the same conformation. For mixtures, deconvolution of data to obtain individual curves for the constituents is only possible in special cases (Karlsen et al. 2015) and will not be treated here. Thus not only is validation required in sample preparation but cross checking the expected values with those observed using SAS is essential to avoid misinterpretation.

### 4.3.1.2 Qualitative Analysis of SAS Curves

Even without any quantitative analysis it is often possible to extract information from SAS curves based on their shape. Figure 4.4a shows the SAXS curves of the proteins K, G and their complex KG, scaled such that their forward scattering matches. Looking at the very small angles, it is obvious that the curve corresponding to G is considerably steeper than those of K and KG, implying that the radius of gyration of this component is actually larger than that of the complex it forms with K. Double logarithmic representation of the results (Fig. 4.4b) highlights some more features: At low angles K flattens off very early whereas as G follows a  $q^{-1}$  power law before leveling off (region I in Fig. 4.4b). This  $q^{-1}$  behaviour is typical for elongated, rod-like particles (Glatter and Kratky 1982). At high angles, K follows a  $q^{-4}$  power law, as expected for well-folded globular proteins (region II in Fig. 4.4b) (Porod 1982), whereas G only decreases as  $q^{-2}$ , indicating at least some extent of flexibility (region III in Fig. 4.4b) (Reyes et al. 2014; Debye 1947). Hence, even without any advanced analysis, one identifies K as a small globular protein and G as an elongated protein, with at least some highly flexible regions. It can be also noted that their complex is less anisotropic than G, as it seems to have a lower radius of gyration than G protein alone.

### 4.3.1.3 Guinier Approximation

The SAS signal of any scatterer at small angles can be described by a Gaussian distribution (Guinier 1938)

$$I(q) = I_0 e^{-\frac{(qR_g)^2}{3}}$$

This allows determination of the forward scattering  $I_0$  as well as the (average) radius of gyration of the scatterer. In the case of a mixture of similarly sized scatterers,  $I_0 = \sum_n f_n I_{0n}$  and  $R_g^2 = \sum_n f_n R_{gn}^2$ , where  $f_n$  is the fraction,  $I_{0n}$  the forward scattering and  $R_{gn}$  the radius of gyration of the  $n$ th component, respectively (Segel et al. 1999). The Guinier approximation is only valid

for small angles, and therefore when fitting  $q \leq 1.3$  needs to be fulfilled for the fit region.

The forward scattering  $I_0$  is proportional to the number of scatterers, the square of their mass and electron density (contrast) compared to the surrounding solvent, and more practically to their concentration (in mass/volume). Hence, if the concentration and chemical composition of the sample and buffer is known (including Hydrogen and Deuterium ratio for neutron scattering), it is possible to estimate its mass, and thereby its oligomeric state (assuming it is monodisperse), directly from the Guinier approximation. In the case of X-rays the proportionality factor for protein in water is  $1.3 \cdot 10^3 \text{ cm}^3/\text{kDa}$ , while for nucleic acids (DNA and RNA) it is  $2.6 \cdot 10^3 \text{ cm}^3/\text{kDa}$  due to their higher electron density and thus contrast  $\Delta\rho$ . For complexes it can be calculated as  $N_A/(\Delta\rho v)^2$ , where is  $N_A$  Avogadro's number,  $\Delta\rho$  the contrast and the  $v$  partial specific volume.

The quality of the Guinier approximation is best examined in the Guinier plot,  $\log I$  vs  $q^2$ . A concave curve in this plot indicates the presence of larger scatterers, often aggregates, while a convex curve indicates repulsion between the scatterers.

