Small angle neutron scattering of RNA-protein complexes

Audrone Lapinaite¹, Teresa Carlomagno^{2,3}, Frank Gabel⁴

¹Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA

²Leibniz University Hannover, Centre for Biomolecular Drug Research, Schneiderberg 38, D-30167 Hannover, Germany

³Helmholtz Centre for Infection Research, Group of Structural Chemistry, Inhoffenstrasse 7, D-38124 Braunschweig, Germany

⁴Univ. Grenoble Alpes, CEA, CNRS, IBS, 71 avenue des Martyrs, F-38000 Grenoble, France. Phone: +33 4 38 78 95 73. E-mail: frank.gabel@ibs.fr

Summary

Small angle neutron scattering (SANS) provides structural information on biomacromolecules and their complexes in dilute solutions at the nanometer length-scale. The overall dimensions, shapes and interactions can be probed and compared to information obtained by complementary structural biology techniques such as crystallography, NMR and EM. SANS, in combination with solvent H₂O/D₂O exchange and/or deuteration, is particularly well-suited to probe the internal structure of RNA-protein (RNP) complexes since neutrons are more sensitive than X-rays to the difference in scattering length densities of proteins and RNA, with respect to an aqueous solvent. In this book chapter we provide a practical guide on how to carry out SANS experiments on RNP complexes, as well as possibilities of data analysis and interpretation.

Key Words

Small Angle Neutron Scattering (SANS); Contrast Variation; Deuterium labeling; Structural Biology; Macromolecular Complex; RNA; Protein; RNP

Running head

SANS of RNA-protein complexes

1. Introduction

Small-angle neutron scattering (SANS) has been used since the 1970s (with the advent of high-flux neutron sources) to study a variety of isolated RNA and RNA-protein (RNP) complexes in solution. Contrary to X-rays, neutrons are scattered by the nuclei of atoms and are sensitive to nuclear scattering length density (SLD) fluctuations in the nanometer range. Therefore, SANS probes the conformational state of solubilized biomacromolecules and their complexes. Pioneering SANS studies on RNP complexes include the internal topology of RNA viruses ¹, the solvent-dependent interaction of tRNA with tRNA-synthetase ^{2, 3} and a mapping of proteins within the bacterial ribosome subunits ⁴⁻⁶, and are based on methods of contrast variation and triangulation, previously developed for X-rays and neutrons ⁷⁻⁹. Remarkably, these early SANS experiments allowed to determine the relative positions and arrangements of protein and RNA subunits within complexes whose atomic-resolution structures were resolved only decades later. In the mid-1970s, SANS was also applied to DNA-protein complexes, in particular in a series of experiments on nucleosome particles, again decades before atomic-resolution models became available ¹⁰⁻¹². A more complete review of early SANS studies can be found elsewhere ^{13, 14}.

The structural information obtained from SANS studies on biomacromolecules in the 1970s and 1980s was mainly analyzed and interpreted in terms of the molecular mass, the radius of gyration, the interaction and stoichiometry of partners, as well as by comparing experimental SANS curves with those of simple geometrical bodies ^{15, 16}. The increase of computing power and software development in the 1990s allowed a direct and rapid calculation of SANS (and SAXS, small angle X-ray scattering) curves from atomic models and their fits against experimental data ^{17, 18}. Concomitantly, so-called *ab initio* programs allowed to propose low-resolution envelopes from (multiple) SANS datasets ^{19, 20}. In recent years, SANS data from RNA-protein complexes have been interpreted by generating docking models of the complex from atomic-resolution structures of building blocks, often in combination with restraints provided by other structural biology techniques, such as NMR ²¹⁻²⁴. Here we provide a short overview of basic SANS theory and of experimental aspects of RNA-protein samples, we illustrate the process of structural analysis on the basis of pedagogical examples, and discuss the state-of-the-art and limitations of current modeling approaches as well as the combination of SANS with complementary techniques.

2. Material

2.1 Neutron sources and instruments

Biological SANS experiments are usually carried out with cold neutrons (wavelengths $\lambda \sim 5\text{-}10\text{ \AA}$), which are well-suited to probe the dimensions of biomacromolecules ($\sim$ several nm) and have a favorable interaction potential with the light atomic nuclei of which biological molecules consist. In contrast to X-ray scattering, neutron “home sources” are not available and experiments need to be carried out at large-scale facilities. Two major types of neutron sources are available worldwide; reactors and spallation sources. These produce high-energy neutrons, which need to be subsequently moderated to yield a maximum flux near the desired wavelength in the Ångstrom range ²⁵. An overview of existing neutron facilities can be found at the following website (<http://www.neutron.anl.gov/facilities.html>). Apart from these major large-scale facilities, a new generation of medium size, high-brilliance neutron sources are currently in development; these sources would provide access to neutrons at a regional level and support local user training ²⁶.

A schematic presentation of a typical SANS instrumental beamline is provided in Fig. 1. An overview of available SANS instruments and their technical characteristics is available under the following link: <https://www.ill.eu/users/scientific-groups/large-scale-structures/more/>. Access to instrumental beamlines usually requires submitting a proposal (typically twice per year) describing the methods and aims of the scientific project. Following a review process carried out by scientific committees, the beamtime is attributed on the basis of the scientific merit of the project, as well as the expected impact of neutrons in it.

2.2 Sample checklist for SANS experiments on RNA-protein complexes

- 1) **RNP in solution:** The reconstituted RNA-protein (RNP) complex should be prepared in the buffer of choice in a concentration range of $\sim 1 - 20\text{ mg/mL}$. There are no general restrictions on solvent, apart from the fact that it should not contain exotic elements that absorb neutrons strongly like ^5B , ^{64}Gd and others ²⁷. A minimum sample volume of $200\text{ }\mu\text{L}$ should be available for the measurement. Sample concentrations should be measured using optical absorbance at 280 and 260 nm with the appropriate calibrations.
- 2) **RNP buffers:** The reference buffer samples should match the buffer containing the RNP complex as faithfully as possible, in particular regarding the $\text{H}_2\text{O}:\text{D}_2\text{O}$ ratio when contrast variation is used. If the sample preparation is finished by a dialysis step, the buffer used during the last dialysis step should be kept

as reference for the SANS measurements. If the sample is prepared using size exclusion chromatography (SEC), an aliquot of the buffer of the chromatography run should be used as reference. If the sample is concentrated (or if the buffer in the sample is exchanged) using molecular weight (MW) cutoff centrifugal filters, the flow-through of the last spin should serve as the reference buffer. The volume of the reference buffer(s) should be the same as the sample volume, and it should be measured in an identical cell as the one used for the RNP sample(s).

- 3) **Sample holder reference:** The sample holder reference (*e.g.* empty quartz cuvette) needs to be measured in the same instrumental conditions and geometry as the sample(s) and buffer(s) in order to account for its contribution to the scattered signal.
- 4) **Detector background measurement:** A strongly neutron-absorbing material (boron composite such as BF₄ or cadmium) is usually measured in order to account for the electronic background signal on the detector once the direct neutron beam is blocked. It is usually provided by the neutron facility.
- 5) **Water (H₂O) reference:** A pure water sample (same volume and geometry as samples and buffers) can be measured in order to calibrate the detector efficiency and put the data on an absolute scale ²⁸. It has to be measured in the same instrumental setup as the other samples and buffers.

The measured neutron scattering intensity is the sum of the scattering intensities of all molecules and atoms that comprise the sample (biomolecules, buffer components, water, air, the sample cuvette etc.) (Fig. 1). Thus, the quality of the structural information obtained for the biomacromolecule in solution depends on the quality of the sample as well as on the accurate correction for background scattering. When preparing samples for SANS experiments two key points must be considered: the quality (purity and monodispersity) and the quantity of the sample. Good general practices of SAXS/SANS experiments on proteins in solution have been reviewed in detail recently ²⁹⁻³¹.

