

Small-angle scattering studies of biological macromolecules in solution

Dmitri I Svergun^{1,2} and Michel H J Koch¹

¹ European Molecular Biology Laboratory, Hamburg Outstation, Notkestraße 85, D-22603 Hamburg, Germany

Received 11 July 2003, in final form 7 August 2003

Published 16 September 2003

Online at stacks.iop.org/RoPP/66/1735

Abstract

Small-angle scattering (SAS) of x-rays and neutrons is a fundamental tool in the study of biological macromolecules. The major advantage of the method lies in its ability to provide structural information about partially or completely disordered systems. SAS allows one to study the structure of native particles in near physiological environments and to analyse structural changes in response to variations in external conditions.

In this review we concentrate on SAS studies of isotropic systems, in particular, solutions of biological macromolecules, an area where major progress has been achieved during the last decade. Solution scattering studies are especially important, given the challenge of the ‘post-genomic’ era with vast numbers of protein sequences becoming available. Numerous structural initiatives aim at large-scale expression and purification of proteins for subsequent structure determination using x-ray crystallography and NMR spectroscopy. Because of the requirement of good crystals for crystallography and the low molecular mass requirement of NMR, a significant fraction of proteins cannot be analysed using these two high-resolution methods. Progress in SAS instrumentation and novel analysis methods, which substantially improve the resolution and reliability of the structural models, makes the method an important complementary tool for these initiatives.

The review covers the basics of x-ray and neutron SAS, instrumentation, mathematical methods used in data analysis and major modelling techniques. Examples of applications of SAS to different types of biomolecules (proteins, nucleic acids, macromolecular complexes, polyelectrolytes) are presented. A brief account of the new opportunities offered by third and fourth generation synchrotron radiation sources (time-resolved studies, coherent scattering and single molecule scattering) is also given.

² Also at: Institute of Crystallography, Russian Academy of Sciences, Leninsky pr. 59, 117333 Moscow, Russia.

Contents

	Page
1. Introduction	1737
2. Basics of SAS	1738
2.1. Scattering of x-rays and neutrons	1738
2.2. Scattering by macromolecular solutions	1739
2.3. Resolution and contrast	1740
2.4. X-ray and neutron scattering instruments	1741
3. Monodisperse systems	1743
3.1. Overall parameters	1744
3.2. Distance distribution function and particle anisometry	1746
3.3. Shannon sampling and information content	1748
3.4. Ab initio analysis of particle shape and domain structure	1748
3.5. Computation of scattering patterns from atomic models	1752
3.6. Building models from subunits by rigid body refinement	1753
3.7. Contrast variation and selective labelling of macromolecular complexes	1754
4. Polydisperse and interacting systems	1756
4.1. Mixtures with shape and size polydispersity	1756
4.2. Interacting systems and structure factor	1757
4.3. Computation of the structure factor from interaction potentials	1759
5. Selected applications	1760
5.1. Analysis of macromolecular shapes	1760
5.2. Quaternary structure of complex particles	1764
5.3. Equilibrium systems and oligomeric mixtures	1767
5.4. Intermolecular interactions and protein crystallization	1769
5.5. Polyelectrolyte solutions and gels	1770
5.6. Time-resolved studies: assembly and (un)folding	1771
5.7. Coherence and single molecule scattering	1773
6. Conclusions	1775
Acknowledgments	1777
References	1777

1. Introduction

Small-angle scattering (SAS) of x-rays (SAXS) and neutrons (SANS) is a fundamental method for structure analysis of condensed matter. The applications cover various fields, from metal alloys to synthetic polymers in solution and in bulk, biological macromolecules in solution, emulsions, porous materials, nanoparticles, etc. First x-ray applications date back to the late 1930s when the main principles of SAXS were developed in the seminal work of Guinier [1] following his studies of metallic alloys. The scattering of x-rays at small angles (close to the primary beam) was found to provide structural information on inhomogeneities of the electron density with characteristic dimensions between one and a few hundred nm. In the first monograph on SAXS by Guinier and Fournet [2] it was already demonstrated that the method yields not just information on the sizes and shapes of particles but also on the internal structure of disordered and partially ordered systems.

In the 1960s, the method became increasingly important in the study of biological macromolecules in solution as it allowed one to get low-resolution structural information on the overall shape and internal structure in the absence of crystals. A breakthrough in SAXS and SANS experiments came in the 1970s, thanks to the availability of synchrotron radiation (SR) and neutron sources, the latter paving the way for contrast variation by solvent exchange ($\text{H}_2\text{O}/\text{D}_2\text{O}$) [3] or specific deuteration [4] methods. It was realized that scattering studies on solutions provide, for a minimal investment in time and effort, useful insights into the structure of non-crystalline biochemical systems. Moreover, SAXS/SANS also made it possible to investigate intermolecular interactions including assembly and large-scale conformational changes, on which biological function often relies, in real time.

SAXS/SANS experiments typically require a homogeneous dilute solution of macromolecules in a near physiological buffer without special additives. The price to pay for the relative simplicity of sample preparation is the low information content of the scattering data in the absence of crystalline order. For dilute protein solutions comprising monodisperse systems of identical particles, the random orientation of particles in solution leads to spherical averaging of the single particle scattering, yielding a one-dimensional scattering pattern. The main difficulty, and simultaneously the main challenge, of SAS as a structural method is to extract information about the three-dimensional structure of the object from these one-dimensional experimental data. In the past, only overall particle parameters (e.g. volume, radius of gyration) of the macromolecules were directly determined from the experimental data, whereas the analysis in terms of three-dimensional models was limited to simple geometrical bodies (e.g. ellipsoids, cylinders, etc) or was performed on an ad hoc trial-and-error basis [5, 6]. Electron microscopy (EM) was often used as a constraint in building consensus models [7, 8]. In the 1980s, progress in other structural methods led to a decline of the interest of biochemists in SAS studies drawing structural conclusions from a couple of overall parameters or trial-and-error models. In contrast, for inorganic and especially polymer systems, integral parameters extracted from SAXS/SANS are usually sufficient to answer most of the structural questions [5, 6]. Introduction of SR for time-resolved measurements during the processing of polymers [9], therefore, also had a major impact.

The 1990s brought a breakthrough in SAXS/SANS data analysis methods, allowing reliable *ab initio* shape and domain structure determination and detailed modelling of macromolecular complexes using rigid body refinement. This progress was accompanied by further advances in instrumentation, and time resolutions down to the sub-ms were achieved on third generation SR sources in studies of protein and nucleic acid folding. This review focuses, after a brief account of the basics of x-ray and neutron SAS theory and instrumentation, on the interpretation of the scattering patterns from macromolecular solutions. Novel data

analysis methods are presented and illustrated by applications to various types of biological macromolecules in solution. New opportunities offered by third and fourth generation SR sources (analysis of fast kinetics, coherent scattering and single molecule scattering) are also discussed.

2. Basics of SAS

2.1. Scattering of x-rays and neutrons

Although the physical mechanisms of elastic x-ray and neutron scattering by matter are fundamentally different, they can be described by the same mathematical formalism. The basics of scattering are therefore presented simultaneously, pointing to the differences between the two types of radiation. X-ray photons with an energy E have a wavelength $\lambda = 1.256/E$, where λ is expressed in nm and E in keV. For structural studies, relatively hard x-rays with energies around 10 keV are used (λ about 0.10–0.15 nm). The neutron wavelength is given by de Broglie's relationship, λ [nm] = $396.6/v$ [m s⁻¹], where v is the (group) velocity of neutrons, and thermal neutrons with wavelengths λ around 0.20–1.0 nm are typically employed. When an object is illuminated by a monochromatic plane wave with wavevector $k_0 = |\mathbf{k}_0| = 2\pi/\lambda$, atoms within the object interacting with the incident radiation become sources of spherical waves. We shall consider only elastic scattering (i.e. without energy transfer) so that the modulus of the scattered wave $k_1 = |\mathbf{k}_1|$ is equal to k_0 . The amplitude of the wave scattered by each atom is described by its scattering length, f . For hard x-rays interacting with electrons the atomic scattering length is $f_x = N_e r_0$ where N_e is the number of electrons and $r_0 = 2.82 \times 10^{-13}$ cm is the Thomson radius. The atomic scattering length does not depend on the wavelength unless the photon energy is close to an absorption edge of the atom. In this case, there is resonant or anomalous scattering, a phenomenon used for experimental phase determination in crystallography [10], and also in some SAS applications [11]. Neutrons interact with the nuclear potential and with the spin and the neutron scattering length consists of two terms $f_n = f_p + f_s$. The last term bears structural information only if the neutron spins in the incident beam and the nuclear spins in the object are oriented [12], otherwise the spin scattering yields only a flat incoherent background. In contrast to the situation with x-rays, f_p does not increase with the atomic number but is sensitive to the isotopic content. Table 1 displays two major differences between the x-ray and neutron scattering length: (i) neutrons are more sensitive to lighter atoms than x-rays; (ii) there is a large difference between the neutron scattering lengths of hydrogen and deuterium. The former difference is largely employed in neutron crystallography to localize hydrogen atoms in the crystal [13]; the latter provides an effective tool for selective labelling and contrast variation in neutron scattering and diffraction [14–17].

The scattering process in the first Born approximation is described by Fourier transformation from the 'real' space of laboratory (object) coordinates \mathbf{r} to the 'reciprocal' space of scattering vectors $\mathbf{s} = (s, \Omega) = \mathbf{k}_1 - \mathbf{k}_0$. Following the properties of the Fourier

Table 1. X-ray and neutron scattering lengths of some elements.

Atom	H	D	C	N	O	P	S	Au
Atomic mass	1	2	12	14	16	30	32	197
N electrons	1	1	6	7	8	15	16	79
$f_x, 10^{-12}$ cm	0.282	0.282	1.69	1.97	2.16	3.23	4.51	22.3
$f_n, 10^{-12}$ cm	−0.374	0.667	0.665	0.940	0.580	0.510	0.280	0.760

transform (i.e. the reciprocity between dimensions in real and reciprocal space implying that the smaller the 'real' size, the larger the corresponding 'reciprocal' size) the neutron scattering amplitudes of atoms can be considered to be constants due to the small (10^{-13} cm) size of the nucleus. The x-ray scattering amplitudes representing the Fourier transform of the electron density distribution in the (spherical) atom are functions $f(s)$ of the momentum transfer $s = 4\pi\lambda^{-1} \sin(\theta)$ where 2θ is the scattering angle, and $f(0) = f_x$. Atomic form factors along with other useful information are now conveniently available on the Web from numerous on-line sources (e.g. [18]).

2.2. Scattering by macromolecular solutions

To describe the scattering from assemblies of atoms, it is convenient to introduce the scattering length density distribution $\rho(\mathbf{r})$ equal to the total scattering length of the atoms per unit volume. The experiments on macromolecules in solutions involve separate measurements of the scattering from the solution and the solvent (figure 1). Assuming that the solvent is a featureless matrix with a constant scattering density ρ_s , the difference scattering amplitude from a single particle relative to that of the equivalent solvent volume, is defined by the Fourier transform of the excess scattering length density $\Delta\rho(\mathbf{r}) = \rho(\mathbf{r}) - \rho_s$

$$A(s) = \mathfrak{F}[\Delta\rho(\mathbf{r})] = \int_V \Delta\rho(\mathbf{r}) \exp(i\mathbf{s}\cdot\mathbf{r}) d\mathbf{r}, \quad (1)$$

where the integration is performed over the particle volume. In a scattering experiment one cannot directly measure the amplitude but only the scattering intensity $I(s) = A(s)A^*(s)$ proportional to the number of photons or neutrons scattered in the given direction s .

If one now considers an ensemble of identical particles, the total scattering will depend on the distribution of these particles and on the coherence properties of the radiation, and for usual sources two major limiting cases should be considered. In the case of an ideal single crystal, all particles in the sample have defined correlated orientations and are regularly distributed in

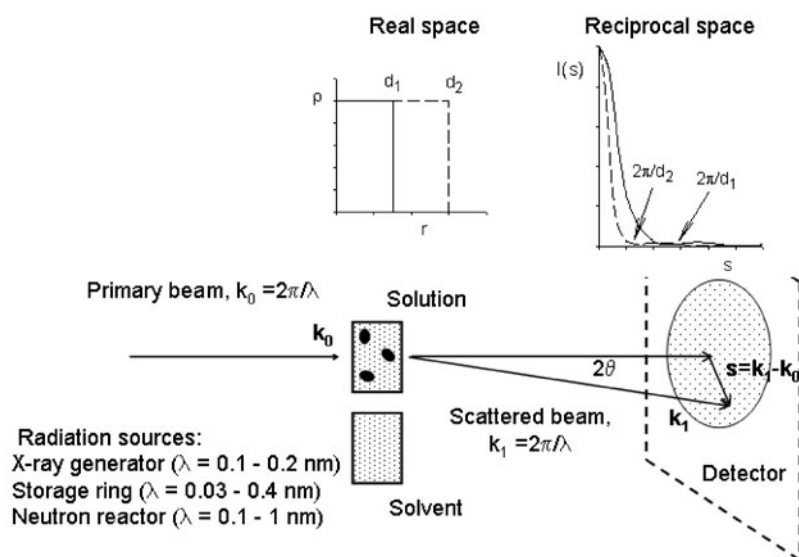


Figure 1. Schematic representation of a SAS experiment and the Fourier transformation from real to reciprocal space.

space, so that scattering amplitudes of individual particles have to be summed up accounting for all interparticle interferences. As a result, the total scattered intensity is redistributed along specific directions defined by the reciprocal lattice and the discrete three-dimensional function $I(s_{hkl})$ measured correspond to the density distribution in a single unit cell of the crystal [19]. If the particles are randomly distributed and their positions and orientations are uncorrelated, their scattering intensities rather than their amplitudes are summed (no interference). Accordingly, the intensity from the entire ensemble is a continuous isotropic function proportional to the scattering from a single particle averaged over all orientations $I(s) = \langle I(s) \rangle_{\Omega}$. Dilute solutions of monodisperse non-interacting biological macromolecules under specific solvent conditions correspond to this second limiting case, which will mostly be considered later. If the particles in solution are randomly oriented but also interact (non-ideal semi-dilute solutions), local correlations between the neighbouring particles must be taken into account. The scattering intensity from the ensemble will still be isotropic and for spherical particles can be written as $I_S(s) = I(s) \times S(s)$, where $S(s)$ is the term describing particle interactions. In the literature, the particle scattering $I(s)$ and the interference term $S(s)$ are called ‘form factor’ and ‘structure factor’, respectively. This is a somewhat misleading terminology, as, for example, $I(s)$ depends not only on the form but also on the internal structure of the particle (to further add to the confusion, in crystallography what is called here ‘structure factor’ is the reciprocal lattice and what is called here ‘form factor’ is called structure factor!). In biological applications, SAS is used to analyse the structure of dissolved macromolecules (based on the particle scattering, section 3) as well as the interactions based on the interference term (section 4.2). Separation of the two terms for semi-dilute solutions is possible by using measurements at different concentrations or/and in different solvent conditions (pH, ionic strength, etc). For systems of particles differing in size and/or shape, the total scattering intensity will be given by the weight average of the scattering from the different types of particles (section 4.1).

2.3. Resolution and contrast

The Fourier transformation of the box function in figure 1 illustrates that most of the intensity scattered by an object of linear size d is concentrated in the range of momentum transfer up to $s = 2\pi/d$. It is therefore assumed that if the scattering pattern is measured in reciprocal space up to s_{\max} it provides information about the real space object with a resolution $\delta = 2\pi/s$. For single crystals, due to the redistribution of the diffracted intensity into reflections, the data can be recorded to high resolution ($d \sim \lambda$). For spherically averaged scattering patterns from solutions, $I(s)$ is usually a rapidly decaying function of momentum transfer and only low resolution patterns ($d \gg \lambda$) are available. It is thus clear that solution scattering cannot provide information about the atomic positions but only about the overall structure of macromolecules in solution.

The average excess density of the particle $\Delta\rho = \langle \Delta\rho(\mathbf{r}) \rangle = \langle \rho(\mathbf{r}) \rangle - \rho_s$, called contrast, is another important characteristic of the sample. The particle density can be represented [20] as $\rho(\mathbf{r}) = \Delta\rho\rho_C(\mathbf{r}) + \rho_F(\mathbf{r})$, where $\rho_C(\mathbf{r})$ is the shape function equal to 1 inside the particle and 0 outside, whereas $\rho_F(\mathbf{r}) = \rho(\mathbf{r}) - \langle \rho(\mathbf{r}) \rangle$ represents the fluctuations of the scattering length density around its average value. Inserting this expression in equation (1), the scattering amplitude contains two terms $A(s) = \Delta\rho A_C(s) + A_F(s)$ so that the averaged intensity is written in terms of three basic scattering functions:

$$I(s) = (\Delta\rho)^2 I_C(s) + 2\Delta\rho I_{CF}(s) + I_F(s), \quad (2)$$

where $I_C(s)$, $I_F(s)$ and $I_{CF}(s)$ are the scattering from the particle shape, fluctuations and the cross-term, respectively [20]. This equation is of general value and the contributions from

the overall shape and internal structure of particles can be separated using measurements in solutions with different solvent density (i.e. for different $\Delta\rho$). This technique is called contrast variation (see section 3.7).

2.4. X-ray and neutron scattering instruments

The concept of resolution allows one to estimate the angular range required for solution scattering experiments. Imagine that one uses radiation with $\lambda = 0.1$ nm to study a particle of characteristic size 10 nm, which requires a resolution range from, say, $d = 20$ to 1 nm. Recalling the resolution relation $2\pi/d = 4\pi \sin \theta/\lambda$, the corresponding angular range will be from about 0.005 to 0.1 rad, i.e. 0.3 – 6° . The entire scattering pattern is thus recorded at very small scattering angles, which gives the generic name for the method: SAS.

Conceptually, SAS measurements are simple (figure 1), but the design of an instrument is a challenging technical task as great care must be taken to reduce the parasitic scattering in the vicinity of the primary beam. Moreover, for biological systems the contrast of the particles $\Delta\rho$ is usually small and the useful signal may be weak compared to the background (see example in figure 2). Here, we shall briefly present the basic elements of SAS cameras; detailed reviews

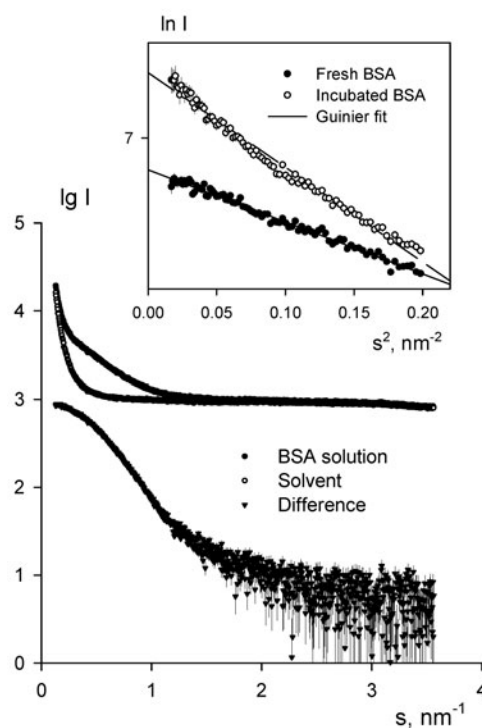


Figure 2. Typical x-ray scattering patterns from a solution of BSA in 50 mM HEPES, pH 7.5, solvent scattering and the difference curve (pure scattering from the protein, scaled for the solute concentration, 5 mg ml^{-1}). Note that the curves are plotted on a semi-logarithmic scale. The inset displays the Guinier plots for the fresh BSA sample (monodisperse solution, linear Guinier plot) and for the same solution after 8 h incubation at room temperature causing unspecific aggregation (no linearity in the Guinier range). The experimental data was recorded at the EMBL beamline X33 (synchrotron DESY, Hamburg); the sample container was a cuvette with two $25 \mu\text{m}$ thick mica windows, sample thickness 1 mm.

on small-angle instrumentation can be found elsewhere (e.g. [21, 22] for SAXS and [23, 24] for SANS).