If concentration-corrected data are scaled to kDa, the forward scattering is identical to the mass of the scatterer. However, some particles may have an inherently high degree of conformational flexibility. An important consequence of the resulting structural heterogeneity is that the movement of the subunits in relation to each other will not be synchronized across all particles in the X-ray beam. Moreover, it can be assumed that all possible relative positions and orientations will be sampled in the scattering data under the assumption of spherical averaging (all possible orientations are present). This gives rise to an increase in the average size of the scatterers and, moreover, to variation in the particle sizes. These effects cause a deviation from the linear expectation of Guinier's law and, as such, are practically indistinguishable in the 1D data from a small amount of aggregation. This artifact is unlikely to depend on concentration in

the dilute concentrations used for SAS experiments. However, in some cases at high concentrations ( $>10$  mg/mL), nearby particles can affect the flexibility (crowding effects), and a concentration dependence can be observed. A convex curve indicated the presence of inter-particle effects (“structure factor”), which typically show a strong concentration dependence and become negligible at sufficiently low concentrations.

Generally, points at very small angles will be ignored for the Guinier analysis. For further analysis, these points should be removed from the curve as they provide no additional information and are prone to be affected by artifacts.

Coming back to our example (Fig. 4.4c), G has the highest radius of gyration and K the lowest, KG one being a bit smaller than that of G. Accordingly, K has the longest Guinier region, going up to over  $0.5 \text{ nm}^{-1}$ . For G, the Guinier region is limited not only by its larger size, but also by its high degree of anisotropy.

For highly anisotropic particles, such as rods or disc additional forms of the Guinier approximation exist. The most relevant for biological macromolecules is the Guinier-approximation for rods, i.e. for particles whose long axis  $L$  is much longer than its cross-sectional diameter (Glatter and Kratky 1982):

$$I(q) = \frac{I_0}{q} e^{(qR_c)^2/2} \text{ for } qL \gg 1 \text{ and } qR_c \leq 1.1.$$

Analogously to globular particles, it can be used to derive the mass-per-length  $M_L$  of a rod and the cross-sectional radius of gyration  $R_c$ . If the available  $q$ -range is large enough to determine both  $R_g$  and  $R_c$ , of a macromolecule, its length can be estimated *via*

$$R_g^2 = R_c^2 + \frac{L^2}{12}$$

#### 4.3.1.4 Qualitative Flexibility Analysis

In contrast to the small angle region that only depends on the particles overall size, the high angle region, which corresponds to small distances in real space, corresponds to the

flexibility. In this region, unfolded or dis-ordered proteins scatter more strongly than globular proteins of the same size.

These differences can be clearly seen when the data is plotted appropriately: In the normalized Kratky plot  $(qR_g)^2 I(q)/I_0$  vs.  $qR_g$ . In this representation, globular proteins display a parabolic peak at  $\sqrt{3}$ . Any anisotropy will move this peak to higher values. On contrast, the signal of completely unfolded proteins will continuously increase in this representation. Flexible proteins can be found between these two extremes (Hammel 2012; Rambo and Tainer 2011).

In the K, G and KG system, the normalized Kratky plot of K displays a symmetric peak exactly at the predicted position, indicating that K is a very globular, well-folded protein with little or no anisotropy (Fig. 4.4d). In contrast, the Kratky plot of G continues to increase well beyond this point, until it finally levels off; a shape corresponding to an elongated, flexible protein. The Kratky plot of the complex KG is found between these two extremes, with its peak shifted towards higher  $qR_g$  values, indicating some anisotropy, and its decrease being slower than for K, indicating some remainder of flexibility.

In the Porod-Debye plot,  $q^4 I(q)$  plotted vs  $q^4$ , the signal of non-flexible, globular proteins level at a plateau while the signal of unfolded proteins continue to increase linearly. Flexible proteins display a decrease in slope, but do not level off completely.

In Fig. 4.4e K shows a typical Porod-plateau, while the more flexible KG complex just levels off. The very flexible G on its own continues to increase.

In addition to being helpful for determining whether a protein is flexible, these representations also can highlight problems with the buffer subtraction due to their emphasis on the higher angles.

The interested reader can find excellent discussions on flexibility assessment in (Hammel 2012; Rambo and Tainer 2011).