2.3 Detailed remarks and recommendations on sample quality

The quality of the structural information depends highly on the quality of the measured sample. The contribution of contaminants to the scattering intensities $I(Q)$ is proportional to the square of the contaminants' volume and to the concentration (see *Basic SANS theory* below). Thus, one must avoid high-molecular weight contaminants and aggregates. The most widely used approaches to assess the purity and monodispersity of the RNA-protein complex in the sample are gel electrophoresis (SDS-PAGE and native

PAGE), size exclusion chromatography (SEC), dynamic light scattering/static light scattering (DLS/SLS), and analytical ultracentrifugation (AUC) ³².

SDS-PAGE is the most common approach to monitor the composition and purity of the RNA-protein complex during preparation. The presence of components with different MWs from the molecule(s) of interest indicates the need for further purification. If the purity of the sample is > 95%, the complex is considered to be suitable for SANS experiments. However, it should be kept in mind that the presence of a single band (or a set of bands corresponding to the sizes of each component of the complex) in an SDS-PAGE is not a guarantee of monodispersity. Native PAGE (without denaturing agent), SEC, AUC, or SLS/DLS should be used to assess the polydispersity of the sample and estimate the size of the RNP of interest. To detect the RNA in reconstituted RNP complexes, the native PAGE can be stained with ethidium bromide (or other types of stains suitable for the detection of nucleic acids), followed by Coomassie blue staining to visualize the protein components. On native PAGE, the band visualized with the nucleic acid stain and the band visualized with Coomassie blue must overlap, as they should both stem from the intact RNP complex.

SEC (also known as gel filtration) is used both for further purification and to assess the monodispersity of the RNP sample. During SEC, RNP complexes are separated based on their MW and shape. The column resin is chosen according to the MW of the species to be separated; column manufacturers specify which resin is suitable for a given MW range. During SEC, absorbance at both 280 nm (highest for protein) and 260 nm (highest for RNA) should be detected to monitor the elution profile of the reconstituted RNP complex. Generally, a single, symmetric peak (ratio of 260/280 > 0.5) at the expected elution volume indicates the presence of a monodisperse RNP complex. When the RNP complex multimerizes or decays to multiple subcomponents, two scenarios can occur: if the rate of exchange between the intact RNP complex and its multimers (or subcomponents) is slower than the rate of separation through the column, the SEC profile shows several peaks, with large aggregates coming off the column in the void volume. In this case the intact RNP can be isolated by fractionation. However, if the rate of exchange is faster than the rate of separation through the column, the SEC profile shows only one peak at higher or lower elution volume with respect to the intact RNP. The presence of fast exchange between the intact RNP and its multimers can be confirmed by varying the concentration of the injected sample. Since the degree of multimerization depends on concentration, the peak resulting from fast exchange between the desired RNP and its

multimers moves towards lower elution volumes (higher percentage of multimers) upon increasing the concentration. In the case of fast exchange, the intact RNP complex cannot be separated from its multimers or subcomponents by fractionation.

The light scattering methods (SLS/DLS) are used to investigate the properties (e.g. MW, size, diffusion) of particles in solution. The scattering intensity is especially sensitive to variations in size and can be exploited to detect aggregates³³. SLS/DLS require small amounts of sample (~ 10 µL) and can be employed to optimize sample conditions by screening multiple conditions (e.g. various buffer compositions, pH, temperature, sample concentrations etc.). Moreover, DLS can be used to guide the storage and handling of RNP samples prior to SANS measurements by assessing the integrity of the samples over time or during freeze-thaw cycles in various buffers and in the presence or absence of stabilizing additives (e.g. glycerol, TCEP, DTT).

The main shortcoming of SLS/DLS methods is their low resolution, which may not allow distinguishing between particles with small differences in shape, such as monomers and dimers. In this case, a high resolution method such as Multiangle Laser Light Scattering (MALLS) may be helpful. MALLS is commonly used in line with SEC (SEC-MALLS) and allows the determination of the absolute mass of particles. This method detects the presence of high molecular weight aggregates or subcomponents of the RNP (as long as they can be separated by SEC) and returns the stoichiometry of the complexes.

As an alternative to light scattering, the homogeneity, MW, and stability of biomacromolecules can be characterized at nearly physiological conditions by analytical ultracentrifugation (AUC). In either equilibrium or velocity mode, AUC relies on observing the distribution of biomacromolecules in real-time during the application of a centrifugal force. The method is very sensitive and it enables the identification of trace quantities of aggregates, oligomers and even multiple co-existing complexes of various stoichiometries. Moreover, through analysis of the hydrodynamic properties of the sample, AUC is able to reveal the presence of different conformations of the same particle.

In general, a good SANS sample should be monodisperse and remain such over the length of the measurement (several hours to 1-2 days). This requisite can be verified by the most appropriate method among those described above.

2.4 Detailed remarks and recommendations on sample quantity

In order to obtain a good signal-to-noise ratio it is important to consider the quantity of the sample. The required volume depends on the sample holder and typically ranges from 200 to 500 μL . The concentration depends on the MW of the RNP of interest and is usually in the range of 1 – 20 mg/mL. For SANS data analysis (see dedicated section below) it is important to estimate the RNP sample concentration as accurately as possible. The estimation of the concentration of large RNP complexes is challenging due to the overlap of the absorption of RNA and proteins. In addition, in the case of very large complexes, light scattering also influences the absorption spectra. We advise to measure the absorptions of RNP samples at 260 and 280 nm (at least three times) and use the following equations ³⁴:

$$C_{RNA} = \frac{A_{260} - \left(\frac{\epsilon_{prot,260}}{\epsilon_{prot,280}} \right) A_{280}}{\left(\epsilon_{RNA,260} - \left(\frac{\epsilon_{prot,260}}{\epsilon_{prot,280}} \right) \epsilon_{RNA,280} \right) L} \quad (\text{Eq. 1a})$$

$$C_{proteins} = \frac{A_{260} - \left(\frac{\epsilon_{RNA,260}}{\epsilon_{RNA,280}} \right) A_{280}}{\left(\epsilon_{prot,260} - \left(\frac{\epsilon_{RNA,260}}{\epsilon_{RNA,280}} \right) \epsilon_{prot,280} \right) L} \quad (\text{Eq. 1b})$$

where L is the cell optical path-length in cm, A_{260} and A_{280} are the absorptions measured at 260 and 280 nm respectively, $\epsilon_{RNA,260}$ is equal to the average extinction coefficient of a single nucleotide at 260 nm (8000 $\text{M}^{-1} \text{cm}^{-1}$) multiplied by the number of nucleotides in a single RNP particle, and $\epsilon_{RNA,280}$ is equal to $\epsilon_{RNA,260} / 2$ (based on the fact that the A_{260}/A_{280} ratio of pure RNA is 2). The extinction coefficient for all the proteins within the complex at 280 nm ($\epsilon_{prot,280}$) should be calculated according to the sequences of all the proteins, while the $\epsilon_{prot,260}$ is estimated using the $\epsilon_{prot,260}/\epsilon_{prot,280}$ ratio of 0.6 for pure proteins.

2.5 Sample preparation and deuteration for contrast variation experiments

Using an appropriate $\text{H}_2\text{O}:\text{D}_2\text{O}$ ratio within the solvent, the scattering signal of proteins or RNA can be masked (see *Basic SANS theory*), so that the resulting SANS curve is dominated by either the protein or the RNA, within the assembled RNP complex. For unlabeled (^1H) proteins, the contrast match point (CMP) is between 40 and 45% (vol/vol) of D_2O in solution, and for unlabeled (^1H) RNA at around 70% (Fig. 2). The precise values depend on the exact amino acid/nucleotide composition, the number of exchangeable protons and the specific volumes.

The desired H₂O:D₂O ratio can be achieved *via* RNP sample dialysis, or buffer exchange by SEC. When using dialysis, the RNP sample is split into two parts: one half is dialyzed overnight (or at least for 8h) against a buffer containing 0% (vol/vol) of D₂O, while the other half is dialyzed against a buffer containing 100% (vol/vol) of D₂O. After the dialysis, the two samples are mixed in the correct proportion to obtain the required H₂O:D₂O ratio. When using SEC, the gel filtration column has to be equilibrated with at least two column volumes of the buffer that contains the appropriate H₂O:D₂O ratio. It is beneficial to use SEC because it also allows monitoring the integrity of the RNP sample during buffer exchange.

