

The basics of small-angle neutron scattering (SANS for new users of structural biology)

Cy M. Jeffries^{1,*}, Zuzanna Pietras², and Dmitri I. Svergun¹

¹European Molecular Biology Laboratory (EMBL), Hamburg Unit, c/o Deutsches Elektronen-Synchrotron, 22607 Hamburg, Germany

²Department of Physics, Chemistry, and Biology, Linköping University, 581 83 Linköping, Sweden

Abstract. Small-angle neutron scattering (SANS) provides a means to probe the time-preserved structural state(s) of bio-macromolecules in solution. As such, SANS affords the opportunity to assess the redistribution of mass, i.e., changes in conformation, which occur when macromolecules interact to form higher-order assemblies and to evaluate the structure and disposition of components within such systems. As a technique, SANS offers scope for ‘out of the box thinking’, from simply investigating the structures of macromolecules and their complexes through to where structural biology interfaces with soft-matter and nanotechnology. All of this simply rests on the way neutrons interact and scatter from atoms (largely hydrogens) and how this interaction differs from the scattering of neutrons from the nuclei of other ‘biological isotopes’. The following chapter describes the basics of neutron scattering for new users of structural biology in context of the neutron/hydrogen interaction and how this can be exploited to interrogate the structures of macromolecules, their complexes and nano-conjugates in solution.

1 Introduction

Structural biologists have a vast array of techniques at their disposal to study the structures of proteins and other bio-macromolecules. The recent ‘resolution revolution’ in cryo-electron microscopy [1-4] has made a significant impact in the field while X-ray crystallography remains as the key method for determining atomic-level structural information [5]. Nuclear magnetic resonance spectroscopy continues to advance [6], with ever-more powerful instruments. In this arguably daunting environment, small-angle neutron scattering (SANS) affords an additional means to determine the low-resolution structures and dispositions of single- and multi-component bio-macromolecular systems [7-16]. Unlike cryo-EM (that, by definition, is performed in a frozen state), or X-ray crystallography (that requires a diffraction-quality crystal), samples for SANS experiments are typically prepared in aqueous buffers [17], although there is no restriction on sample-type or environment. Gels [18, 19], conjugates [20], phases [21]; trapped/confined/emulsion states [22, 23]; shear or pressure states [24] are all amenable to SANS. The reason for this diversity is due to the way

* Corresponding author: cy.jeffries@embl-hamburg.de

in which neutrons interact with matter: Neutrons scatter predominantly from atomic nuclei and do so depending on the isotope-specific neutron-nuclei interaction. As such, neutrons are both deeply-penetrating and, unlike X-rays that are scattered or absorbed from electrons, do not directly affect electronic transitions, i.e., are relatively chemically inert (although their absorption into materials and the production of secondary α - β and γ radiation over time will result in deleterious biological effects). Additional physical parameters of the neutron are summarized in Figure 1.

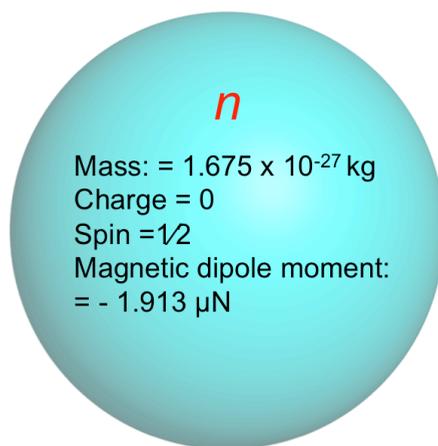


Fig. 1. *General properties of the neutron.*

Neutrons have a mass, zero charge, a spin-state of $1/2$ and a magnetic dipole moment and may be described as particles or as waves with a wavelength λ , i.e., display ‘wave-particle duality’. The magnetic dipole moment may be utilized for magnetic neutron scattering, although this is rarely employed in biological investigations. For most structural biology experiments, SANS focuses on measuring the elastic (i.e. without energy transfer) and coherent (i.e. direction-dependent) scattering of neutrons (Figure 2) from the atomic nuclei of the ‘biological isotopes’ found in proteins, polynucleotides, carbohydrates, lipids, etc.

gyration, R_g [25, 26], and maximum particle dimension, D_{max} , as well as the probable frequency of scattering-pair distances in real space, or $p(r)$ [27, 28], that may be used to build low-resolution models of macromolecules, their complexes and higher-order assemblies[29, 30]. All of this largely depends on the unique way neutrons scatter from the most abundant isotope in the biological/universal sphere, regular light hydrogen, ^1H [15].

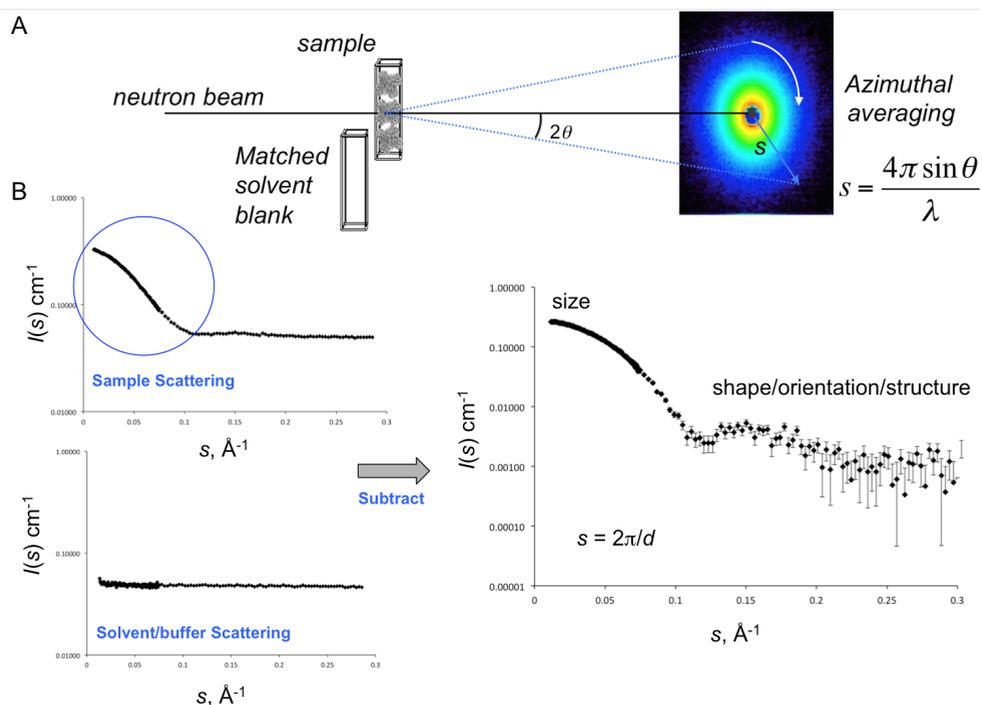


Fig. 3. SANS data measurement and reduction. **A.** A neutron hits a dilute solution of isotopically tumbling macromolecules in solution. In very rare instances, momentum is transferred to the neutron resulting in a scattering event at a particular scattering angle 2θ . The intensity of the scattering is recorded on a 2D-detector vs. the momentum transfer-vector with a modulus s , in units of inverse-distance (i.e., the scattering is recorded in momentum- or ‘reciprocal-space’). The intensities recorded for each value of s are azimuthally averaged (a.k.a. radial averaging) to reduce the 2D-data to 1D-data, i.e., the neutron scattering intensities are plotted as $I(s)$ vs. s . **B.** The sample scattering consists of scattering contributions from the macromolecules in the sample *plus* neutron scattering contributions from the solvent. Preserved spatial correlations within the volume boundary of the macromolecules of the sample, give rise to a summed-coherent SANS signal ‘over-and-above’ the solvent scattering. The buffer does scatter, but for the distances probed in the SANS regime there are no preserved long-distance spatial correlations in the solvent, yielding ‘flat scattering’. Solvent scattering contributions are subtracted from the sample scattering after the measurement of an exactly-matched solvent blank. After background subtraction and transmission corrections are applied (to take into account any differences in neutron absorption between the sample and the solvent) what remains are the coherent scattering intensities derived from the population of macromolecules in the sample. The intensity of coherent scattering signal encodes the frequency of preserved spatial correlations between scattering centres internal to the macromolecules, i.e., the size, shape and structure.

2 Neutrons

The detailed physics and mathematics underpinning neutron scattering are interesting but generally impenetrable for the non-physicist [31]. Since the discovery of the neutron by

James Chadwick in 1932 [32], and the subsequent pioneering work of Lise Meitner, Otto Frisch, Otto Hahn, and Fritz Straßmann [33, 34], nearly a century of highly dedicated theoretical and experimental research has taken place to describe what a neutron is and what it does when interacting with a nucleus. For the purposes of this chapter, the descriptions are overly-simplified and are provided to illustrate link between the underlying principles of neutron scattering and how it applies to structural biology.

There are three main outcomes when illuminating a sample with a neutron: 1) Nothing happens, i.e., the neutron travels through the sample and is transmitted; 2) The neutron may be absorbed by a nucleus that may result in secondary fissile events and the production of both ionizing and non-ionizing radiation and 3) Scatter (Figure 4). Neutron scattering events can be further sub-divided into:

- i) Elastic scattering, where the neutron does not loose energy after scattering from a nucleus;
- ii) Inelastic scattering, where the neutron changes energy after the scattering event;
- iii) Nuclear-potential scattering where the scattered neutron from the sample relates to the spatial correlations between nuclei within the sample. If the distances between nuclei are preserved over time, this gives rise to a coherent SANS signal;
- iv) Nuclear-spin scattering, where the scattered neutron does not necessarily relate to distance correlations between atom-pairs per-se, but relates to nuclear-spin orientations/correlations and internal-structural/time-dynamics of nuclear spin-transitions.
- v) Combinations of the above and;
- vi) ‘Magnetic scattering’ from unpaired-electron or exotic-magnetic lattice systems.

The focus of this chapter is points i) and iii), that is, utilizing the *elastic-coherent neutron scattering component* from which structure of macromolecules in solution or the disposition of macromolecular components within fully-formed complexes can be investigated.

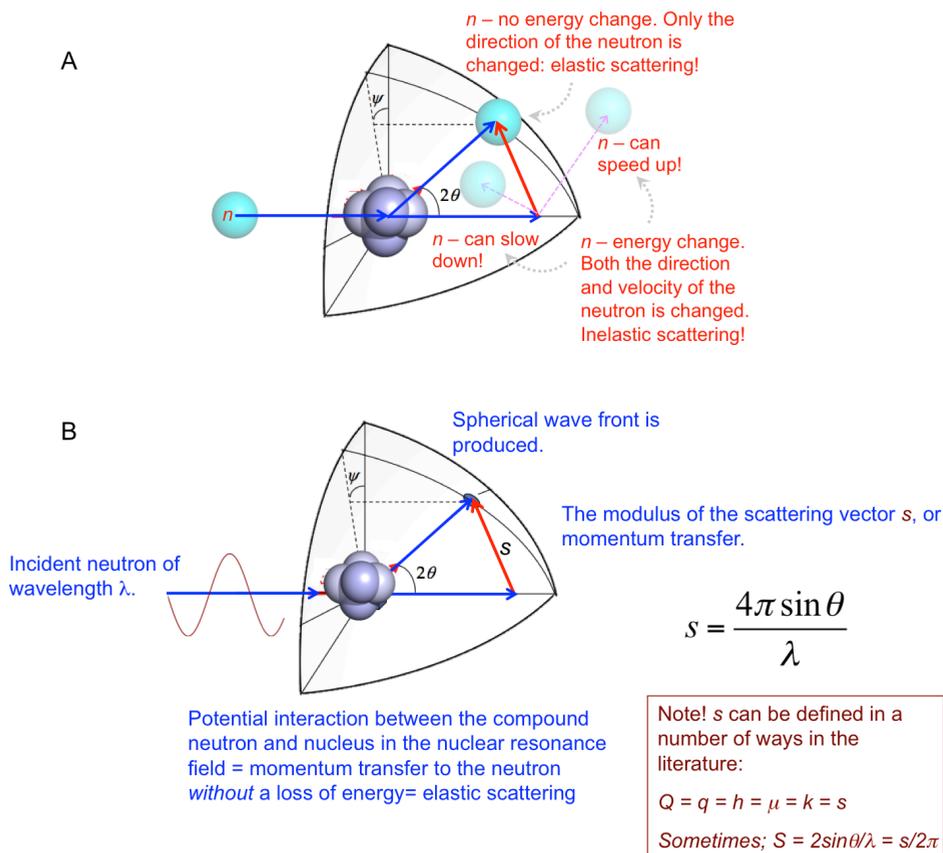


Fig. 4. Neutron scattering interactions with a nucleus. **A.** Neutrons can scatter from a nucleus, i.e., the transfer of momentum, such that they can speed up, slow down, and change direction. Elastic neutron scattering events yield in a change in angular momentum, without a loss in energy, i.e., a change in direction. Inelastic scattering events result in both a change in direction and speed of the neutron. **B.** SANS for structural biology relies on measuring the elastic scattering events arising from a nucleus-neutron interaction. When the incident neutron is viewed as a wave, the probability that a neutron will scatter at any point in any direction in any given time can, when viewed over all possible scattering angles 2θ , be mapped in terms of the wave-amplitude into a space that traces out a ‘spherical wave front’ where 2θ is described in terms of any point on this spherical wave by the momentum transfer, s .