### 4.3.1.5 Indirect Fourier Transform

In most cases, the calculation of pair distance distribution function  $p(r)$  of a SAS curve is ambiguous, as there are several free parameters (Svergun 1992; Semenyuk and Svergun 1991):

- The region of the curve used for its calculation: Typically, its lower limit is given by the Guinier approximation, whereas the upper limit depends on what the  $p(r)$  is calculated for: For bead modeling with DAMMIF, DENFERT etc. normally  $8/R_g$  is sufficient, while for GASBOR modeling the range should extend to at least  $3.5 \text{ nm}^{-1}$ . In addition, most algorithms work best when the cut-off lies in a region of decreasing intensity.
- The maximum distance  $D_{max}$ . A good starting point for  $D_{max}$  is the value of  $3R_g$ . From there, it needs to be adjusted in such a way that the resulting distribution approaches zero non-abruptly without strong oscillations, a trailing tail or even negative data points. In addition, a wrong  $D_{max}$  results in a  $p(r)$  based radius of gyration that deviates significantly from the one determined *via* the Guinier approximation.
- The smoothing factor. It needs to be decreased if the fit no longer matches the data and increased if either the  $p(r)$  function of the fit show oscillations.

In the case of flexible proteins, it can be difficult to find a suitable  $p(r)$  function and to determine  $D_{max}$ . Similar problems are observed if the background is not well corrected.

Figure 4.4f shows how an incorrect  $D_{max}$  affects the  $p(r)$  function. A too small value results in a sharp down-turn of the curve and often a near perpendicular approach towards 0 (green curve), while a too large value results in a long and extended tail (blue curve). Figure 4.4g shows how different particle shapes affect the  $p(r)$  function: For the small globular protein K (blue curve), it has a sharp, nearly symmetric peak. The more anisotropic shape of the KG complex (orange curve) is reflected in the asymmetric shape of the peak, with a slower

descent towards larger distances. For G alone (green), the asymmetric shape of the peak is even more pronounced. The shoulder at larger distances indicates the possibility of two (or more) separated domains.

For elongated, rod-like objects, the cross-sectional pair-distribution function  $p_c(r)$  can be determined instead.

### 4.3.1.6 Porod Analysis

For globular scatterers, another helpful SAS invariant is the Porod invariant

$$Q = \int_0^{\infty} I(q)q^2 dq = 2\pi I_0/v_p$$

which allows to determine the Porod volume  $V_p$  of the scatterer (Porod 1982). In both cases of proteins and nucleic acids, the Porod volume (in units of  $\text{nm}^3$ ) corresponds to about 1.5–2 times the molecular weight (in kDa) (Petoukhov et al. 2007). This determination does not depend on the absolute scaling of the SAS curve, therefore this method for mass determination can be applied even if the concentration of the sample or/and the absolute scaling of the data are unknown.

Note that as the above integral only converges for mostly globular proteins, the Porod volume tends to deviate strongly from volume expected based on the mass if the sample is highly flexible or even disordered. This can be easily understood when one considers that  $Q$  matches the volume under the peak in the Kratky plot, which is only finite for globular objects.

For calculating the Porod volume, one needs to extrapolate the data to infinity by fitting the higher angle data around the Porod-Plateau in the Porod-Debye plot by a power law, whose exponent (the *Porod* exponent) needs to be smaller than  $-3$ . Based in this extrapolation,  $Q$  and thereby the volume can be easily calculated.

In this example, the Porod volume of the flexible protein G is not well defined, as its SAXS curve only decreases as  $q^{-2}$  (as noticed in its Kratky plot of Fig. 4.4d). On the other hand,

the globular protein K shows a well-defined Porod plateau (Fig. 4.4e), which permits to determine its Porod volume. In the case of the KG complex, the power law fit of the higher  $q$  region gives a Porod exponent of  $-3$ , prohibiting the determination of the Porod volume.

Some algorithms for indirect Fourier transform provide the Porod volume as an additional result. It should also be noted, that most software tools will provide a result for the Porod volume, even when the conditions for determination are not met. These results are typically not related to the actual volume of the macromolecule in question.