SANS data recorded from RNP samples where the scattering of proteins is matched, and only the RNA component contributes to the scattering curve, often has a low signal-to-noise ratio. In this case, the quality of RNA scattering data can be improved either by increasing the sample concentration or by increasing the contrast through deuteration of the RNA (Fig. 2). The RNA can be uniformly deuterated by *in vitro* transcription using T7 RNA polymerase, ²H-NTPs and synthetic DNA template ³⁵.

In the case of multi-protein RNP complexes, where the CMPs of all proteins are very close, an artificial contrast between proteins can be achieved *via* selective deuteration (²H-labeling) of individual proteins. Uniform deuteration of recombinant proteins is achieved using the *Escherichia coli* bacterial expression host cultured in M9 medium containing D₂O and a deuterated carbon source (usually glucose). *E. coli* can tolerate deuterated medium, despite at the expense of reduced growth rates and lower protein yields. To optimize the amount of protein produced by a definite amount of cell culture, the bacteria are adapted to grow in deuterated medium by repeated subculturing in progressively higher amounts of D₂O. The resulting D₂O-tolerant cells are then used to express the protein of interest in M9 medium containing 99% D₂O and ²H-glucose (all M9 components must be dissolved in 100% D₂O) ³⁶. The uniformly deuterated protein expressed in these conditions is purified following the protocol optimized for the unlabeled protein, using H₂O containing buffers. The deuteration levels can be confirmed by ¹H-NMR or mass spectrometry (e.g. LC/electrospray ionization-TOF-MS).

It has been reported that deuteration may affect the protein biochemical characteristics. It may decrease the hydrophobicity of the molecule and thus change the optimal storage and handling conditions ³⁷. Moreover, the presence of D₂O in the solvent promotes protein aggregation and may change the protein-ligand interactions ³⁸. Thus, when preparing RNP samples for contrast matching and variation experiments one must investigate the integrity of the deuterated RNP complex. First of all, we advise to perform a small-

scale ^2H -protein expression and purification to assess its solubility and stability, and to optimize purification, storage and handling conditions using the previously described SDS-PAGE, SEC, DLS/SLS and AUC approaches. Next, it is important to assess whether the deuteration of subunits or the presence of D_2O in the buffer affects the dissociation constant (K_d) of the RNP complex. Electrophoretic mobility shift assay (EMSA), double-filter binding assay, and isothermal titration calorimetry (ITC) are the most popular methods to determine the K_d of RNA-protein interactions ³⁹⁻⁴¹. We advise to use either EMSA or filter binding assay with radioactively (^{32}P) labeled RNA, because these approaches are very sensitive and do not require large amounts of sample. It is important to keep in mind that, even if there is no significant effect on the K_d of the RNP complex, the k_{on} and k_{off} rates may be affected by the deuteration of subunits and/or presence of D_2O in the buffer. Thus, it may be necessary to estimate the k_{on} and k_{off} of the RNP complex in these specific conditions. The off rate can be easily determined *via* EMSA-based competition binding assay, where the RNP complex, reconstituted using the radioactively labeled RNA, is incubated over time in the presence of an excess of unlabeled RNA. Finally, using SDS-PAGE, SEC, DLS/SLS or AUC, one must ensure that the deuterated RNP sample in the D_2O containing buffer is monodisperse and stable throughout the data acquisition.

2.6 Sample for ligand titration SANS experiments

The concentration for SANS samples (1 – 20 mg/mL) is above the K_d of most ligand binding events ($K_d < 10\ \mu\text{M}$). Thus, upon ligand binding the complexes exist in a single, fully-bound state (holo state). Under these conditions, SANS can be useful to investigate the conformational changes of complexes upon interaction with ligands ⁴². The measurements can be performed on two independently prepared apo and holo RNP samples or on a single sample by titrating the ligand to the apo RNP complex after SANS data acquisition. If the second setup is chosen, there are several points to consider: first, the ligand must be dissolved in the same buffer as the sample and reference buffer. Second, the ligand must be dissolved at as high concentration as possible, to be able to add only small volumes ($\sim 1 - 10\ \mu\text{L}$) to the initial solution and avoid excessive dilution. If the affinity of the ligand for the RNP is moderate ($K_d > 1\ \mu\text{M}$), the ligand must be added in excess, to ensure that the majority of the RNP complex is in holo state. In this case one must consider the contribution of the free ligand to the scattering curve. In alternative, one can use an appropriate $\text{H}_2\text{O}:\text{D}_2\text{O}$ ratio within the sample solvent to mask the ligand scattering.

3. Methods

3.1 Schematic representation of a SANS experiment

A SANS experiment records the scattered neutrons from a biological sample in solution (Fig. 1). To this end, a well-defined incoming beam is prepared by a series of monochromators and optical elements that select neutrons of a specific speed (i.e. wavelength λ) and direction (wave vector $\vec{k}$). After interaction with the sample, the neutrons are scattered in a direction $\vec{k}'$. The diffraction pattern is recorded on a neutron-sensitive detector as a function of the momentum transfer $Q = |\vec{k}' - \vec{k}| = (4\pi/\lambda)\sin\theta$ (where 2θ is the scattering angle). For isotropic solutions (i.e. the orientation of one biomacromolecule does not influence the orientation of others), the two-dimensional scattering pattern on the detector is symmetric around the direct beam. Given the typical fluxes ($10^6 - 10^8 \text{ n}\cdot\text{cm}^{-2}\cdot\text{s}^{-1}$) and wavelengths (some Å) of cold neutrons, sample-detector distances ($\approx 1 - 40 \text{ m}$), detector dimensions ($\approx 1 \times 1 \text{ m}^2$), and exposure times (minutes to hours), the sensitivity of SANS experiments from dilute ($< 1 \% \text{ v/v}$) biomacromolecular solutions ranges from a few kDa to several MDa, and the structural length-scales from a few Å to several 1000 Å.

Figure 1: Schematic representation of a SANS experiment on a RNA-protein complex. The small white rectangle in the centre of the detector represents the beam-stop, composed of a strongly neutron-absorbing material, and protects the detector from the direct neutron beam.

3.2 Instrumental setup for RNA-protein complexes

The instrumental setup (in particular the sample-detector distance) depends on the linear dimensions of the RNP complex to be measured. The resulting Q -ranges need to cover both low angles in order to seize the accurate maximum dimension of the complex, as well as higher angles for accurate solvent (buffer) subtraction. Often, both requirements cannot be fulfilled with a single sample-detector setup but require at

least two. The lowest Q -value, $Q_{\min}$, which is limited by the size of the beam-stop (Fig. 1), has to allow a proper extraction of the radius of gyration R_g via the Guinier approximation (see *Basic data analysis*). This requires fulfilling the condition $Q_{\min} \approx 0.3/R_g$ ⁴³. R_g can be estimated from PDB models of RNP complexes of comparable size by using programs such as CRYSON¹⁸. Alternatively, if the maximum dimension $D_{\max}$ of the complex is known, e.g. from (Cryo-)EM, a first rough estimate of the radius of gyration can be obtained from $R_g \approx 0.3 \cdot D_{\max}$. For example, a small RNP complex (PAZ domain in complex with a nine nucleotide long RNA, PDB ID 1si3, 20 kDa) has an R_g of 15.2 Å (in H₂O) and requires $Q_{\min} < 0.02 \text{ Å}^{-1}$ for the proper application of the Guinier approximation. On the other hand, a very large complex (large ribosomal subunit of *Haloarcula marismortui*, PDB ID 1ffk, 1.6 MDa) with an $R_g \approx 66 \text{ Å}$ (in H₂O) requires a $Q_{\min} < 0.005 \text{ Å}^{-1}$. The upper limit of the Q -range, $Q_{\max}$, should ensure a proper analysis of short-distance information and accurate solvent-subtraction. In practice, the decrease of signal:noise at higher angles limits $Q_{\max}$ to $0.5 - 0.7 \text{ Å}^{-1}$ on most SANS instruments.