For SAXS, laboratory cameras based on x-ray generators are available (e.g. the NanoSTAR camera from the Bruker Group, or Kratky camera from Hecus M Braun, Austria) but most challenging projects rely on the much higher brilliance of SR. Modern SR beamlines are generally equipped with a tunable fixed exit double monochromator and mirrors to reject higher harmonics ($\lambda/2$, $\lambda/3$, ...). The parasitic scattering around the beam is reduced by using several pairs of guard slits made from highly absorbing material like tungsten or tantalum. The example in figure 3 is that of the BioCAT undulator beamline at the third generation advanced photon source (APS, Argonne National Laboratory, USA [25]) and most modern SAXS beamlines at large-scale facilities (ESRF in Grenoble, Spring-8 in Himeji, LNS in Brookhaven, SSRL in Stanford, Elettra in Trieste, etc) have a similar design. The design of some of the beamlines on bending magnets at second generation sources relied on bent monochromators to obtain a sufficiently small focus and large photon flux (EMBL Outstation in Hamburg, SRS in Daresbury, LURE-DCI in Orsay, Photon factory at Tsukuba). The monochromatic beam (bandpass $\Delta\lambda/\lambda \sim 10^{-4}$) of the BioCAT beamline contains about 10^{13} photons \times s $^{-1}$ on the sample over the range from 3.5 to 39 keV (wavelength λ from 0.34 to 0.03 nm). Double focusing optics provides focal spot sizes of about $150 \times 40 \mu\text{m}^2$ (FWHM) at $\lambda = 0.1$ nm with a positional beam stability of a few μm within the time interval of an experiment. Exchangeable vacuum chambers allow sample-to-detector distances from 150 to 5500 mm covering the s range from ~ 0.001 to $\sim 30 \text{ nm}^{-1}$. Low concentration ($\sim 1 \text{ mg ml}^{-1}$) protein solutions can be measured using short exposure times (~ 1 s). The BioCAT beamline employs a high-sensitivity charge coupled device (CCD) detector with a $50 \times 90 \text{ mm}^2$ working area and $50 \mu\text{m}$ spatial resolution. CCD detectors with [26] or without [27] image intensifiers are increasingly being used on high flux beamlines but special experimental procedures are required to reduce the effects of dark current [28]. These effects may lead to systematic deviations in the intensities recorded at higher angles, which makes the buffer subtraction yet more difficult (cf figure 2). On lower flux beamlines, position sensitive gas proportional detectors with delay line readout are still used, which, although only tolerating lower count rates, are free from such distortions [29]. Pixel detectors, which are fast readout solid state counting devices [30], do not yet have sufficient dimensions to be useful for SAXS applications.

Neutron scattering beamlines on steady-state sources (e.g. ILL in Grenoble, NIST in Gaithersburg, FZJ in Jülich, ORNL in Oak Ridge, ANSTO in Menai, LLB in Saclay,

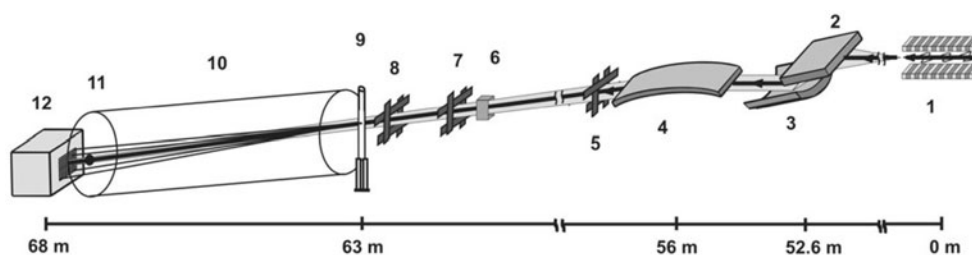


Figure 3. Schematic representation of the synchrotron x-ray scattering BioCAT-18ID beamline at the APS, Argonne National Laboratory, USA: (1) primary beam coming from the undulator, (2) and (3) flat and sagittally focusing Si (111) crystal of the double-crystal monochromator, respectively, (4) vertically focusing mirror, (5) collimator slits, (6) ion chamber, (7) and (8) guard slits, (9) temperature-controlled sample-flow cell, (10) vacuum chamber, (11) beamstop with a photodiode, (12) CCD detector (T Irving, personal communication).

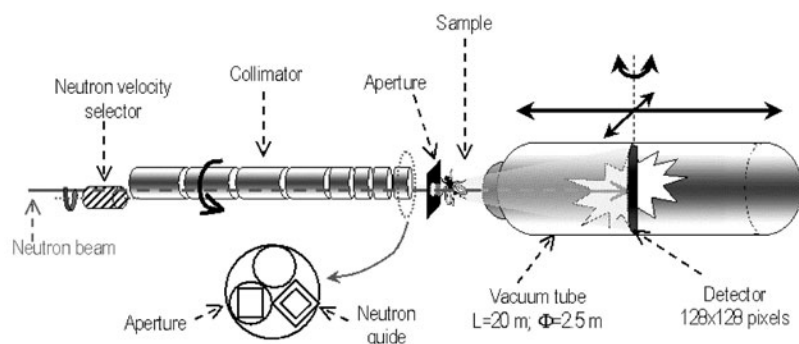


Figure 4. Schematic representation of the D22 neutron scattering instrument at the Institut Laue-Langevin, Grenoble, France. Adapted with permission from <http://www.ill.fr/YellowBook/D22/>.

PSI in Villigen) are conceptually similar to the synchrotron SAXS beamlines (see figure 4). The main difference with most x-ray instruments is that to compensate for the low neutron flux a relatively broad spectral band ($\Delta\lambda/\lambda \sim 0.1$ full-width at half-maximum (FWHM)) is selected utilizing the relation between the velocity of neutrons and their de Broglie wavelength by a mechanical velocity selector with helical lamellae. Position sensitive gas proportional detectors filled with ^3He are used for neutron detection, but the requirements for spatial resolution and count rate are much lower than in the case of x-rays, due to the much lower spectral brilliance and large beam sizes of neutron sources. Even on the D22 SANS camera at the ILL schematically presented in figure 4, currently the best neutron scattering instrument, the flux on the sample does not exceed $10^8 \text{ neutrons} \times \text{cm}^{-2} \times \text{s}^{-1}$.

On pulsed reactors or spallation sources (e.g. JINR in Dubna, ISIS in Chilton, IPNS in Argonne, KEK in Tsukuba) ‘white’ incident beam is used. The scattered radiation is detected by time-of-flight methods again using the relation between neutron velocity and wavelength. The scattering pattern is recorded in several time frames after the pulse, and the correlation between time and scattering angle yields the momentum transfer for each scattered neutron. Time-of-flight techniques allow one to record a wide range of momentum transfer in a single measurement without moving the detector. Current SANS instruments using cold sources on steady-state reactors still outperform existing time-of-flight stations. However, because of potential hazards associated with steady-state reactors, new generation of pulsed neutron spallation sources may be the future for neutron science and SANS in particular.

The optimum sample thickness—determined by the sample transmission—is typically about 1 mm for aqueous solutions of biomolecules, both for SAXS and SANS. The sample containers are usually thermostated cells with mica windows or boron glass capillaries for SAXS (typical sample volume about 50–100 μl) and standard spectroscopic quartz cuvettes for SANS (volume 200–300 μl). Using high flux SR, radiation damage is a severe problem, and continuous flow cells are used to circumvent this effect (figure 3). As thermal neutrons have much lower energies than x-rays, there is virtually no radiation damage during SANS experiments. To reduce the contribution from the sample container to parasitic scattering near the primary beam, evacuated sample chambers can be used. Special purpose containers (e.g. stopped-flow cells) are required for most time-resolved experiments (see section 5.6).

3. Monodisperse systems

This section is devoted to data analysis from monodisperse systems assuming the ideal case of non-interacting dilute solutions of identical particles. In other words, it will be assumed that

an isotropic function $I(s)$ proportional to the scattering from a single particle averaged over all orientations is available. The main structural task in this case is to reconstruct the particle structure (i.e. its excess scattering length density distribution $\Delta\rho(\mathbf{r})$) at low resolution from the scattering data.

3.1. Overall parameters

Using expression (1) for the Fourier transformation one obtains for the spherically averaged single particle intensity

$$I(s) = \langle A(s)A^*(s) \rangle_{\Omega} = \left\langle \int_V \int_V \Delta\rho(\mathbf{r})\Delta\rho(\mathbf{r}') \exp\{i\mathbf{s}(\mathbf{r}-\mathbf{r}')\} d\mathbf{r} d\mathbf{r}' \right\rangle_{\Omega} \quad (3)$$

or, taking into account that $\langle \exp(i\mathbf{s}\mathbf{r}) \rangle_{\Omega} = \sin(sr)/sr$ and integrating in spherical coordinates,

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

where

$$\gamma(r) = \left\langle \int_{\omega} \Delta\rho(\mathbf{u})\Delta\rho(\mathbf{u}+\mathbf{r}) d\mathbf{u} \right\rangle_{\omega} \quad (5)$$

is the spherically averaged autocorrelation function of the excess scattering density, which is obviously equal to zero for distances exceeding the maximum particle diameter D_{\max} . In practice, the function $p(r) = r^2 \gamma(r)$ corresponding to the distribution of distances between volume elements inside the particle weighted by the excess density distribution is often used. This distance distribution function is computed by the inverse transformation

$$p(r) = \frac{r^2}{2\pi^2} \int_0^{\infty} s^2 I(s) \frac{\sin sr}{sr} ds. \quad (6)$$

The behaviour of the scattering intensity at very small ($s \rightarrow 0$) and very large ($s \rightarrow \infty$) values of momentum transfer is directly related to overall particle parameters. Indeed, near $s = 0$ one can insert the McLaurin expansion $\sin(sr)/sr \approx 1 - (sr)^2/3! + \dots$ into (4) yielding

$$I(s) = I(0)[1 - \frac{1}{3}R_g^2 s^2 + O(s^4)] \cong I(0) \exp(-\frac{1}{3}R_g^2 s^2), \quad (7)$$

where the forward scattering $I(0)$ is proportional to the squared total excess scattering length of the particle

$$I(0) = \int_V \int_V \Delta\rho(\mathbf{r})\Delta\rho(\mathbf{r}') d\mathbf{r} d\mathbf{r}' = 4\pi \int_0^{D_{\max}} p(r) dr = (\Delta\rho)^2 V^2 \quad (8)$$

and the radius of gyration R_g is the normalized second moment of the distance distribution of the particle around the centre of its scattering length density distribution

$$R_g = \left[\frac{\int_0^{D_{\max}} r^2 p(r) dr}{2 \int_0^{D_{\max}} p(r) dr} \right]^{-1}. \quad (9)$$

Equation (7), derived by Guinier [1], has long been the most important tool in the analysis of scattering from isotropic systems and continues to be very useful at the first stage of data analysis. For ideal monodisperse systems, the Guinier plot ($\ln(I(s))$ versus s^2) should be a linear function, whose intercept gives $I(0)$ and the slope yields the radius of gyration R_g . Linearity of the Guinier plot can be considered as a test of the sample homogeneity and deviations indicate attractive or repulsive interparticle interactions leading to interference effects (see example in figure 2, and also section 4.2). One should, however, always bear

in mind that the Guinier approximation is valid for very small angles only, namely in the range $s < 1.3/R_g$, and fitting a straight line beyond this range is unphysical.

Whereas the radius of gyration R_g characterizes the particle size, the forward scattering $I(0)$ is related to its molecular mass (MM). Indeed, the experimentally obtained value of $I(0)$ is proportional to the squared contrast of the particle, the number of particles in the illuminated volume, and to the intensity of the transmitted beam. The contrast can be computed from the chemical composition and specific volume of the particle, the number of particles from the beam geometry and sample concentration c , and the beam intensity can be obtained directly (using, e.g. an ionization chamber or a photodiode) or indirectly using a standard scatterer. Equations exist to compute the MM of the solute from the absolute SAXS or SANS measurements using primary or secondary (calibrated) standards like lupolen [31]. In practice, the MM can often be readily estimated by comparison with a reference sample (for proteins, lysozyme or bovine serum albumin (BSA) solution). In SANS, calibration against water scattering is frequently used [14], and a similar procedure exists for SAXS [32]. In practice, the accuracy of MM determination is often limited by that of the protein concentration required for normalization.

Equation (7) is valid for arbitrary particle shapes. For very elongated particles, the radius of gyration of the cross-section R_c can be derived using a similar representation plotting $sI(s)$ versus s^2 , and for flattened particles, the radius of gyration of the thickness R_t is computed from the plot of $s^2I(s)$ versus s^2 :

$$sI(s) \cong I_C(0) \exp(-\frac{1}{2}R_c^2s^2), \quad s^2I(s) \cong I_T(0) \exp(-R_t^2s^2). \quad (10)$$

In some cases it is possible to extract the cross-sectional or thickness information in addition to the overall parameters of the particle. However, for biological filaments like actin, myosin, chromatin, which may be hundreds of nm long, it may not be possible to record reliable data in the Guinier region ($s < 1.3/R_g$). Clearly, in these cases only cross-sectional parameters are available and correspondingly less structural information can be obtained by SAXS or SANS than for more isometric particles.

To analyse the asymptotic behaviour of $I(s)$ at large angles, let us integrate equation (4) twice by parts. Taking into account that $\gamma(D_{\max}) = 0$, one can write

$$I(s) \cong 8\pi s^{-4} \gamma'(0) + O_1 s^{-3} + O_2 s^{-4} + o(s^{-5}), \quad (11)$$

where O_1, O_2 are oscillating trigonometric terms of the form $\sin(sD_{\max})$. The main term responsible for the intensity decay at high angles is therefore proportional to s^{-4} , and this is known as Porod's law [33]. Moreover, for homogeneous particles, $\gamma'(0)$ is equal to $-(\Delta\rho)^2 S/4$, where S is the particle surface. To eliminate the particle contrast, one can use the so-called Porod invariant [33]

$$Q = \int_0^\infty s^2 I(s) ds = 2\pi^2 \int_V (\Delta\rho(\mathbf{r}))^2 d\mathbf{r} \quad (12)$$

(the reciprocal and real space integrals are equal due to Parseval's theorem applied to equation (3)). For homogeneous particles, $Q = 2\pi^2(\Delta\rho)^2 V$, and, taking into account that $I(0) = (\Delta\rho)^2 V^2$, the excluded particle (Porod) volume is $V = 2\pi^2 I(0) Q^{-1}$. Hence, the normalized asymptote allows to estimate the particle specific surface as $S/V = (\pi/Q) \lim_{s \rightarrow \infty} [s^4 I(s)]$ (note that, thanks to the Porod invariant, both parameters can be obtained from the data on relative scale). In practice, internal inhomogeneities lead to deviations from the Porod asymptote, which, for single-component macromolecules with a large MM (>40 kDa) at sufficiently high contrasts, can usually reasonably be taken into account by simply subtracting a constant term from the experimental data. The data at high angles are assumed to follow a linear plot in $s^4 I(s)$ against s^4 coordinates: $s^4 I(s) \approx Bs^4 + A$, and subtraction of the constant B from $I(s)$ yields an approximation to the scattering of the corresponding homogeneous body.

3.2. Distance distribution function and particle anisotropy

In principle, the distance distribution function $p(r)$ contains the same information as the scattering intensity $I(s)$, but the real space representation is more intuitive and information about the particle shape can often be deduced by straightforward visual inspection of $p(r)$ [5]. Figure 5 presents typical scattering patterns and distance distribution functions of geometrical bodies with the same maximum size. Globular particles (curve 1) display bell-shaped $p(r)$ functions with a maximum at about $D_{\max}/2$. Elongated particles have skewed distributions with a clear maximum at small distances corresponding to the radius of the cross-section (curve 2). Flattened particles display a rather broad maximum (curve 3), also shifted to distances smaller than $D_{\max}/2$. A maximum shifted towards distances larger than $D_{\max}/2$ is usually indicative of a hollow particle (curve 4). Particles consisting of well-separated subunits may display multiple

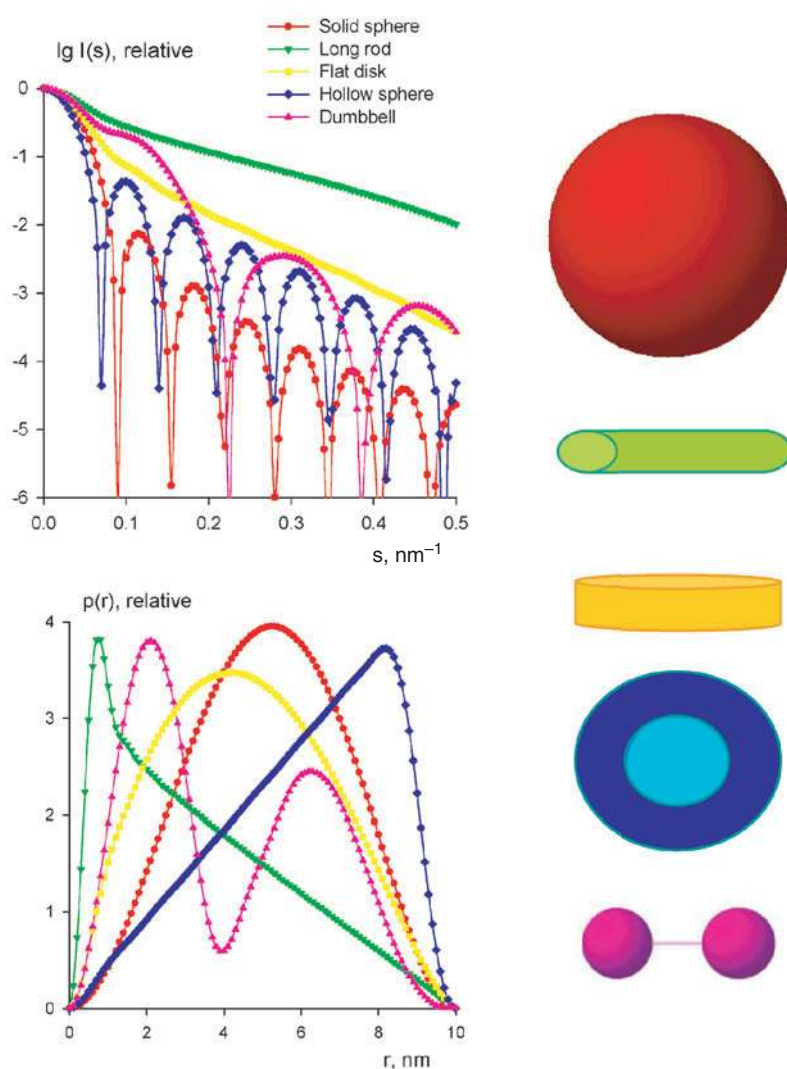


Figure 5. Scattering intensities and distance distribution functions of geometrical bodies.

maxima, the first corresponding to the intrasubunit distances, the others yielding separation between the subunits (curve 5). The differences in the scattering patterns themselves allow one to easily detect spherically symmetric objects which give scattering patterns with distinct minima. Very anisometric particles yield featureless scattering curves which decay much more slowly than those of globular particles. Most frequently occurring distances manifest themselves as maxima or shoulders in the scattering patterns (note the shoulder at $s = 0.1$ nm in the dumbbell scattering (curve 5)). In general, however, the scattering curves are somewhat less instructive than the $p(r)$ functions.

Even for simple geometrical bodies there are only a few cases where $I(s)$ and/or $p(r)$ functions can be expressed analytically. The best known are the expressions for a solid sphere of radius R : $I(s) = A(s)^2$, $A(s) = (4\pi R^3/3)[\sin(x) - x \cos(x)]/x^3$ where $x = sR$, and $p(r) = (4\pi R^3/3)r^2(1 - 3t/4 + t^3/16)$, where $t = r/R$. Semi-analytical equations for the intensities of ellipsoids, cylinders and prisms were derived by Mittelbach and Porod in the 1960s, and later, analytical formulae for the $p(r)$ function of some bodies were published (e.g. of a cube [34]). A collection of analytical and semi-analytical equations for $I(s)$ of geometrical bodies can be found in [6].

Reliable computation of $p(r)$ is a necessary prerequisite for further analysis in terms of three-dimensional models. Direct Fourier transformation of the experimental data using equation (6) is not possible, as the exact intensity $I(s)$ is not available. Instead, the experimental data $I(s)$ is only measured at a finite number of N points (s_i) in the interval $[s_{\min}, s_{\max}]$ rather than $[0, \infty]$. The precision of these measurements is determined by the corresponding statistical errors (σ_i) but there are also always some systematic errors. In particular, especially in laboratory x-ray or in neutron scattering experiments, smearing due to instrumental effects (finite beam size, divergence and/or polychromaticity) may occur so that the measured data deviate systematically from the ideal curve. One could, in principle, desmear the data and extrapolate $I(s)$ to zero (using a Guinier plot) and infinity (using the Porod asymptote) but this procedure, although often used in the past, is cumbersome and not very reliable. It is more convenient to use indirect Fourier transformation based on equation (4), the technique first proposed in [35]. Representing $p(r)$ on $[0, D_{\max}]$ by a linear combination of K orthogonal functions $\varphi_k(r)$

$$p(r) = \sum_{k=1}^K c_k \varphi_k(s_i), \quad (13)$$

the coefficients c_k can be determined by fitting the experimental data minimizing the functional

$$\Phi_\alpha(c_k) = \sum_{i=1}^N \left[\frac{I_{\text{exp}}(s_i) - \sum_{k=1}^K c_k \psi_k(s_i)}{\sigma(s_i)} \right]^2 + \alpha \int_0^{D_{\max}} \left[\frac{dp}{dr} \right]^2 dr, \quad (14)$$

where $\psi_k(q)$ are the Fourier transformed and (if necessary) smeared functions $\varphi_k(r)$. The regularizing multiplier $\alpha \geq 0$ controls the balance between the goodness of fit to the data (first summand) and the smoothness of the $p(r)$ function (second summand).