#### 4.3.1.7 Correlated Volume

Another approach for estimating the mass of non-flexible macro-molecules is provided by the so-called volume-of-correlation (Rambo and Tainer 2013) given by

$$V_c = \frac{I_0}{\int_0^\infty I(q)q dq}$$

The advantage of this approach is that the above integral usually converges well in the available data range and no extrapolation of the data is necessary. Additionally, it converges as long as the Porod exponent is smaller than  $-2$ . In our example, this implies that while the Porod volume of the KG complex is not well defined, the correlated volume can be determined.

In this case, the mass (in kDa) of proteins is roughly equal to  $8V_c^2/R_g$  and of RNA to  $(107V_c^2/R_g)^{0.8}$  if the scattering vector is provided in  $\text{nm}^{-1}$ . For DNA no data are available to our knowledge.

#### 4.3.2 Comparison to Predicted Scattering Curves of Atomistic Models

The calculation of theoretical scattering from known atomic coordinates is a direct problem with no inherent ambiguity i.e. one always obtains the same SAS curve for any set of known atomic coordinates. However,

macromolecules in solution are surrounded by a hydration shell that contributes to the SAS signal. Atomistic models generally do not account for these highly dynamic structures (Zhang et al. 2007). Different implementations are available to account for these water molecules and their effect on the SAS signal. Some programs use molecular dynamics to describe the hydration shell (Chen and Hub 2015; Knight and Hub 2015), but more often it is modeled as a fixed-width shell around the macromolecule. In order to perform comparisons with actual data, the parameters of the hydration shell are generally adjusted to provide the best fit between model and data (Svergun et al. 1995; Schneidman-Duhovny et al. 2010). Particular attention should be paid to the values used for the hydration layer when comparing the resulting fits of multiple structures to the same data to avoid misinterpretation.

---

## 4.4 Building and Interpretation of 3D Models

If no or incomplete structural information is available, it is often possible to build 3D models of the macromolecule based on the SAS data. Approaches to model building span from the construction of bead models to fully fledged molecular dynamics simulations. Here, we will only discuss a few of them and refer the readers to Chap. 7 for more details.

### 4.4.1 *Ab-initio* Modeling

*Ab-initio* modeling techniques allow construction of three dimensional bead models optimized to the SAS signal. This is also mathematically possible in the case of corrupted data (mixture, non-matching buffer, etc.), therefore the data needs to be validated before attempting a reconstruction.

Most algorithms assume a monodisperse system and uniform contrast (i.e. all beads are identical) (Svergun et al. 2001; Svergun 1999; Franke and Svergun 2009), but some specialized

programs can model oligomeric mixtures (Petoukhov et al. 2007), DNA-protein complexes (Petoukhov et al. 2007; Svergun 1999), hydration layers (Koutsioubas and Perez 2013) etc. Other programs can include known partial structures in the modeling (Petoukhov and Svergun n.d.).

As the reconstruction of three dimensional models from the one dimensional SAS curve is intrinsically ambiguous and the high number of free parameters makes stochastic approaches to model building necessary, it is essential to repeat the modeling several times and to compare the results. The similarity between two resulting models can be quantified by the normalized spatial discrepancy (NSD). If two models systematically differ from each other, their NSD exceeds 1, for identical objects, it is 0. Therefore, large NSD values are indicative of ambiguity in the modeling.

#### 4.4.2 Rigid-Body Modeling

In the case of complexes or multi-domain proteins, the structures of the individual components or homologues thereof are often known. The relative positions of the individual “rigid bodies” can be modeled to fit the SAS data (Petoukhov et al. 2007; Petoukhov and Svergun n.d.). Inclusion of additional constraints to conserve known connectivity between domains is also often required especially if the domain is symmetric. However, inclusion of any constraint will bias the model thus if the constraint is false, being based on an incorrect assumption the resulting model may still fit the data and lead to false conclusions. As for *ab-initio* modeling, the results are often ambiguous and repeated reconstructions are necessary. As for *ab initio* modeling, the robustness of the reconstruction can be assessed by calculating the NSD between the resulting models. If different rigid-body modeling approaches are applied,  $\chi^2$  can be used to identify which model describes the available data best. As the absolute value of  $\chi^2$  depends mostly on the signal quality it cannot be used to assess the absolute quality of the resulting fit.