3.3 Basic SANS theory, solvent contrast variation and deuteration

In the case of an ideal (i.e. isotropic and non-interacting) solution of N macromolecules, the two-dimensional scattering pattern is centrosymmetric around the direct beam and the scattered intensity I can be described as a function of the modulus of the momentum transfer $\vec{Q}$ (Fig. 1)^{15, 43, 44}:

$$I(Q) \propto N \left| \int_V \Delta\rho(\vec{r}) e^{i\vec{Q} \cdot \vec{r}} dV \right|^2 \quad (\text{Eq. 2})$$

The integral runs over the entire (solvent-excluded) volume V of a particle with $\vec{r}$ being a vector pointing to a volume element dV of scattering contrast $\Delta\rho(\vec{r})$, within V :

$$\Delta\rho(\vec{r}) = \frac{\sum_i b_i}{dV} - \rho_s \quad (\text{Eq. 3})$$

In the case of neutron scattering, b_i are the (coherent) scattering lengths (Fig. 2B) of the atomic nuclei within a volume element dV of the particle, and ρ_s is the average scattering length density (SLD) of the solvent, e.g. of a certain H₂O:D₂O ratio (Fig. 2A). Values b_i of individual atoms can be retrieved from literature¹³. Fig. 2B lists the ones of the most common atoms found in proteins and RNA, and compares them to the respective values for X-rays. Average contrasts $\Delta\rho$ of proteins or RNA at a given H₂O:D₂O

ratio (Fig. 2A) can be calculated by summing up scattering lengths of all protein residues ¹³ (or RNA nucleotides ⁴⁵) and by dividing by the solvent-excluded volume V of the molecules. Importantly, exchangeable hydrogens need to be taken into account in SANS. By using appropriate software tools, it is possible to calculate the contrast of RNP complexes directly from their primary sequences ^{46, 47}.

Figure 2: Neutron scattering length densities, and X-ray and neutron scattering lengths of selected atoms. **A)** Scattering length densities (SLDs) for hydrogenated (¹H) and perdeuterated (²H), i.e. fully deuterated, proteins and RNA. Contrasts ($\Delta\rho$) are shown for a ¹H-RNA in 20% D₂O and for a ¹H-protein in 80% D₂O. Please note that the sign of contrast (with respect to solvent) can be positive or negative. **B)** The circles illustrate the relative size of the X-ray and neutron cross sections (i.e. a measure of the size of atoms as they appear to the respective radiation), calculated from the respective scattering lengths of light atoms composing proteins and RNAs. The red color of ¹H indicates a change of sign of the scattering length with respect to other atoms or isotopes and is the basis of solvent contrast variation and deuterium-labeling in biological SANS.

Fig. 2A illustrates that deuteration (exchange of hydrogen, ^1H , against deuterium, ^2H) in proteins or RNA modifies the contrast. The artificial contrast conditions created by deuteration allow distinguishing two chemically similar macromolecules by SANS, e.g. a ^1H - from a ^2H -protein (or two RNAs) within multi-RNA-protein complexes^{42, 48}. Several methods have been developed to bio-synthetically deuterate proteins or RNA (see *Sample preparation and deuteration for contrast variation experiments*). The interested reader can find more detailed information in recent reviews^{36, 49}, and references therein.

3.4 Raw data reduction

The raw data in a SANS experiment of solubilized, isotropic biomacromolecular complexes are two-dimensional, pixelated centrosymmetric intensity images, recorded by a neutron detector system (Fig. 1). Both sample and buffer images need to be corrected for detector background signal. The background signal (originating from electronic, terrestrial and cosmic as well as parasitic radiation from the direct neutron beam) is determined using a strongly neutron-absorbing sample (*e.g.* boron or cadmium) that blocks the direct neutron beam. The empty sample holder (often a quartz cuvette such as those used for optical density measurements) and a water (H_2O) sample are measured to correct for either sample holder signal and detector efficiency or absolute scaling, respectively.

All samples and references are also measured in transmission mode to determine the respective weight with which they contribute to the overall signal. The beam center is usually determined from the transmission experiment of the empty beam. Several software programs have been developed at the respective neutron centers to reduce the raw data. They allow defining appropriate masks to cut away parasitic scattering close to the beam center and defective pixels or artifacts at the edges of the detector. The two-dimensional scattering patterns are reduced to one-dimensional scattering curves $I(Q)$, without a loss of information for isotropically oriented protein-RNA complexes in solution, by integrating over concentric rings around the beam center.

In a final step, the one-dimensional buffer intensities need to be subtracted from the respective sample intensities using available software⁵⁰. Due to a highly variable incoherent background signal⁵¹, it may be necessary to align the buffer and sample levels at the highest available Q -values, prior to the subtraction. A scaling factor $> 2\%$ needed for alignment can be indicative of differences in sample and buffer

preparations, possibly due to a slightly different H₂O:D₂O ratios; in this case, the respective samples should be tested again, e.g. by comparing their transmission values, which are very sensitive to the H₂O:D₂O ratio

28

3.5 Basic data analysis and model-free parameters

For ideal (i.e. isotropic orientation and no inter-particle effects; see also comments in the *Notes* section) and monodisperse (no mixture of oligomeric states or excess of free partners) solutions, the one-dimensional SANS data can be directly related to the (conformationally averaged) structure of the RNP complex.

The two most important (and model-free) structural parameters are the radius of gyration, R_g , and the intensity scattered in the forward direction, $I(0)$. Both can be extracted by a so-called Guinier approximation, i.e. by a linear fit in a $\ln I(Q)$ vs. Q^2 plot ⁵²:

$$\ln[I(Q)] = \ln[I(0)] - 1/3 R_g^2 Q^2 \quad (\text{Eq. 4})$$

The Guinier approximation is valid up to $Q_{\max} R_g < 1.0 \dots 1.3$ ^{43, 44}. As discussed above, the lower limit of the fit range ($Q_{\min} R_g$) should be < 0.3 .

R_g is a measure for the normalized distribution of scattering contrast volume elements around the center of scattering mass of the solubilized complex:

$$R_g^2 = \int \Delta\rho(\vec{r}) r^2 dV / \int \Delta\rho(\vec{r}) dV \quad (\text{Eq. 5})$$

r is the distance between a scattering element (e.g. protein residue or RNA nucleotide) and the center of contrast of the complex. It is important to note that R_g integrates over all volume elements that display a different SLD than the bulk solvent and can therefore account for solvent/solute elements in the vicinity of the surface of the complex, if they display a different scattering length density than the bulk ^{53, 54}.

The forward scattered $I(0)$ intensity can be related to the volume $V = V_{\text{prot}} + V_{\text{RNA}}$ of a given RNA-protein complex at a concentration C (in Mol·cm⁻³), if the SANS data are recorded in absolute units (cm⁻¹):

$$I(0) = C \cdot N_A \cdot [\Delta\rho_{\text{prot}} \cdot V_{\text{prot}} + \Delta\rho_{\text{RNA}} \cdot V_{\text{RNA}}]^2 \quad (\text{Eq. 6})$$

N_A is the Avogadro constant, V_{prot} and V_{RNA} are the solvent-excluded volumes of the protein and RNA molecules (in cm³) that compose the complex, and $\Delta\rho_{\text{prot}}$ and $\Delta\rho_{\text{RNA}}$ are their respective scattering length densities (SLDs), in cm⁻². Both solvent-excluded volumes and SLDs of proteins and RNA can be

calculated from literature, as a function of their primary sequence^{45, 55}. Equation 6 represents a quality check, with respect to whether the RNP complex is in the expected oligomeric state or there are excess partners in solution. Given the accuracy of the solvent-excluded volumes, SLDs, and measured concentrations, a relative deviation of 15-20%, with respect to the theoretical $I(0)$, can be expected for a well-assembled and monodisperse complex. Deviations $> 20\%$ may indicate either systematic errors in the measured concentration or problems with sample monodispersity and/or stoichiometry.