There exist several implementations of the indirect transform approach, differing in the type of orthogonal functions used to represent $p(r)$ and in numerical detail [35–37]. The method of Moore [36] using few sine functions does not require a regularization term but may lead to systematic deviations in the $p(r)$ of anisometric particles [38]. The other methods usually employ dozens of parameters c_k and the problem lies in selecting the proper value of the regularizing multiplier α . Too small values of α yield solutions unstable to experimental errors, whereas too large values lead to systematic deviations from the experimental data. The program GNOM [37, 39] provides the necessary guidance using a set of perceptual criteria

describing the quality of the solution. It either finds the optimal solution automatically or signals that the assumptions about the system (e.g. the value of D_{\max}) are incorrect.

The indirect transform approach is usually superior to other techniques as it imposes strong constraints, namely boundedness and smoothness of $p(r)$. An approximate estimate of D_{\max} is usually known *a priori* and can be iteratively refined. The forward scattering and the radius of gyration can be readily derived from the $p(r)$ functions following equations (8) and (9) and the use of indirect transformation yields more reliable results than the Guinier approximation, to a large extent because the calculation using $p(r)$ is less sensitive to the data cut-off at small angles. Indeed, with the indirect methods, the requirement of having a sufficient number of data points for $s < 1.3/R_g$ for the Guinier plot is relaxed to the less stringent condition $s_{\min} < \pi/D_{\max}$ (see next section).

3.3. Shannon sampling and information content

Following the previous section, some overall particle parameters can be computed directly from the experimental data without model assumptions (R_g , MM, D_{\max}) and a few more can be obtained under the assumption that the particle is (nearly) homogeneous (V , S). This raises the general question about the maximum number of independent parameters that can in principle be extracted from the scattering data. A measure of information content is provided by Shannon's sampling theorem [40], stating that

$$sI(s) = \sum_{k=1}^{\infty} s_k I(s_k) \left[\frac{\sin D_{\max}(s - s_k)}{D_{\max}(s - s_k)} - \frac{\sin D_{\max}(s + s_k)}{D_{\max}(s + s_k)} \right]. \quad (15)$$

This means that the continuous function $I(s)$ can be represented by its values on a discrete set of points (Shannon channels) where $s_k = k\pi/D_{\max}$, which makes $I(s)$ a so-called analytical function [41]. The minimum number of parameters (or degrees of freedom) required to represent an analytical function on an interval $[s_{\min}, s_{\max}]$ is given by the number of Shannon channels ($N_S = D_{\max}(s_{\max} - s_{\min})/\pi$) in this interval.

The number of Shannon channels does provide a very useful guidance for performing a measurement, in particular, the value of s_{\min} should not exceed that of the first Shannon channel ($s_{\min} < \pi/D_{\max}$). This obviously puts some limits on the use of indirect transformation methods described in the previous section. In practice, solution scattering curves decay rapidly with s and they are normally recorded only at resolutions below 1 nm, so that the number of Shannon channels typically does not exceed 10–15. It would, however, be too simple to state, that N_S limits the number of parameters that could be extracted from the scattering data. The experimental SAS data are usually vastly oversampled, i.e. the angular increment in the data sets is much smaller than the Shannon increment $\Delta s = \pi/D_{\max}$. As known from optical image reconstruction [41], this oversampling in principle allows one to extend the data beyond the measured range (so-called 'superresolution') and thus to increase the effective number of Shannon channels. The level of detail of models, which can be deduced from solution scattering patterns depends not only on the actual value N_S but also on other factors, like the accuracy of the data or the available *a priori* information.

3.4. Ab initio analysis of particle shape and domain structure

It is clear that reconstruction of a three-dimensional model of an object from its one-dimensional scattering pattern is an ill-posed problem. To simplify the description of the low-resolution models that can legitimately be obtained data interpretation is often performed in terms of homogeneous bodies (the influence of internal inhomogeneities for single component particles

can largely be eliminated by subtracting a constant as described in section 3.1). In the past, shape modelling was done by trial-and-error, computing scattering patterns from different shapes and comparing them with the experimental data. The models were either three-parameter geometrical bodies like prisms, triaxial ellipsoids, elliptical or hollow circular cylinders, etc, or shapes built from assemblies of regularly packed spheres (beads). The scattering patterns of these models was computed using analytical or semi-analytical formulae (see section 2.2), except for the bead models where Debye's formula [42] was used

$$I(s) = \sum_{i=1}^K \sum_{j=1}^K f_i(s) f_j(s) \frac{\sin(sr_{ij})}{sr_{ij}}, \quad (16)$$

where K is the number of beads, $f_i(s)$ is the scattering amplitude from the i th bead (usually, that of a solid sphere) and $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ is the distance between a pair of spheres. This type of modelling allowed to construct complicated models but had to be constrained by additional information (e.g. from EM or hydrodynamic data).

Historically, the first and very elegant *ab initio* shape determination method was proposed in [43]. The particle shape was represented by an angular envelope function $r = F(\omega)$ describing the particle boundary in spherical coordinates (r, ω) . This function is economically parametrized as

$$F(\omega) \approx F_L(\omega) = \sum_{l=0}^L \sum_{m=-l}^l f_{lm} Y_{lm}(\omega), \quad (17)$$

where $Y_{lm}(\omega)$ are spherical harmonics, and the multipole coefficients f_{lm} are complex numbers. For a homogeneous particle, the density is

$$\rho_c(\mathbf{r}) = \begin{cases} 1, & 0 \leq r < F(\omega), \\ 0, & r \geq F(\omega) \end{cases} \quad (18)$$

and the shape scattering intensity is expressed as [44]

$$I(s) = 2\pi^2 \sum_{l=0}^{\infty} \sum_{m=-l}^l |A_{lm}(s)|^2, \quad (19)$$

where the partial amplitudes $A_{lm}(s)$ are readily computed from the shape coefficients f_{lm} using recurrent formulae based on 3j-Wigner coefficients [45]. These coefficients are determined by non-linear optimization starting from a spherical approximation to minimize the discrepancy χ between the experimental and the calculated scattering curves

$$\chi^2 = \frac{1}{N-1} \sum_{j=1}^N \left[\frac{I_{\text{exp}}(s_j) - \eta I(s_j)}{\sigma(s_j)} \right]^2, \quad (20)$$

where η is a scaling factor. The truncation value L in equation (17) defines the number of independent parameters N_p , which, for low-resolution envelopes, is comparable with the number of Shannon channels in the data. In the general case, $N_p = (L+1)^2 - 6$, i.e. one requires 10–20 parameters for $L = 3$ –4, and this number is further reduced for symmetric particles [46]. The method—implemented in the program SASHA [47]—was the first publicly available shape determination program for SAS.

The envelope function approach contributed substantially to the progress of the methods for solution scattering data interpretation. The spherical harmonics formalism proved to be extremely useful for the analysis of SAS data and its formalism was employed in many later methods. Thanks to the small number of parameters, the envelope method yielded unique

solutions in most practical cases and its successful applications demonstrated that the SAS curves did contain information, enabling one to reconstruct three-dimensional shapes at low resolution. Use of the angular envelope function was, however, limited to relatively simple shapes (in particular, without holes inside the particle). A more comprehensive description is achieved in the bead methods [48, 49], which use the vastly increased power of modern computers to revive the ideas of trial-and-error Debye modelling. A (usually) spherical volume with diameter D_{\max} is filled by M densely packed beads (spheres of much smaller radius r_0). Each of the beads may belong either to the particle (index = 1) or to the solvent (index = 0), and the shape is thus described by a binary string X of length M . Starting from a random distribution of 1s and 0s, the model is randomly modified using a Monte Carlo-like search to find a string X fitting the experimental data. As the search models usually contain thousands of beads, the solution must be constrained. In the simulated annealing procedure implemented in the program DAMMIN [49], an explicit penalty term $P(X)$ is added to the goal function $f(X) = \chi^2 + P(X)$ to ensure compactness and connectivity of the resulting shape. Instead of using Debye's formula, the intensity is computed with spherical harmonics to speed up the computation. Further acceleration is achieved by not recomputing the model intensity after each modification, but only updating the contribution from beads changing their index. The original bead method (program DALAI_GA [48]) using a genetic algorithm did not impose explicit constraints, although the solution was implicitly constrained by gradually decreasing r_0 during minimization, but in its later version [50] explicit connectivity conditions were also added. Monte Carlo based *ab initio* approaches also exist, which do not restrain the search space. A 'give-n-take' procedure [51] implemented in the program SAXS3D places beads on a hexagonal lattice, and, at each step, a new bead is added, removed or relocated to improve the agreement with the data. The SASMODEL program [52] does not use a fixed grid but represents the model by a superposition of interconnected ellipsoids and employs a Monte Carlo search (or, in the later implementation, a genetic algorithm [53]) of their positions and sizes to fit the experimental data. Tests on proteins with known structure demonstrated the ability of the above methods to satisfactorily restore low-resolution shapes of macromolecules from solution scattering data (for practical applications, see section 5.1).

A principal limitation of the shape determination methods, the assumption of uniform particle density, limits the resolution to 2–3 nm and also the reliability of the models, as only restricted portions of the data can be fitted. In the simulated annealing procedure [49], the beads may belong to different components so that the shape and internal structure of multi-component particles can be reconstructed. This can be done, e.g. using neutron scattering by simultaneously fitting curves recorded at different contrasts (see example of ribosome study in section 5.2). For single component particles and a single scattering curve, the procedure degenerates to *ab initio* shape determination as implemented in DAMMIN. A more versatile approach to reconstruct protein models from SAXS data has recently been proposed [54], where the protein is represented by an assembly of dummy residues (DR). The number of residues M is usually known from the protein sequence or translated DNA sequence, and the task is to find the coordinates of M DRs fitting the experimental data and building a protein-like structure. The method, implemented in the program GASBOR, starts from a randomly distributed gas of DRs in a spherical search volume of diameter D_{\max} . The DRs are randomly relocated within the search volume following a simulated annealing protocol, but the compactness criterion used in shape determination is replaced by a requirement for the model to have a 'chain-compatible' spatial arrangement of the DRs. In particular, as C_α atoms of neighbouring amino acid residues in the primary sequence are separated by ≈ 0.38 nm it is required that each DR would have two neighbours at a distance of 0.38 nm.

Compared to shape determination, DR-modelling substantially improves the resolution and reliability of models and has potential for further development. In particular, DR-type modelling is used to add missing fragments to incomplete models of proteins (program suite CREDO [55]). Inherent flexibility and conformational heterogeneity often make loops or even entire domains undetectable in crystallography or NMR. In other cases parts of the structure (loops or domains) are removed to facilitate crystallization. To add missing loops/domains, the known part of the structure—high- or low-resolution model—is fixed and the rest is built around it to obtain a best fit to the experimental scattering data from the entire particle. To complement (usually, low-resolution) models, where the location of the interface between the known and unknown parts is not available, the missing domain is represented by a free gas of DRs. For high-resolution models, where the interface is known (e.g. C- or N-terminal or a specific residue) loops or domains are represented as interconnected chains (or ensembles of residues with spring forces between the C_α atoms), which are attached at known position(s) in the available structure. The goal function containing the discrepancy between the experimental and calculated patterns and relevant penalty terms containing residue-specific information (e.g. burial of hydrophobic residues) is minimized by simulated annealing. With this approach known structures can be completed with the degree of detail justified by the experimental data and available *a priori* information.

It is clear that different random starts of Monte Carlo based methods yield multiple solutions (spatial distributions of beads or DRs) with essentially the same fit to the data. The independent models can be superimposed and averaged to analyse stability and to obtain the most probable model, which is automated in the program package DAMAVER [56]. The package employs the program SUPCOMB [57], which aligns two (low or high resolution) models represented by ensembles of points and yields a measure of dissimilarity of the two models. All pairs of independent models are aligned by SUPCOMB, and the model giving the smallest average discrepancy with the rest is taken as a reference (most probable model). All other models except outliers are aligned with the reference model, a density map of beads or DRs is computed and cut at a threshold corresponding to the excluded particle volume. The DAMAVER package can be used for models derived by any *ab initio* method, but a similar (more or less automated) average is also mentioned by other authors [50, 51, 53]. The diversity of the *ab initio* models and the results of the averaging procedure are illustrated in section 5.1.

The reliability of *ab initio* models can be further improved if additional information about the particle is available. In particular, symmetry restrictions permit to significantly speed up the computations and reduce the effective number of model parameters. In the programs SASHA, DAMMIN and GASBOR, symmetry restrictions associated with the space groups P2–P10 and P222–P62 can be imposed.

An example of application of different *ab initio* methods is presented in figure 6, which displays the reconstructed models of yeast pyruvate decarboxylase (PDC) superimposed on its atomic structure in the crystal taken from the protein data Bank (PDB) [58], entry 1pvd [59]. PDC is a large tetrameric enzyme consisting of four 60 kDa subunits, and the *ab initio* reconstructions were performed assuming a P222 point symmetry group. In the synchrotron x-ray scattering pattern in figure 6(a) [60, 61] the contribution from the internal structure dominates the scattering curve starting from $s = 2 \text{ nm}^{-1}$. The models restored by the shape determination programs SASHA and DAMMIN (figure 6(b), left and middle columns) are only able to fit the low angle portion of the experimental scattering pattern, but still provide a fair approximation of the overall appearance of the protein. The DR method (program GASBOR) neatly fits the entire scattering pattern and yields a more detailed model in figure 6(b), right column. The DR modelling brings even clearer advantages over the shape determination methods for proteins with lower MM; the example in figure 6 was selected because the envelope

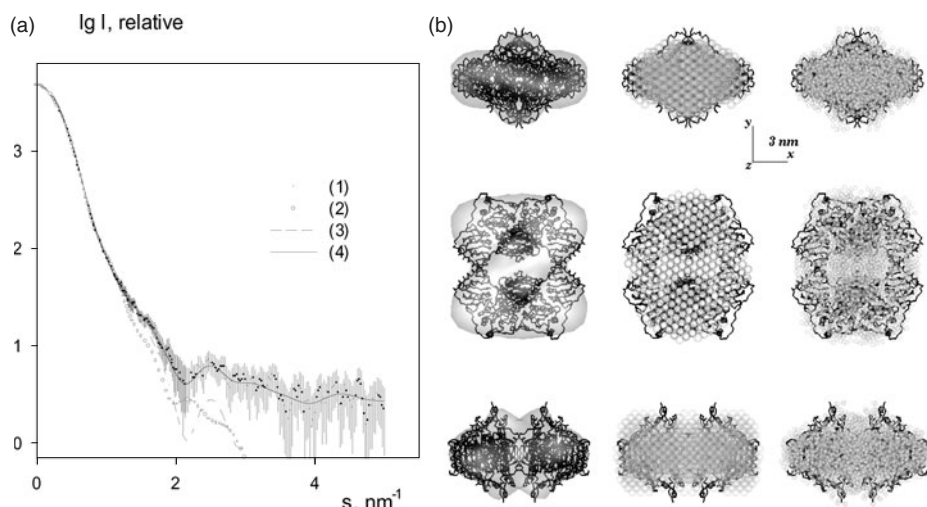


Figure 6. (a) Synchrotron x-ray scattering from PDC (1) and scattering from the *ab initio* models: (2) envelope model (SASHA); (3) bead model (DAMMIN); (4) DR model (GASBOR). (b) Atomic model of PDC [59] displayed as C_{α} chain and superimposed to the models of PDC obtained by SASHA (left column, semi-transparent envelope), DAMMIN (middle column, semi-transparent beads) and GASBOR (right column, semi-transparent DRs). The models superimposed by SUPCOMB [57] were displayed on an SGI Workstation using ASSA [78]. The middle and bottom rows are rotated counterclockwise by 90° around X and Y , respectively.

model (left column) had been constructed and published [61] before the crystal structure [59] became available.

3.5. Computation of scattering patterns from atomic models

The previous section described the situation where no information about the structure of the particle is available. If the high-resolution model of the entire macromolecule or of its individual fragments is known (e.g. from crystallography or NMR) a more detailed interpretation of SAS data is possible. A necessary prerequisite for the use of atomic models is accurate evaluation of their scattering patterns in solution, which is not a trivial task because of the influence of the solvent, more precisely of the hydration shell. In a general form, the scattering from a particle in solution is

$$I(s) = \langle |A_a(s) - \rho_s A_s(s) + \delta\rho_b A_b(s)|^2 \rangle_{\Omega}, \quad (21)$$

where $A_a(s)$ is the scattering amplitude from the particle in vacuum, $A_s(s)$ and $A_b(s)$ are, respectively, the scattering amplitudes from the excluded volume and the hydration shell, both with unit density. Equation (21) takes into account that the density of the bound solvent ρ_b may differ from that of the bulk ρ_s leading to a non-zero contrast of the hydration shell $\delta\rho_b = \rho_b - \rho_s$. Earlier methods [62–65] differently represented the particle volume inaccessible to the solvent to compute $A_s(s)$, but did not account for the hydration shell. It was pointed out in several studies [66–69] that the latter should be included to adequately describe the experimental scattering patterns. The programs CRY SOL [70] for x-rays and CRY SON [71] for neutrons surround the macromolecule by a 0.3 nm thick hydration layer with an adjustable density ρ_b . These programs utilize spherical harmonics to compute partial amplitudes $A_{lm}(s)$ for all terms in equation (21) so that the spherical averaging can be done analytically (see equation (19)). The partial amplitudes can also be used in rigid body modelling

(see next section). Given the atomic coordinates, e.g. from the PDB [58], these programs either fit the experimental scattering curve using two free parameters, the excluded volume of the particle and the contrast of the hydration layer $\delta\rho_b$, or predict the scattering pattern using the default values of these parameters.

Analysis of numerous x-ray scattering patterns from proteins with known atomic structure indicated that the hydration layer has a density of 1.05–1.20 times that of the bulk. Utilizing significantly different contrasts between the protein and the solvent for x-rays and neutrons in H₂O and D₂O it was demonstrated that the higher scattering density in the shell cannot be explained by disorder or mobility of the surface side chains in solution and that it is indeed due to a higher density of the bound solvent [71], a finding corroborated by molecular dynamics calculations [72].

3.6. Building models from subunits by rigid body refinement

Comparisons between experimental SAXS and SANS patterns and those evaluated from high-resolution structures have long been used to verify the structural similarity between macromolecules in crystals and in solution, and also to validate theoretically predicted models [62, 63, 73, 74]. Moreover, structural models of complex particles in solution can be built from high-resolution models of individual subunits by rigid body refinement against the scattering data. To illustrate this, let us consider a macromolecule consisting of two domains with known atomic structures. If one fixes domain A while translating and rotating domain B, the scattering intensity of the particle is

$$I(s, \alpha, \beta, \gamma, \mathbf{u}) = I_a(s) + I_b(s) + 4\pi^2 \sum_{l=0}^{\infty} \sum_{m=-l}^l \text{Re}[A_{lm}(s)C_{lm}^*(s)], \quad (22)$$

where $I_a(s)$ and $I_b(s)$ are the scattering intensities from domains A and B, respectively. The $A_{lm}(s)$ are partial amplitudes of the fixed domain A, and the $C_{lm}(s)$ those of domain B rotated by the Euler angles α, β, γ and translated by a vector \mathbf{u} . The structure and the scattering intensity from such a complex depend on the six positional and rotational parameters and these can be refined to fit the experimental scattering data. The algorithms [75, 76] allow to rapidly evaluate the amplitudes $C_{lm}(s)$ and thus the intensity $I(s, \alpha, \beta, \gamma, \mathbf{u})$ for arbitrary rotations and displacements of the second domain (the amplitudes from both domains in reference positions must be pre-computed using CRY SOL or CRYSON). Spherical harmonics calculations are sufficiently fast to employ an exhaustive search of positional parameters to fit the experimental scattering from the complex by minimizing the discrepancy in equation (20). Such a straightforward search may, however, yield a model that perfectly fits the data but fails to display proper intersubunit contacts. Relevant biochemical information (e.g. contacts between specific residues) can be taken into account by using an interactive search mode. Possibilities for combining interactive and automated search strategies are provided by programs ASSA for major UNIX platforms [77, 78] and MASSHA for Wintel-based machines [79], where the main three-dimensional graphics program is coupled with computational modules implementing equation (22). The subunits can be translated and rotated as rigid bodies while observing corresponding changes in the fit to the experimental data and, moreover, an automated refinement mode is available for performing an exhaustive search in the vicinity of the current configuration. Alternative approaches to rigid body modelling include the ‘automated constrained fit’ procedure [80], where thousands of possible bead models are generated in the exhaustive search for the best fit, and the ellipsoidal modelling [15, 81], where the domains are first positioned as triaxial ellipsoids following by docking of the atomic models using information from other methods, molecular dynamics and energy minimization [82].