## 4.5 Presentation of SAS Data for Publication

When presenting BioSAS data for publication, it is generally recommended to follow the IUCr guidelines (Jacques et al. 2012)

- Scattering data should be presented either in logarithmic intensity scaling or in double-logarithmic representation. Linear representation of the intensity hides key features of the curves and should be avoided.
- Guinier fits should be shown in the Guinier representation, showing a sufficient data range to evaluate the quality of the fit.
- If  $p(r)$  functions are used for further modelling, these must be shown. When comparing different functions, it is not uncommon to scale their respective maxima to 1.
- The (normalized) Kratky plot should be shown to allow assessment of flexibility.
- When presenting models, the variability of the results needs to be illustrated. In the case of bead models, this means that averaged and filtered models need to be shown.
- The instrument and conditions used for data acquisition as well as numeric primary processing results have to be presented (see an example in Table 4.1).
- Data and models should be made publicly available *via* an appropriate venue such as the SAS Biological Data Bank (SASBDB).

---

## 4.6 Conclusions

The analysis and interpretation of small angle scattering data rely on correctly collected and reduced data as well as on sample properties such as monodispersity or compactness.

This chapter describes the necessary steps of data reduction and the most common approaches to data analysis. While the majority of these steps can be automatized, it is still necessary to understand the underlying assumptions and possible sources of error so that one can verify the validity

**Table 4.1** Data-collection and scattering-derived parameters. The data in this table belongs to the data shown in Fig. 4.2i. Software listed in *italics* provides examples for its class, but was not used in this analysis

Data-collection parameters	
Instrument:	ESRF BM29
Beam geometry	0.7 mm × 0.7 mm
Wavelength (Å)	0.99
q-range (Å <sup>-1</sup> )	0.0032–0.49
Exposure time (sec)	1 per frame, 10 frames
Concentration range (mg/ml)	0.22–19.22
Temperature (K)	293
Structural parameters	
I <sub>0</sub> (cm <sup>-1</sup> ) [from P(r)]	0.01184
R <sub>g</sub> (Å <sup>-1</sup> ) [from P(r)]	15.6
I <sub>0</sub> (cm <sup>-1</sup> ) [from Guinier]	0.01152 ± 0.0004
R <sub>g</sub> (Å <sup>-1</sup> ) [from Guinier]	15.4 ± 2.5
D <sub>max</sub> (Å)	58
Porod volume V <sub>p</sub> (Å <sup>3</sup> )	22·10 <sup>3</sup>
Molecular-mass determination	
Partial specific volume (cm <sup>3</sup> g <sup>-1</sup> )	0.724
Contrast (Δρ × 10 <sup>12</sup> cm <sup>-2</sup> )	3.047
Molecular mass M <sub>r</sub> [from I <sub>0</sub> ]	14.4 ± 0.10
Calculated monomeric M <sub>r</sub> from sequence (kDa)	14.3
Software employed	
Primary data reduction	BM29 online data analysis, pyFAI, Primus
1D data processing	Primus, Gnom
<i>Ab initio</i> analysis	n.a. (e.g. <i>DAMMIF</i> )
Validation and averaging	n.a. (e.g. <i>SUPALM</i> , <i>DAMAVER</i> )
Rigid-body modeling	n.a. (e.g. <i>SASREF</i> )
Computation of model intensities	n.a. (e.g. <i>WAXSiS</i> )
Three-dimensional graphics representations	n.a. (e.g. <i>PyMOL</i> )

of results and does not need to blindly trust them. Any conclusions on the shape and behaviour of a sample in solution drawn from SAS data should take into account the quality and reliability of the data.

Data analysis should be viewed as part of an exploratory, interactive process designed to test hypotheses and learn more about the system under study. Cross checking with complementary information is a highly valuable part of the analysis procedure enabling any differences to be highlighted and investigated. In this way artifacts and biased conclusions can be avoided, novel insights can be discovered and more in depth interpretation of the data with greater

confidence can be obtained compared to using SAS alone.