3.6 Sophisticated data analysis: low resolution *ab initio* envelopes

Several approaches have been developed to interpret SANS data beyond model-free parameters (molecular weight and radius of gyration). If atomic models are available from other techniques, such as crystallography, NMR, cryo-EM or even homology modelling, for the same RNP complex, the SANS data calculated from this model can be compared and scored against the experimental SANS data¹⁸.

If no atomic-resolution model is available, shapes of proteins and RNAs within the assembled complex can be reconstructed using *ab initio* shape analysis²⁰. We illustrate the capacity and the limits of this approach by employing the program MONSA¹⁹ on a pedagogical example (Fig. 3): using back-calculated SANS curves at multiple contrasts, endowed with errors, we retrieve the shapes of the proteins SRP19 and SRP68, as well as the SRP RNA, within the assembled signal recognition particle (SRP) complex⁵⁶ (PDB ID 4p3e), and compare them to the original atomic model from which the SANS curves were generated.

If the SANS data originate from a monodisperse, properly assembled complex (Fig. 3A), the *ab initio* approach retrieves accurately the topology of the RNP: the densities and distances of the two protein partners, as well as the elongated shape of the RNA molecule, generated from the simulated SANS data, are in good agreement with the original atomic model. However, the orientation of the proteins, as well as the RNA secondary structure elements, remain undetermined, even though the loose RNA density suggests an open helical structure. In the case of contamination, simulated by addition of the SANS signal of 50% free SRP68 to the original SANS data, the reconstructed *ab initio* shapes contain artifacts (Fig. 3B).

This simulated example underlines the paramount importance of monodispersity in SANS samples. In our example, the excess of protein SRP68 leads to an erroneous model, i.e. an additional protein density (Fig. 3B, left). Monodispersity should be verified before the SANS experiment, and importantly for all contrasts conditions (i.e. H₂O:D₂O ratios), as explained in the *Sample quality* section above.

Figure 3: MONSA *ab initio* models of the SRP complex as a function of data simulated for various experimental conditions (cyan: RNA densities, green: protein densities). **A)** SANS data generated from a monodisperse solution of the complex, and *ab initio* shapes derived from a typical MONSA run. **B)** SANS data generated from a solution of the SRP complex and the free protein SRP68 in a 2:1 stoichiometric ratio and *ab initio* shapes derived from a typical MONSA run. Note the appearance of artifacts in the *ab initio* model in the presence of excess protein SRP68, indicated by a red arrow. Filled circles on the right represent the calculated SANS curves at multiple contrasts, including artificially added noise. Continuous lines denote the back-fits of the *ab initio* models to the calculated data at each contrast.

3.7 Sophisticated data analysis II: Combined approaches using SANS and NMR data

In the past two decades, the application of small angle scattering (SAS) to biological samples has seen a revival. The combination of NMR and SAXS is especially popular, probably due to the similarity of the sample conditions and to the complementarity of the structural information provided by the two techniques. NMR distance restraints cover a range from a few Ångstroms (Nuclear Overhauser Effect data, NOE) to 20 – 30 Å (Paramagnetic Relaxation Enhancement data, PRE). Depending on the size of the

complex, this structural information reports on local interactions (NOEs) or at most on the relative orientation of domains (PREs). SAS data inform on the global shape of the particle in solution up to a range of 1000 Å, while it rapidly loses sensitivity for detailed structural features (< 15 Å). Thus, NMR and SAS can be considered fully complementary. SAXS and SANS data are commonly used to validate structures calculated by NMR, particularly in terms of domain orientation in multi-domain proteins, nucleic acids and complexes thereof^{24, 57, 58}. Other more sophisticated approaches use SAS data readily during structure calculation, either by refinement against the scattering curve⁵⁹ or by restraining the conformational space available to the particle to the shape envelope calculated *ab initio* from the SAS data^{42, 60}.

A novel application of SANS in combination with NMR has been developed in our laboratories to study the structure of a large RNP complex, the Box C/D RNP from Archaea⁴², consisting of twelve proteins and either six or two RNAs, in its holo and apo forms, respectively (Fig. 4A). This RNP is responsible for methylation of ribosomal RNA (rRNA) at the 2'-O-ribose. The substrate sequence in the rRNA is recognized by base pairing with guide sequences comprised in the guide-RNA molecule (gRNA). The gRNA also serves as binding platform for proteins L7Ae, Nop5 and fibrillarin, among which fibrillarin catalyzes the methylation reaction. Each particle contains two copies of gRNA, four substrate RNAs (only in the holo form) and four copies of each of the three proteins, making up a MW of 390 kDa. At this molecular size, NMR focuses on the observation of methyl groups of proteins, while all other NMR signals are too broad to be detected. Methyl chemical shift perturbation (CSP) data allow detection of the interaction interfaces between proteins, while PRE effects measured on the methyl groups are used to determine inter-protein distances.

NMR does not provide any structural information on the relative position of RNA secondary structure elements, as all RNA signals are broad beyond detection in this MW range. In addition, for this complex, the NMR resonances of the four copies of each protein mostly overlap, making it impossible to distinguish one copy from another. Both shortcomings of NMR could be compensated by SANS applied with contrast variation. Firstly, a SANS curve measured for the Box C/D RNP containing ²H-gRNA in a buffer of 42%:58% D₂O:H₂O, where the scattering length density of the proteins is matched by the solvent (Fig. 2A), was used to calculate a low resolution RNA shape (Fig. 4B). This, together with information on RNA secondary structure elements derived from biochemical data and from the structure of subcomplexes,

allowed us to calculate the structure of the gRNA in the complex. The SANS-derived RNA envelope was instrumental in this task. Secondly, the four copies of each proteins could be nicely distinguished in *ab initio* MONSA²⁰ envelopes calculated from SANS curves measured for the Box C/D RNP containing only one deuterated protein species and dissolved in a 42%:58% D₂O:H₂O buffer (Fig. 4B). This is illustrated particularly well for fibrillarin, whose four copies in the apo form of the RNP (without the four substrate RNA molecules) are well separated in space. The *ab initio* MONSA envelope, integrating several SANS curves from differentially ²H-labeled Box C/D RNP complexes at multiple solvent (D₂O:H₂O) contrasts, shows the presence of four distinct protein densities (Fig. 4C), confirming the stoichiometry of the complex and allowing an approximate estimation of the distances between the fibrillarin copies.

We then developed a structure calculation protocol that actively uses both NMR and SANS restraints to calculate the structure of the complex in two steps. The high-resolution structures of subcomponents of the complex, as well as RNA and protein secondary structure elements, were docked together under the guidance of the experimental data. In the first calculation step, the RNA envelope and an approximate estimate of the distances between copies of L7Ae and fibrillarin, derived from *ab initio* modeling, were used together with the NMR restraints to restrict the conformational space of the search. Subsequently, the fit of the models to one SAXS and seven SANS curves, measured in different contrast conditions, was used to select a subset of models that was subjected to the structure refinement step. At the end of the second step, the final structure ensemble was selected again based on the fit of the individual models to the eight SAS curves.

The combination of the NMR and SAS data, together with the structure calculation protocol developed to incorporate both data types early on in the calculation, allowed to solve the structure of a 390 kDa RNP in solution (Fig. 4D). This is a remarkable achievement that demonstrates the enormous potential of SANS, in the implementation with contrast variation, to determine the structure of multicomponent RNP complexes in solution. The intrinsic low resolution of the structural information provided by SANS is alleviated by: 1. Use of the high resolution structure of stable protein domains and RNA secondary structure elements in the docking protocol; 2. Use of distance restraints derived by NMR; 3. Measurement of several SANS curves with contrast variation. We believe that this combination of NMR and SANS, despite being laborious, opens the way to study various RNP complexes that do not crystallize or are too fragile to be imaged by electron microscopy⁶¹.