2.1 Neutrons as waves...that can flip

It is perhaps easiest to understand of what a SANS pattern encodes if neutrons are conceived as waves. As with all waves, neutrons have a wavelength (that is inversely proportional to their energy), an amplitude, and a phase. In addition, the nucleus of each individual isotope has a probability to elastically scatter an incoming neutron through any given solid angle in a given time. In effect, when viewing the probabilistic scattering event over all angles, each nucleus in the sample effectively acts as a point source over time, with the scattered neutron propagating and mapping to a spherical wave front (or ‘S-wave scattering’). The actual probability of a nucleus to scatter a neutron through any given solid angle per unit time is referred to as a differential cross section, that can be conceptualized as a circle, where the radius relates to what is called the scattering length (in cm) of that specific isotope [35]. The larger the circle, the longer the scattering length and a corresponding

increased probability of an elastic neutron-nucleus scattering event (Figure 5). A cross section is essentially a measure of the probability of a nucleus-neutron interaction; there are cross sections that describe the elastic nuclear-potential scattering component, or coherent cross-section – the primary focus of this chapter – as well as elastic spin-scattering (incoherent cross section) and absorption components (indeed, any type of probabilistic interaction of a neutron with a nucleus may be described by a cross section). In the case of a standard SANS experiment, the isotope-specific scattering length composition of a sample per unit volume, i.e., the scattering length density (Figure 5), and the resulting summed convolution of all elastically-scattered neutron amplitudes emanating from each nucleus within the sample is what generates scattering intensities. If the sample has components within it that have preserved point-to-point distance correlations between atomic nuclei internal to their structure, e.g., due to the macromolecule tumbling in solution, the subsequent development of summed elastic-coherent S-wave amplitudes caused by the spatial correlation between all time-preserved scattering centres within the macromolecule is what eventually yields structural information.

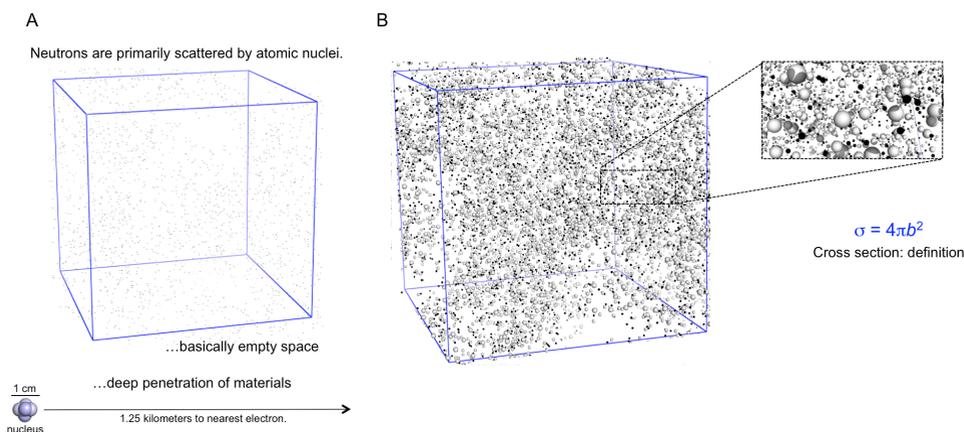


Fig. 5. What neutrons ‘experience’ on approaching a block of regular matter. **A.** Neutrons primarily interact with the nuclei of a material that on an atomic scale are extremely far apart from each other. For example, if a single atom nucleus were scaled up to be 1 cm across, the nearest electron of the same atom would be 1.25 km away. Matter is basically empty space. This is the reason why neutrons are deeply penetrating. **B.** However, each atomic isotope has a different probability to interact with a neutron that can be viewed as ‘conceptual circles’ occupying the block of matter. These circles are more formally described by the cross section of an isotope, which simply quantifies an atomic nucleus’s capacity to interact with a neutron. There are cross sections that describe elastic and inelastic scattering interactions as well as absorption. For neutron scattering, the cross section relates to the square of the scattering length, b (cm), where b is, in effect, the scaled radius of the ‘conceptual circles’. Different isotopes have different scattering lengths and therefore different probabilities to interact with an incoming neutron.

One key advantage of SANS is that it is possible to experimentally control the magnitude of the SANS interference pattern. Somewhat intuitively, the SANS intensities can be changed by simply changing the isotopic composition of the sample, i.e., the isotopic composition of the sample-solvent or the macromolecules themselves. Swapping out one isotope for another should, and does, alter both the coherent-scattering S-wave front as well as incoherent spin-scattering contributions. Coincidentally, deuterium (^2H) and the nuclei of commonly occurring isotopes found in biological macromolecules (^{12}C , ^{16}O , ^{14}N , ^{31}P , and mainly ^{32}S) interact with the nuclear potential component to coherently scatter neutrons with a similar

magnitude, whereas neutrons scattered from ^1H undergo the ‘inverse’ of what happens compared to the nuclei of ^2H and the other biological-elements.

Many of the isotopes encountered in biology have what is termed a positive coherent neutron scattering length, while simple ^1H has a *negative* scattering length. A negative scattering length is a terminology that means ^1H effectively experiences an attractive energy potential when interacting with neutrons so that the out-going scattered neutron maintains the same phase as it was prior to scattering, i.e., the amplitude of the scattered neutron does not undergo a phase shift relative to the incident neutron (Figure 6). For the nuclei of other biological isotopes and deuterium, a repulsive energy potential interaction with neutrons is set up such that the out-going scattered neutron amplitude undergoes a 180° phase shift (or phase inversion) relative to an incident neutron, i.e., have a positive scattering length. So, neutrons experiencing nuclear-potential scattering from ^1H are 180° out of phase with neutrons scattered from most other biological isotopes. Importantly, ^2H also has a positive scattering length and therefore the isotopic substitution of ^1H for ^2H , which alters the average ^1H per unit volume of a sample, will systematically alter the overall magnitude of the summed scattering amplitudes: the negative scattering length of ^1H is effectively cancelled by the positive scattering length of ^2H . SANS can be conceptualized as a ‘sum of the waves-game’ where the scattering amplitudes can be experimentally controlled by ^1H - ^2H substitution, for example in the supporting background solvent, or buffer (adjusting $^1\text{H}_2\text{O}$: $^2\text{H}_2\text{O}$ ratios) or through the substitution of ^1H with ^2H at non-exchangeable hydrogen positions within a macromolecule. Note that for exchangeable hydrogens, e.g. those belonging to the hydrophilic groups in the macromolecule, ^1H is replaced by ^2H proportionally to the percentage of heavy water in solution. Isotopic substitution also affects the magnitude of the incoherent neutron spin-scattering contributions, but in this instance all of the biological isotopes and ^2H have, either a positive scattering length, or lack the capacity to undergo spin scattering due to the spin-state of the particular isotope. For biological SANS performed on dilute macromolecular samples in aqueous solution the isotopic substitution of ^1H with ^2H will result in a decrease in overall spin-scattering contributions as the incoherent scattering length of ^1H is enormous in comparison to ^2H .

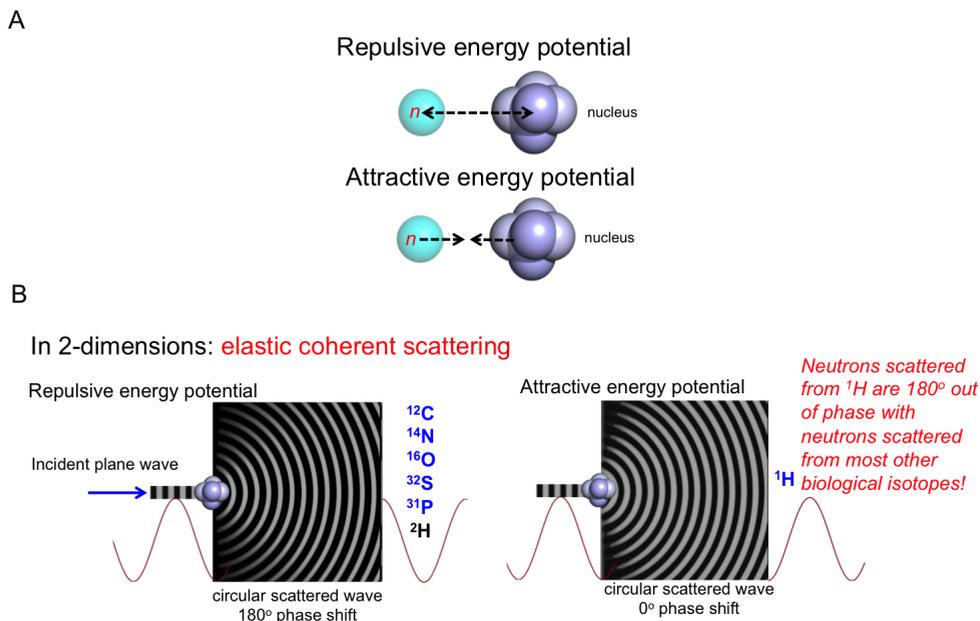


Fig.6. Nuclear potential interactions and the development of elastic-coherent neutron scattering amplitudes. **A.** A neutron may experience either repulsive or negative energy potentials when interacting with a specific type of atomic nucleus that is dependent on the isotope. **B. Left:** For most of the biological isotopes and deuterium, ^2H , neutrons generate a repulsive-potential with the nucleus that results in a 180° ‘flip’ in the phase of the elastically scattered circular-wave, or S-wave, neutron amplitude relative to the amplitude of the incident neutron. This effect is termed a positive scattering length. **Right:** Neutrons scattered from regular light hydrogen, ^1H , experience a repulsive-potential such that the phase-amplitude of the incident neutron is maintained, i.e., the scattered neutron does not undergo an amplitude ‘phase-flip’ like the other biological isotopes. Therefore, ^1H has what is termed a *negative* coherent scattering length, i.e., the scattering amplitudes from ^1H are 180° out of phase relative to the neutrons scattered from most of the other biological isotopes.

2.2 The sum of the waves game

The constructive and destructive interference of the scattered neutron amplitudes from each nucleus of a sample gives rise to a convoluted wave front. For a single nucleus, the scattering amplitude remains constant over all angles, only changing in magnitude depending on the scattering length. In other words, the coherent-scattering neutron atomic form factors of the isotopes are effectively angle-independent. However, if *two* scattering centres, related by a time-preserved distance r , are present (Figure 7), then lines of constructive and destructive interference are set up in the combined coherent wave front caused by the amplitudes of the two S-waves from each point source interacting and summing together. The most intense line of constructive interference occurs at zero angle, while as the angle increases, the amplitudes undergo angular-dependent ‘highs and lows’ as the amplitudes arising from both scattering centres constructively and destructively interfere, weighted by the magnitude of the scattering length. The convoluted zero angle amplitudes relate to the combined sum of the individual scattering amplitudes from each scattering centre weighted by the magnitude of the scattering length, while the ‘highs and lows’ in the amplitudes of the scattered neutron waves at ever-increasing angle relate to the distance separation between the centres, weighted by the scattering lengths. Of course, macromolecules in solution that are tumbling about have many scattering pair distance correlations within the extent of their volume boundary and therefore the coherent wave front from them is derived from the combined sum of the scattering lengths

and the scattered wave amplitudes arising from these correlations – time and rotationally averaged – as a function of angle. The lower the angle (lower s), the longer the correlated distances d , such that s and d share a reciprocal relationship, $s = 2\pi/d$, while the summed zero-angle amplitudes reflects the total summed scattering from all scattering pairs within the macromolecular volume, and therefore relates to the molecular mass. It must also be remembered that the solution, i.e., the supporting solvent of the sample, also scatters. As the solvent (hopefully) does not have any time-preserved long-range distances correlations at similar length scales as the macromolecule (e.g., greater than the solvent hydrogen bonding network of 1.5–3 Å), its scattering contributions add as a spatially uncorrelated set of amplitudes that generates a 'flat background' which has to be subtracted to yield the coherent scattering of the macromolecule in solution. Therefore, two measurements have to be performed; on the sample and a corresponding exactly matched buffer to remove the scattering contributions made by the supporting solvent [17].