#### Tips

- Show the data: scattering curve in log-log or log-lin is the minimum requirement. For models, show the fitted curves!
- Garbage in = garbage out: Most tools will give you an answer even if the prerequisites are not met at all – always check before!
- Units matter: In particular, the scattering vector  $q$  can be reported in either nm<sup>-1</sup> or Å<sup>-1</sup>

(continued)



- Crosschecks: Utilise information from complimentary techniques wherever possible to allow validation
- Easy to fit noisy data: always be critical
- Don't over-interpret models: Results are not atomic models, just a filled volume!  
Your solution might not be unique!

**Acknowledgments** The authors would like to thank Dr. Bárbara Calisto Machado for providing the SAXS data of the K and G proteins and KG complex.

## References

- Ashiotis G, Deschildre A, Nawaz Z, Wright JP, Karkoulis D, Picca FE, Kieffer J (2015) The fast azimuthal integration python library. *J Appl Crystallogr* 48(Pt 2):510–519. doi:[10.1107/S1600576715004306](https://doi.org/10.1107/S1600576715004306)
- Benecke G, Wagermaier W, Li C, Schwartzkopf M, Flucke G, Hoerth R, Zizak I, Burghammer M, Metwalli E, Muller-Buschbaum P, Trebbin M, Forster S, Paris O, Roth SV, Fratzl P (2014) A customizable software for fast reduction and analysis of large X-ray scattering data sets: applications of the new DPDAK package to small-angle X-ray scattering and grazing-incidence small-angle X-ray scattering. *J Appl Crystallogr* 47(5):1797–1803. doi:[10.1107/S1600576714019773](https://doi.org/10.1107/S1600576714019773)
- Bernado P, Svergun DI (2012) Structural analysis of intrinsically disordered proteins by small-angle x-ray scattering. *Mol BioSyst* 8(1):151–167. doi:[10.1039/c1mb05275f](https://doi.org/10.1039/c1mb05275f)
- Bonneté F, Finet S, Tardieu A (1999) Second virial coefficient: variations with lysozyme crystallization conditions. *J Cryst Growth* 196(2–4):403–414. doi:[http://dx.doi.org/10.1016/S0022-0248\(98\)00826-4](https://doi.org/10.1016/S0022-0248(98)00826-4)
- Brennich ME, Kieffer J, Bonamis G, De Maria AA, Hutin S, Pernot P, Round A (2016) Online data analysis at the ESRF bioSAXS beamline, BM29. *J Appl Crystallogr* 49:203–212. doi:[10.1107/S1600576715024462](https://doi.org/10.1107/S1600576715024462)
- Brian Richard P (2013) Everything SAXS: small-angle scattering pattern collection and correction. *J Phys Condens Matter* 25(38):383201
- Brookes E, Pérez J, Cardinali B, Profumo A, Vachette P, Rocco M (2013) Fibrinogen species as resolved by HPLC-SAXS data processing within the UltraScan Solution Modeler (US-SOMO) enhanced SAS module. *J Appl Crystallogr* 46(6):1823–1833
- Chen PC, Hub JS (2015) Interpretation of solution x-ray scattering by explicit-solvent molecular dynamics. *Biophys J* 108(10):2573–2584. doi:[10.1016/j.bpj.2015.03.062](https://doi.org/10.1016/j.bpj.2015.03.062)