Figure 4: SANS in contrast variation applied to the 390 kDa Box C/D RNP complex **A)** Schematic representation of the apo form of the Box C/D RNP consisting of two copies of the gRNA and four copies of each protein L7Ae (green), Nop5 (grey) and fibrillarin (Fib, blue) **B)** *Ab initio* envelope of the gRNA from a single contrast, illustrating a connected and extended topology. **C)** *Ab initio* envelope calculated with MONSA from multiple SANS curves from differentially labeled Box C/D RNPs. The fibrillarin molecules from the final model are represented in ribbon and overlap with the envelope. **D)** Structure of the apo Box C/D RNP calculated by a combination of NMR and SANS data as described in the text ⁴².

3.8 Concluding remarks and perspectives

In this chapter we have presented the application of SANS to RNP complexes, we have given recommendations on how to prepare high-quality RNP samples for SANS measurements, how to decide on the instrumental set-up and how to process the data. For RNP complexes, we have put a special emphasis on the contrast variation approach, which allows visualizing selected combinations of RNA and protein components in the context of the intact RNP. A set of several SANS curve, measured in contrast variation

with different combinations of ^1H - and ^2H -components, provides enough data to determine the relative positions of these components in the RNP. We have illustrated this concept on the example of the 390 kDa Box C/D RNP, whose structure was obtained through the joint use of NMR and SANS data.

In principle, the concept of contrast variation can be applied also to structured elements of the same molecule. ^2H -labelled structural elements, as part of a molecule, can be studied by the SANS contrast variation approach using a solvent that matches the scattering length density of the ^1H -elements. This can be useful to determine the distance between domains in multi (> 2)-domain proteins or between stems in RNA molecules. Two conditions are necessary for the approach to be successful: first, each ^2H -labeled element must have a significant scattering mass ($> \sim 10$ nucleotides for an RNA); second, the ^2H -labelled elements must not be randomly distributed in the RNA, i.e. their overall distribution must be distinguishable from the global RNA envelope. A number of methods are available to achieve segmental deuterium labeling of both proteins⁶²⁻⁶⁵ and RNAs⁶⁶⁻⁶⁸.

X-ray crystallography, NMR, EM, mass spectrometry (MS), SAS, fluorescence (FRET) and Electron Paramagnetic Resonance spectroscopy (EPR) provide structural data on biomolecules at different levels of resolution and completeness. Despite the availability of so many experimental approaches, the determination of the structure of large nucleic acids and their complexes with proteins remains a challenging task. X-ray crystallography and EM may fail with highly flexible molecules such as RNA and RNPs. NMR can cope very well with dynamic processes; however, its applicability as a stand-alone structure determination method is limited to particles of ca. 50 kDa, due to the rapid decrease of the quality of the NMR signal upon increasing molecular weight. SAS, FRET, MS and EPR deliver either low-resolution or incomplete structural information, which is insufficient for *de novo* structure determination. The rapid discovery of large and dynamic RNP complexes with paramount biological relevance demands a rapid answer to the challenges posed by the structure determination of RNPs. We believe that integrative structure biology methods may provide such an answer, especially if applied in solution, where the RNP complex preserves its functional dynamics. The combination of several data types allows overcoming the limitations of each technique, while providing a reciprocal validation of the datasets. In this context, SANS in combination with techniques that are able to measure interatomic distances (NMR, EPR, FRET, MS) is destined to play a prominent role.

4. Notes

This section contains a list of typical problems that may be encountered during a SANS experiment and indications on how to avoid them or minimize their impact.

- 1) *Water condensation on quartz cuvettes*: A potential source of error can be the condensation of water vapour on quartz cuvettes. This phenomenon can happen when samples (or cuvettes) are transferred from cold temperature (ice or cold room) to room temperature. It can falsify SANS results, in particular for D₂O samples, by changing transmission values and the incoherent background. Therefore, the quartz cuvettes should be sealed (e.g. with parafilm) in order to minimize sample evaporation and condensation inside the sample holder. Condensed water on the outside surface of the sample holder should be carefully wiped off, and a sufficient time for temperature equilibration should be allowed prior to the SANS experiments.
- 2) *Oligomerization/aggregation in solvents containing D₂O*: The physico-chemical properties of D₂O differ from those of H₂O, reinforcing the hydrophobic interaction between macromolecules⁶⁹. While the use of heavy water is popular to increase contrast (for hydrogenated proteins) and to reduce the incoherent background, it has been experimentally observed that it increases the tendency of biomacromolecules to oligomerize, aggregate or precipitate in solution. While very big precipitates (all linear dimensions large with respect to the RNP complex of interest) may be tolerated, lower oligomeric states impede a proper data analysis and should be avoided. Therefore, the oligomeric state of the RNP complex should be probed prior to the SANS experiments with biophysical techniques²⁹ in **H₂O and in D₂O**.
- 3) *Repulsive inter-particle interactions*: Highly charged biomolecules such as RNA repel each other in solution, if not properly screened by an appropriate buffer. The effects of repulsive interactions can be observed in SANS experiments at low angles and influence the values of radii of gyration and molecular weight³¹. Usually, the effects tend to increase with increasing concentration and should be avoided. The presence of inter-particle repulsion can be probed by a concentration series, typically in the range 1 to 20 mg/mL. If all curves superpose (after normalization by concentration), interaction effects can in general be excluded. If interaction effects are present, two strategies can be explored to reduce them: a) lowering the sample concentration as far as possible in order to minimize the effects and merging a low-concentration SANS curve at low angles with a high-concentration curve at higher angles^{31, 53}. b) Changing the buffer properties (e.g. ionic strength, pH) in order to eliminate the long-range electrostatic effects.

4) *Solvent subtraction*: Even in a carefully prepared SANS experiment the scattering intensity of the sample might not align with the corresponding buffer at large angles. Since the background level at high angles originates mainly from incoherent scattering of hydrogen atoms, this discrepancy is in most cases due to slightly different levels of H₂O:D₂O in the sample and buffer. A minor scaling of the buffer level (< 2%) can remedy this situation and can be tolerated. If significantly higher discrepancies are observed, the transmission values (which depend strongly on the H₂O:D₂O ratio) of the samples and buffers should be compared.

Concluding remarks: At all stages of a SANS experiment, it is advisable to discuss with a local contact at the neutron facility (instrument responsible or experienced user) that can help out with i) successful proposal writing and submission, ii) strategies of sample preparation, iii) measuring the SANS samples and iv) data analysis and interpretation, in particular assessing the uniqueness of a structural model.

Abbreviations

AUC: Analytical Ultracentrifugation

CMP: Contrast Match Point

CSP: Chemical Shift Perturbation

DLS: Dynamic Light Scattering

DNA: Deoxyribonucleic Acid

DTT: Dithiothreitol

EM: Electron Microscopy

EMSA: Electrophoretic Mobility Shift Assay

EPR: Electron Paramagnetic Resonance

FRET: Fluorescence Resonance Energy Transfer

gRNA: Guide Ribonucleic Acid

kDa: kilo Dalton

LC-MS: Liquid Chromatography Mass Spectrometry

MALLS: Multiangle Laser Light Scattering

MDa: Mega Dalton

MS: Mass Spectrometry

MW: Molecular Weight

NOE: Nuclear Overhauser Effect

NMR: Nuclear Magnetic Resonance

NTP: Nucleoside Triphosphate

PAGE: Polyacrylamide Gel Electrophoresis

PDB: Protein Data Bank

PRE: Paramagnetic Relaxation Enhancement

RNA: Ribonucleic Acid

RNP: RNA-protein

rRNA: Ribosomal Ribonucleic Acid

SANS: Small Angle Neutron Scattering

SAS: Small Angle Scattering

SAXS: Small Angle X-ray Scattering

SDS-PAGE: Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

SEC: Size Exclusion Chromatography

SLD: Scattering Length Density

SLS: Static Light Scattering

SRP: Signal Recognition Particle

TCEP: Tris(2-carboxyethyl)phosphine

TOF-MS: Time-Of-Flight Mass Spectrometry

tRNA: Transfer Ribonucleic Acid

Acknowledgements

FG acknowledges funding from the French Agence Nationale de la Recherche (Grant ANR-15-CE11-0026-01 “PROTstretch”). TC acknowledges funding from the DFG (grants CA294/3-2 and CA294/11-1).