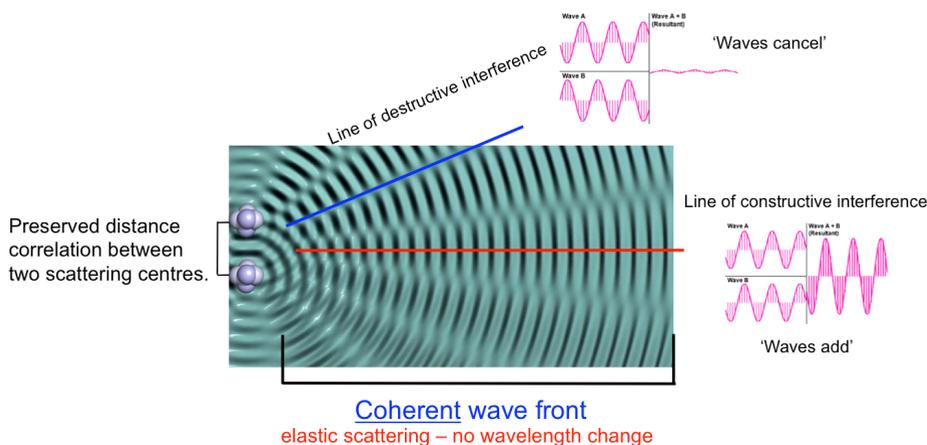


Fig. 7. *The sum of the waves game: preserved distance correlations and amplitude convolution.* If two scattering centres are spatially correlated through time by a distance d , the resulting S-wave scattering amplitudes emanating from each point interfere with each other. The scattering amplitudes, weighted by the scattering length of the isotope at each point, simply add together. The result is a combined coherent spherical wave front of scattering amplitudes that undergo constructive or destructive interference as a function of angle. The 'peaks and troughs' in the summed amplitude wave relate to the initial distance between the scattering centres. Unfortunately, it is not possible to access these scattering amplitudes experimentally: SANS records the intensity of the scattered neutrons that manifest as the square of the scattering amplitudes.

2.3 From nuclei to macromolecules: From reciprocal space amplitudes to real-space distances.

More formally, an incident neutron that makes up a collimated beam that travels toward a sample in a direction z , has what is termed a wave vector, \mathbf{k} , that also manifests in the same direction and relates to the neutron energy, i.e., its wavelength. Consequently, the incident neutron that is in effect collimated as a plane wave can be described by a complex-valued wave function:

$$\varphi_{incident} = e^{ikz} \quad (1).$$

When a nucleus gets in the way of this incident neutron and an elastic scattering event occurs, (i.e., the magnitude of the scattering vector, \mathbf{k} , does not change), then the resulting scattered neutron also follows a wave function weighted by the probability of a nucleus to scatter the incident neutron in any direction r , through a given time that when viewed over all scattering angles can be conceptualized as a ‘probabilistic spherical/circular wave’ (or the ‘ e^{ikr} ’ component in the following expression):

$$\varphi_{scattered} = -\frac{b}{r}e^{ikr} \quad (2).$$

The scattered neutron wavefunction is nominally negative to reflect the fact that most isotopes cause a phase inversion in the amplitudes of the incident neutron, i.e., have a positive scattering length, b , arising from the repulsive potential interaction between the incident neutron and nucleus. In turn, the scattering length of an isotope relates to the total scattering cross section, σ_c , which is a measure of the total number of neutrons scattered per second by that nucleus, normalized to the incident neutron flux, or perhaps more simply: The effective probability of a particular nucleus to scatter a neutron in any given time in any direction is:

$$\sigma_c = 4\pi b^2 \quad (3).$$

The magnitude of the scattering length, which is isotope specific, will then reflect the neutron ‘scattering ability’ for each nucleus and the production of the resulting spherical/circular-scattered wave amplitudes. Intuitively, if there are multiple isotopes in a sample, then the sum of their different scattering abilities, and the corresponding wavefunctions, will add together – for example, substituting ^1H with ^2H , i.e., from a negative b to a positive b , will radically alter the combined $\varphi_{scattered}$ i.e., the magnitude of the amplitudes of the scattered wave front.

As mentioned above, depending on the isotope in question, their initial energy states, the energy of the incident neutrons, energy transfer and spin interactions that occur within the compound neutron-nucleus, nuclear resonance states develop that lead to:

- Scattering (the real component of b).
- Absorption (imaginary components of b).

For the purposes of this chapter, absorption will not be described in detail, but basically involves the interaction of a neutron with a particular incident energy and a nucleus leading to the formation of a compound-neutron-nucleus that is commensurate in energy with an excited nuclear resonance-state (i.e., the compound-nucleus has a large absorption cross section). For most applications in biological SANS, the imaginary absorption component of the scattering length is small, i.e., the interactions within the compound-nucleus are away from the excited resonance states. The scattering events therefore arise from the real component of b . The total neutron scattering length, in turn, consists of two terms:

$$b_n = b_p + b_s \quad (4),$$

where the b_p represents the probabilistic interaction of a neutron with a nuclear potential – the coherent scattering length – and b_s represents its interaction with the nuclear spin (Figure 8).

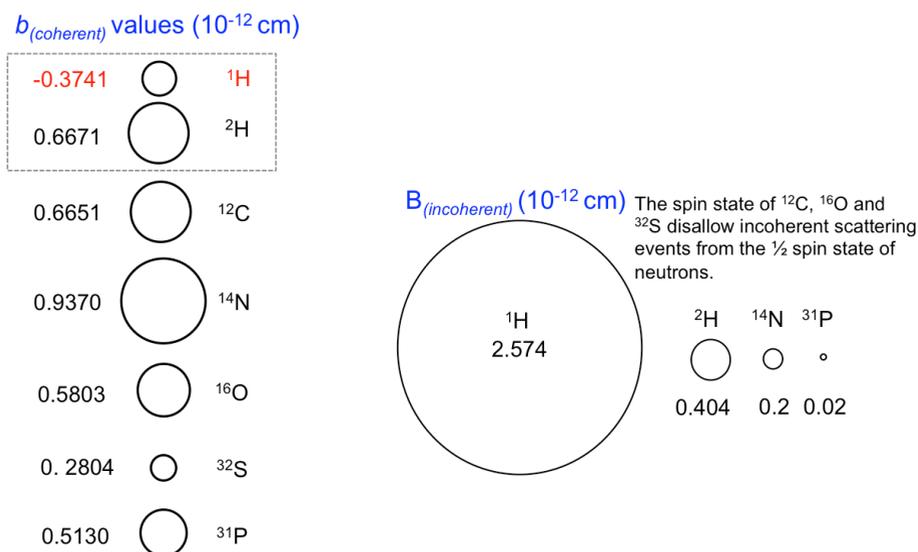


Fig. 8. Neutron scattering lengths of common biological isotopes. The magnitude of neutron scattering cross-sections represented as circles.

For SANS, the neutron-nuclear potential interaction, b_p , is important as this gives rise to elastic coherent scattering events that can be used to determine the atom-pair separation within a macromolecule. The spin scattering term, b_s , adds to this signal and can be expressed as:

$$b_s = 2B(IS) \quad (5),$$

where B is the spin scattering length of an isotope (I and S are spin operators of the neutron and nucleus). This additional spin-interaction and the resulting elastic scattering produced by it can cause a problem for biological SANS experiments. If the spins of the atoms comprising the sample and of the incoming neutrons are not ordered, neutrons will scatter incoherently. In other words, correlations between the wave-amplitudes of the spin-scattered neutrons and the distances between the spin-scattering centers no longer exist. Incoherent elastic spin-scattering produces additional, and sometimes significant, background scattering contributions depending on the isotope, and in particular light-hydrogen. In simple terms, the incoherent scattering amplitudes ‘ripple across’ the entire coherent amplitude wave front, convoluting with them across all angles to generate statistical fluctuations in the coherent amplitudes that manifest as noise in the measured SANS intensities. As most biological samples are weakly-scattering, are dominated by ^1H and are typically measured in a dilute-aqueous regime, without a spin-polarized neutron source or oriented nuclear spins, incoherent scattering ‘noise’ can come to dominate a scattering profile necessitating long exposure times to improve signal-to-noise ratios. As can be seen in Figure 8, regular ^1H is unusual compared to the other biological-isotopes in that its coherent scattering length is negative, and its incoherent scattering length is enormous. The practical consequences are that solvent-subtracted SANS patterns measured from macromolecules in solution in buffers containing high percentages of ^1H are, when measured for the same amount of time, quite noisy

compared to samples in high percentages of $^2\text{H}_2\text{O}$, i.e., where ^1H has been swapped out for ^2H in the buffer.

If the distance vectors, \vec{r} , between any two atoms are spatially correlated, then the magnitude of the amplitudes, A , of the convoluted S-wave amplitudes as a function of s (Figure 7) are proportionate to the sum of the individual atomic scattering factors for each isotopic nucleus (i.e., probability to scatter):

$$A(s) = \sum_{i=1}^N b_i e^{i\vec{s}\vec{r}} \quad (6).$$

The issue is that it is not possible to access the scattering amplitudes experimentally. SANS measures the intensity of the scattered neutrons, that is, the scattering amplitudes multiplied by their complex conjugate, or more simply, the absolute magnitude of the wave amplitudes squared:

$$I(s) = A(s)A(s)^* = |A(s)|^2 \quad (7).$$

Fundamentally, at zero angle, $I(0)$, the amplitudes from both nuclei will have a maximum-level of ‘in-phase constructive interference’. For macromolecules that contain numerous spatially-correlated scattering pairs – and after background scattering contributions have been subtracted – the combined sum of the squared amplitudes that manifest experimentally as $I(s)$ vs. s , will depend on the number and type of nuclei per unit volume, i.e., the scattering length density, ρ , as well as the real-space distribution of distances, \vec{r} , between all scattering pairs. Different isotopes will have different scattering lengths resulting in different weighted contributions to $I(s)$, while the real-space distance-separation frequency within the volume boundary of a macromolecule will determine the angular dependence in the intensities, i.e., the constructive and destructive interference pattern of the squared scattering amplitudes as a function of angle. For a continuous *excess* neutron scattering length density distribution, $\rho(\vec{r})$, bounded by the total particle volume, V_r , the resulting form factor, $F(s)$, of a macromolecule describes the magnitude and subsequent convolution of the scattering amplitudes (eq.6) caused by time-preserved real-space nuclei-pair separations (eq.7) and can be expressed as:

$$I(s) = \langle F(s)F(s)^* \rangle_{\Omega} = \langle |F(s)|^2 \rangle_{\Omega}, \quad \text{where } F(s) = \langle \int_{V_r} \rho(\vec{r}) e^{i\vec{s}\vec{r}} dV_r \rangle_{\Omega} \quad (8),$$

The $\langle \rangle_{\Omega}$ in the above expression represents that for a pure, infinitely-dilute macromolecule tumbling in solution (Ω), the resulting form factor is rotationally averaged over all orientations, yielding an isotropic scattering pattern. Eq.8 shows that the $F(s)$ of a macromolecule and the distribution of the excess scattering length density $\rho(\vec{r})$ are related by a Fourier transform. It is possible to express the distribution of $\rho(\vec{r})$ in terms of an autocorrelation function, $\gamma(r)$, that is essentially the spatial convolution of the excess scattering length density of an object with itself taking all pair-wise distances within the volume boundary into account over all orientations calculated from a common origin (Figure 9). The $I(s)$ relates to this autocorrelation function via:

$$I(s) = \langle \int_{V_r} \gamma(r) e^{i\vec{s}\vec{r}} dV_r \rangle_{\Omega} \quad (9).$$

The coherent S-wave amplitude components of the relation, $\langle e^{isr} \rangle$, can be re-formulated as:

$$\langle e^{isr} \rangle = \frac{\sin(sr)}{sr} \quad (10),$$

or an ‘integrated $\sin(x)/x$ Fourier term’, while the integral of all scattering-pair volume elements, dV_r , can be expanded and then simplified to;

$$dV_r = 4\pi r^2 dr \quad (11).$$

The substitution of eq. 10 and 11 into eq. 9 produces the relation between $I(s)$ and $\gamma(r)$:

$$I(s) = \int_0^\infty 4\pi\gamma(r) \frac{\sin(sr)}{sr} r^2 dr \quad (12).$$

What does all of this mean? Fundamentally what these relationships describe that it is possible to obtain the probable frequency of all scattering pair-distances distances, r , within a macromolecule from $I(s)$ vs. s , i.e., to transform between the measured intensities in momentum- or ‘reciprocal-space’ into real-space (Figure 9). Therefore, the measured coherent SANS intensities should reflect the time-preserved excess scattering length density distribution within a single macromolecule, i.e., the structure. The relationship between $I(s)$ and the r distance correlations can then be expressed as a probable real-space scattering-pair distance distribution, or $p(r)$. As:

$$p(r) = \gamma(r)r^2 \quad (13),$$

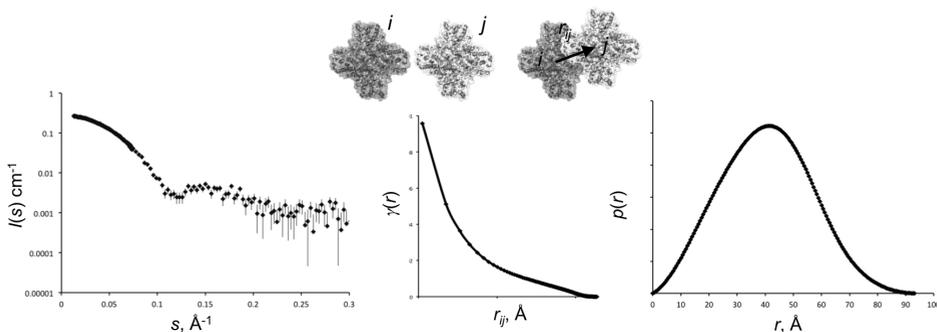
Then substitution into eq. 12 yields $I(s)$ in terms of $p(r)$:

$$I(s) = 4\pi \int_0^\infty p(r) \frac{\sin(sr)}{sr} dr \quad (14),$$

Which on inverse transformation leads to the expression of the real-space distance distribution in terms of $I(s)$:

$$p(r) = \frac{r^2}{2\pi^2} \int_0^\infty s^2 I(s) \frac{\sin(sr)}{sr} ds \quad (15).$$

The scattering intensities – and thus the associated form factor amplitudes of the entire macromolecule in reciprocal space – relate to the frequency of atom-pair distances in the macromolecule in real-space by a Fourier transform.



$$I(s) = 4\pi \int_0^{D_{max}} p(r) \frac{\sin(sr)}{sr} dr \quad p(r) = \frac{r^2}{2\pi^2} \int_0^\infty s^2 I(s) \frac{\sin(sr)}{sr} ds$$

Fig. 9. Scattering intensities in relation to real-space scattering-pair frequencies. If a macromolecule is pure, monodisperse, and is sufficiently dilute so as does not experience any interparticle interactions, the resulting background-corrected scattering profile recorded in reciprocal space, $I(s)$ vs. s , relates to the probable frequency of preserved scattering-pair distances, r , in real space, $p(r)$, via a Fourier transform. In effect, the resulting $p(r)$ represents the overall structure of the macromolecule expressed in terms of a distance distribution and can be derived from the excess scattering length density distribution internal to the particle, here represented as $\gamma(r)$. The $\gamma(r)$ can be viewed as the spatial correlation of the particle, i , with itself, j , in terms of the autocorrelation of all r_{ij} vector-pairs determined from a set origin and over all particle orientations, the maximum extent of which is determined by the maximum particle dimension ($\gamma(D_{max}) = 0$).