Similarly to *ab initio* methods, information about the particle symmetry reduces the number of free parameters for rigid body modelling and speeds up the computations [46, 79]. Interestingly, rigid body modelling of the tetrameric PDC (see figure 6) in terms of movements and rotations of the crystallographic dimers [83] demonstrated that the structure in solution is somewhat more compact and that the two dimers are tilted, as one could have already expected from the *ab initio* models. Differences between the quaternary structure in the crystal and in solution, apparently caused by the crystal packing forces, are often observed for multi-subunit proteins (see also examples in section 5.2).

Further useful constraints can be provided by incorporating NMR data from partially oriented samples [84]. Analysis of the main-chain N–H residual dipolar couplings yields information on relative orientation of the secondary structure elements in the protein, which significantly reduces the rotational degrees of freedom during rigid body modelling.

3.7. Contrast variation and selective labelling of macromolecular complexes

All previous considerations in this section referred to the case of a single scattering curve (for shape determination, also measured at sufficiently high contrast, so that the intensity at low angles was dominated by the first term in equation (2)). Below we discuss the additional information, which can be obtained from a series of measurements at different solvent densities ρ_s . Clearly, all structural parameters computed from the scattering curves are functions of the contrast. In particular, recalling the expression for the forward intensity (8), the plot $[I(0)]^{1/2}$ versus ρ_s should yield a straight line [20] intercepting zero at the matching point of the particle (i.e. the point of zero contrast where the solvent density equals to the average density and the scattering is solely due to the internal structure). The sign of the square root is taken as positive for positive contrasts and negative for negative contrasts, and the slope of this plot yields the particle volume. For the radius of gyration, one can write [3]:

$$R_g^2 = R_c^2 + \frac{\alpha}{\Delta\rho} - \frac{\beta}{(\Delta\rho)^2}, \quad (23)$$

$$\alpha = \frac{1}{V} \int \Delta\rho(\mathbf{r})\mathbf{r}^2 d\mathbf{r}, \quad \beta = \frac{1}{V^2} \iint \Delta\rho(\mathbf{r}_1)\Delta\rho(\mathbf{r}_2)\mathbf{r}_1\mathbf{r}_2 d\mathbf{r}_1 d\mathbf{r}_2. \quad (24)$$

Here, R_c is the radius of gyration of the particle shape, whereas α is the second moment of the internal structure. A zero value of α corresponds to a homogeneous particle, a positive one to a particle with a higher scattering density in its outer part and a negative one to a higher scattering density closer to the centre. The non-negative parameter β describes the displacement of the centre of the scattering length density distribution with the contrast (if $\beta = 0$, the centre is not displaced). These parameters, evaluated from the plot of R_g versus $(\Delta\rho)^{-1}$ (Stuhrmann plot), provide overall information about the density distribution within the particle. Whereas it is straightforward to obtain accurate values of α this is not the case for β , which depends entirely on measurements at low contrast and low angles where parasitic scattering tends to be important.

Equations (2) and (23), (24) demonstrate that contrast variation enables one to separate information about the particle shape and internal structure. For multi-component particles it is further possible to extract information about individual components. Thus, for a system with two homogeneous components, the scattering intensity as a function of contrast can, alternatively to equation (2), be written as

$$I(s) = (\Delta\rho_1)^2 I_1(s) + 2\Delta\rho_1\Delta\rho_2 I_{12}(s) + (\Delta\rho_2)^2 I_2(s), \quad (25)$$

where $\Delta\rho_1$, $I_1(s)$, $\Delta\rho_2$, $I_2(s)$, denote the contrast and scattering from the first and second components, respectively, and $I_{12}(s)$ is the cross-term. It follows that if one measures such

a particle at the matching point of one component, the scattering data exclusively yield information about the other one. If the components are inhomogeneous, equation (25) holds, strictly speaking, only at $s = 0$, but can still be applied for the entire scattering patterns as long as the scattering density difference between the components is much larger than the density fluctuations inside the components (e.g. neutron scattering from nucleoprotein complexes). The experimental radius of gyration of a two-component system is

$$R_g^2 = w_1 R_{g1}^2 + (1 - w_1) R_{g2}^2 + w_1(1 - w_1) L^2, \quad (26)$$

where R_{g1} and R_{g2} are the radii of gyration of the components, $w_1 = \Delta\rho_1 V_1$ is the fraction of the total scattering length of the first component in the particle, and L denotes the separation between the centres of the scattering length distributions of two components. This approach was used, in particular, to estimate the separation between components in nucleoprotein complexes like ribosomes [85, 86].

Contrast variation is most often used in SANS studies, where it relies on the remarkable difference in the scattering between hydrogen and deuterium. Neutron contrast variation in H_2O/D_2O mixtures allows to reach matching points of all major components of biological macromolecules as illustrated in table 2. Moreover, specific deuteration is a very effective method for highlighting selected structural fragments in complex particles. The scattering length density of deuterated protein or nucleic acid significantly different from that of the protonated material (table 2) and contrast variation on selectively deuterated hybrid particles allows to establish positions of the labelled fragments. The classical example illustrating the power of selective deuteration is that of the selective labelling of protein pairs in the 30S ribosomal subunit *Escherichia coli* which led to a complete three-dimensional map of the positions of ribosomal proteins by triangulation [87]. Comparison of this map, which predicted the centres of mass of 21 proteins with the high-resolution structure of the small ribosomal subunit from *Thermus thermophilus* determined 15 years later [88] indicated that the positions of only five smaller proteins were significantly different from those in the crystallographic model. The discrepancy can be attributed to poor signal-to-noise ratio of the scattering data from small labels and possibly also to imperfection of the reconstitution procedure employed for the labelling. The overall agreement should in general be considered very good especially given the fact that the crystal structure corresponds to another species (containing 19 instead of 21 proteins).

In x-ray scattering, the solvent density can be changed by addition of various contrasting agents (like sucrose, glycerol or salts) [89] and the labelling can be done by isomorphous

Table 2. X-ray and neutron scattering length densities of biological components.

Component	X-rays		Neutrons		
	ρ (electrons nm ⁻³)	Matching solvent	ρ in H ₂ O (10 ¹⁰ cm ⁻²)	ρ in D ₂ O (10 ¹⁰ cm ⁻²)	Matching % D ₂ O
H ₂ O	334	—	−0.6	—	—
D ₂ O	334	—	6.4	—	—
50% sucrose	400	—	1.2	—	—
Lipids	300	—	0.3	−6.0	≈10–15%
Proteins	420	65% sucrose	1.8	3.1	≈40%
D-proteins	420	65% sucrose	6.6	8.0	—
Nucleic acids	550	—	3.7	4.8	≈ 70%
D-nucleic acids	550	—	6.6	7.7	—

For x-rays, the scattering length densities are often expressed in terms of electron density, i.e. the number of electrons per nm³; 1 electron nm⁻³ = 2.82 × 10⁸ cm.

replacement using heavy-atom labels [90] but these studies are experimentally difficult and their range of application is therefore limited. ‘Physical’ contrast variation employing anomalous SAXS on specific types of atoms [11] is also technically complicated for biological samples because of usually small anomalous signals but may become easier on the high brilliance SR sources [91].

4. Polydisperse and interacting systems

In the previous section, ideal monodisperse systems for which the measured intensity is directly related to the single particle scattering and the aim of data analysis is to obtain information about the particle structure, have been considered. In practice one often has to deal with non-ideal cases when the particles differ in size and/or shape, and/or interparticle interactions cannot be neglected. Analysis of such systems is driven by different types of questions and, accordingly, different data interpretation tools are required, which will be considered below.

4.1. Mixtures with shape and size polydispersity

Let us consider a system consisting of different types of non-interacting particles with arbitrary structures. The scattering pattern from such a mixture can be written as a linear combination

$$I(s) = \sum_{k=1}^K v_k I_k(s), \quad (27)$$

where $v_k > 0$ and $I_k(s)$ are the volume fraction and the scattering intensity from the k th type of particle (component), respectively, and K is the number of components. It is clear that, given only the experimental scattering from the mixture, one cannot reconstruct the structures of the individual components, and the amount of useful information, which can be extracted depends on the availability of independent additional information. If the number of components and their scattering patterns are known *a priori*, one can determine the volume fractions in linear combination (27) simply by non-negative linear least squares [92] minimizing the discrepancy in equation (20). This approach is useful to characterize well-defined systems like oligomeric equilibrium mixtures of proteins (see examples in section 5.3).

If the number of components and their scattering patterns are unknown but a series of experiments has been performed for samples with different volume fractions v_k , useful information about the system can still be obtained from singular value decomposition (SVD) [93]. The matrix $A = [A_{ik}] = [I^{(k)}(s_i)]$, ($i = 1, \dots, N$, $k = 1, \dots, K$ where N is the number of experimental points) is represented as $A = U * S * V^T$, where the matrix S is diagonal, and the columns of the orthogonal matrices U and V are the eigenvectors of the matrices $A * A^T$ and $A^T * A$, respectively. The matrix U yields a set of so-called left singular vectors, i.e. orthonormal base curves $U^{(k)}(s_i)$, that spans the range of matrix A , whereas the diagonal of S contains their associated singular values in descending order (the larger the singular value, the more significant the vector). Physically, the number of significant singular vectors (non-random curves with significant singular values) yields the minimum number of independent curves required to represent the entire data set by their linear combinations. Non-random curves can be identified by a non-parametric test due to Wald and Wolfowitz [94] and the number of significant singular vectors provides an estimate of the minimum number of independent components in equilibrium or non-equilibrium mixtures. SVD, initially introduced in the SAXS analysis in the early 1980s [95], has become popular in the analysis of titration and time-resolved experiments [96–98]. One should keep in mind that SVD imposes only a lower limit, and the actual number of components (e.g. the number of intermediates in (un)folding or

assembly of proteins) may of course be larger. Programs for the linear least squares analysis of mixtures and SVD are publicly available (e.g. [99]).

Another type of mixtures results from systems with size polydispersity, where particles have similar shapes and differ only in size. Such systems are conveniently described by the volume distribution function $D(R) = N(R)V(R)$, where $N(R)$ is the number of particles with characteristic size R and $V(R)$ is the volume of the particles of this size. The scattering intensity is given by the integral

$$I(s) = (\Delta\rho)^2 \int_{R_{\min}}^{R_{\max}} D(R)V(R)i_0(sR) dR, \quad (28)$$

where $i_0(sR)$ is the normalized scattering intensity of the particle ($i_0(0) = 1$), and R_{\min} and R_{\max} are the minimum and maximum particle sizes, respectively. Protein or nucleic acid solutions rarely display the kind of size polydispersity described by equation (28) but this equation is often applicable to micelles, microemulsions, block copolymers or metal nanoparticles. In most practical cases one assumes that the particle form factor is known (in particular, for isotropic systems, the particles can usually be considered spherical) and equation (28) is employed to obtain the volume distribution function $D(R)$. This can be done with the help of the indirect transformation method described in section 3.2 (the function $D(R)$ is expanded into orthogonal functions as in equation (13) on the interval $[R_{\min}, R_{\max}]$). The structural parameters of polydisperse systems do not correspond to a single particle but are obtained by averaging over the ensemble. Thus, for a polydisperse system of solid spheres $R_g = (3\langle R^2 \rangle_z / 5)^{1/2}$, where the average sphere radius is expressed as

$$\langle R^2 \rangle_z = \int_{R_{\min}}^{R_{\max}} R^5 D(R) dR \left[\int_{R_{\min}}^{R_{\max}} R^3 D(R) dR \right]^{-1}. \quad (29)$$

4.2. Interacting systems and structure factor

Interactions between macromolecules in solution may be specific or non-specific [100] and they involve the macromolecular solute and co-solutes (salts, small molecules, polymers), the solvent and, where applicable, co-solvents. Specific interactions usually lead to the formation of complexes involving cooperative interactions between complementary surfaces. This case is effectively considered in the previous section dealing with mixtures and equilibria. In contrast non-specific interactions can usually be described by a generic potential such as the DLVO potential [101] initially proposed for colloidal interactions. This potential takes into account the mutual impenetrability of the macromolecules, the screened electrostatic repulsions between charges at the surfaces of the macromolecules and the longer-ranged Van der Waals interactions. Non-specific interactions essentially determine the behaviour at larger distances whereas in the case of attractive interactions leading to, e.g. crystallization specific interactions dominate at short range.

In general the spherically averaged scattering from a volume of a solution of anisotropic objects like macromolecules that is coherently illuminated is given by:

$$I(s, t) = \left\langle \sum_{i=1}^N \sum_{j=1}^N A_i(\mathbf{s}, \mathbf{u}_i, \mathbf{v}_i, \mathbf{w}_i) \cdot A_j(\mathbf{s}, \mathbf{u}_j, \mathbf{v}_j, \mathbf{w}_j) \exp(i\mathbf{s} \cdot \mathbf{r}_{ij}(t)) \right\rangle_{\Omega}, \quad (30)$$

where A is the scattering amplitude of the individual particles computed as in equation (1) and $\mathbf{u}, \mathbf{v}, \mathbf{w}$ are unit vectors giving their orientation relative to the reference coordinate system in which the momentum transfer vector \mathbf{s} is defined. If the particles can be described as spheres on the scale of their average separation, the general expression in equation (30) simplifies to

the product of the square of the form factor of the isolated particles and of the structure factor of the solution which reflects their spatial distribution. This is valid for globular proteins and weak or moderate interactions in a limited s -range [102, 103]. The structure factor can then readily be obtained from the ratio of the experimental intensity at a concentration c to that obtained by extrapolation to infinite dilution or measured at a sufficiently low concentration c_0 where all correlations between particles have vanished:

$$S(s, c) = \frac{c_0 I_{\text{exp}}(s, c)}{c I(s, c_0)}. \quad (31)$$

Interparticle interactions thus result in a modulation of the scattering pattern of isolated particles by the structure factor which reflects their distribution and to a much lesser extent their relative orientation in solution.

If separation of the structure factor and the form factor using equation (31) is straightforward in the case of monodisperse solutions and repulsive interactions this is no longer the case when the interactions are attractive and the polydispersity of the solution depends on its concentration. For spherical particles the generalized indirect Fourier transformation (GIFT) has been proposed, which is a generalization of the indirect transformation technique described in section 3.2. The structure factor is also parametrized similarly to the characteristic function and non-linear data fitting is employed to find both the distribution function and the structure factor. For non-spherical (e.g. rod-like) particles the method yields an effective structure factor [104, 105].

The interaction of rod-like molecules has been studied in detail [106] and a pair potential of the form $V(r, \mathbf{u}_1, \mathbf{u}_2)$ can be used to describe the interactions between molecules where r is the distance between the centre of mass and \mathbf{u}_1 and \mathbf{u}_2 denote the orientation of their axis. Unfortunately, for filaments SAXS usually only yields cross-section information and an effective structure factor must be used.

For thin rods like DNA at low ionic strength, the length distribution has little influence on the effective structure factor [107]. In the dilute regime the position of its first maximum, determined by the centre to centre separation between rigid segments, varies like the square root of the concentration. The length distribution has, however, a strong influence on the relaxation times observed in electric field scattering [107, 108] and on the slow mode observed in dynamic light scattering [109].

For mixtures of different types of particles with possible polydispersity and interactions between particles of the same component, the scattering intensity from a component entering equation (27) can be represented as

$$I_k(s) = S_k(s) \cdot \int_0^\infty D_k(R) \cdot V_k(R) \cdot [\Delta\rho_k(R)]^2 \cdot i_{0k}(s, R) dR, \quad (32)$$

where $\Delta\rho_k(R)$, $V_k(R)$ and $i_{0k}(s, R)$ denote the contrast, volume and form factor of the particle with size R (these functions are defined by the shape and internal structure of the particles, and $i_{0k}(0, R) = 1$), whereas $S_k(s)$ is the structure factor describing the interference effects for the k th component. It is clear that quantitative analysis of such systems is only possible if assumptions are made about form and structure factors and about the size distributions. A parametric approach was proposed [110] to characterize mixtures of particles with simple geometrical shapes (spheres, cylinders, dumbbells). Each component is described by its volume fraction, form factor, contrast, polydispersity and, for spherical particles, potential for interparticle interactions. The functions $D_k(R)$ are represented by two-parametric monomodal distributions characterized by the average dimension R_{0k} and dispersion ΔR_k . The structure factor for spherical particles $S_k(s)$ is represented in the Percus–Yevick approximation using the sticky hard sphere potential [111] described by two parameters, hard sphere interaction

radius R_k^{hs} and ‘stickiness’ τ_k . The approach was developed in the study of AOT water-in-oil microemulsions and applied for quantitative description of the droplet to cylinder transition in these classical microemulsion systems [110]. A general program MIXTURE based on this method is now publicly available [99].

4.3. Computation of the structure factor from interaction potentials

The above methods were aiming at experimental determination of the structure factor, but in many practically important cases the latter can at least be approximately modelled based on the thermodynamic and physico-chemical parameters of the system. The relationship between the value of the static structure factor of a monodisperse solution at the origin to its osmotic compressibility or to the osmotic pressure Π is given by:

$$S(c, 0) = \left(\frac{RT}{M} \right) \left(\frac{\partial \Pi}{\partial c} \right)^{-1}, \quad (33)$$

where R is the gas constant and M the molecular mass of the solute (here we do not consider the dynamic (time dependent) structure factor, which would result in speckle (see section 5.7)). For weakly interacting molecules at sufficiently low concentrations the osmotic pressure can be linearly approximated by series expansion which yields the second virial coefficient A_2 :

$$\frac{\Pi}{cRT} = \frac{1}{M} + A_2c + O(c^2) \quad (34)$$

and

$$[S(c, 0)]^{-1} = 1 + 2MA_2c. \quad (35)$$

Compared to the equivalent ideal solution for which $S(c, 0) = 1$ the osmotic pressure is higher ($A_2 > 0$) when the interactions are repulsive and the particles evenly distributed and lower ($A_2 < 0$) when attractive interactions lead to large fluctuations in the particle distribution.

With appropriate modelling based on methods developed for the liquid state [112, 113] the s -dependence of the structure factor yields more information. In recent studies [114] the different interactions in the potential are each represented by a Yukawa potential defined by a hard-sphere diameter σ , a depth J and a range d (k_B is Boltzmann’s constant)

$$\frac{u(r)}{k_B T} = J \left(\frac{s}{r} \right) \exp \left(-\frac{r-s}{d} \right). \quad (36)$$

These parameters are determined by trial-and-error calculating the structure factor for various combinations of values.

The structure factor can be calculated from the number density of particles in the solution (ρ) and the pair distribution function of the macromolecules at equilibrium $g(r)$ obtained on the basis of the Ornstein–Zernicke (OZ) and the hypernetted chain (HNC) integral equations

$$S(c, s) = 1 + \rho \int_0^\infty 4\pi^2 (g(r) - 1) \frac{\sin sr}{sr} dr. \quad (37)$$

The OZ relationship between Fourier transforms of the total and direct correlation functions $h(r) = g(r) - 1$ and $c(r)$, which can be solved iteratively is given by:

$$\{1 + \mathfrak{F}[h(r)]\} \{1 - \mathfrak{F}[c(r)]\} = 1 \quad (38)$$

and for an interaction potential $u(r)$ the HNC equation is:

$$g(r) = \exp \left[-\frac{u(r)}{k_B T} + h(r) - c(r) \right]. \quad (39)$$

For small proteins, like lysozyme at high ionic strength or γ -crystallins near the isoelectric point, two cases where the Coulomb repulsions can be neglected, it was shown that the interactions were satisfactorily described by a purely attractive Yukawa potential, corresponding to the Van der Waals interactions. The liquid–liquid phase separation, which occurs at low temperatures, as well as the structure factor of the solutions could be explained using values for the range (d) and depth (J) close to 0.3 nm and -2.7 kT, respectively, and a value of σ corresponding to the dry volume of the protein [115], in agreement with simulations [116].