- David G, Pérez J (2009) Combined sampler robot and high-performance liquid chromatography: a fully automated system for biological small-angle X-ray scattering experiments at the Synchrotron SOLEIL SWING beamline. *J Appl Crystallogr* 42(5):892–900. doi:[10.1107/S0021889809029288](https://doi.org/10.1107/S0021889809029288)
- Debye P (1947) Molecular-weight determination by light scattering. *J Phys Colloid Chem* 51(1):18–32. doi:[10.1021/j150451a002](https://doi.org/10.1021/j150451a002)
- Franke D, Svergun DI (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J Appl Crystallogr* 42(2):342–346. doi:[10.1107/S0021889809000338](https://doi.org/10.1107/S0021889809000338)
- Franke D, Kikhney AG, Svergun DI (2012) Automated acquisition and analysis of small angle X-ray scattering data. *Nucl Instrum Methods Phys Res Sect A* 689:52–59. doi:<http://dx.doi.org/10.1016/j.nima.2012.06.008>
- Franke D, Jeffries CM, Svergun DI (2015) Correlation Map, a goodness-of-fit test for one-dimensional X-ray scattering spectra. *Nat Methods* 12(5):419–422. doi:[10.1038/nmeth.3358](https://doi.org/10.1038/nmeth.3358)
- Glatter O (1977) A new method for the evaluation of small-angle scattering data. *J Appl Crystallogr* 10(5):415–421
- Glatter O, Kratky O (1982) Small angle x-ray scattering. Academic, London
- Graewert MA, Franke D, Jeffries CM, Blanchet CE, Ruskule D, Kuhle K, Flieger A, Schäfer B, Tartsch B, Meijers R, Svergun DI (2015) Automated pipeline for purification, biophysical and x-ray analysis of biomacromolecular solutions. *Sci Rep* 5:10734. doi:[10.1038/srep10734](https://doi.org/10.1038/srep10734)
- Grant TD, Luft JR, Wolfley JR, Tsuruta H, Martel A, Montelione GT, Snell EH (2011) Small angle x-ray scattering as a complementary tool for high-throughput structural studies. *Biopolymers* 95(8):517–530
- Guinier A (1938) La diffusion des rayons X sous les très faibles angles appliquée à l'étude de fines particules et de suspensions colloïdales. *Compte Rendu de l'Académie des Sciences de Paris* 206:1374–1376
- Hammel M (2012) Validation of macromolecular flexibility in solution by small-angle X-ray scattering (SAXS). *Eur Biophys J EBJ* 41(10):789–799. doi:[10.1007/s00249-012-0820-x](https://doi.org/10.1007/s00249-012-0820-x)
- Hammersley A (1997) FIT2D: an introduction and overview. European synchrotron radiation facility internal report ESRF97HA02T 68
- Hosemann R (1939) Neues röntgenographisches Verfahren zur Bestimmung des submikroskopischen Feinbaues eines Stoffes. *Diskussion des Zellulosemodells. Z Phys* 114(3–4):133–169
- Jacques DA, Trehwella J (2010) Small-angle scattering for structural biology – expanding the frontier while avoiding the pitfalls. *Protein Sci* 19(4):642–657. doi:[10.1002/pro.351](https://doi.org/10.1002/pro.351)