It is of note here that the conversion between reciprocal-space scattering and a real-space distance distribution is not a trivial exercise using the relationship above (eq.15) due to the discrete properties of the experimental data that are otherwise measured in Δs increments within a limited experimental s -range ($s_{min}-s_{max}$), and the fact that macromolecules have a maximum particle dimension, D_{max} , that provides an upper limit of r . Through these restrictions into the ‘ $\sin(x)/x$ Fourier term’ as well as experimental noise, i.e., the scattering variance, the resulting solution(s) of $p(r)$ can become unstable. The nature of the SANS measurement does not satisfy continuous integral/mathematical expressions. Therefore, it is necessary to employ indirect inverse Fourier transform methods to obtain $p(r)$, for example implemented in the programs GNOM[27] or GIFT[28]. Here, $p(r)$ is described in terms of a linear combination of coefficient-weighted functions, $\phi(r)$ across the interval of $r = 0-D_{max}$ whose individual Fourier transforms $\psi(s)$ are known. The experimental data and the resulting $p(r)$ are represented as a coefficient-weighted sum of these functions:

$$p(r) = \sum_{k=1}^K c_k \phi_k(r) \quad (16).$$

$$I_{calc}(s) = \sum_{k=1}^K c_k \psi_k(s) \quad (16a).$$

The derived $p(r)$ profile is optimized by the coefficients, c_k , that multiply each $\phi(r)$ so as to minimize the reduced χ^2 discrepancy between the experimental data points, $I_{\text{exp}}(s_j)$, and the calculated scattering intensities, $I_{\text{calc}}(s_j)$, taking into account the experimental errors, $\sigma_{\text{exp}}(s_j)$, while also maintaining a smooth $p(r)$ in real-space by incorporating a penalty term $P(p)$. The function to be minimized is:

$$\Psi = \chi^2 + \alpha P(p) \quad (17),$$

where:

$$\chi^2 = \frac{1}{N-1} \sum_{j=1}^N \left[\frac{I_{\text{exp}}(s_j) - cI_{\text{calc}}(s_j)}{\sigma(s_j)} \right]^2 \quad (18),$$

(c is a scaling constant, N is the number of points), and:

$$P(p) = \int_0^{D_{\text{max}}} \left[\frac{dp}{dr} \right]^2 dr \quad (19).$$

Relationship (19) describes a smoothing term applied to $p(r)$ in real-space where the regularization parameter α in eq. 17 acts as a balance between the stability of the computed $p(r)$ and the goodness of fit to the experimental data in reciprocal-space. For example, improving χ^2 , so as to minimize eq.18, would mean the calculated fit to the data starts to trace every experimental point, i.e., the resulting $p(r)$ in real-space would start to become increasingly affected by the sampling variance in the experimental intensities, resulting in an unphysical $p(r)$ profile with many maxima and minima (i.e., it would be overly-affected by experimental noise). In this case α is too low. Conversely, attempting to limit the influence of experimental noise by increasing α , i.e., increasing the magnitude of the smoothing term, may generate a smooth $p(r)$ but a poor fit to the data. The programs GNOM[27] or GIFT[28] help in the selection of an appropriate stable solution of $p(r)$ in real-space and the corresponding optimized fit to the data in reciprocal space. These relationships and the subsequent ‘discrete/indirect approach’ to what is otherwise an ill-posed problem of the inverse-Fourier transform described by the continuous integrals of eq.14 and 15, simply show that it is possible to extract real-space scattering-pair distance information within the constraints imposed by the experimental data.

2.4 Solutions, macromolecules and contrast.

The resultant sum of the coherent neutron scattering amplitudes and subsequent scattering intensities produced from the macromolecules making up a sample are generated because of the spatial correlation between scattering pairs within each macromolecule. In the small-angle regime, it is often the case that the buffer/solvent employed – and unlike the macromolecules themselves – lack these time-preserved long-range distance correlations. Consequently, there is no effective pair-wise ‘self-amplification’ of the scattering amplitudes arising from the solvent and, as a result, the solvent scattering scale in intensity across all angles only based on the average scattering length density, i.e., the isotopic composition per unit volume and the summed ‘scattering power’ of each isotope (eq.4). This results in a ‘flat’ scattering contribution made by the solvent across a SANS profile that can be subtracted from the sample scattering by measuring an equivalent solvent blank (Figure 3). The background-

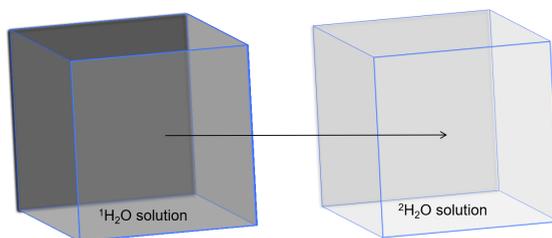
corrected $I(s)$ vs. s profile will then represent the form factor of the population of macromolecules in solution.

Importantly, $I(s)$ in the small-angle region only arises if there is an excess scattering length density within the particle volume relative to the solvent, that is, there is a difference between the average scattering length density of the solvent, $\bar{\rho}_s$, and the average scattering length density of the particles of interest, $\bar{\rho}$. This difference is called contrast ($\Delta\rho$; Figure 10):

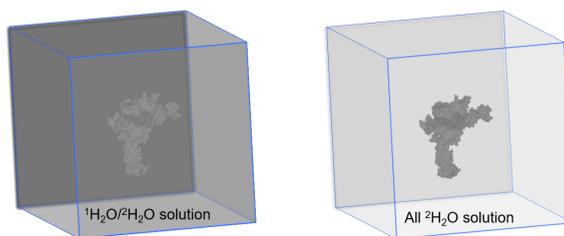
$$\Delta\rho = V_r^{-1} \int_{V_r} (\bar{\rho}(\vec{r}) - \bar{\rho}_s) dV_r = \bar{\rho} - \bar{\rho}_s \quad (20).$$

It is reasonably straightforward to see that if the scattering amplitudes arising from the solvent are of a similar magnitude as the macromolecule floating around in it, the resulting scattering intensities will be ‘drowned out’ by the solvent scattering, i.e., where $\Delta\rho$ limits to zero. However, for SANS, the ability to alter the contrast is the key to obtain structural information between components of complexes and assemblies of *different* average scattering length density because it is possible to match-out the scattering contributions of selected components by altering the scattering length density of the solvent.

- A** Substituting ^1H with ^2H , e.g., replacing regular light water ($^1\text{H}_2\text{O}$) with heavy water ($^2\text{H}_2\text{O}$) results in an alteration in the average neutron scattering length density of the solvent.



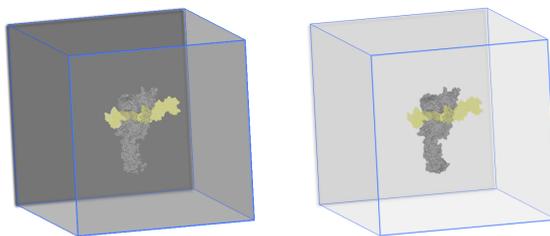
- B** If a macromolecule with an average scattering length density, $\bar{\rho}$, is placed into two different $\bar{\rho}_s$ solutions...



...the scattering intensity will be proportionate to the difference in average scattering length density, or the contrast, squared:

$$I(s) \propto \Delta\rho^2$$

- C** If a macromolecular complex of two *different* regions of average scattering length density is placed into different $\bar{\rho}_s$ solutions...



It is possible to access the coherent scattering contributions from the individual components of the complex and those contributions arising from between the components while bound together.

Fig. 10. Solutions macromolecules and contrast. **A.** Altering the isotopic composition of a solvent changes the average neutron scattering length density. For example, as the scattering length of ^1H is negative, swapping ^1H with ^2H in the solvent, that has a positive scattering length, changes the scattering length density considerably. **B.** Placing a macromolecule into a solvent with a similar average neutron scattering length density as the solvent results in poor contrast, $\Delta\rho$. Changing the scattering length density of the solvent or macromolecule via ^1H - ^2H substitution radically alters the contrast and the resulting magnitude of the scattering intensities as $I(s) \propto \Delta\rho^2$. **C.** For macromolecular complexes with two regions of average scattering length density, altering the contrast of the supporting solvent enables components to be ‘matched-in’ and ‘matched out’ of a SANS profile yielding structural information from the complex, the individual components and scattering contributions from between the components (the cross-term).

If we take a solution, we can sum the coherent scattering lengths for each isotope of each atom per unit volume to obtain the mean scattering length density [36]. Intuitively, as the scattering amplitudes of ^1H are 180° out of phase with ^2H , then substituting ^1H with ^2H in the solvent e.g., replacing regular light water ($^1\text{H}_2\text{O}$) with heavy water ($^2\text{H}_2\text{O}$), will alter the average scattering density of that solvent; the positive scattering length of ^2H and the other isotopes in the solution will cancel-out the negative scattering length of ^1H as the concentration of $^2\text{H}_2\text{O}$ is increased (Figure 10). If we take a macromolecule, that typically has reasonably stable level of non-exchangeable ^1H per unit volume (i.e., ^1H covalently bound to functional groups that are not in short-time scale exchange with the solvent) and place it into a solution with a different average scattering length density, the magnitude of scattering intensities will be proportional to the contrast squared, i.e.,

$$I(s) \propto \Delta\rho^2 \quad (21).$$

Consequently, when systematically swapping out $^1\text{H}_2\text{O}$ with $^2\text{H}_2\text{O}$ there will come a point where the average ^1H per unit volume in the solvent is such that the resulting scattering length density will, on average, be equivalent to the average scattering length density of the macromolecule (i.e., the scattering amplitudes arising from non-exchangeable ^1H plus all of the other biological isotopes making up the macromolecule will be cancelled by the scattering amplitudes from the solvent). As a result, the contrast will equal zero, eliminating the scattering signal at zero angle (note that at non-zero angles there will be non-zero scattering due to particle structure, see Section 4). The concept of contrast matching applies irrespective of the size of the macromolecule under investigation. If an enormous macromolecule (with a large volume) is placed in a $^1\text{H}_2\text{O}/^2\text{H}_2\text{O}$ solution such that $\Delta\rho = 0$, the forward scattering intensities will limit to zero, as it applies to small macromolecules in solution.

Contrast matching becomes a very useful property for any SANS investigation. Different classes of macromolecule have different average isotopic compositions per unit volume i.e., have different average scattering length densities due to their different chemical compositions. For example, protein, DNA, carbohydrates and lipids are comprised of different $^1\text{H}:^{12}\text{C}:^{16}\text{O}:^{14}\text{N}:^{31}\text{P}$: and ^{32}S ratios. Consequently, proteins typically have coherent neutron scattering match points between 40–43% v/v $^2\text{H}_2\text{O}$, whereas lipids match out between 2–15% v/v $^2\text{H}_2\text{O}$ and DNA matches out at ~60-70% v/v $^2\text{H}_2\text{O}$ (Figure 11). Therefore, if hetero-macromolecular complex with two regions of average scattering length density is placed into the appropriate % v/v $^1\text{H}_2\text{O}/^2\text{H}_2\text{O}$ solvents, it is possible to extract the scattering contributions for the whole complex (e.g., 100% v/v $^1\text{H}_2\text{O}$ buffers) and from the individual components. For example, a protein/DNA complex will have three contrast match points of around 43% v/v $^2\text{H}_2\text{O}$ for the protein component (whereby the scattering signal is dominated by contributions from the DNA) and around 65% v/v $^2\text{H}_2\text{O}$ for the DNA component (where the scattering signal will be dominated by the protein) as well as the a whole-complex match point (whereby the entire complex is matched out from the scattering signal). Therefore, using either contrast matching, or contrast variation, it should be possible to extract structural information from complexes with different regions of average scattering length density and obtain distance distributions of the individual components as well as the distances between those components of the complex (called the ‘cross-term’). SANS is one of the few techniques that enables the determination of the orientation and relative disposition of components within a full-formed macromolecular complex in solution and to assess whether, for example, any major structural alterations occur in one or more the components on forming a complex.