5. Selected applications

Modern SAXS and SANS are characterized by an increasing proportion of publications utilizing advanced data analysis methods. Below, we shall review some recent applications of the methods to the study of biological macromolecules in solution, and also consider potential novel applications of SAXS utilizing the high brilliance and coherence of new and forthcoming x-ray sources.

5.1. Analysis of macromolecular shapes

During the last few years, *ab initio* methods have become one of the major tools for SAS data analysis in terms of three-dimensional models. Several programs are publicly available on the Web and the users may test them before applying them to their specific problems. The performance of *ab initio* shape determination programs DALI-GA, DAMMIN, and SAXS3D was compared in two recent papers. In one of them [117], all three methods allowed the authors to reliably reconstruct the dumbbell-like shape of troponin-C (its high-resolution structure in solution has been determined earlier by NMR) from the experimental data. In another paper [118], the methods were tested on synthetic model bodies and yielded similar results in the absence of symmetry. Optional symmetry and anisometry restrictions, absent in other programs, lead to a better performance of DAMMIN on symmetric models. Although all of these methods had been extensively tested by the authors in the original papers, these independent comparative tests are very important.

The validity of models generated *ab initio* from solution scattering data can also be assessed by *a posteriori* comparison with high-resolution crystallographic models that became available later. In all of the few known cases there is good agreement between the *ab initio* models and the later crystal structures. One example (tetrameric yeast PDC) was presented in section 3.4 (figure 6), another is the study of dimeric macrophage infectivity potentiator (MIP) from *Legionella pneumophila* (the *ab initio* model was published in 1995 [119] and the crystal structure reported six years later [120]. For the 50 kDa functional unit of *Rapana venosa* haemocyanin the low-resolution model was published in 2000 [121], whereas the crystal structure of the homologous *Octopus* haemocyanin unit, although reported in 1998 [122], only became available in 2001 under PDB accession number 1js8.

Practical applications of *ab initio* shape determination range from individual macromolecules to large macromolecular complexes. In the study of MutS protein (MM about 90 kDa), a component of the mismatch–repair system correcting for mismatched DNA base-pairs [123], low-resolution models were built for the nucleotide-free, adenosine di-phosphate (ADP)- and adenosine tri-phosphate (ATP)-bound protein. ATP provides energy to all cellular processes through hydrolysis to ADP and inorganic phosphate (P_i) ($\text{ATP} + \text{H}_2\text{O} \rightarrow \text{ADP} + P_i$) yielding about 12 kcal mol^{-1} . The hollow *ab initio* models displayed remarkably good agreement with the crystal structure of MutS complexed with

DNA, but also revealed substantial conformational changes triggered by both binding and hydrolysis of ATP. In another study related to ATP hydrolysis [124] the structure of eucaryotic chaperonin TRiC (MM \sim 960 kDa) was analysed. Chaperonins are large complexes promoting protein folding inside their central cavity. Comparison of *ab initio* models of TRiC in different nucleotide- and substrate-bound states with available crystallographic and cryo-EM models and with direct biochemical assays suggested that ATP binding is not sufficient to close the folding chamber of TRiC, but the transition state of ATP hydrolysis is required. Further studies of structural transitions include conformational changes of calpain from human erythrocytes triggered by Ca^{2+} binding [125], dramatic loosening of the structure of human ceruloplasmin upon copper removal [126], effect of the phosphorylation on the structure of the FixJ response regulator [127], major structural changes in the *Manduca sexta* midgut V1 ATPase due to redox-modulation [128]. In the latter study, a three-fold symmetry was used for *ab initio* reconstruction to enhance the resolution. Symmetry restrictions were also successfully applied to study DNA- and ligand-binding domains of nuclear receptors, proteins regulating transcription of target genes [129], yielding U-shaped dimeric and X-shaped tetrameric molecules. Available crystallographic models of monomeric species (MM of the monomer about 37 kDa) positioned inside the *ab initio* models suggested a possible explanation for the higher affinity of dimers in target gene recognition. In another study of nuclear receptors [130], *ab initio* methods are combined with rigid body modelling to study various oligomeric species revealing the conformational changes induced by ligand binding.

SAXS is often used in combination with methods like circular dichroism, structure prediction to assess the secondary structure and analytical ultracentrifugation to further validate size and anisometry. *Ab initio* reconstructions yielded elongated shape models of extracellular domain of the human amyloid precursor protein (MM = 69 kDa), linked to the genesis of Alzheimer's disease [131] and of haemoglobin protease (MM = 110 kDa), a principal component of the iron acquisition system in a pathogenic *E. coli* strain [132]. A variety of biophysical and bioinformatics-based methods was incorporated in the low-resolution structural study of tetrameric ES-62, a 231 kDa multi-functional properties protein secreted by filarial nematodes [133]. A tri-lobed model was generated for the C-propeptide trimer (MM \approx 90 kDa) from human procollagen III, directing chain association during intracellular assembly of procollagen molecules [134], and this study was complemented by the elongated shape model of the procollagen C-proteinase enhancer (MM \approx 50 kDa) [135]. In the latter study, predicted models of the three domains of the C-proteinase enhancer were positioned inside the *ab initio* shape and the missing loops were added following the method described above [55]. *Ab initio* models of tropomodulin, a 39 kDa capping protein of the actin-tropomyosin filament, and of its C-terminal fragment [136] were constructed to propose a model of tropomodulin association with tropomyosin and actin. In the study of hydrophobins, highly surface-active proteins specific to filamentous fungi, scattering data from dissolved proteins were combined with x-ray diffraction from crystallites [137] and from Langmuir–Blodgett films [138], to derive consistent low-resolution models.

Shape determination methods are also applicable to nucleic acids. The *ab initio* model of free *Thermus flavus* 5S ribosomal RNA in solution [139] displayed an elongated molecule with a compact central region and two projecting arms. *A posteriori* comparison with the 5S RNA inside the ribosome [140] indicated that it becomes essentially more compact upon complexation with ribosomal proteins. Recently, *ab initio* analysis was also applied to study gels of native sulfated polysaccharides (carrageenans). In complex with an oppositely charged surfactant, the gel collapses and the carrageenan/surfactant bilayers form well-defined clusters [141]. Tentative *ab initio* models of these clusters display hollow nanostructures consisting of bent worm-like bilayer substructures.

The DR-modelling [54] fits the experimental data to higher resolution and thus allows one to build yet more detailed and reliable *ab initio* models than the shape determination methods. We shall illustrate the potential of the DR approach by the study of complexes between titin fragments and telethonin [142]. The giant multi-domain muscle protein titin (MM up to 4 MDa) acts as a molecular ruler within the muscle sarcomere. Its N-terminus is located within the so-called Z-disc, which functions as a terminal anchor for a number of sarcomere filament systems. Telethonin, a small 18 kDa protein, interacts specifically with the two Z-disc domains of titin (Z1Z2, MM = 22 kDa; the structure of the two domains was predicted by [143]). The molecular basis of the interaction between the N-terminus of titin and telethonin is a key for understanding the anchoring mechanism of titin. Samples of purified Z1Z2 and its complexes with telethonin were measured, and were reconstructed *ab initio* from the scattering patterns in figure 7(a). The typical variation of independent DR reconstructions of Z1Z2 is displayed in figure 7(b). Although the local arrangement of DRs differs from one model to the other, they all display the same overall appearance with a well-defined two-domain structure, even more pronounced in the average model (figure 7(c), cyan). Comparison with an independently obtained model of Z1Z2 construct containing a seven residues long poly-histidine tag (his-Z1Z2) permits one to localize the tag at the tip of the Z1 domain (figure 7(c)). To our knowledge, this is the first time that such a small fragment is located using solution scattering. The *ab initio* shapes of the complex between the telethonin construct lacking the C-terminus and Z1Z2 (TE(90)-Z1Z2) reconstructed by DAMMIN and GASBOR indicate a 2 : 1 stoichiometry with antiparallel association of two Z1Z2 molecules and telethonin acting as a central linker (figure 7(d)). The complex of full-length telethonin with Z1Z2 (TE(167)-Z1Z2) appears to also have a 1 : 2 stoichiometry at concentrations below 1 mg ml⁻¹, but dimerizes at higher concentrations. These results suggest a cross-linking function for telethonin, which connects two titin molecules at their N-termini leading to a telethonin-mediated auto-anchoring of titin dimers in the Z-disc.

Other recent applications of the DR method include the study of the N-terminal extension of rusticyanin [144], where the *ab initio* models of monomeric wild type protein (MM \approx 17 kDa) and of the hexameric mutant lacking 35 residues at the N-terminus were constructed (this last model assuming P3 symmetry). The obtained molecular shapes can be reconciled with the crystallographic structure of the monomer indicating that the N-terminus of rusticyanin is not responsible for its acid stability, as previously believed. The solution conformation of cellulase Cel45 from *Humicola insolens*, a protein of MM \approx 36 kDa containing a catalytic and a cellulose-binding domain separated by a 36 residue long glycosylated linker peptide, was also analysed [145]. *Ab initio* models of several constructs with different linker length and flexibility were generated and mutations leading to higher rigidity of the linker were introduced. In the study of a 105 kDa glycoprotein β -mannosidase from *T. reesei*, the model obtained using GASBOR was further enhanced using available low-resolution crystallographic data phased by molecular replacement [146]. Another example of combined use of SAXS and crystallography data was presented in the study of the *Pseudomonas aeruginosa* TolA protein [147]. The crystal structure of the C-terminal domain III (MM \approx 15 kDa) was determined and SAXS was employed to further establish the overall shape of the entire periplasmic portion of the protein including the partially unfolded domain II (MM \approx 40 kDa).

In some publications, several *ab initio* methods are used to ensure reproducibility of the results. The models of β_2 -Glycoprotein, a phospholipid-binding plasma protein (MM = 36 kDa and 19% (w/w) carbohydrates) consisting of four domains were generated by DAMMIN, DALAI-GA and GASBOR consistently yielding elongated S-like shapes with side arms [148]. These models suggest re-orientation of the middle flexible domains compared to the structure of the protein in the crystal. A more complicated case was reported in the

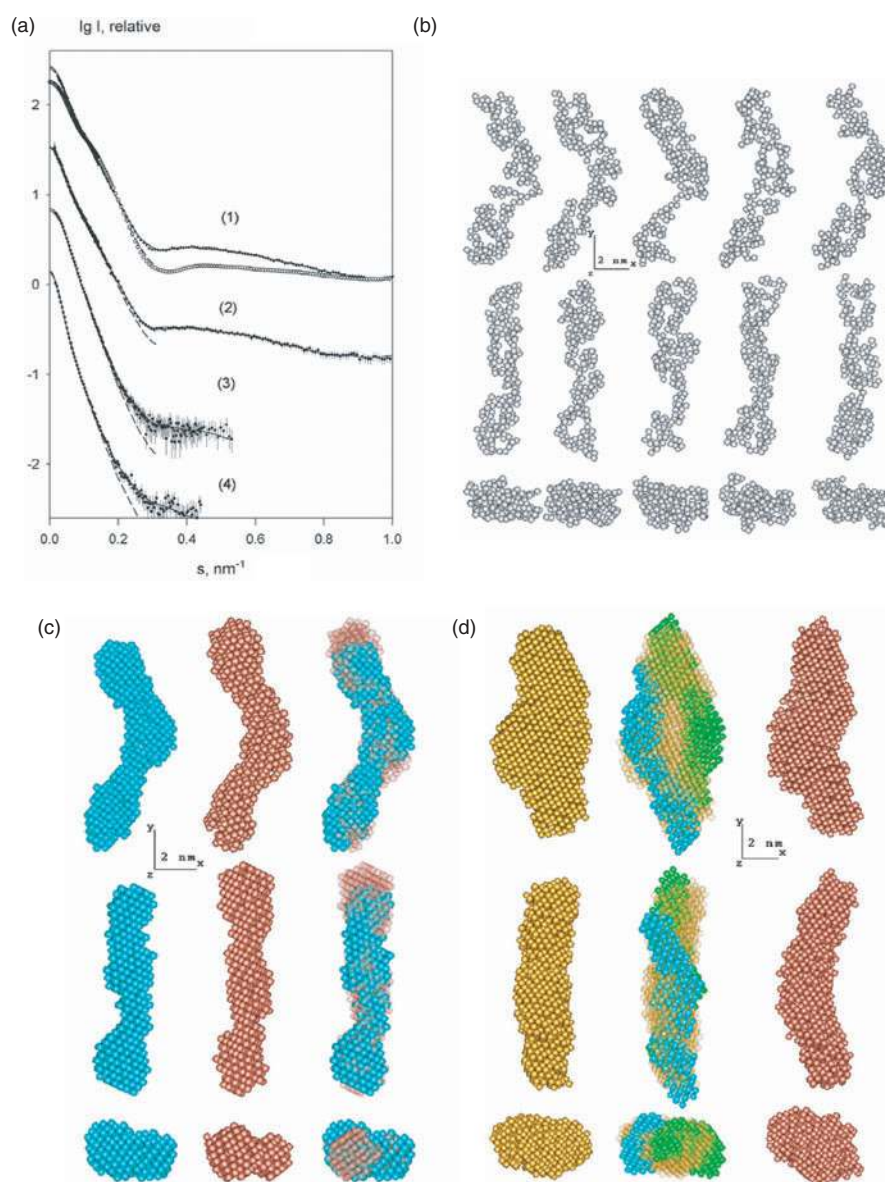


Figure 7. *Ab initio* low-resolution models of Z1Z2 and its complexes with telethonin. (a) X-ray scattering patterns from Z1Z2 (1) His-Z1Z2 (2), TE(90)-Z1Z2 (3) and TE(167)-Z1Z2 (4). The experimental data are displayed as dots with error bars, DAMMIN fits as dashed lines, GASBOR fits as full lines. The scattering from the homology model of Z1Z2 [143] is displayed as open circles. The scattering patterns are displaced by one logarithmic unit for better visualization. (b) DR models of Z1Z2 obtained *ab initio* in five independent GASBOR runs (from left to right). (c) Averaged DR models of Z1Z2 (cyan beads) and His-Z1Z2 (brown beads), and their overlap (the extra seven residues due to the His-tag correspond to the extra volume on the top of the molecule in upper and middle rows). (d) The low-resolution shape of TE(90)-Z1Z2 obtained by averaging 12 DAMMIN models (yellow beads, left panel) and this model as semi-transparent beads superimposed with two antiparallel DAMMIN models of Z1Z2 (cyan and green beads, middle panel). The right panel displays the model of TE(90)-Z1Z2 (brown beads) obtained by averaging 12 GASBOR models. In all panels, the middle and bottom rows are rotated counter clockwise by 90° around the Y- and X-axis, respectively.

study of α -crustacyanin, the blue carotenoprotein of lobster carapace [149]. Based on the scattering data, it was established that the protein contains eight heterodimeric subunits of β -crustacyanin (MM = 42 kDa), the crystallographic structure of which is available [150]. *Ab initio* reconstruction methods yielded elongated models compatible with a zigzag or helical arrangement of the dimeric β -crustacyanins, but their exact positions could not be established unequivocally. It is conceivable that rigid body modelling and/or symmetry restrictions could have provided more definitive conclusions about the quaternary structure in this case.

5.2. Quaternary structure of complex particles

Rigid body modelling is the most popular approach in the analysis of the structure of complexes. Here, comparisons between the structures in the crystal and in solution continue to be important. Striking differences between crystal and solution have been found even for a textbook illustration of allosteric enzymes, *E. coli* aspartate transcarbamylase (ATCase), a dodecameric assembly with MM = 306 kDa [151] (see more on ATCase in section 5.6). More recently, a systematic study of differences between the quaternary crystal and solution structures of five thiamine diphosphate-dependent enzymes was performed [83]. For all enzymes except the very compact tetrameric PDC from *Z. mobilis*, differences were observed between the experimental profiles and those calculated from the available crystal structures. For tetrameric pyruvate oxidase from *L. plantarum* and dimeric transketolase from *S. cerevisiae*, which have tight intersubunit contacts in the crystal, relatively small rigid body modifications of the quaternary structure were sufficient to fit the experimental data. For the enzymes with looser contacts (the native and activated forms of yeast PDC, both tetrameric), much larger modifications of the crystallographic models were required. The magnitude of the distortions induced by the crystal environment was thus correlated with the interfacial area between subunits. In general, quaternary structure formation of multi-subunit proteins involves only low-energy non-covalent interactions, and the crystal packing forces, which also originate from non-covalent bonds between neighbouring molecules, may easily distort these subtle architectures, especially in the case of loose intersubunit contacts.

In an increasing number of publications, new crystal structures are validated against solution scattering data. A 0.34 nm resolution crystal structure of a 'prokaryotic proteasome' HslUV from *H. influenzae* composed of the HslV protease and the HslU ATPase (MM about 750 kDa) provides a good fit to the experimental SAXS data [152]. The fit can further be improved by adding about 50 residues missing in the crystallographic model of HslUV. It is interesting that the scattering computed from the earlier atomic model of *E. coli* HslUV in the crystal (showing different arrangement of subunits) fails to fit the SAXS pattern. The crystal structure of phosphoenolpyruvate carboxykinase from *Trypanosoma cruzi* at 0.2 nm resolution has been reported [153] and solution scattering demonstrates that the enzyme is dimeric in solution (total MM \approx 115 kDa). The best fitting dimer was selected among the crystallographic contacts, and rigid body refinement was employed to find the conformation of the enzyme in solution. A similar approach was used to establish the quaternary structure of dimeric *Thermotoga maritima* α -Glucosidase AglA [154], and to build a coiled-coil superhelical model of the dimerization domain of *E. coli* ATP synthase *b* subunit [155].

An example of validation of a theoretically predicted model is given by the study of the dimeric α -crystallin domain of α B-crystallin [156]. α B-crystallin, a member of the small heat-shock protein (sHSP) family is a major eye lens protein and can act as a molecular chaperone. A deletion mutant from the human α B-crystallin (α B57-157) is a dimeric protein that comprises the α -crystallin domain of the α B-crystallin and retains some chaperone-like activity. The high-resolution crystallographic model of homologous sHSP from *M. janaschii* (MjHSP16.5)

[157] was employed [158] as a template to build a model of dimeric α B-crystallin domain. The dimeric interface of the homology model is virtually identical to that of MjHSP16.5 (figure 8(a), left panel) and does not display contacts between residues 114–118 in the two monomers reported by spin-labelling [159]. The scattering computed from the homology model significantly deviates from the experimental scattering by the α B-crystallin (figure 8(b), curve (1)). A new dimerization interface has been proposed and rigid body modelling was employed to refine the angle between the monomers (figure 8(b), curves (2)–(6)). The final model (figure 8(a), right panel) neatly fits the SAXS data, accounts for the spin-labelling results and suggests that the α B-crystallin is composed of flexible building units with an extended surface area which may be important for its chaperone activity.

Applications of the ‘automated constrained fit’ procedure include the study of monomeric factor H of human complement, a protein comprising 20 short consensus/complement repeat (SCR) domains (each having MM about 7 kDa), where a folded-back model was generated [160]. In the study of rat complement receptor-related protein (rCrry) containing five SCR domains, a family of extended rCrry structures was established yielding best agreement with scattering and ultracentrifugation data [161]. Further, a mouse Crry (mCrry) construct was analysed containing two rCrry-like complexes attached to immunoglobulin G1 (IgG1) molecule. The two SCR antennae were found to extend from the Fc fragment of the IgG1, but no preferred orientation could be identified suggesting that the accessibility of the antennae for their molecular targets was not affected by the covalent link to IgG1.

For protein complexes, additional information can be obtained with neutron scattering using selective deuteration of subunits. As the scattering length density of perdeuterated protein differs significantly from that of the native one, measurements at different D₂O concentrations allow to separate information about the structure of subunits and of their relative positions (cf equations (25) and (26)).

Interactions of procaryotic molecular chaperone GroEL, a hollow protein complex with 14 identical subunits of MM 57 kDa each, with perdeuterated subtilisin were studied [162] to determine the position and shape of the latter inside the chaperonin. A series of SAXS and SANS studies of the cAMP-dependent protein kinase (PKA) is reviewed in [15]. PKA was the first discovered kinase and serves as a prototype for understanding kinase structure–function relationship. Analysis of scattering from samples with selectively deuterated catalytic and regulatory subunits (MM \approx 39 kDa and 44 kDa, respectively) ultimately led to the structural model of the catalytic subunit–regulatory subunit dimeric complex built of homology models of the subunits. The latter were docked constrained by the scattering data, mutagenesis data, and the side chain packing at the heterodimer interface was further refined using molecular dynamics and energy minimization [82]. A similar strategy was used by this group in a SAXS/SANS analysis of the Ca²⁺ mediated interactions between calmodulin and myosin light chain kinase [81, 163].