- Jacques DA, Guss JM, Svergun DI, Trehwella J (2012) Publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution. *Acta Crystallogr D Biol Crystallogr* 68 (6):620–626
- Karlsen ML, Thorsen TS, Johnner N, Ammendrup-Johnsen I, Erlendsson S, Tian X, Simonsen JB, Hoiberg-Nielsen R, Christensen NM, Khelashvili G, Streicher W, Teilum K, Vestergaard B, Weinstein H, Gether U, Arleth L, Madsen KL (2015) Structure of dimeric and tetrameric complexes of the BAR domain protein PICK1 determined by small-angle X-ray scattering. *Structure* 23(7):1258–1270. doi:10.1016/j.str.2015.04.020
- Knight CJ, Hub JS (2015) WAXSiS: a web server for the calculation of SAXS/WAXS curves based on explicit-solvent molecular dynamics. *Nucleic Acids Res* 43 (W1):W225–W230. doi:10.1093/nar/gkv309
- Koutsoubas A, Perez J (2013) Incorporation of a hydration layer in the ‘dummy atom’ ab initio structural modelling of biological macromolecules. *J Appl Crystallogr* 46(6):1884–1888. doi:10.1107/S0021889813025387
- Lambright D, Malaby AW, Kathuria SV, Nobrega RP, Bilsel O, Matthews CR, Muthurajan U, Luger K, Chopra R, Irving TC (2013) Complementary techniques enhance the quality and scope of information obtained from SAXS. *ACA Trans*:1–12
- Mathew E, Mirza A, Menhart N (2004) Liquid-chromatography-coupled SAXS for accurate sizing of aggregating proteins. *J Synchrotron Radiat* 11 (4):314–318. doi:10.1107/S0909049504014086
- Mylonas E, Svergun DI (2007) Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. *J Appl Crystallogr* 40(s1):s245–s249. doi:10.1107/S002188980700252X
- Orthaber D, Bergmann A, Glatter O (2000) SAXS experiments on absolute scale with Kratky systems using water as a secondary standard. *J Appl Crystallogr* 33(2):218–225. doi:10.1107/S0021889899015216
- Pernot P, Round A, Barrett R, De Maria AA, Gobbo A, Gordon E, Huet J, Kieffer J, Lentini M, Mattenet M (2013) Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. *J Synchrotron Radiat* 20 (4):660–664. doi:10.1107/S0909049513010431
- Petoukhov MV, Svergun DI Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys J* 89(2):1237–1250. doi:10.1529/biophysj.105.064154
- Petoukhov MV, Konarev PV, Kikhney AG, Svergun DI (2007) ATSAS 2.1-towards automated and web-supported small-angle scattering data analysis. *Applied crystallography*
- Porod G (1982) General theory. In: Glatter O, Kratky O (eds) *Small angle X-ray scattering*. Academic, London
- Rambo RP, Tainer JA (2011) Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers* 95 (8):559–571. doi:10.1002/bip.21638
- Rambo RP, Tainer JA (2013) Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* 496(7446):477–481. doi:10.1038/nature12070
- Reyes FE, Schwartz CR, Tainer JA, Rambo RP (2014) Methods for using new conceptual tools and parameters to assess RNA structure by small-angle X-ray scattering. *Methods Enzymol* 549:235–263. doi:10.1016/B978-0-12-801122-5.00011-8
- Rodriguez-Navarro A (2006) XRD2DScan: new software for polycrystalline materials characterization using two-dimensional X-ray diffraction. *J Appl Crystallogr* 39(6):905–909. doi:10.1107/S0021889806042488
- Round A, Brown E, Marcellin R, Kapp U, Westfall CS, Jez JM, Zubieta C (2013) Determination of the GH3.12 protein conformation through HPLC-integrated SAXS measurements combined with X-ray crystallography. *Acta Crystallogr D Biol Crystallogr* 69 (10):2072–2080. doi:10.1107/S0907444913019276
- Schneidman-Duhovny D, Hammel M, Sali A (2010) FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res* 38(suppl 2):W540–W544. doi:10.1093/nar/gkq461
- Segel DJ, Bachmann A, Hofrichter J, Hodgson KO, Doniach S, Kiefhaber T (1999) Characterization of transient intermediates in lysozyme folding with time-resolved small-angle X-ray scattering. *J Mol Biol* 288(3):489–499
- Semenyuk AV, Svergun DI (1991) GNOM – a program package for small-angle scattering data processing. *J Appl Crystallogr* 24(5):537–540. doi:10.1107/S002188989100081X
- Svergun D (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Crystallogr* 25(4):495–503
- Svergun DI (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* 76 (6):2879–2886. doi:10.1016/S0006-3495(99)77443-6
- Svergun D, Barberato C, Koch M (1995) CRYSOLE – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr* 28(6):768–773
- Svergun DI, Petoukhov MV, Koch MH (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* 80(6):2946–2953. doi:10.1016/S0006-3495(01)76260-1
- Watanabe Y, Inoko Y (2009) Size-exclusion chromatography combined with small-angle X-ray scattering optics. *J Chromatogr A* 1216(44):7461–7465. doi:10.1016/j.chroma.2009.02.053
- Zhang L, Wang L, Kao YT, Qiu W, Yang Y, Okobiah O, Zhong D (2007) Mapping hydration dynamics around a protein surface. *Proc Natl Acad Sci U S A* 104 (47):18461–18466. doi:10.1073/pnas.0707647104