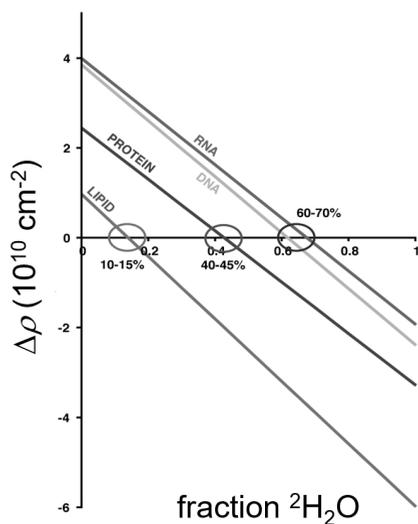


Fig. 11. Biomacromolecule classes and their average neutron scattering contrasts in $^2\text{H}_2\text{O}$. Proteins, lipids, and polynucleotides have different average neutron scattering length density due to different chemical/isotopic compositions. These contrast in volume fractions of $^2\text{H}_2\text{O}$ can be altered by isotopically labelling biomolecules with different levels of non-exchangeable ^2H , e.g., using biodeuteration. This will shift the match-point of the macromolecule, i.e., where the contrast limits to zero.

The capacity to alter the contrast in a SANS experiment via ^1H - ^2H substitution opens up a vast array of experiments. Not only is it possible to alter the scattering length density of the solvent, but it is also possible to control the average scattering length density of macromolecules using, for example, recombinant protein expression combined with non-exchangeable ^2H isotopic labelling [17]. Such experiments allow for the interrogation of unlabelled/labelled protein-protein complexes and their assemblies. In addition, lipids or detergents that are used in membrane protein investigations, can be chemically synthesized with different scattering length densities to effectively render them invisible in a contrast-matching/variation SANS experiments, affording a means through which to determine the low-resolution structure of a proteins and complexes embedded in detergent micelles or multi-component lipid nanodiscs [37, 38] or bicelles [55-56]. Nanoparticle conjugates can also be investigated, e.g., the covalent attachment of proteins, lipids or organic polymers to heavy-metal nanoparticles [39, 40]. The coherent neutron scattering length of, for example gold ($0.760 \cdot 10^{-12}$ cm) or iron ($0.945 \cdot 10^{-12}$ cm) is of a similar magnitude as ^2H ($0.667 \cdot 10^{-12}$ cm), and therefore these elements can be contrast-matched in high concentrations of ^2H -containing solvents, leaving behind SANS signal derived from any ^1H -organic conjugate that will otherwise have high-contrast due to the negative scattering length of ^1H ($-0.374 \cdot 10^{-12}$ cm).

2.5 Solution SANS in summary.

To obtain structural information from macromolecules SANS, the solvent scattering contributions must be subtracted to yield the excess scattering length density of the macromolecule in solution, due to a difference in contrast. As would be expected, the larger the volume of a macromolecule, the greater number of scattering-pair distances will be present within the volume boundary, giving rise to a more intense contrast-weighted

scattering profile at low angles. This leads to a generalized expression that relates the magnitude of the measured scattering intensities to the bulk properties of a sample with the form-factors of each individual macromolecule, i , of a population:

$$I(s) = S(s) \sum_i^n [(\Delta\rho_i V_i)^2 \langle |F(s)|^2 \rangle_{\Omega_i}] \quad (22).$$

The measured scattering intensities simply the summed contribution of all individual macromolecule form factors, weighted by the product of the volume squared and contrast squared relative to the solvent. The first term, called the structure factor, $S(s)$, relates to the contributions to $I(s)$ made by correlated distances of closest approach between the macromolecules in the sample. As long as a solution SANS experiment is performed using a sufficiently dilute sample, the $S(s)$ (also called interparticle interference) limits to 1. Therefore, and on the condition that each and every molecule within the population is identical, i.e., is pure, monodisperse and non-interacting, then eq. 22 simplifies to:

$$I(s) = N(\Delta\rho V)^2 \langle |F(s)|^2 \rangle_{\Omega} \quad (23),$$

such that the measured $I(s)$ profile after background correction will represent the time and rotationally averaged scattering from a single particle in solution (be it a single macromolecule, a complex, higher-order assembly, nanoconjugate, etc.). By extension, for a multi-component and monodisperse complex or assembly that contains *two* regions of average scattering length density (e.g. a protein/nucleic acid complex), the contributions made to the buffer-subtracted scattering profile by the two regions follow:

$$I(s) = N[(\Delta\rho_1 V_1)^2 \langle |F_1(s)|^2 \rangle_{\Omega} + (\Delta\rho_2 V_2)^2 \langle |F_2(s)|^2 \rangle_{\Omega} + 2\Delta\rho_1 V_1 \Delta\rho_2 V_2 \langle |F_{12}(s)|^2 \rangle_{\Omega}] \quad (24)$$

that is, the total (coherent) scattering intensities incorporate the contrast-weighted scattering from: Component 1 ($\Delta\rho_1 V_1$); component 2 ($\Delta\rho_2 V_2$) and; the component 1-2 scattering cross-term resulting from scattering events between the different scattering-length density regions ($\Delta\rho_1 V_1 \Delta\rho_2 V_2$). In effect, what this relationship shows is that by controlling the scattering length density of the sample by taking advantage of ^1H 's 'negative phasing power' either via $^1\text{H}_2\text{O}/^2\text{H}_2\text{O}$ substitution in the solvent and/or swapping non-exchangeable ^1H with non-exchangeable ^2H in the macromolecule, will yield structural parameters relating to the form factor of the assembly ($F_1(s) + F_2(s) + F_{12}(s)$), the individual components of the assembly ($F_1(s)$ or $F_2(s)$) and the between component cross-term ($F_{12}(s)$). All of this structural information relies on those conditions where $\Delta\rho = 0$ at the requisite component match points.

3 Structural Parameters.

There are two key structural parameters that can be extracted from SANS data that relate to the contrast times volume of a macromolecule and the scattering length distribution within this volume, expressed as the radius of gyration, R_g . If an experiment is performed on a complex consisting of two components of different average scattering length density and contrast variation is employed, these structural parameters vary systematically as the coherent scattering contributions of the components are 'matched out' ($\Delta\rho = 0$) or 'matched in' ($\Delta\rho \neq 0$) as the $^1\text{H}_2\text{O}/^2\text{H}_2\text{O}$ ratio in the supporting solvent is adjusted.

3.1 Forward scattering, $I(0)$.

At zero angle, the magnitude of $I(s)$ primarily depends on the total number of scattering centres within the squared-volume of a macromolecule – independent of the shape – weighted by the number density, N , (or sample concentration) and the contrast squared:

$$I(0) \sim N(\Delta\rho V)^2 \quad (25).$$

The intensity at zero angle, also called forward scattering $I(0)$, is the total contrast-weighted scattering derived from the constructive interference of the coherent scattering amplitudes arising from all scattering pairs within the volume boundary of the particles under the assumption that there are no interparticle interference effects. Of course, $I(0)$ is not accessible experimentally as the zero-angle position is occupied by the incident neutron beam. However, $I(0)$ can be calculated using the Guinier approximation (see 3.2 below) or from the area under $p(r)$, from eq.15:

$$I(0) = 4\pi \int_0^{D_{max}} p(r) dr \quad (26).$$

If the system under study experiences a change in contrast by altering the $^1\text{H}_2\text{O}/^2\text{H}_2\text{O}$ ratio of the solvent, then the magnitude of $I(0)$ will also change. The square-root of $I(0)$, when normalized to the sample concentration, should decrease linearly as the fraction of $^2\text{H}_2\text{O}$ increases to the point where $\Delta\rho = 0$ and a whole particle scattering contribution is matched out from the SANS profile. As the fraction of $^2\text{H}_2\text{O}$ increases even further, the contrast term goes negative, resulting in an increase in $I(0)$ once again (Figure 12). As $I(0)$ is linked to the V^2 of a macromolecule, and as macromolecules have an average scattering length density dependent on the isotopic composition, then it is possible to obtain the molecular weight (MW) of the macromolecule in solution using the following relationship:

$$MW = \frac{I(0)N_A}{c(\Delta\rho v)^2} \quad (27),$$

where N_A is Avogadro's number, v is the partial specific volume (cm^3/g , [36]) and c the concentration in g/cm^3 . Importantly, for this relation to hold, the SANS scattering intensities need to be measured on an absolute scale, cm^{-1} , which is routine practice at SANS beam lines. Deviations in the experimentally determined MW compared to an expected MW of a macromolecule (e.g., calculated from the atomic composition), or deviations from linearity in the $\sqrt{I(0)}$ vs fraction of $^2\text{H}_2\text{O}$ leading up to or away from the whole-particle match-point[52,53] are key indicators that the macromolecules in a sample are being perturbed, for example may aggregate in high volume fractions of $^2\text{H}_2\text{O}$, or in the case of multi-component systems, experience component disassociation.

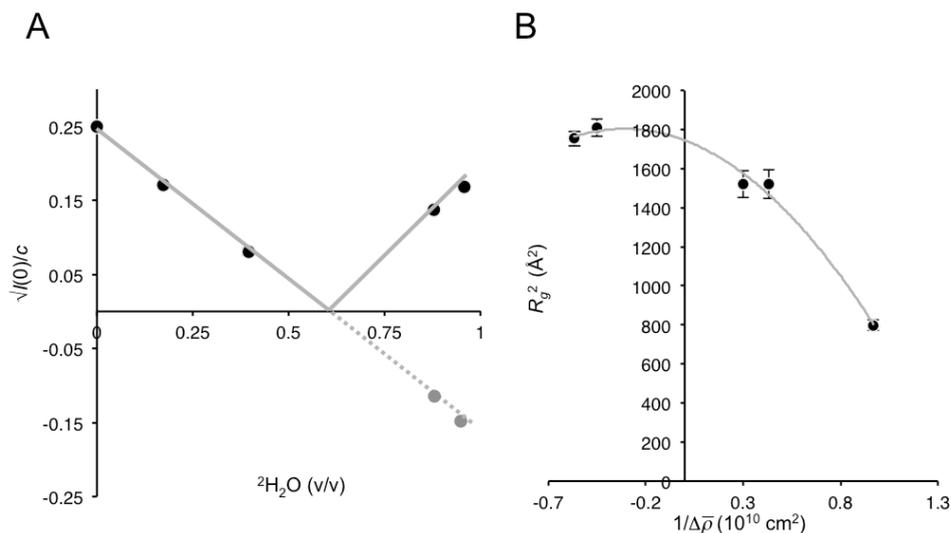


Fig. 12. $I(0)$ and R_g through a contrast variation series. **A.** The forward scattering $I(0)$, when normalized to sample concentration, c , systematically changes as the average neutron scattering length density of the solvent is altered by increasing the $^2\text{H}_2\text{O}$ concentration. As $I(s) \propto \Delta\rho^2$, a plot of $\sqrt{I(0)}/c$ vs. the volume fraction of $^2\text{H}_2\text{O}$ in the solvent should generate a straight line leading up to and away from the match point of the entire macromolecule ($\Delta\rho = 0$), in this instance where the volume fraction of $^2\text{H}_2\text{O} = 0.55$. To assess whether high concentrations of $^2\text{H}_2\text{O}$ perturb the sample, it is common practice to multiply the experimental $\sqrt{I(0)}/c$ obtained above $\Delta\rho = 0$ by -1 to generate a straight line (dashed-grey) and estimate the linearity of the resulting plot across all contrast points. Deviations from linearity suggest something is happening in the sample, e.g., aggregation or the disassociation of that can be cross-checked against concentration-dependent molecular weight estimates of the sample at each contrast point (eq. 27). **B.** Stuhmann plot showing the relationship between the inverse of the contrast and the R_g^2 and as described in eq. 30.

3.2 The Guinier approximation, radius of gyration, and Stuhmann relationship.

The R_g is a measure of the contrast weighted root mean square (or quadratic mean) distance of all volume elements occupied by scattering centres with respect to the centre of the scattering length density of a particle. This centre may *or may not* coincide with the centre of mass of a macromolecule or complex. The magnitude of the R_g is a useful overall parameter influenced by the size, shape and internal structure of the macromolecule. For example, an observed change in R_g of 2 \AA for a compact-globular particle equates to, either a significant change in shape, or, alternatively, a significant change in volume corresponding to a sizeable increase in mass.