Selective deuteration is also extremely useful in contrast variation studies of multi-component, e.g. nucleoprotein complexes. An example is given by the study of the 70S *E. coli* ribosome, the first practical application of the multi-phase simulated annealing bead modelling [49]. Ribosomes are supramolecular complexes (MM \sim 2.5 MDa) responsible for protein synthesis in all organisms, and each of the two unequal ribosomal subunits is a complex assembly of proteins and nucleic acids. A total of 42 x-ray and neutron solution scattering curves from reconstituted ribosomes were collected, where the proteins and rRNA moieties in the subunits were either protonated or deuterated in all possible combinations. The search volume defined by a cryo-EM model of the ribosome [164] is divided into almost 8000 densely packed 0.5 nm radius spheres. Each sphere is assigned either to solvent, to protein or to ribosomal RNA (rRNA) moieties to simultaneously fit all the scattering curves. The resulting

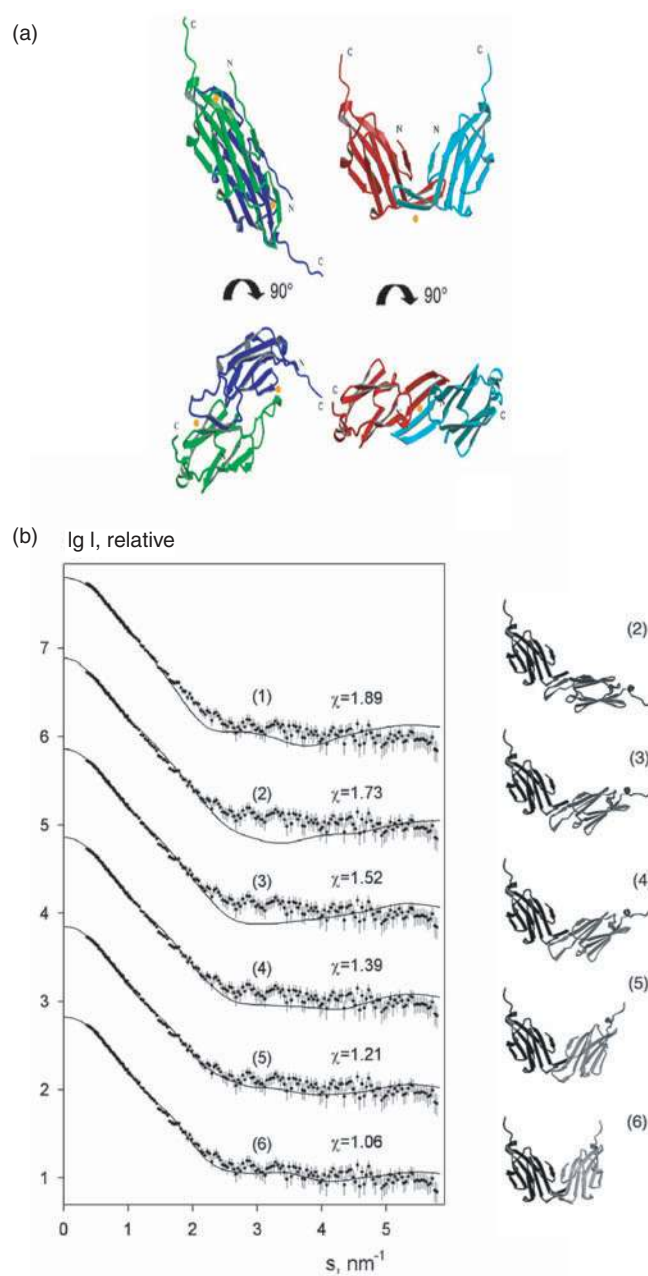


Figure 8. (a) Crystallographic model of the MjHSP16.5 dimer (left panel) and the model of the dimeric α -crystallin domain obtained by rigid body refinement (right panel) in two perpendicular views. The monomers coloured green (left) and red (right) are in the same orientation. The residues, which should be in contact according to spin-labelling data are indicated by orange spheres. (b) Experimental scattering from α -crystallin domain and the fits calculated from the atomic models. (1) the crystallographic dimer of MjHSP16.5; (2)–(6) scattering from the dimeric homology models with increasing compactness in displayed on the right. The curves (1)–(6) are displaced down by one logarithmic unit for clarity and the discrepancies with the experimental data are presented.

3 nm resolution model represents the volumes occupied by the rRNA and protein moieties in the entire ribosome [165]. The predicted protein–rRNA map is in remarkably good agreement with the later high-resolution crystallographic models of the ribosomal subunits from other species [88, 140, 166]. It is interesting that the map obtained from solution scattering reveals peripheral proteins in the large ribosomal subunit (so-called L1 and L7/L12), which cannot be seen in the crystal structure of the 50S subunit from *Haloarcula marismortui* [140], apparently because of their flexibility, but they are revealed in the crystallographic model of the complete 70S ribosome from *T. thermophilus* [167]. The multi-phase modelling approach of contrast variation data developed for the ribosome should be useful in future studies of macromolecular complexes.

5.3. Equilibrium systems and oligomeric mixtures

SAS is one of the most useful techniques to quantitatively characterize mixtures of different macromolecules or of different conformations/aggregation states of the same macromolecule. Time-resolved studies of the evolution of non-equilibrium systems will be considered in the next section. For equilibrium macromolecular systems, a typical task is to analyse the oligomeric compositions in protein mixtures. Specific oligomerization of proteins can sometimes be observed at the first stages of crystallization but often also has biochemical implications as some enzymes are catalytically active only as oligomers.

When high- or low-resolution models of the oligomers are available, their scattering curves can be computed and the volume fractions determined using equation (27). A pH-dependent dimer–tetramer equilibrium of yeast PDC was established [60] based on low-resolution models. The enzyme was found to be predominantly tetrameric but to dissociate into dimers at alkaline pH. Comparison with the results of biochemical methods and of NMR signals from the free substrate clearly indicated that tetramers are the catalytically active species. The effectors, thiamin diphosphate and phosphate, shifted the equilibrium towards tetramers at higher pH values without altering the structure of the oligomers [168]. Further analysis using also structure of the enzyme in the crystal [59] confirmed the results obtained using low-resolution models and also allowed comparison of the oligomeric behaviour of PDC from different organisms [169].

Concentration dependent monomer–dimer equilibrium was observed for conventional kinesin from *Drosophila melanogaster* [170]. Kinesins belong to the superfamily of microtubule-associated motor proteins and dimerize via a coiled-coil tail domain. The high-resolution model of the functional dimer [171] displays an asymmetric structure (MM = 80 kDa) where the two heads of the dimer are related by a rotation of about 120° with respect to the coiled-coil axis. The constructs with the tail domain removed remained monomeric in solution; those with a long tail (44 or 53 residues) were always dimeric. The construct with an intermediate tail length (28 residues) is monomeric at a concentration of 1 mg ml⁻¹, but the volume fraction of tetramers gradually increases with concentration, to 100% at 10 mg ml⁻¹.

A similar approach can also be employed to study stoichiometry of protein complexes. In the study of ferredoxin-dependent glutamate synthase (GltS) [172], the crystal structures of GltS (MM = 162 kDa) and its physiological electron donor ferredoxin (Fd, MM = 10 kDa) are reported and SAXS patterns from solutions of free GltS, free Fd, and their 1 : 1 and 1 : 2 molar mixtures are analysed. The scattering from the 1 : 1 mixture can be well fitted by the scattering from a tentative model of a complex of Fd and GltS. In contrast, the scattering from the 1 : 2 mixture is neatly fitted by a linear combination of the scattering from Fd and from the 1 : 1 complex. This clearly demonstrated that GltS formed a (1 : 1) complex with Fd, and has implications for the understanding of GltS-mediated catalytic process.

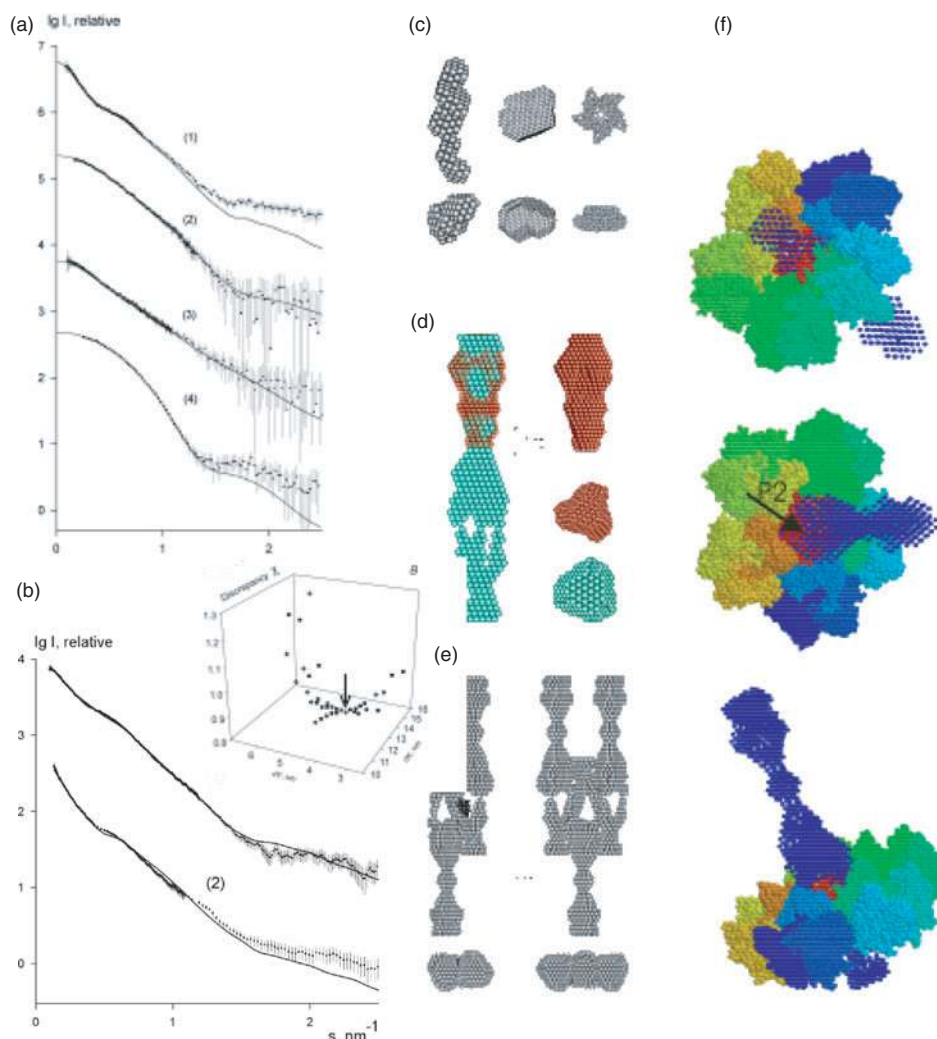


Figure 9. (a) Experimental SAXS curves (dots with error bars) and fitted curves corresponding to the *ab initio* models (solid curves) of trimeric P5₃ (1), its C-terminal domain P5C₃ (2), P2 (3) and pentameric P31 (4). (b) SAXS curves and the fits calculated using OLIGOMER (solid lines) from the mixtures containing P5₆ : P31, P5₃ and P31 (1) and P5₃ and P5₉ (2). The discrepancy is displayed as a function of the separation between the centres of mass of the two P5₃ complexes forming P5₆ : P31 in the inset. (c) Low-resolution shapes of monomeric P2 (left column) and of P31₅ without symmetry restrictions (middle column) and assuming five-fold symmetry (right column). The bottom row represents the same models rotated counterclockwise by 90° around the X-axis. (d) *Ab initio* models of the complexes P5₃ (green) and P5C₃ (orange) restored independently applying 3-fold symmetry. Orientation as in (c). (e) The best models of complexes P5₃ : P31 and P5₉ (left column) built interactively using the program MASSHA [79]. A monomer of P31 in P5₃ : P31 is displayed as a darker body. Orientation as in (c). (f) Structural model of PRD1 vertex assembly in the context of the viral shell. Top: view from the inside of the virion. Middle: top view, the putative attachment site of P2 is indicated by an arrow. Bottom: side view. Only a layer of major coat protein (P3) trimers immediately adjacent to the vertex (peripentonal trimers, individual subunits shown in green, blue and orange) was considered. Atomic coordinates corresponding to these trimers were obtained from the pseudo-atomic model of the P3 shell [266]. The SAXS model of the P31 pentamer (red) was visually aligned with P3 trimers to minimize steric clashes. The SAXS model of the P5 trimer (dark blue) was placed within the P31 to reproduce the P5-P31 position found in the SAXS model of P5₆ : P31 as in (e).

In the work devoted to the crystal growth of bovine pancreatic trypsin inhibitor (BPTI; MM = 6.5 kDa), SAXS data was recorded from saturated solutions during crystallization of BPTI at pH = 4.5 [173]. Monomeric, dimeric, pentameric and decameric BPTI molecules were identified within the crystal structure, and the experimental curves at different salt concentrations were analysed as linear combinations of the theoretical patterns from these entities. The results, confirmed by gel filtration experiments, unequivocally demonstrated that the BPTI solution contained only monomers and decamers, without other intermediates. The fraction of decamers increased with salt concentration suggesting that crystallization of BPTI at acidic pH was a two-step process where formation of decamers in saturated solutions was followed by the growth of 'BPTI decamer' crystals.

Several modelling techniques were employed in the study on bacteriophage PRD1 vertex complex [174]. Bacteriophage PRD1 is a prototype of viruses with an internal membrane, and its icosahedral capsid is built from 240 copies of the major coat protein, P3 trimer. The arrangement of P3 in PRD1 and its fold are similar to those of the adenovirus hexon. The spike structure of bacteriophage PRD1 consists of monomeric P2, trimeric P5₃ and pentameric P31₅ proteins (MM = 66 kDa, 103 kDa and 69 kDa, respectively) mediating host cell binding and controls dsDNA delivery via a membrane conduit. To study the spike structure, quaternary structures and interactions of purified spike proteins were studied by SAXS. Low-resolution models of the vertex proteins P2, P5₃, and P31₅ were reconstructed *ab initio* from the SAXS data (figures 9(a), (c) and (d)). The receptor-binding protein P2 is a 15.5 nm long and thin monomer that is anchored to the vertex base. Protein P5₃ is a 27 nm long trimer that resembles the adenovirus Ad2 spike protein pIV. P31₅ forms a globular, pentameric base with a maximum diameter of 8.5 nm. Being smaller than its adenovirus counterpart (pIII), P31₅ most likely contributes only to the external densities of the five-fold vertex in the virion. P5₃ further polymerizes into a nonameric form (P5₉), and, in the presence of P31₅, into a P5₆ : P31₅ complex. Solutions of these assemblies were always mixtures so that direct shape determination was not possible. Instead, tentative models of these aggregates were constructed interactively followed by fitting the experimental scattering data by linear combination of assembled and free P5 trimers yielding information on possible interactions in the vertex complex (figures 9(b) and (e)). The low-resolution structures were combined in a plausible model of the vertex within the context of the virion (figure 9(f)). This model displays similarity to the adenovirus spike assembly but also reveals clear structural differences, conceivably related to the modes of DNA delivery adopted by PRD1 and the adenovirus, thus providing an example of divergent evolution. The study involving *ab initio* methods, rigid body modelling, analysis of mixtures, and also accounting for the results of hydrodynamics, EM and crystallography, yields an example of a comprehensive analysis of a complicated protein assembly.

5.4. Intermolecular interactions and protein crystallization

The best known example of the influence of intermolecular interactions between globular proteins in physiology is that of the transparency of the eye lens [103, 175]. Modifications of the interactions leads to various pathological conditions like cataract (see, e.g. [176]). Similarly, transparency of the cornea, which is effectively a long pass interference filter, is controlled by the interactions between collagen fibrils [177, 178].

Crystallization of biological macromolecules is obviously an area where a better understanding of the interactions is of great importance. Extensive SAXS and light scattering studies on simple systems like lysozyme [179–181] indicate that crystallization induced by addition of salt corresponds to a change in the interactions between macromolecules from repulsive to attractive. Weakly attractive interactions define a 'crystallization slot'

between regions of undersaturation and precipitation [182]. Correspondingly, more repulsive interactions lead to a higher solubility [183–185]. Although for most practical purposes light scattering readily allows detection of crystallization (or precipitation!) it does not give information about the species involved in crystallization. This information can be obtained using SAS (see the study of BPTI in the previous section). Note that crystallization is a highly cooperative condensation polymerization process and that no intermediates are found between the basic unit and the crystals [186]. The above example of BPTI clearly illustrates the subtle interplay between long and short range (specific) interactions involved in crystallization.

The approach to the modelling of the structure factor in section 4.3 was extended to studies of the temperature, ionic strength and protein concentration dependence of interactions between small proteins in solution by introducing a Yukawa potential, with a depth defined by the effective charge of the protein [114]. The most obvious consequence of the fact that the effective charge tends to be significantly smaller than that calculated from the composition and the pH is that, at least for the foreseeable future, interparticle interactions can only be measured. Matters are more complicated for large (multimeric) proteins or their complexes with nucleic acids, lipids or detergents, which tend to have an excluded volume much larger than their dry volume. This is even more pronounced for very anisotropic macromolecules. It has also been observed but remains to be explained that the Van der Waals contribution seems to become negligible in the case of large, less compact, macromolecules [114].

Some of the limitations of the present approach to the quantitative description of the structure factor certainly lie in the fact that the interaction potentials and the structure factor calculations neglect the shape of the macromolecules and the charge distribution on their surface [187]. Salt induced interactions are always short range (< 1 nm) but at higher salt concentrations the effects also become specific. Indeed, salt concentrations below 0.2 M essentially only result in screening of coulombic repulsions but at higher ionic strength, the anion-dependent attractive contribution defined by the Hofmeister series [188, 189], which is not yet understood, but is probably associated with the ion distribution around the macromolecule [190], becomes important.

In practice other cosolutes than salt such as neutral polymers (e.g. polyethylene glycol (PEG)) are often used to make the interactions weakly attractive, in view of crystallization, especially in the case of large macromolecules or assemblies [134, 191–193]. The attractive interaction between macromolecules upon addition of neutral polymers like PEG finds its origin in depletion forces. In a solution of hard spheres containing an excess number of polymer chains, maximization of the translational entropy of these chains implies the minimization of the volume between the spheres that is inaccessible to the polymer [194]. Simulations indicate that the depth and range of this depletion potential depends on the size and concentration of the polymer [195]. Depletion can lead to structuring of the solution as observed in the case of ATCase [191], where the characteristic features of attractive and repulsive interactions simultaneously appear in the scattering patterns and result from protein-rich, PEG-depleted domains with small average distances between proteins surrounded by protein-depleted, PEG-enriched areas.

5.5. Polyelectrolyte solutions and gels

Polyelectrolytes, such as polysaccharides (carrageenans, hyaluronic acid) or DNA in pure form or in their interactions with proteins present the next level of complexity, and a review of this field would fall outside the scope of this paper. It should, however, be mentioned that in those systems intermolecular interactions often lead to rich phase diagram with a number

of different solution, gel or even liquid crystalline phases depending on ionic conditions, pH and temperature. Some of the phase transitions like the condensation of DNA (for a review see [196]) or nucleosomes [197] or the gelation of carrageenans are controlled by ionic conditions and play an important role in fundamental biological processes as well as in practical applications. Beside SAS and light scattering few methods are available for studying solutions of polyelectrolytes and anomalous SAXS provides a unique way to probe the counterion distribution around rod-like polyelectrolytes [198] like DNA [91].

The influence of ionic effects on the ordering and association phenomena in dilute and semi-dilute of hyaluronic acid [199] and iota and kappa carrageenan solutions [200, 201] have been studied by SAS but, given the range and timescale of some the ordering phenomena, light scattering studies (see, e.g. [202]) are often preferable. For DNA a number of neutron and dynamic light scattering [203] and x-ray scattering studies are available [107].

The results of these studies indicate that at low ionic strength the strong repulsive interactions between polyelectrolyte chains result in an upturn in the intensity at low angles and a strong interference maximum. The interpretation of these features is not entirely clear. The presence of weaker maxima indicates, however, that the solutions are more ordered than the concept of correlation hole developed for synthetic polyelectrolytes [204] would suggest. The existence and nature of the long range attractive forces between equally charged molecules which have been proposed [205] is also still being investigated. Ordering and long distance correlations are also revealed by slow modes in light scattering [203].