References

- [1] Jacrot, B., Chauvin, C., and Witz, J. (1977) Comparative neutron small-angle scattering study of small spherical RNA viruses, *Nature* 266, 417-421.
- [2] Giegé, R., Jacrot, B., Moras, D., Thierry, J.-C., and Zaccai, G. (1977) A neutron investigation of yeast valyl-tRNA synthetase interaction with tRNAs, *Nucleic Acids Res* 7, 2421-2427.
- [3] Dessen, P., Blanquet, S., Zaccai, G., and Jacrot, B. (1978) Antico-operative Binding of Initiator Transfer RNAMet to Methionyl-transfer RNA Synthetase from Escherichia coli: Neutron Scattering Studies *Journal of Molecular Biology* 126, 293-313.
- [4] Hoppe, W., May, R., Stoeckel, P., Lorenz, S., Erdmann, V. A., Wittmann, H. G., Crespi, H. L., Katz, J. J., and Ibel, K. (1975) Neutron Scattering Measurements with the Label Triangulation Method on the 50S Subunit of E. Coli Ribosomes, In *Neutron Scattering for the Analysis of Biological Structures* (Schoenborn, B. P., Ed.), pp IV38-IV48, Brookhaven National Laboratory, Brookhaven.
- [5] Stuhrmann, H. B., Haas, J., Ibel, K., Wolf, B., Koch, M. H., Parfait, R., and Crichton, R. R. (1976) New low resolution model for 50S subunit of Escherichia coli ribosomes, *Proc Natl Acad Sci U S A* 73, 2379-2383.
- [6] Engelman, D. M., Moore, B. P., and Schoenborn, B. P. (1975) Neutron scattering measurements of separation and shape of proteins in 30S ribosomal subunit of Escherichia coli: S2-S5, S5-S8, S3-S7., *Proc Natl Acad Sci U S A* 72, 3888-3892.
- [7] Engelman, D. M., and Moore, P. B. (1972) A new method for the determination of biological quaternary structure by neutron scattering, *Proceedings of the National Academy of Sciences USA* 69, 1997-1999.
- [8] Hoppe, W. (1972) A New X-Ray Method for the Determination of the Quaternary Structure of Protein Complexes, *Isr J Chem* 10, 321-333.
- [9] Stuhrmann, H. B., and Kirste, R. G. (1965) Elimination der intrapartikulären Untergrundstreuung bei der Roentgenkleinwinkelstreuung an kompakten Teilen (Proteinen), *Zeitschrift fuer Physikalische Chemie* 46, 247-250.
- [10] Pardon, J. F., Worcester, D. L., Wooley, J. C., Tatchell, K., Van Holde, K. E., and Richards, B. M. (1975) Low-angle neutron scattering from chromatin subunit particles, *Nucleic Acids Res* 2, 2163-2176.
- [11] Baldwin, J. P., Boseley, P. G., Bradbury, E. M., and Ibel, K. (1975) The subunit structure of the eukaryotic chromosome, *Nature* 253, 245-249.
- [12] Bram, S., Butler-Browne, G., Baudy, P., and Ibel, K. (1975) Quaternary Structure of Chromatin, *Proc Natl Acad Sci U S A* 72, 1043-1045.
- [13] Jacrot, B. (1976) The study of biological structures by neutron scattering from solution, *Reports on Progress in Physics* 39, 911-953.
- [14] Zaccai, G., and Jacrot, B. (1983) Small Angle Neutron Scattering, *Annual Review of Biophysics and Biomolecular Structure* 12, 139-157.
- [15] Glatter, O., and Kratky, O. (1982) *Small Angle X-ray Scattering*, Academic Press, London, New York, Paris, San Diego, San Francisco, Sao Paolo, Sydney, Tokyo, Toronto.
- [16] Kratky, O. (1972) Recent advances and applications of diffuse X-ray small-angle scattering on biopolymers in dilute solutions, *Quarterly Reviews of Biophysics* 5, 481-537.
- [17] Svergun, D. I., Barberato, C., and Koch, M. H. J. (1995) CRY SOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates, *Journal of Applied Crystallography* 28, 768-773.
- [18] Svergun, D. I., Richard, S., Koch, M. H. J., Sayers, Z., Kuprin, S., and Zaccai, G. (1998) Protein hydration in solution: experimental observation by x-ray and neutron scattering, *Proceedings of the National Academy of Sciences USA* 95, 2267-2272.
- [19] Svergun, D. I., and Nierhaus, K. H. (2000) A Map of Protein-rRNA Distribution in the 70S Escherichia coli Ribosome, *J Biol Chem* 275, 14432-14439.
- [20] Svergun, D. I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing, *Biophysical Journal* 76, 2879-2886.
- [21] Mertens, H. D. T., and Svergun, D. I. (2017) Combining NMR and small angle X-ray scattering for the study of biomolecular structure and dynamics, *Archives of biochemistry and biophysics* 628, 33-41.

- [22] Hennig, J., and Sattler, M. (2014) The Dynamic Duo: Combining NMR and Small Angle Scattering in Structural Biology, *Protein Science*.
- [23] Carlomagno, T. (2014) Present and future of NMR for RNA-protein complexes: A perspective of integrated structural biology, *Journal of Magnetic Resonance* 241, 126-136.
- [24] Madl, T., Gabel, F., and Sattler, M. (2011) NMR and small-angle scattering-based structural analysis of protein complexes in solution, *J Struct Biol* 173, 472-482.
- [25] Svergun, D. I., Koch, M. H. J., Timmins, P. A., and May, R. P. (2013) *Small Angle X-ray and Neutron Scattering from Solutions of Biological Macromolecules*, Oxford University Press, Oxford.
- [26] Gutberlet, T., Rücker, U., and Brückel, T. (2017) Towards Compact Accelerator Driven Neutronsources for Europe, *Neutron News* 28, 20-25.
- [27] Rauch, H., and Waschkowski, W. (2001) Neutron Scattering Lengths, In *Neutron Data Booklet* (Dianoux, A.-J., and Lander, G., Eds.), Institut Laue-Langevin, Grenoble.
- [28] Jacrot, B., and Zaccai, G. (1981) Determination of Molecular Weight by Neutron Scattering, *Biopolymers* 20, 2413-2426.
- [29] Jeffries, C. M., Graewert, M. A., Blanchet, C. E., Langley, D. B., Whitten, A. E., and Svergun, D. I. (2016) Preparing monodisperse macromolecular samples for successful biological small-angle X-ray and neutron-scattering experiments, *Nat Protoc* 11, 2122-2153.
- [30] Jacques, D. A., and Trewella, J. (2010) Small-angle scattering for structural biology - Expanding the frontier while avoiding the pitfalls, *Protein Science* 19, 642-657.
- [31] Putnam, C. D., Hammel, M., Hura, G. L., and Tainer, J. A. (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution, *Quarterly Reviews of Biophysics* 40, 191-285.
- [32] Zaccai, N. R., Serdyuk, I. N., and Zaccai, J. (2017) *Methods in Molecular Biophysics: Structure, Dynamics, Function for Biology and Medicine*, 2 ed., Cambridge University Press, Cambridge.
- [33] Philo, J. S. (2006) Is any measurement method optimal for all aggregate sizes and types?, *The AAPS journal* 8, E564-E571.
- [34] Porterfield, J. Z., and Zlotnick, A. (2010) A simple and general method for determining the protein and nucleic acid content of viruses by UV absorbance, *Virology* 407, 281-288.
- [35] Milligan, J. F., Groebe, D. R., Witherell, G. W., and Uhlenbeck, O. C. (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates, *Nucleic acids research* 15, 8783-8798.
- [36] Haertlein, M., Moulin, M., Devos, J. M., Laux, V., Dunne, O., and Trevor Forsyth, V. (2016) Chapter Five - Biomolecular Deuteration for Neutron Structural Biology and Dynamics, In *Methods in Enzymology* (Kelman, Z., Ed.), pp 113-157, Academic Press.
- [37] Hattori, A., Crespi, H. L., and Katz, J. J. (1965) Effect of Side-Chain Deuteration on Protein Stability, *Biochemistry* 4, 1213-1225.
- [38] Makhatadze, G. I., Clore, G. M., and Gronenborn, A. M. (1995) Solvent isotope effect and protein stability, *Nat Struct Biol* 2, 852-855.
- [39] Rio, D. C. (2012) Filter-Binding Assay for Analysis of RNA-Protein Interactions, *Cold Spring Harbor Protocols* 2012, pdb.prot071449.
- [40] Ryder, S. P., Recht, M. I., and Williamson, J. R. (2008) Quantitative analysis of protein-RNA interactions by gel mobility shift, *Methods in molecular biology (Clifton, N.J.)* 488, 99-115.
- [41] Recht, M. I., Ryder, S. P., and Williamson, J. R. (2008) Monitoring assembly of ribonucleoprotein complexes by isothermal titration calorimetry, *Methods in molecular biology (Clifton, N.J.)* 488, 117-127.
- [42] Lapinaite, A., Simon, B., Skjaerven, L., Rakwalska-Bange, M., Gabel, F., and Carlomagno, T. (2013) The structure of the box C/D enzyme reveals regulation of RNA methylation, *Nature* 502, 519-523.
- [43] Feigin, L. A., and Svergun, D. I. (1987) *Structure Analysis by Small-Angle X-Ray and Neutron Scattering*, Plenum Press, New York and London.
- [44] Guinier, A., and Fournet, G. (1955) *Small Angle Scattering of X-rays*, John Wiley & Sons, New York.