Guinier [25, 26] showed that the scattering intensity at very-low angles is dependent on the radius of gyration:

$$I(s) = I(0)e^{\frac{-s^2 R_g^2}{3}} \quad (28).$$

For monodisperse macromolecules in solution, and after the solvent scattering contributions have been subtracted, a plot of the natural logarithm of the experimental $I(s)$ vs s^2 – or Guinier plot – will produce a negative linear relationship in the range $s < s_{\min}$ where $s_{\min} R_g < 1.3$. When

extrapolated to the $\ln I(s)$ intercept at $s = 0$, this yields $I(0)$ and the slope of the plot is proportional to R_g^2 . The R_g may also be calculated from the second moment of $p(r)$, eq. 15):

$$R_g^2 = \frac{\int_0^{D_{max}} p(r)r^2 dr}{2 \int_0^{D_{max}} p(r) dr} \quad (29).$$

The two independent R_g estimates from the Guinier approximation and $p(r)$ may be compared and cross-checked for consistency. However, the R_g as determined from the Guinier approximation, as opposed to $p(r)$, may become increasingly difficult to assess for structurally anisotropic inhomogeneous systems consisting of distinct regions of neutron scattering length density under conditions of varying contrast, i.e., when a complex has internal scattering-length density heterogeneity.

The proposition arises that for macromolecular complexes with two regions of average scattering length density, the R_g must be affected by $\Delta\rho$ and it should be possible to evaluate the R_g of a whole complex and the R_g of the individual components of the complex as the $\Delta\rho$ is altered. In addition, it should also be possible to assess the systematic change in R_g as a function of contrast that ultimately relates to the disposition between the two region centres of mass internal to the complex as the component scattering contributions are systematically ‘phased-in’ and ‘phased-out’ of a SANS profile as a function of % v/v $^2\text{H}_2\text{O}$ in the solvent.

Stuhrmann [52,53] showed that a relationship exists between the R_g^2 and the inverse of the contrast for a scattering object with an internal heterogeneous scattering length density (Figure 12):

$$R_g^2 = R_m^2 + \frac{\alpha}{\Delta\rho} - \frac{\beta}{\Delta\rho^2} \quad (30),$$

where R_m is the radius of gyration of the particle at ‘an infinite contrast’. At infinite contrast, the last two terms of the Stuhrmann relationship limit to zero. Hypothetically, at this point, the difference in scattering length density between heterogeneous regions would, in principle, also limit to zero (i.e., for an n -component particle, $\Delta\rho_n = \infty$ for all regions). Thus, the R_m is an extrapolated measure of the radius of gyration of a hypothetical particle with a *homogeneous* scattering length density and therefore relates to, and is affected by, the overall volume and shape occupied by the combined scattering length density regions. The term α , or the second moment of the internal density fluctuations within the scattering object, is a term that is sensitive to the radial distribution of the scattering length densities of the two regions relative to the object’s centre of mass. The β coefficient relates to the displacement between the heterogeneous regions, i.e., the distance between the centre of mass of the particle and the centres of the different of scattering length density regions. Assuming that scattering arising from non-uniform $^1\text{H}/^2\text{H}$ exchange, incoherent scattering and interparticle interference do not unduly ‘warp’ the estimation of R_g , then the magnitudes of α and β provide insights into the contrast-weighted disposition between the centres of scattering length density of the individual components within a complex. For example, if $\beta = 0$, the components of a complex share the same centres of scattering length density; a positive α implies the higher scattering density (on average) is located more toward the outside of the particle, while a negative α , places the higher scattering density (on average) more toward the inside of the particle. A zero α implies a homogeneous scattering particle. What the

Stuhrmann relationship elegantly captures is that the R_g is affected by the size and shape of the whole particle as well as its individual components and varies systematically as a function of the contrast depending on the disposition between the heterogeneous regions of scattering length density within a complex. The analysis of the contrast- R_g dependency can be analyzed using the R_g module of MULCh [36] (<http://smb-research.smb.usyd.edu.au/NCVWeb/>).

4 $\Delta\rho = 0$, experimental design and sample preparation.

As mentioned in section 2.4, different classes of biological macromolecules – protein, polynucleotides, lipids, carbohydrates, etc – have different average neutron scattering length densities. Although all atoms have the probability to scatter neutrons, the match points of each macromolecular class are strongly influenced by their average amount of ^1H per unit volume, stemming from the very different interactions between neutrons and ^1H compared to the other biological isotopes. As a result, $\Delta\rho = 0$ occurs at different percent volume fractions of $^2\text{H}_2\text{O}$ in a supporting solvent for each class of biomacromolecules. Two popular experimental approaches are widely used for interrogating macromolecular structures using SANS based on accessing $\Delta\rho = 0$ condition: SANS with contrast matching or SANS with contrast variation. The choice of experiment often boils down to how much sample is available and an understanding what signal is measured at a component match point.

4.1 Contrast matching.

If a macromolecule is covalently bound to, or is in complex with, another molecule that has a different average scattering-length density then the coherent scattering profile obtained at the match point for the first molecule is derived almost-exclusively from the second component. That is, at the match point of macromolecule x ($\Delta\rho_x = 0$), it will be possible to obtain shape information from macromolecule y and vice versa [57]. This type of experiment is called contrast matching and typically requires measuring the SANS profile of the whole complex (e.g., at 0% v/v $^2\text{H}_2\text{O}$, i.e., in regular light water) and at least one component match point, or preferably two, i.e., where $\Delta\rho_x = 0$ and/or where $\Delta\rho_y = 0$. The % v/v $^2\text{H}_2\text{O}$ in the solvent where the expected component and whole-complex match points occur can be calculated *a priori* from the atomic composition of the solvent and the component macromolecules of a complex, for example using the *Contrast* module of MULCh [36]. However, experimentally, when an individual component is selectively matched out at a certain % v/v $^2\text{H}_2\text{O}$ in the supporting solvent, the resulting scattering profile – although dominated by the coherent scattering of the bound partner – may still be influenced by additional scattering contributions made by the matched-out component not taken into account by the ‘expected’ calculations. This includes ^1H incoherent scattering and weak scattering arising from $^1\text{H}/^2\text{H}$ solvent-exchange or any other internal particle scattering length density fluctuations. Biological macromolecules are not homogeneous objects and are in dynamic $^1\text{H}/^2\text{H}$ exchange with the solvent and ‘unexpected’ contributions to the scattering intensities may arise at a component match-point by solvent exchange or internal structural inhomogeneity. These imperfections may become proportionately more influential toward $\Delta\rho = 0$, i.e., where the scattering contribution made to the SANS profile becomes exceedingly weak, but remains nonetheless, and could measurably influence the interpretation of the data.

Differences in scattering between macromolecules of *the same class* may manifest under ‘near-but-not-quite $\Delta\rho=0$ ’ conditions, that is, the average scattering length density approximation, which otherwise holds at points away from the match point where the coherent signal is strong, may not hold-up experimentally. For example, proteins may on *average* have a match point in conditions of 43% v/v $^2\text{H}_2\text{O}$, but experimentally and depending

on the protein and its exact chemical composition, the match point could be between 40-45% v/v $^2\text{H}_2\text{O}$ due to protein-specific internal/localized structure, hydrogen exchange rates/accessibility, etc. If a component is not precisely matched out, then coherent scattering contributions internal to that component and contributions arising from *between* the components of a complex, i.e., the cross-term, could remain. This effect becomes especially acute near a large component match point. As $I(s)$ is proportional to the square of the volume, deviations from the match point of a large component could have quite deleterious effects on the interpretation of a smaller component scattering pattern (eqs.23, 24, and 25). Experimentally, it may be necessary to perform a set of SANS measurements using % v/v $^2\text{H}_2\text{O}$ in-and-around the expected match point of a component (e.g., at and +/- 3% v/v $^2\text{H}_2\text{O}$ either side of the expected match point) and then apply a careful analysis of $I(0)$, MW and R_g . The advantage of contrast matching is that: 1) it works if the samples are accurately prepared and; 2) consumes less sample compared to a full contrast variation series.

4.2 Contrast variation.

SANS with contrast variation is similar to contrast matching, but typically requires more material and more beam time. For this, contrast variation allows one to extract the scattering functions from individual components of a complex from the set of experimental data, i.e., extract the full scattering curves from the components. A full SANS contrast variation series typically requires preparing both samples and buffers using incremental ratios of $^2\text{H}_2\text{O}$ in the solvent. The scattering profiles measured at these contrast points will relate to each other as described in eq.24 under conditions of varying contrast, setting up a set of related expressions. As small-angle scattering is always a combination of sums, the eq.24 relationship derived for each contrast point can be interpolated so that the component scattering functions can be derived from the contrast-set even if the experimental conditions did not exactly satisfy the $\Delta\rho=0$ condition of the individual components, as in, it is not strictly necessary to measure data from the sample at the exact contrast match-points. SANS with contrast variation allows for the deduction of the distance separating the components' centres of contrast (that is reflected in the mass separation between components) as well as the atom pair-distance distribution of the individual components *and* the distance distribution between the individual components of a complex. Therefore, as SANS with contrast variation generates multiple related datasets, the approach effectively 'boosts' the information content of a scattering experiment: The extrapolated component scattering functions and the cross-term yield both the structure and disposition of components within complexes[41] and assemblies [42].

A contrast series can be visualized along the lines using the following example of a 1:1 w/w protein/DNA complex. At 0% v/v $^2\text{H}_2\text{O}$ (i.e., pure regular light water) the coherent scattering data will encode shape information about the whole complex, but not the disposition of the individual components within the complex. On increasing the % v/v $^2\text{H}_2\text{O}$ in the solvent from 0% v/v up to near the protein match point (near 43 % v/v $^2\text{H}_2\text{O}$ in the solvent), the coherent scattering contributions will be increasingly and proportionately dominated by the DNA component until the scattering from the protein component has been almost solvent-matched. Consequently, at the protein match point, the magnitude of the coherent scattering will predominately reflect contributions to the scattering made by the DNA. When increasing the % v/v $^2\text{H}_2\text{O}$ in the solvent beyond the protein match point, the coherent scattering signals reduce toward zero until the match point of the *entire* complex is reached. Any registered signal at the entire-complex match point will be very weak and be dominated by incoherent spin-scattering and residual coherent scattering arising from correlated $^1\text{H}/^2\text{H}$ exchange with the solvent and internal-particle density fluctuations. Increasing the % v/v $^2\text{H}_2\text{O}$ beyond the entire-complex match point results in the re-emergence of coherent scattering intensities until

at ~60% v/v $^2\text{H}_2\text{O}$, or at the DNA match point, the coherent scattering will be dominated by the protein component alone. Above the DNA match point, DNA scattering contributions begin to add back into the measured profile, reflecting the proportionate scattering contributions from both the protein and DNA components of the complex. Therefore, using SANS with contrast variation, the overall shape of the protein/DNA complex as well as the shape and disposition of its subunits, in principle, can be determined from the coherent scattering data from the whole complex as well as the individual components at their respective % v/v $^2\text{H}_2\text{O}$ match points. The *Compost* module of MULCh [36] can be used to solving the set of equations derived from a SANS with contrast variation series (eq.24) and the extraction of the requisite cross-term scattering functions.

4.3 What concentration?

The ability to ‘match in’ and ‘match out’ components using SANS influences how samples are prepared [17]. First, there is selecting an appropriate sample concentration. The consequence of contrast variation is that the magnitude of the coherent scattering derived from any one individual component of a complex will be less than the combined scattering intensities derived from both components of the whole complex. In other words, when matching out components, the net SANS intensities will be reduced in relative to the ‘start concentration’ of the original complex and this will be proportionate to the volume squared of the components. Therefore, it is advisable to begin an experiment at a reasonable sample concentration to ensure that the coherent scattering measured from an individual component is above the background noise derived from the incoherent spin-scattering contributions arising from the solvent. Incoherent scattering contributions can be an issue, especially in the early stages of a contrast variation series (0–50% v/v $^2\text{H}_2\text{O}$) due to the large incoherent-spin scattering length and subsequent cross section of ^1H . But what concentration does one choose? This relates to the volume squared of the whole complex and the volume squared of the individual components that in effect relates to the mass ratios of components. For larger complexes (50 kDa and above e.g., at 5–7 mg.ml $^{-1}$), incoherent noise from the solvent may not be of concern, but for smaller complexes (50 kDa and below) even at relatively high concentrations (e.g. 6–10 mg ml $^{-1}$) ^1H incoherent spin-scattering contributions may ‘drown out’ the coherent scattering. The obvious solution is to simply increase sample concentrations, but then there is a risk of introducing interparticle interference effects. Another option is to increase exposure times (quadrupling collection time should result in a two-fold improvement in counting statistics) but sometimes even this may not be effective, especially if one of the components of a complex is significantly smaller than the other. The solution: separate the match points so that larger components of a complex match in? at high % v/v $^2\text{H}_2\text{O}$ to enable data collection from smaller components in a solvent background with low incoherent scattering, i.e., in high % v/v $^2\text{H}_2\text{O}$ in the solvent.