Clearly, there are still many challenging and biologically relevant problems related to intermolecular interactions, especially those involving polyelectrolytes. They can only be solved by combining careful experimentation with simulations.

5.6. Time-resolved studies: assembly and (un)folding

Biophysical investigations of assembly, large scale conformational changes or folding of biological macromolecules are part of the broader search for organizing principles on the mesoscopic scale [206]. The kinetics of transformations of matter during biologically relevant self-organizing processes like assembly or protein folding can usefully be investigated at the spatial and time resolutions accessible by SAS.

The assembly properties of several protein systems have been studied by SAS about 20 years ago but recently efforts have concentrated on viruses and/or their capsids, which are built from a limited number of different types of protein subunits—in the simplest cases only one—forming highly symmetrical assemblies. Their formation and maturation mainly involves relatively slow steps (50 ms to min) which are amenable to stopped-flow mixing studies. Assembly of virus capsids or filaments can usually be interpreted in terms of condensation polymerization mechanisms involving a (slow) nucleation step followed by a more rapid growth phase (see [207]). This process may proceed through a number of intermediates, which can usually easily be detected [208–210]. Assembly studies are useful because although *in vivo*, processes of, e.g. filament formation, certainly differ from those *in vitro* in the nucleation steps, it can be assumed that subsequent steps are rather similar.

As far as they can be investigated by SAS most assembly processes appear as a relaxation to equilibrium except microtubule assembly, which clearly is a dissipative process requiring the input of energy in the form of guanosine triphosphate (GTP), where more complex phenomena like oscillations [211] can occur. Completely consistent mechanisms for assembly and self-organization [212] of microtubules are, however, not yet available. Assembly phenomena are akin to phase transitions such as liquid–solid (e.g. crystallization) [213] or liquid–liquid [214] phase separations in proteins, which can also be addressed by time-resolved SAXS.

Monitoring conformational changes is more difficult than studying assembly as the MM of the molecules, and hence their scattering power, remains constant in the process. Large scale conformational changes occur in the quaternary structure (i.e. the relative arrangement of the subunits) of allosteric enzymes. These enzymes undergo transitions between two (or more) of states, called T (tense) and R (relaxed) under the influence of highly specific small molecules—the effectors—and thereby play a key role in the regulation of metabolism. Much experimental effort has been spent trying to discriminate between the classical MWC [215] (strictly two-state), and KNF [216] (incremental) models of allostery. Under appropriate experimental conditions (cooling and increased solvent viscosity) the time resolutions required to study the $T \leftrightarrow R$ transitions are in the range of 100 ms. The most thoroughly investigated case by SAS, as well as by a variety of other methods, is that of ATCase, where apparent rate constants of $0.05\text{--}3\text{ s}^{-1}$ were found for the $T \leftrightarrow R$ transition [217], and no intermediates were detected. More recently the ATP-induced transition in the molecular chaperone GroEL, was investigated in a combined equilibrium and time-resolved SAXS and fluorescence study [218]. The two heptameric rings forming the protein complex undergo independent ($T \leftrightarrow R$) transitions with rate constants of $3\text{--}5\text{ s}^{-1}$ and the three allosteric states (TT, TR and RR) could clearly be detected, again without intermediates. The results for these two examples can thus be explained using a time-dependent version of the MWC model, but more local incremental KNF-type effects have also been observed in ATCase [219]. The available measurements set clear limits on the spatial and temporal resolution required for further progress in this area.

The understanding of protein folding has not only provided an intellectual challenge in areas ranging from physical biochemistry to statistical mechanics for the last decades but obviously also has an immense potential for application. In particular, the solution of the folding problem is seen as an important step in protein engineering and in understanding some diseases (amyloidoses) associated with misfolding of proteins and formation of amyloid fibrils (for a review see [220]).

At the fundamental level the considerable interest that the folding of proteins is attracting has often been justified by Levinthal's paradox [221], expressing that the folded state of a protein cannot be reached by random search in typical folding times. Since it was first stated, the paradox may no longer be one (see, e.g. [222]) as it has become increasingly clear that secondary and tertiary structure interactions form simultaneously and that the unfolded state may already be partially structured. This also implies that folding occurs over several decades in time and that it is thus best studied by combining several experimental and theoretical techniques. The problem has recently been reviewed from a SAS perspective [223]. SAS can contribute to characterize the overall features of the different states and together with SVD analysis determine the minimum number of states required to describe the observations. This is hitherto limited to the 0.1 ms–min time range. To cover this broad time window various continuous flow ($>0.1\text{ ms}$) [224, 225], stopped flow ($>20\text{ ms}$) [226] and manual mixing systems have been used. Most of the slower ($>50\text{ ms}$) experiments have been done with gas detectors. The mixing devices used for time resolutions below 20 ms and down to 0.15 ms are microfabricated continuous flow devices [224, 227], which are translated relative to the fixed x-ray beam to record scattering patterns at different points along the reaction coordinate. The detectors used in these rapid kinetic experiments are CCD [227] or CCD/image intensifier [26] systems. In the first case corrections must be made for non-uniformity of response and image distortion, whereas in the second case an additional correction for contrast reduction is necessary [228]. The validity of the corrections can be verified by comparison with data obtained with a position sensitive gas detector, corrected for its non-uniformity.

The current view of *in vitro* folding is that proteins and nucleic acids fold on a funnelled landscape (for a review see [229]). Other types of landscapes can, however, not be

excluded [230]. Folding is envisioned as starting from unfolded states at the top of the funnel followed by a fast early collapse to a molten globule states at intermediate levels and a restricted search for the native state at the bottom of the funnel.

Many—but not all—single domain proteins fold on a smooth energy landscape and folding without early collapse [231] or without stable intermediates [232] has been observed. A first quantitative model—the topomer search model—has recently been proposed for this type of protein [233, 234]. Some caution is required in the interpretation of SAS data, even for single domain proteins as illustrated by the detailed equilibrium study on the small all- β protein neocarzinostatin [98]. SVD detects only one intermediate state but a more detailed thermodynamic analysis indicates that at least one more intermediate state or an ensemble of states with temperature-dependent populations, must also exist. A folding intermediate differing essentially in hydration from the native state was found both in equilibrium and time-resolved SAS studies on guadinium hydrochloride denatured α -lactalbumin [228].

These two examples illustrate that the variety of results in the literature may also partly depend on the type of perturbation used to induce (un)folding. Temperature and pH jumps are particularly useful and facilitate interpretation of the data compared to pressure or denaturant concentration jumps where solvation and contrast effects are more likely to be superimposed to the folding processes. To deal with the variety a classification of unfolded states has been proposed [235] and volume changes during pressure-induced protein unfolding have been reviewed [236]. Detailed equilibrium studies thus remain an indispensable complement to the time-resolved studies.

The folding of acid denatured cytochrome *c* by pH-jump [224] and of the engrailed homeodomain of *D. melanogaster* [237] have been investigated over several decades in time. Whereas the first study is purely experimental, the second one relies on a combination of simulations and experimental data from various techniques including SAS.

There are far fewer studies on the folding of RNA. They include compaction of yeast tRNA^{Phe} and of the catalytic domain of RNase P RNA from *Bacillus subtilis* [238]. The most extensive studies so far have been done on the Mg⁺⁺-induced folding of the Tetrahymena group I ribozyme. The results have confirmed that the folding landscapes for RNA are more rugged than those of proteins with large barriers separating distinct pathways, including some leading to kinetic traps [239]. Most processes thus also occur on much longer timescales than in the case of proteins. The Tetrahymena ribozyme is the only case where folding, albeit through a misfolded intermediate, was followed over five decades in time [240]. In this case two kinetic phases reveal the existence of at least one intermediate even if the scattering patterns can be described as a linear combination of only two components.

Folding of proteins and RNA is clearly an area where some systematic work is still required before it becomes possible to, even qualitatively, link experimental observations to specific theoretical models. It should also be kept in mind that the physiological significance of *in vitro* studies is limited by the fact that *in vivo* folding and processing conditions differ significantly from *in vitro* conditions, especially for proteins, due among others to the fact that the polypeptide chains fold during their synthesis on the ribosome and/or with intervention of other proteins (chaperones).

5.7. Coherence and single molecule scattering

Although coherence is one of the most important prerequisites for scattering it has attracted less attention for x-rays because of its limited extent on traditional x-ray sources. The high brilliance of third generation SR sources is rapidly changing this situation. As is well known from dynamic light scattering (see, e.g. [241]) coherent scattering leads to a characteristic

speckle pattern containing information about the large scale structure and dynamics of the sample. In applications of coherent scattering, like x-ray photon correlation spectroscopy (XPCS), the sample volume must be coherently illuminated and the maximum path length difference (PLD) of the scattered waves originating from it must be smaller than the longitudinal coherence length of the beam.

The transverse one-sigma coherence area of an incoherent quasi-monochromatic source with mean wavelength λ_0 is $S = \lambda_0^2 R^2 / 4\pi \sigma_x \sigma_y$, where σ_x and σ_y are the horizontal and vertical one-sigma source sizes and R the source-sample distance. Even for third generation sources these areas are not much larger than a few hundred μm^2 and the photon flux through the sample is correspondingly low. This can partly be compensated by lowering the longitudinal coherence length $\Lambda = \lambda_0^2 / \Delta\lambda$, by increasing the bandpass determined by $\Delta\lambda$, the spectral FWHM of the radiation. At a scattering angle 2θ for a sample thickness t , $\text{PLD} = 2t \sin^2 \theta$ in transmission geometry. Hence for wavelengths around 0.15 nm and PLDs close to 10 nm one can afford a bandpass of 1–3% as provided by multilayer monochromators [242] or pink radiation from an undulator [243].

The coherent wide-bandpass SAXS undulator beamline (8-ID) at the APS [243] has a one-sigma coherence area with horizontal (H) and vertical (V) dimensions of $(7(\text{H}) \times 50(\text{V}) \mu\text{m}^2)$ at the sample position ($R = 55 \text{ m}$), reflecting the differences in source dimensions $(350(\text{H}) \times 50(\text{V}) \mu\text{m}^2)$. This instrument delivers 10^{12} photons s^{-1} in a pink beam with $\Delta\lambda/\lambda = 2.6 \times 10^{-2}$ at $\lambda_0 = 0.162 \text{ nm}$ (i.e. $\Lambda = 6 \text{ nm}$). For organic samples the optimal thickness is around 1 mm and with this value of Λ the maximum spatial resolution at which coherent effects can still be observed around is 50 nm ($s = 0.13 \text{ nm}^{-1}$). This is often too restrictive and more recently static and dynamic scattering experiments on this instrument have been made with a $\Delta\lambda/\lambda = 3 \times 10^{-4}$ ($\Lambda \sim 0.5 \mu\text{m}$) giving a flux of 10^{10} photons s^{-1} . The largest scattering vector compatible with coherent scattering ($s_{\text{max}} \sim 1.2 \text{ nm}^{-1}$) corresponds to a maximum spatial resolution of 5 nm, which represents a significant improvement over photon correlation spectroscopy with visible light ($s_{\text{max}} \sim 4 \times 10^{-2} \text{ nm}^{-1}$), given also that the method is applicable to opaque samples.

It has, however, been observed in all cases so far that the speckle contrast with x-rays was systematically lower than expected and this has been attributed to the reduction of beam coherence due to windows and imperfections in optical elements [243–245]. The effects have been investigated in some detail [246] but still present a real technological challenge, especially in the context of future sources.

The normalized time autocorrelation function of the scattered intensity $I(s, t)$ in the speckle pattern (see, e.g. [241]):

$$g_2(s, \tau) \equiv \frac{\langle I(s, 0)I(s, \tau) \rangle}{\langle I(s) \rangle^2} \quad (40)$$

yields the characteristic relaxation times (τ) of the sample related to the diffusion coefficient. For sufficiently concentrated systems the diffusion coefficient is time and wavevector dependent, as shown by XPCS measurements on latex spheres (diameter 142 nm) as models of complex fluids in the time range (30 ms to 100 s) [247]. These results also suggest that XPCS may in the not too distant future also be applicable to the study of phenomena that result in the formation of gels by large objects such as cylindrical viruses [248] or nucleosomes [249] and are related to properties like self-assembly and the functioning of some biological systems. Note that this represents a significant challenge since the scattering power of a typical spherical virus is still an order of magnitude lower than that of the latex particles used hitherto in these studies.

The possibility of investigating atomic correlations associated with protein dynamics on a sub- μs timescale, using the coherent radiation which short wavelength x-ray free electron lasers (XFELs) promise to deliver during the coming decade, remains entirely to be explored

[250]. The expected extraordinary brilliance of XFELs has already resulted in proposals for single macromolecule scattering, which have been put forward as part of the scientific cases for the construction of these machines [251]. These proposals are largely based on algorithms for structure determination using direct phase retrieval by the oversampling technique [252]. These algorithms have been successfully tested on model systems and recently also in the practical imaging at 30 nm resolution of bacterial cells fixed and stained with KMnO_4 using a $20\text{ }\mu\text{m}$ diameter spot of 0.2 nm wavelength coherent x-rays with a longitudinal coherence length of $7\text{ }\mu\text{m}$ [253]. This two-dimensional imaging application provides an important proof of principle but also illustrates that there still is a long way to go. Indeed, scattering is proportional to the square of the volume for particles with similar contrasts and the scattering intensities of an unstained bacterial cell (*E. coli*), a 35 nm diameter spherical virus (e.g. TBSV: tomato bushy stunt virus) and a 40 kDa protein are in the ratio $10^{13} : 10^4 : 1$. As in the case of SAXS, the degree of coherence of the beam, which depends on the quality of optical elements and windows [246] determines the signal to background ratio and hence also the quality of the reconstructions. The effect of partial coherence on imaging and scattering has been investigated [254, 255]. Experimental and numerical methods for recombining a set of two dimensional patterns into a three-dimensional one exist at least in principle [252], even if substantial developments are still necessary.

The main foreseeable effect of very intense beams is radiation damage and Coulomb explosion of the object. As radiation damage depends on dose as well as on dose rate it is hoped that very high intensity pulses will allow collection of meaningful data before detectable damage sets in at the required resolution. The few theoretical studies on such effects of XFEL-radiation [251, 256] differ in their assumptions of radiated XFEL power by several orders of magnitude and some of the calculations probably represent a best case scenario. The effect of pulse length is also largely unknown and facilities like the SPPS [257] should provide much needed data in this area. First experimental data on the Coulomb explosion of xenon clusters with 98 nm soft XFEL radiation corresponding to approximately a thousand times lower power densities than those expected at XFELs suggest that absorption is strongly enhanced in these clusters [258].

The length of the pulse and the velocity of light ($\text{max. } 30\text{ nm fs}^{-1}$) determine the maximum thickness of the sample if partial images are to be avoided. This also sets an upper limit to the duration of observation on such systems with an XFEL. Clearly, a considerable research and development effort is required before these techniques reach the level of routine that are characteristic of life science applications.

6. Conclusions

During the last decade, SAXS and SANS have become increasingly important tools in studies of biological macromolecules. This process has been accelerated by the broader accessibility of instruments in large scale facilities and novel and powerful data analysis software to a larger user community. Several third generation synchrotron sources, where the scattering patterns can be collected in less than a second, have recently become available (ESRF, Grenoble; Spring-8, Himeji; APS, Argonne), and several additional low emittance synchrotrons are being constructed or planned (Soleil, France; Diamond, UK; Petra, Germany). All these sources possess or plan biology/soft condensed matter SAXS beamlines. Most of large scale facilities maintain user support programs helping users to collect and sometimes analyse their data, whereas many modelling programs are publicly available on the Web. Methods for the analysis of equilibrium structures especially for monodisperse solutions and mixtures are now well developed, whereas there is still much to be done before routine applications become possible

for more complicated cases involving solutions of interacting particles and time-resolved studies. A comprehensive set of programs for data processing, *ab initio* analysis, rigid body refinement, three-dimensional visualization and characterization of mixtures for biological SAS can be downloaded from www.embl-hamburg.de/ExternalInfo/Research/Sax/. There are also sites for other *ab initio* programs, e.g. DALAI_GA (akilonia.cib.csic.es/DALAI_GA2/) and SAXS3D (www.cmpharm.ucsf.edu/~walther/saxs/). The program FISH for non-linear parametric modelling of form and structure factors is available from http://www.isis.rl.ac.uk/largescale/LOQ/FISH/FISH_intro.htm.

SAS is often employed in combination with other structural and biochemical methods to generate consistent models. Even without further analysis, SAXS patterns can be used to correct the Fourier amplitudes in cryo-EM [259–261] and to constrain the models of protein fold built by structure prediction algorithms [262]. *Ab initio* envelopes have been employed for phasing low-resolution reflections in protein crystallography [263, 264], and more detailed DR-models should be even more useful to this end [146, 265]. As seen from the applications presented in this review, SAS not only provides low-resolution models of particles in solution, but in many cases delivers answers to significant functional questions. Most importantly, the method allows to study native particles in nearly physiological solutions and to analyse structural changes in response to variations in external conditions thus yielding valuable information on the relationship between these structural changes and the operation of these systems. A comparison of SAXS/SANS with the other major methods employed in the analysis of macromolecular structures (table 3) indicates that, despite its low resolution, the method is a general tool for studying the structure, kinetics and interactions of biological macromolecules in solution.

Current trends in the life sciences point to a phase of integrative systems biology aiming at understanding living systems across the different levels of biological organization. Here,

Table 3. Advantages and limitations of major methods for structure analysis of biological macromolecules.

Method	Samples	Advantages	Limitations
Crystallography	Crystals	Very high resolution (up to 0.1 nm) revealing fine detail of atomic structure (e.g. of the active centres)	Crystals required. Flexible portions are not seen. Structure may be influenced by crystal packing forces
NMR	Dilute solutions (~5–10 mg ml ⁻¹)	High-resolution (0.2–0.3 nm) in solution	Hardly applicable for MM exceeding ~50 kDa
Cryo-EM	Frozen very dilute solutions (<1 mg ml ⁻¹)	Low amount of material. Direct visualization of particle shape and symmetry	Low (~1 nm) resolution. Hardly applicable for MM less than ~200 kDa
SAXS/SANS	Dilute and semi-dilute solutions (from ~1 to ~100 mg ml ⁻¹)	Analysis of structure, kinetics and interactions in nearly native conditions. Study of mixtures and non-equilibrium systems. Wide MM range (few kDa to hundreds MDa)	Low (~1–2 nm) resolution. Requires additional information to resolve ambiguity in model building
Static and dynamic light scattering, ultracentrifugation	Very dilute solutions (<1 mg ml ⁻¹)	Non-destructive. Low amount of material. Simplicity of the experiments	Yield overall parameters only

SAS and EM should provide the structural framework for the analysis of complex structures in which the high-resolution models from crystallography and NMR can meaningfully be fitted. SAS is applicable to macromolecules with very different sizes: only in this review, examples of proteins as small as 6.5 kDa (BPTI) and complexes as large as ribosome (2.3 MDa) have been presented. Moreover, for complexes with MMs of a few hundred kDa—too large for NMR and too small for cryo-EM—SAS remains effectively the only method to obtain their shape in solution. The momentum gained in the last decade suggests that in future SAS will further strengthen its position as a mainstream method for the analysis of macromolecular solutions.

Acknowledgments

The authors thank T Irving for providing the schematic drawing of the BioCAT beamline. This work was supported in part by the International Association for the Promotion of Cooperation with Scientists from New Independent States of the former Soviet Union (INTAS Project 02-243).