- [45] Voss, N. R., and Gerstein, M. (2005) Calculation of Standard Atomic Volumes for RNA and Comparison with Proteins: RNA is Packed More Tightly, *Journal of Molecular Biology* 346, 477-492.
- [46] Sarachan, K. L., Curtis, J. E., and Krueger, S. (2014) Small-angle scattering contrast calculator for protein and nucleic acid complexes in solution, *Journal of Applied Crystallography* 46, 1889-1893.
- [47] Whitten, A. E., Cai, S., and Trehwella, J. (2008) MULCh: modules for the analysis of small-angle neutron contrast variation data from biomolecular assemblies, *Journal of Applied Crystallography* 41, 222-226.
- [48] Hennig, J., Militti, C., Popowicz, G. M., Wang, I., Sonntag, M., Geerlof, A., Gabel, F., Gebauer, F., and Sattler, M. (2014) Structural basis for the assembly of the Sxl-Unr translation regulatory complex, *Nature* 515, 287-290.
- [49] Mahieu, E., and Gabel, F. (2018) Biological small-angle neutron scattering: recent results and development, *Acta Crystallographica Section D* 74, 715-726.
- [50] Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., and Svergun, D. I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis, *Journal of Applied Crystallography* 36, 1277-1282.
- [51] Gabel, F., Bicout, D., Lehnert, U., Tehei, M., Weik, M., and Zaccai, G. (2002) Protein dynamics studied by neutron scattering, *Quarterly Reviews of Biophysics* 35, 327-367.
- [52] Guinier, A. (1939) La diffraction des Rayons X aux Tres Faibles Angles: Applications a l'Etude des Phenomenes Ultra-microscopiques, *Annales de Physique (Paris)* 12, 161-236.
- [53] Kim, Henry S., Martel, A., Girard, E., Moulin, M., Härtlein, M., Madern, D., Blackledge, M., Franzetti, B., and Gabel, F. (2016) SAXS/SANS on Supercharged Proteins Reveals Residue-Specific Modifications of the Hydration Shell, *Biophysical Journal* 110, 2185-2194.
- [54] Kim, H. S., and Gabel, F. (2015) Uniqueness of quasi-atomic bio-macromolecular models derived from small angle scattering data, *Acta Crystallogr D Biol Crystallogr*.
- [55] Jacrot, B. (1976) The study of biological structures by neutron scattering from solution, *Reports on Progress in Physics* 39, 911-953.
- [56] Grotwinkel, J. T., Wild, K., Segnitz, B., and Sinning, I. (2014) SRP RNA Remodeling by SRP68 Explains Its Role in Protein Translocation, *Science* 344, 101.
- [57] Wang, Y.-X., Zuo, X., Wang, J., Yu, P., and Butcher, S. E. (2010) Rapid global structure determination of large RNA and RNA complexes using NMR and small-angle X-ray scattering, *Methods (San Diego, Calif.)* 52, 180-191.
- [58] Falb, M., Amata, I., Gabel, F., Simon, B., and Carlomagno, T. (2010) Structure of the K-turn U4 RNA: a combined NMR and SANS study, *Nucleic Acids Res* 38, 6274-6285.
- [59] Grishaev, A., Wu, J., Trehwella, J., and Bax, A. (2005) Refinement of Multidomain Protein Structures by Combination of Solution Small-Angle X-ray Scattering and NMR Data, *Journal of the American Chemical Society* 127, 16621-16628.
- [60] Trabuco, L. G., Villa, E., Mitra, K., Frank, J., and Schulten, K. (2008) Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics, *Structure (London, England : 1993)* 16, 673-683.
- [61] Karaca, E., Rodrigues, J. P. G. L. M., Graziadei, A., Bonvin, A. M. J. J., and Carlomagno, T. (2017) M3: an integrative framework for structure determination of molecular machines, *Nature Methods* 14, 897.
- [62] Sonntag, M., Jagtap, P. K. A., Simon, B., Appavou, M.-S., Geerlof, A., Stehle, R., Gabel, F., Hennig, J., and Sattler, M. (2017) Segmental, Domain-Selective Perdeuteration and Small-Angle Neutron Scattering for Structural Analysis of Multi-Domain Proteins, *Angewandte Chemie International Edition* 56, 9322-9325.
- [63] Michel, E., and Allain, F. H. T. (2015) Chapter Sixteen - Selective Amino Acid Segmental Labeling of Multi-Domain Proteins, In *Methods in Enzymology* (Kelman, Z., Ed.), pp 389-422, Academic Press.
- [64] Freiburger, L., Sonntag, M., Hennig, J., Li, J., Zou, P., and Sattler, M. (2015) Efficient segmental isotope labeling of multi-domain proteins using Sortase A, *Journal of Biomolecular NMR* 63, 1-8.

- [65] Michel, E., Skrisovska, L., Wüthrich, K., and Allain, F. H. (2013) Amino acid-selective segmental isotope labeling of multidomain proteins for structural biology, *Chembiochem* 14, 457-466.
- [66] Duss, O., Lukavsky, P. J., and Allain, F. H. (2012) Isotope labeling and segmental labeling of larger RNAs for NMR structural studies, *Adv Exp Med Biol* 992, 121-144.
- [67] Nelissen, F. H. T., van Gammeren, A. J., Tessari, M., Girard, F. C., Heus, H. A., and Wijmenga, S. S. (2008) Multiple segmental and selective isotope labeling of large RNA for NMR structural studies, *Nucleic acids research* 36, e89-e89.
- [68] Tzakos, A. G., Easton, L. E., and Lukavsky, P. J. (2007) Preparation of large RNA oligonucleotides with complementary isotope-labeled segments for NMR structural studies, *Nature Protocols* 2, 2139.
- [69] Nemethy, G., and Scheraga, H. A. (1962) Structure of Water and Hyrdophobic Bonding in Proteins. I. A model for the Thermodynamic Properties in Liquid Water, *The Journal of Chemical Physics* 36, 3382-3417.