4.4 Improving match-point separation using biodeuteration.

Incoherent scattering from the solvent may become increasingly problematic if the mass ratios of the components of a complex become increasingly disproportionate. As the mass ratio tends toward the extreme, not only do the total coherent scattering contributions from a smaller component become proportionately less ($I(0) \propto V^2$), but the separation of the match points between the large component and the entire-complex (where scattering intensities are negligible) become less and less distinguishable. The combined effect complicates the collection of data from the small component. For example, a 75 kDa protein in complex with a 20 kDa piece of DNA will have protein match-point and entire-complex match point that are very close (~42% v/v $^2\text{H}_2\text{O}$ and ~45% v/v $^2\text{H}_2\text{O}$). Such a scenario makes it difficult to

perform a SANS experiment in terms of accurately making up solvents to the exact % v/v $^2\text{H}_2\text{O}$ as well as measuring data above the background noise for the DNA component at the protein match point where incoherent scattering from ^1H at these % v/v $^2\text{H}_2\text{O}$ remains significant. The issue of poor match point separation can be solved by altering the $^1\text{H}:^2\text{H}$ per unit volume on one of the components, i.e., partially or completely deuterating a component with non-exchangeable ^2H .

The level and extent of ^2H incorporation into a macromolecule can be controlled using biodeuteration, e.g. the expression of a target in *Escherichia coli* B strains in $^2\text{H}_2\text{O}$ media [16, 17, 43, 44]. Biodeuteration alters the non-exchangeable ^1H per unit volume on a component and consequently, when a complex is formed between the ^2H -labelled material and an 'all- ^1H ' binding partner, the contrast against the solvent and between the components of a complex are radically altered. This causes the component and whole complex match points to separate. Match point separation using biodeuteration opens up opportunities to structurally interrogate otherwise difficult macromolecular complexes, such as the analysis of complexes comprised of subunits with disproportionate masses. In the example above, if the protein component of the protein/DNA complex were labelled with $\sim 72\%$ non-exchangeable ^2H , then the match points separate to $\sim 100\%$ (protein), 91% (whole complex) and $\sim 60\%$ (DNA) v/v $^2\text{H}_2\text{O}$, respectively, making it possible to measure the scattering from the protein and the DNA in high % v/v $^2\text{H}_2\text{O}$ buffers. Furthermore, and of extreme significance, biodeuteration allows the extraction shape information from ^1H -protein- ^2H -protein complexes, that is otherwise almost-impossible to achieve from all- ^1H protein systems (Figure 13).

Assessing what average level of non-exchangeable ^2H should be incorporated into a component to obtain decent match point separations within a sample can be predicted *a-priori* using the *Contrast* module of MULCh [36]. The module takes primary protein, RNA or DNA sequences and the atomic composition of the solvent and calculates the match points for a system. It also calculates the contrasts for a hypothetical contrast variation experiment, considering the extent of ^2H -labelling on a component and the percentage of acidic protons likely to be in exchange between a complex and the solvent (usually around 90–95%). These calculations provide an excellent guide for setting up biodeuteration runs (e.g., deciding on what % v/v $^2\text{H}_2\text{O}$ to grow cell cultures to obtain the desired level of non-exchangeable ^2H incorporation) and evaluating how SANS signal intensities will change for each component at different contrast values.

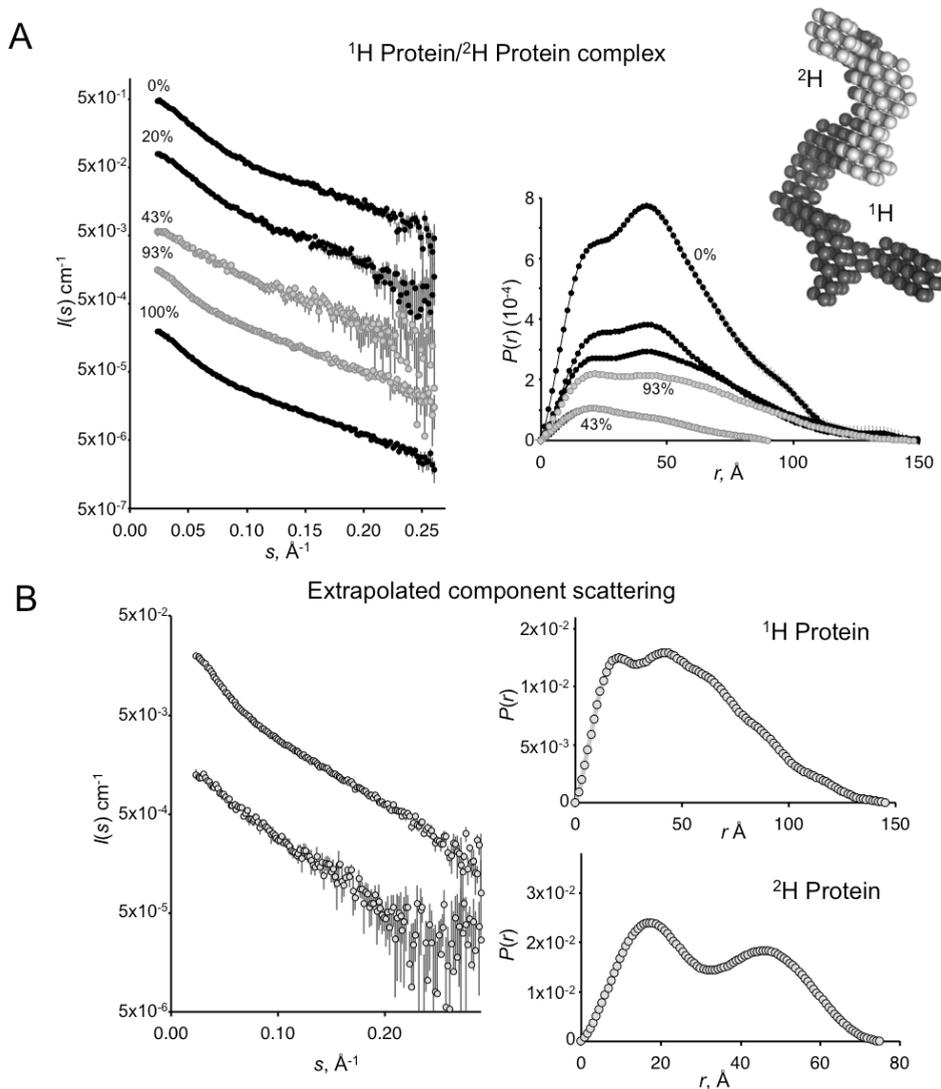


Fig. 13. A full contrast variation series measured from a ^1H and ^2H labelled protein-protein complex. **A.** SANS profiles measured at different %v/v $^2\text{H}_2\text{O}$ in the solvent. The near, but not exact, match point scattering data for the ^1H -component (43% v/v $^2\text{H}_2\text{O}$) and the ^2H -component (93% v/v $^2\text{H}_2\text{O}$) are coloured grey. The corresponding $p(r)$ profiles obtained each contrast variation point/dataset are shown. **B.** The individual component scattering functions of the ^1H - and ^2H -protein components extrapolated from the contrast variation data obtained from eq.24 and using the *Compost* module of MULCh, and the corresponding $p(r)$ profiles of each individual protein. The resulting real-space low-resolution representation of the protein-protein complex is shown as a space-filling dummy atom model.

4.5 Is your complex a complex in $^2\text{H}_2\text{O}$?

There are several aspects to consider when preparing samples for a SANS experiment that are provided in detail in [17]. This includes, but is not limited to, assessing sample quality prior to a SANS measurement and the quantities of material required for contrast matching

or contrast variation experiments. For example, five or more samples are typically required to span a SANS with contrast variation series that, for a ^1H -protein- ^2H -protein complex, might translate into setting up dialysis against 0%, 20%, 43%, 90%, 100% v/v $^2\text{H}_2\text{O}$ using a total of ~ 2 – 2.5 ml of sample in a concentration range somewhere between 5 – 10 mg.ml $^{-1}$. The concentration used for the final experiment has to balance between maximizing signal that greatly depends on the V^2 of the components, versus introducing unwanted interparticle interference effects $S(s)$ terms in the scattering at too-high concentration, versus the disassociation constant, K_d , of the components of the complex. What must be reiterated here, is that it is absolutely necessary to evaluate whether samples remain fully associated and soluble in solutions containing high % v/v $^2\text{H}_2\text{O}$. The strength of ^2H -hydrogen bonds is different to ^1H -hydrogen bonds, as reflected, for example, in the difference between pH and p ^2H (termed pD, where pD = pH + 0.4). In addition, the solvation layer around a macromolecule has different properties compared to the bulk solvent. The cumulative effects of these differences are that $^2\text{H}_2\text{O}$ has the potential to affect the solubility, stability and structural dynamics of macromolecules[45-51] that could result in significant shifts in K_d , resulting in complexes falling apart in $^2\text{H}_2\text{O}$, or, perhaps more commonly, lead to non-specific aggregation. Consequently, it is absolutely required to assess the effects of $^2\text{H}_2\text{O}$ on the stability of a sample in order to satisfy conditions where a complex is both stable, soluble and monodisperse as a function $^2\text{H}_2\text{O}$ concentration. Indeed, the biggest challenge facing a biomacromolecular SANS experiment, aside from producing sufficient material, is maintaining the integrity of the sample, be it ^2H labelled or not, in high % v/v $^2\text{H}_2\text{O}$ solutions. As scattering intensity scales to the volume squared, even trace levels of aggregation can ruin SANS experiment, destroying the relationships between the contrast points as described in eq.24 and the subsequent evaluation of accurate structural information.

5 Questions and answers: a student's perspective.

- *What is the first aspect of SANS – be it theory or experiment – that a student of structural biology should be aware of before they begin?*

It is very important to evaluate what pre-existing data is available on your system and to obtain as much information as possible about the properties of the sample, especially in terms of stability, overall size, molecular weight, and the propensity of a sample to aggregate. The aim of the wet lab work is to obtain as pure and monodisperse sample as possible and characterize it with available methods. Beforehand, it is beneficial to perform small angle X-ray scattering (SAXS) experiments as SAXS helps to determine two major factors: i) The optimal sample conditions, e.g., to evaluate the aggregation propensity of the sample and; ii) To estimate the s -range, and in particular the experimental s_{\min} required to capture scattering data that spans the longest vector lengths in the sample (as a rule of thumb it is good practice to measure to where $s_{\min} = \pi/D_{\max}$, or better yet, $s_{\min} = 1/D_{\max}$). This information is useful to select what sample-detector positions are chosen for the SANS measurements before the experiment takes place. In addition, and what is important for structural biologists, is that SAXS also provides insights into low resolution structure of the sample. For two component complexes, the low-resolution structure obtained from SAXS can be used to complement the SANS data. With the help of neutrons, a more detailed structural picture then emerges as the SANS yields additional information on how the components are oriented inside this low-resolution structure.

The other important consideration is the amount of sample required for SANS. Compared to SAXS (20–50 μl), the quantity of sample required for a full SANS experiment can be a lot more (100s μl to ml) simply because of the size of the sample cells used at a neutron facility

(due to the large area of the neutron beam). There are several neutron facilities offering deuteration of biomolecules based on a proposal application system. The use of bio-fermenters results in production of high-density bacterial cell paste with overexpressed ^1H and ^2H -labeled proteins that can really help produce enough material for the experiment compared to lab-based protein over expression. Generally, the least 200–500 μl of a sample at 5–10 mg ml^{-1} are required for a *single* point of SANS contrast series, therefore it might be advantageous to access the resources on offer at biodeuteration facilities.

It is also very important to optimize and characterize samples on a small scale as much as possible before to leaping into a full SANS with contrast variation experiment due to the often heavy sample preparation demands of the experiments. However, one of the advantages of SANS is that (and unlike X-rays) the technique is generally non-destructive which is beneficial even when sample quantity might be limited. In cases where sample amounts did not quite reach planned expectations it is possible, after measuring and checking for radiation activity, to re-use the samples by mixing them together to obtain different contrast points and repeat the procedure with corresponding buffer.

- *What is the most difficult/complicated aspect of SANS for structural biology? How do you overcome these difficulties?*

The most difficult part is optimizing conditions where the samples are, and remain, soluble and stable in high percentage $^2\text{H}_2\text{O}$ buffers. The regular buffer composition in $^1\text{H}_2\text{O}$ solution might not always be the best for your system when going for 100% v/v $^2\text{H}_2\text{O}$. In some cases, it is necessary to change the base of the buffer, consider using potassium chloride instead/or with sodium chloride, adding deuterated glycerol or a changing to a more stable reducing agent (e.g., tris(2-carboxyethyl)phosphine), or TCEP, instead of dithiothreitol, DTT).

Sample mono-dispersity is essential during solution studies and its assessment should be done at different stages of the sample preparation. For this, dynamic light scattering, DLS, is an easy and efficient technique available in most molecular biology laboratories, which allows evaluation of sample quality. There are several things to check:

- 1) The sample's behavior in high percentage $^2\text{H}_2\text{O}$ solutions.
- 2) Time-induced aggregation, as SANS measurement may requires long time exposures, especially near the component match points.
- 3) Aggregation formation upon thawing/freezing, which is important to consider when transporting samples to a facility.
- 4) Sample stability at different temperatures. Depending on the specific beam line, sample holder temperatures below 10°C might not be available.

Of course, and aside from DLS, SAXS may be used to determine the optimal sample conditions and concentrations to assess appearance of interparticle interactions, be they repulsive or attractive (i.e., to assess aggregation).

Furthermore, overexpression of deuterated biomolecules in ^2H minimal media might result in the formation of insoluble inclusion bodies or, in general, lower amounts of soluble protein. It is a good idea to perform a small-scale solubility screen to test different extraction protocols to obtain higher yields of material.

- *Are there aspects of the technique that are easily overlooked/forgotten or misunderstood – theory, instrument, and sample preparation?*

Before setting up a SANS experiment there are several theoretical steps which must be done *in silico*, e.g., using MULCh. It is not only important to obtain the estimate of the components

match points, but also to calculate differences in the magnitude of scattering intensities of molecules in different $^1\text{H}_2\text{O}/^2\text{H}_2\text{O}$ ratios so as to help choose what contrast points to measure. The major factor is to decide prior to the experiment on what level of non-exchangeable deuteration is required to obtain the best match point separation of the system components.

Another aspect to consider is the ‘wet-lab’ perspective on sample preparation and how this may, or may not, be compatible with a SANS instrument/experiment. Typically, the majority of purification steps and sample handling procedures in the wet-lab are performed in the cold room or on ice in order to preserve the sample over the often-extended number of steps and time required to isolate a sample. It is really important to assess the temperature stability of the sample after it has been purified, i.e., can it maintain its integrity for an extended period of time (e.g., up to 2–3 hrs) out of the cold-room? Such information it is important to know. Dilute macromolecule samples are *very*-weakly scattering and although it might be possible to measure SANS at colder-than-room-temperature, colder temperature measurements run the risk of spontaneous water-vapor condensation from the atmosphere on the large sample cells over extended neutron exposure times. Such condensation on the outside of the SANS sample cell can ruin the measurement. So it is important to check what temperature range is available at a beam line and, very importantly, consult with your local beam line facility contact about any difficulties with water condensation at or below 10 °C. There is little point spending time and finding out in the wet-lab that your sample is stable at 4 °C only to discover at the last minute you cannot measure it at 4 °C.

When travelling to a facility it is a good idea to devise an experimental plan and to take some of your own consumables, for example, pack your own pipettes and tips. It can take a while to familiarize yourself with a facility user lab and to figure out where everything is. Having some of your own consumables at hand simply relieves the stress of running around looking for particular items. In regards to the experimental plan, have a well-prepared/thought-out and systematic dialysis protocol, that limits the number of steps required to complete the dialysis and, hopefully, the stress involved with the sample handling. Write the protocol down, having already pre-established the quantity of buffer components required, sample concentrations, etc, prior to setting up the dialysis at the facility. A written protocol helps set a routine and clarifies what to do at 4 am in the morning after having worked for 15 hrs straight setting up the samples. It is also easier to travel with team members, where members have been allocated into ‘instrument responsables’ – who learn how to load and drive the instrument – and others as ‘sample responsables’ – who deliver the samples to the instrument.

Considerations for the dialysis of samples at a facility include: i) Take your own dialysis equipment; ii) When dialyzing against different $^2\text{H}_2\text{O}$ solutions it is crucial to avoid formation of air pockets in the sample to ensure proper proton-deuterium exchange and inadvertent changes in sample concentration; iii) Even when invisible to a naked eye, $^2\text{H}_2\text{O}$ buffers, especially cold buffers, tend to release bubbles over time (especially when warming up) due to dissolved gasses. These microbubbles scatter like crazy. It may take a few hours to fully remove them by sonication or degassing equipment. Dialyzing at room temperature, or warming up the samples to room temperature in advance of the SANS beam time, limits the dissolved-gas problem; iv) Always centrifuge the post dialysis samples and buffers at full speed in a bench-top centrifuge (e.g., 30 000 x g) to help get rid of bubbles and potential small insoluble or large soluble aggregates. Do this immediately prior to loading into the SANS sample cuvette; v) It is good practice to measure accurate concentrations of both the pre- and post-dialysis samples and the post-measured samples (after passing radiation safety). These concentration measurements may indicate if something unwanted happened during dialysis i.e. aggregates formation or precipitation, or during the measurement; vi) If possible, apply for simultaneous access to synchrotron SAXS beam line or bench-top SAXS instrument during your SANS experiment to get additional data from the post-SANS sample.

- *In what situations do you think SANS is useful for structural biology investigations, that is, what types of questions do you ask that SANS can answer that other techniques cannot?*

High-resolution structure determination techniques are the key to reveal the atomic details of biomolecular interactions. However, the beauty of implementing SANS is that it can be used to analyze almost-any biological system in solution covering wide molecular-weight range across almost-any sample environment. With new advances in production of deuterium labelled biological components, SANS can be used to measure very complex systems inaccessible to many standard structural biology techniques, including membrane proteins, in more biologically relevant environment(s). Additionally, SANS is an exciting tool to probe conformational changes caused by interactions, e.g., upon binding of intrinsically disordered proteins. Structural flexibility or intrinsic disorder of macromolecules is not a limitation. The other advantage of SANS is to push structural biology investigations beyond the more traditional ‘structures of monodispersed systems’ type of experiment, e.g., to ask questions relating to fibrillation, aggregation, gelation, the formation of ordered assemblies, etc. Finally, and just point out, neutrons are relatively non-destructive compared to X-rays so are great for studying samples that are highly susceptible to X-ray radiation damage.

6 Summary.

Small angle neutron scattering affords a unique method to interrogate the size, structure and spatial disposition of macromolecules, complexes, assemblies, nanoconjugates, etc, due to the unique way neutrons scatter from ^1H compared to other biological isotopes. The ability to control the coherent neutron scattering contrast in such systems via the isotopic substitution of ^1H with ^2H affords a way of assessing the internal structure and arrangement of individual component in higher-order systems. SANS is not confined to a specific sample environment, allowing one to access diverse experimental samples under diverse experimental conditions including particles in near-native solutions, flexible macromolecules, particles under shear, pressure or extremes in temperature and pH, phase-changes, etc. This chapter outlines the most basic explanation of SANS that will hopefully be of use for structural biology students and, therefore, helping structural biology going forward.

Acknowledgements

The authors wish to acknowledge the European Union’s Horizon 2020 research and innovation programme ‘iNEXT Discovery’ grant agreement No 871037. ZP is funded by the the Swedish Foundation for Strategic Research (SSF) within the Swedish national graduate school in neutron scattering (SwedNess). We would also like to thank Patrice Vachette of the Institut de Biologie Intégrative de la Cellule, Université Paris-Sud (for general inspiration) as well as the Société Française de la Neutronique, including the travel support/resources made available during the 2018 ‘Neutrons and Biology’ School, Carqueiranne, France.

References

1. Z.H. Zhou, Adv. Protein Chem. Struct. Biol., **82**, 1 (2011)
2. M.A. Herzik, Jr., M. Wu, G.C. Lander, Nat. Commun., **10**, 1032 (2019)
3. M.G. Campbell, et al., Structure, **20**, 1823 (2012)
4. A.F. Brilot, et al., J. Struct. Biol., **177**, 630 (2012)

5. J.R. Helliwell, *Acta. Crystallogr. D Struct. Biol.*, **76**, 87 (2020)
6. T. Sugiki, N. Kobayashi, T. Fujiwara, *Comput. Struct. Biotechnol. J.*, **15**, 328 (2017)
7. J.K. Krueger and G.D. Wignall, *Small-Angle Neutron Scattering from Biological Molecules*, in *Neutron Scattering in Biology*. 2006, Springer Berlin Heidelberg. p. 127.
8. J.H. Lakey, *Journal of the Royal Society, Interface / the Royal Society*, **6 Suppl 5**, S567 (2009)
9. W.T. Heller and K.C. Littrell, *Methods Mol. Biol.*, **544**, 293 (2009)
10. S. Krueger, *Physica B: Condensed Matter*, **241–243**, 1131 (1997)
11. E. Mahieu and F. Gabel, *Acta Crystallogr D Struct Biol*, **74**, 715 (2018)
12. C. Breyton, et al., *Eur. Phys. J. E Soft Matter*, **36**, 71 (2013)
13. D.I. Svergun, et al., *Small Angle X-Ray and Neutron Scattering from Solutions of Biological Macromolecules*. 1 ed. 2013: Oxford University Press.
14. *Neutron scattering in biology techniques and applications*, ed. J. Fitter, T. Gutberlet, and J. Katsaras. 2006, Berlin Heidelberg: Springer-Verlag.
15. G. Zaccai, *Eur. Biophys. J.*, **41**, 781 (2012)
16. N.R. Zaccai, et al., *Methods Enzymol.*, **566**, 159 (2016)
17. C.M. Jeffries, et al., *Nat. Protoc.*, **11**, 2122 (2016)
18. A.D. Martin, et al., *Sci Rep*, **7**, 43947 (2017)
19. M. Martínez-Sanz, M.J. Gidley, E.P. Gilbert, *Soft Matter*, **12**, 1534 (2016)
20. P. J.K. and M.J.A. Hore, *Curr. Opin. Colloid Interface Sci.*, **42**, 157 (2019)
21. I. Yadav, et al., *Langmuir*, **33**, 1227 (2017)
22. R.F. Tabor, J. Eastoe, I. Grillo, *Soft Matter*, **5**, 2125 (2009)
23. J. Siefker, et al., *J. Am. Chem. Soc.*, **140**, 12720 (2018)
24. B. E., et al., *J. App. Cryst*, **42**, 461 (2009)
25. A. Guinier, *Ann. Phys. (Paris)*, **12**, 161 (1939)
26. A. Guinier and G. Fournet, *Small Angle Scattering of X-Rays*. 1955, New York: Wiley.
27. D.I. Svergun, *J. Appl. Crystallogr.*, **25**, 495 (1992)
28. O. Glatter, *J. Appl. Cryst.*, **10**, 415 (1977)
29. L.A. Feigin and D.I. Svergun, *Structure analysis by small-angle x-ray and neutron scattering*. 1987, New York: Plenum Press. xiii.
30. D.I. Svergun and K.H. Nierhaus, *J. Biol. Chem.*, **275**, 14432 (2000)
31. G.L. Squires, *Introduction to the theory of thermal neutron scattering*. 1996, Mineola, New York: Cambridge University Press.
32. J. Chadwick, *Proceedings of the Royal Society A*, **136**, 692 (1932)
33. L. Meitner and O.R. Frisch, *Nature*, **143** 239 (1939)
34. O. Hahn and F. Strassmann, *Naturwissenschaften*, **27**, 11 (1939)
35. V.F. Sears, *Neutron News*, **3**, 26 (1992)
36. A.E. Whitten, S. Cai, J. Trehwella, *J. Appl. Crystallogr.*, **41**, 222 (2008)
37. S. Maric, et al., *Acta. Crystallogr. D Struct. Biol.*, **70**, 317 (2014)
38. T.E.t. Cleveland, et al., *Protein science*, **27**, 780 (2018)
39. Z. Luo, et al., *Nat. Commun.*, **9**, 1343 (2018)
40. M. Moglianetti, et al., *Chem. Sci.*, **5**, 1232 (2014)
41. A.E. Whitten, et al., *J. Mol. Biol*, **368**, 407 (2007)
42. A.E. Whitten, et al., *Proc. Natl. Acad. Sci. USA*, **105**, 18360 (2008)
43. B. Leitng, F. Marsilio, J.F. O'Connell, *Anal. Biochem*, **265**, 351 (1998)
44. X. Chen, et al., *Food and Bioproducts Processing*, **90**, 563 (2012)
45. P.A. Baghurst, L.W. Nichol, W.H. Sawyer, *J. Biol. Chem.*, **247**, 3199 (1972)
46. L. Dougan, et al., *Chemphyschem*, **9**, 2836 (2008)
47. M. Tehei, et al., *Proc. Natl. Acad. Sci. USA*, **98**, 14356 (2001)

48. M. Jasnin, et al., *Eur. Biophys. J.*, **37**, 613 (2008)
49. D.I. Svergun, et al., *Proc. Natl. Acad. Sci. USA.*, **95**, 2267 (1998)
50. A.S. Goryunov, *Gen. Physiol. Biophys.*, **25**, 303 (2006)
51. S.Y. Sheu, et al., *J. Phys. Chem. A*, **112**, 797 (2008)
52. H.B. Stuhmann, *J. App. Cryst.*, **7**,173 (1974)
53. H.B. Stuhmann, A. Miller, *J. App. Cryst.*, **11**, 325 (1978)
54. J. Trehwella et al., *Acta Crystallogr D*, **73**, 710 (2017)
55. R. Dos Santos Morais, et al., *Langmuir*, **33**, 6572 (2017)
56. R. Dos Santos Morais et al., *Biophys. J.*, **115**, 1231 (2018)