References

- [1] Guinier A 1939 *Ann. Phys. (Paris)* **12** 161–237
- [2] Guinier A and Fournet G 1955 *Small Angle Scattering of X-Rays* (New York: Wiley)
- [3] Ibel K and Stuhmann H B 1975 *J. Mol. Biol.* **93** 255–65
- [4] Engelman D M and Moore P B 1972 *Proc. Natl Acad. Sci. USA* **69** 1997–9
- [5] Glatter O and Kratky O 1982 *Small Angle X-ray Scattering* (London: Academic)
- [6] Feigin L A and Svergun D I 1987 *Structure Analysis by Small-Angle X-Ray and Neutron Scattering* (New York: Plenum)
- [7] Pilz I, Glatter O and Kratky O 1972 *Z. Naturf. b* **27** 518–24
- [8] Tardieu A and Vachette P 1982 *EMBO J.* **1** 35–40
- [9] Koch M J H *et al* 1979 *Polym. Bull.* **1** 709–14
- [10] Hendrickson W A and Ogata C M 1997 *Methods in Enzymology* vol 276, ed C W Carter and R M Sweet (New York: Academic) pp 494–523
- [11] Stuhmann H B 1981 *Quart. Rev. Biophys.* **14** 433–60
- [12] Stuhmann H B *et al* 1986 *Eur. Biophys. J.* **14** 1–6
- [13] Shu F, Ramakrishnan V and Schoenborn B P 2000 *Proc. Natl Acad. Sci. USA* **97** 3872–7
- [14] Zaccai G and Jacrot B 1983 *Ann. Rev. Biophys. Bioengng.* **12** 139–57
- [15] Wall M E, Gallagher S C and Trewella J 2000 *Ann. Rev. Phys. Chem.* **51** 355–80
- [16] Snijder H J *et al* 2003 *J. Struct. Biol.* **141** 122–31
- [17] Chatake T *et al* 2003 *Proteins* **50** 516–23
- [18] Chantler C T *et al* 2003 National Institute of Standards and Technology, Gaithersburg, MD, <http://physics.nist.gov/ffast>
- [19] Giacovazzo C *et al* 1992 *Fundamentals of Crystallography* vol 2 (New York: Oxford University Press)
- [20] Stuhmann H B and Kirste R G 1965 *Zeitschr. Physik. Chem. Neue Folge* **46** 247–50
- [21] Koch M H J 1988 *Makromolekulare Chem., Macromol. Symp.* **15** 79–90
- [22] Bras W and Ryan A J 1998 *Adv. Colloid Interface Sci.* **75** 1–43
- [23] Lindner P, May R P and Timmins P A 1992 *Physica B* **180–181** 967–72
- [24] Heenan R K, Penfold J and King S M 1997 *J. Appl. Crystallogr.* **30** 1140–7
- [25] Irving T C 1998 *Nat. Struct. Biol.* **5** (suppl) 648–50
- [26] Amemiya Y *et al* 1995 *Rev. Sci. Instrum.* **66** 2290–4
- [27] Lumma D, Lurio L B and Mochrie S G J 2000 *Rev. Sci. Instrum.* **71** 3274–89
- [28] Pontoni D, Narayanan T and Rennie A R 2002 *J. Appl. Crystallogr.* **35** 207–11
- [29] Petrascu A-M, Koch M H J and Gabriel A 1998 *J. Macromol. Sci. Phys. B* **37** 463–83
- [30] Berar J F *et al* 2002 *J. Appl. Crystallogr.* **35** 471–6
- [31] Kratky O and Pilz I 1972 *Q. Rev. Biophys.* **5** 481–537
- [32] Orthaber D, Bergmann A and Glatter O 2000 *J. Appl. Crystallogr.* **33** 218–25
- [33] Porod G 1982 *Small-Angle X-Ray Scattering* ed O Glatter and O Kratky (London: Academic) pp 17–51
- [34] Goodisman J 1980 *J. Appl. Crystallogr.* **13** 132–4

- [35] Glatter O 1977 *J. Appl. Crystallogr.* **10** 415–21
- [36] Moore P B 1980 *J. Appl. Crystallogr.* **13** 168–75
- [37] Svergun D I, Semenyuk A V and Feigin L A 1988 *Acta Crystallogr. A* **44** 244–50
- [38] Svergun D I and Pedersen J S 1994 *J. Appl. Crystallogr.* **27** 241–8
- [39] Svergun D I 1992 *J. Appl. Crystallogr.* **25** 495–503
- [40] Shannon C E and Weaver W 1949 *The Mathematical Theory of Communication* (Urbana: University of Illinois Press)
- [41] Frieden B R 1971 *Progress in Optics* vol 9, ed E Wolf (Amsterdam: North-Holland) pp 312–407
- [42] Debye P 1915 *Ann. Physik* **46** 809–23
- [43] Stuhmann H B 1970 *Zeitschr. Physik. Chem. Neue Folge* **72** 177–98
- [44] Stuhmann H B 1970 *Acta Crystallogr. A* **26** 297–306
- [45] Svergun D I and Stuhmann H B 1991 *Acta Crystallogr. A* **47** 736–44
- [46] Svergun D I *et al* 1997 *J. Appl. Crystallogr.* **30** 798–802
- [47] Svergun D I *et al* 1996 *Acta Crystallogr. A* **52** 419–26
- [48] Chacon P *et al* 1998 *Biophys. J.* **74** 2760–75
- [49] Svergun D I 1999 *Biophys. J.* **76** 2879–86
- [50] Chacon P *et al* 2000 *J. Mol. Biol.* **299** 1289–302
- [51] Walther D, Cohen F E and Doniach S 2000 *J. Appl. Crystallogr.* **33** 350–63
- [52] Vigil D *et al* 2001 *Biophys. J.* **80** 2082–92
- [53] Heller W T *et al* 2002 *Biochemistry* **41** 15654–63
- [54] Svergun D I, Petoukhov M V and Koch M H J 2001 *Biophys. J.* **80** 2946–53
- [55] Petoukhov M V *et al* 2002 *Biophys. J.* **83** 3113–25
- [56] Volkov V V and Svergun D I 2003 *J. Appl. Crystallogr.* **36** 860–4
- [57] Kozin M B and Svergun D I 2001 *J. Appl. Crystallogr.* **34** 33–41
- [58] Berman H M *et al* 2000 *Nucl. Acids Res.* **28** 235–42
- [59] Arjunan P *et al* 1996 *J. Mol. Biol.* **256** 590–600
- [60] Koenig S *et al* 1992 *Biochemistry* **31** 8726–31
- [61] Koenig S *et al* 1993 *Eur. Biophys. J.* **22** 185–94
- [62] Langridge R *et al* 1960 *J. Mol. Biol.* **2**
- [63] Ninio J, Luzzati V and Yaniv M 1972 *J. Mol. Biol.* **71** 217–29
- [64] Pavlov M Y and Fedosov B A 1983 *Biopolymers* **22** 1507–22
- [65] Lattman E E 1989 *Proteins* **5** 149–55
- [66] Hubbard S R, Hodgson K O and Doniach S 1988 *J. Biol. Chem.* **263** 4151–8
- [67] Grossmann J G *et al* 1993 *Biochemistry* **32** 7360–6
- [68] Perkins S J *et al* 1993 *Biochem. J.* **295** 87–99
- [69] Fujisawa T *et al* 1994 *J. Biochem. (Tokyo)* **115** 875–80
- [70] Svergun D I, Barberato C and Koch M H J 1995 *J. Appl. Crystallogr.* **28** 768–73
- [71] Svergun D I *et al* 1998 *Proc. Natl Acad. Sci. USA* **95** 2267–72
- [72] Merzel F and Smith J C 2002 *Proc. Natl Acad. Sci. USA* **99** 5378–83
- [73] Pavlov M *et al* 1986 *Biopolymers* **25** 1385–97
- [74] Mueller J J, Damaschun G and Schrauber H 1990 *J. Appl. Crystallogr.* **23** 26–34
- [75] Svergun D I 1991 *J. Appl. Crystallogr.* **24** 485–92
- [76] Svergun D I 1994 *Acta Crystallogr. A* **50** 391–402
- [77] Kozin M B, Volkov V V and Svergun D I 1997 *J. Appl. Crystallogr.* **30** 811–5
- [78] Kozin M B and Svergun D I 2000 *J. Appl. Crystallogr.* **33** 775–7
- [79] Konarev P V, Petoukhov M V and Svergun D I 2001 *J. Appl. Crystallogr.* **34** 527–32
- [80] Boehm M K *et al* 1999 *J. Mol. Biol.* **286** 1421–47
- [81] Krueger J K *et al* 1998 *Biochemistry* **37** 13997–4004
- [82] Tung C S, Walsh D A and Trewella J 2002 *J. Biol. Chem.* **277** 12423–31
- [83] Svergun D I *et al* 2000 *J. Biol. Chem.* **275** 297–302
- [84] Mattinen M L *et al* 2002 *Biophys. J.* **83** 1177–83
- [85] Serdyuk I N and Grenader A K 1975 *FEBS Lett.* **59** 133–6
- [86] Koch M H J and Stuhmann H B 1979 *Methods Enzymol.* **59** 670–706
- [87] Capel M S *et al* 1987 *Science* **238** 1403–6
- [88] Brodersen D E *et al* 2002 *J. Mol. Biol.* **316** 725–68
- [89] Stuhmann H B *et al* 1976 *J. Mol. Biol.* **100** 399–413
- [90] Vainshtein B K *et al* 1980 *FEBS Lett.* **116** 107–10
- [91] Das R *et al* 2003 *Phys. Rev. Lett.* **90** 188103

- [92] Lawson C L and Hanson R J 1974 *Solving Least Squares Problems* (Englewood Cliffs, NJ: Prentice-Hall)
- [93] Golub G H and Reinsh C 1970 *Numer. Math.* **14** 403–20
- [94] Larson H J 1975 *Statistics: An Introduction* (New York: Wiley)
- [95] Fowler A G *et al* 1983 *J. Biochem. Biophys. Methods* **7** 317–29
- [96] Chen L, Hodgson K O and Doniach S 1996 *J. Mol. Biol.* **261** 658–71
- [97] Bilgin N *et al* 1998 *Biochemistry* **37** 8163–72
- [98] Perez J *et al* 2001 *J. Mol. Biol.* **308** 721–43
- [99] Konarev P V *et al* 2003 *J. Appl. Crystallogr.* **36** 1277–82
- [100] Leckband D and Israelachvili J 2001 *Quart. Rev. Biophys.* **34** 105–267
- [101] Verwey E J W and Overbeek J T G 1948 *Theory of the Stability of Lyophobic Colloids* (Amsterdam: Elsevier)
- [102] Tardieu A 1994 Neutron and synchrotron radiation for condensed matter studies *Applications to Soft Condensed Matter and Biology* vol III (Les editions de Physique (France)) (Berlin: Springer) pp 145–60
- [103] Veretout F, Delaye M and Tardieu A 1989 *J. Mol. Biol.* **205** 713–28
- [104] Brunner-Popela J and Glatter O 1997 *J. Appl. Crystallogr.* **30** 431–42
- [105] Weyerich B, Brunner-Popela J and Glatter O 1999 *J. Appl. Crystallogr.* **32** 197–209
- [106] Weyerich B *et al* 1990 *Faraday Discussions* **90** 245–59
- [107] Koch M H J *et al* 1995 *Macromolecules* **28** 4904–7
- [108] Koch M H J *et al* 1988 *Science* **240** 194–6
- [109] Skibinska L *et al* 1999 *J. Chem. Phys.* **110** 1794–800
- [110] Svergun D I *et al* 2000 *J. Chem. Phys.* **113** 1651–65
- [111] Baxter R J 1968 *J. Chem. Phys.* **49** 2270–3
- [112] Belloni L 1991 *Interacting Monodisperse and Polydisperse Spheres* (Amsterdam: Elsevier)
- [113] Hansen J P and McDonald I R 1986 *Theory of Simple Liquids* (New York: Academic)
- [114] Tardieu A *et al* 1999 *J. Cryst. Growth* **196** 193–203
- [115] Malfois M *et al* 1996 *J. Chem. Phys.* **105** 3290–300
- [116] Lomakin A, Asherie N and Benedek G B 1996 *J. Chem. Phys.* **104** 1646–56
- [117] Takahashi Y, Nishikawa Y and Fujisawa T 2003 *J. Appl. Crystallogr.* **36** 549–52
- [118] Zipper P and Durchschlag H 2003 *J. Appl. Crystallogr.* **36** 509–14
- [119] Schmidt B *et al* 1995 *FEBS Lett.* **372** 169–72
- [120] Riboldi-Tunnicliffe A *et al* 2001 *Nat. Struct. Biol.* **8** 779–83
- [121] Dainese E *et al* 2000 *Arch. Biochem. Biophys.* **373** 154–62
- [122] Cuff M E *et al* 1998 *J. Mol. Biol.* **278** 855–70
- [123] Kato R *et al* 2001 *J. Mol. Biol.* **309** 227–38
- [124] Meyer A S *et al* 2003 *Cell* **113** 369–81
- [125] Dainese E *et al* 2002 *J. Biol. Chem.* **277** 40296–301
- [126] Vachette P *et al* 2002 *J. Biol. Chem.* **277** 40823–31
- [127] Birk C *et al* 2002 *J. Mol. Biol.* **321** 447–57
- [128] Grueber G *et al* 2000 *J. Biol. Chem.* **275** 30082–7
- [129] Fischer H *et al* 2003 *J Biol Chem* **278** 16030–8
- [130] Egea P F *et al* 2001 *J. Mol. Biol.* **307** 557–76
- [131] Gralle M *et al* 2002 *Biophys. J.* **83** 3513–24
- [132] Scott D J *et al* 2002 *J. Mol. Biol.* **315** 1179–87
- [133] Ackerman C J *et al* 2003 *Biophys. J.* **84** 489–500
- [134] Bernocco S *et al* 2001 *J. Biol. Chem.* **276** 48930–6
- [135] Bernocco S *et al* 2003 *J. Biol. Chem.* **278** 7199–205
- [136] Fujisawa T, Kostyukova A and Maeda Y 2001 *FEBS Lett.* **498** 67–71
- [137] Torkkeli M *et al* 2002 *Biophys. J.* **83** 2240–7
- [138] Paananen A *et al* 2003 *Biochemistry* **42** 5253–8
- [139] Funari S S *et al* 2000 *J. Biol. Chem.* **275** 31283–8
- [140] Ban N *et al* 2000 *Science* **289** 905–20
- [141] Shtykova E V *et al* 2003 *J. Appl. Crystallogr.* **36** 669–73
- [142] Zou P *et al* 2003 *J. Biol. Chem.* **278** 2636–44
- [143] Fraternali F and Pastore A 1999 *J. Mol. Biol.* **290** 581–93
- [144] Grossmann J G *et al* 2002 *Biochemistry* **41** 3613–9
- [145] Receveur V *et al* 2002 *J. Biol. Chem.* **277** 40887–92
- [146] Aparicio R *et al* 2002 *Biochemistry* **41** 9370–5
- [147] Witty M *et al* 2002 *EMBO J.* **21** 4207–18

- [148] Hammel M *et al* 2002 *J. Mol. Biol.* **321** 85–97
- [149] Dellisanti C D *et al* 2003 *FEBS Lett.* **544** 189–93
- [150] Cianci M *et al* 2002 *Proc. Natl Acad. Sci. USA* **99** 9795–800
- [151] Svergun D I *et al* 1997 *Proteins* **27** 110–7
- [152] Sousa M C *et al* 2000 *Cell* **103** 633–43
- [153] Trapani S *et al* 2001 *J. Mol. Biol.* **313** 1059–72
- [154] Lodge J A *et al* 2003 *J. Biol. Chem.* **278** 19151–8
- [155] Shilton B H *et al* 2002 *Eur. J. Biochem.* **269** 3801–9
- [156] Feil I K *et al* 2001 *J. Biol. Chem.* **276** 12024–9
- [157] Kim K K, Kim R and Kim S H 1998 *Nature* **394** 595–9
- [158] Muchowski P J *et al* 1999 *J. Mol. Biol.* **289** 397–411
- [159] Berengian A R, Parfenova M and McHaourab H S 1999 *J. Biol. Chem.* **274** 6305–14
- [160] Aslam M and Perkins S J 2001 *J. Mol. Biol.* **309** 1117–38
- [161] Aslam M *et al* 2003 *J. Mol. Biol.* **329** 525–50
- [162] Krueger S *et al* 2003 *J. Struct. Biol.* **141** 240–58
- [163] Trehwella J and Krueger J K 2002 *Methods Mol. Biol.* **173** 137–59
- [164] Frank J *et al* 1995 *Nature* **376** 441–4
- [165] Svergun D I and Nierhaus K H 2000 *J. Biol. Chem.* **275** 14432–9
- [166] Schlutzenzen F *et al* 2000 *Cell* **102** 615–23
- [167] Yusupov M M *et al* 2001 *Science* **292** 883–96
- [168] Koenig S *et al* 1993 *Eur. Biophys. J.* **22** 185–94
- [169] Koenig S *et al* 1998 *Biochemistry* **37** 5329–34
- [170] Kozielski F *et al* 2001 *J. Biol. Chem.* **276** 1267–75
- [171] Kozielski F *et al* 1997 *Cell* **91** 985–94
- [172] Van Den Heuvel R H *et al* 2003 *J. Mol. Biol.* **330** 113–28
- [173] Hamiaux C *et al* 2000 *J. Mol. Biol.* **297** 697–712
- [174] Sokolova A *et al* 2001 *J. Biol. Chem.* **27** 27
- [175] Benedek G B 1971 *Appl. Opt.* **10** 459–73
- [176] Suarez G, Oronsky A L and Koch M H J 1993 *J. Biol. Chem.* **268** 17716–21
- [177] Sayers Z *et al* 1982 *J. Mol. Biol.* **160** 593–607b
- [178] Meek K M *et al* 1981 *J. Mol. Biol.* **149** 477–88
- [179] Guilloteau J P 1991 *PhD Thesis* Université Louis Pasteur, Strasbourg
- [180] Ducruix A *et al* 1996 *J. Cryst. Growth* **168** 28–39
- [181] Muschol M and Rosenberger F 1995 *J. Chem. Phys.* **103** 10424–32
- [182] George A and Wilson W W 1994 *Acta Crystallogr. D* **50** 361–5
- [183] Gripon C *et al* 1997 *J. Cryst. Growth* **178** 575–84
- [184] Bonnete F, Finet S and Tardieu A 1999 *J. Cryst. Growth* **196** 403–14
- [185] Velev O D, Kaler E W and Lenhoff A M 1998 *Biophys. J.* **75** 2682–97
- [186] Finet S *et al* 1998 *Eur. Biophys. J.* **27** 263–71
- [187] Lomakin A, Asherie N and Benedek G B 1999 *Proc. Natl Acad. Sci. USA* **96** 9465–8
- [188] Hofmeister F 1888 *Arch. Exp. Pathol. Pharmacol.* **24** 247–60
- [189] Cacace M G, Landau E M and Ramsden J J 1997 *Quart. Rev. Biophys.* **30** 241–77
- [190] Ninham B W and Yaminsky V 1997 *Langmuir* **13** 2097–108
- [191] Budayova M *et al* 1999 *J. Cryst. Growth* **196** 210–9
- [192] Casselyn M *et al* 2001 *Acta Crystallogr. D: Biol. Crystallogr.* **57** 1799–812
- [193] Bonnete F *et al* 2001 *J. Cryst. Growth* **232** 330–9
- [194] Lekkerkerker H N W 1997 *Physica A* **244** 227–37
- [195] Vivares D *et al* 2002 *Eur. Phys. J. E—Soft Matter* **9** 15–25
- [196] Livolant F and Leforestier A 1996 *Prog. Polym. Sci.* **21** 1115–64
- [197] Mangenot S *et al* 2002 *Biophys. J.* **82** 345–56
- [198] Guillaume B *et al* 2002 *Eur. Phys. J. E* **8** 299–309
- [199] Rinaudo M *et al* 1993 *Polymer* **34** 3710–5
- [200] Deneff B *et al* 1996 *Int. J. Biol. Macromol.* **18** 151–9
- [201] Mischenko N *et al* 1996 *Int. J. Biol. Macromol.* **19** 185–94
- [202] Cuppo F, Reynaers H and Paoletti S 2002 *Macromolecules* **35** 539–47
- [203] Borsali R, Nguyen H and Pecora R 1998 *Macromolecules* **31** 1548–55
- [204] Hayter J *et al* 1980 *J. Phys. Lett.* **41** L451–L4
- [205] Ise N 1986 *Angewandte Chemie, International Edition in English* **25** 323–34

- [206] Laughlin R B *et al* 2000 *Proc. Natl Acad. Sci. USA* **97** 32–7
- [207] Oosawa F and Asakura S 1975 *Thermodynamics of the Polymerization of Proteins* (London: Academic)
- [208] Perez J *et al* 2000 *Cell. Mol. Biol. (Noisy-le-grand)* **46** 937–48
- [209] Canady M A, Tsuruta H and Johnson J E 2001 *J. Mol. Biol.* **311** 803–14
- [210] Lata R *et al* 2000 *Cell* **100** 253–63
- [211] Marx A and Mandelkow E 1994 *Eur. Biophys. J.* **22** 405–21
- [212] Tabony J *et al* 2002 *Langmuir* **18** 7196–207
- [213] Casselyn M *et al* 2002 *Acta Crystallogr. D* **58** 1568–70
- [214] Annunziata O *et al* 2002 *Proc. Natl Acad. Sci. USA* **99** 14165–70
- [215] Monod J, Wyman J and Changeux J P 1965 *J. Mol. Biol.* **12** 88–118
- [216] Koshland J D E, Nemethy G and Filmer D 1966 *Biochemistry* **5** 365–85
- [217] Tsuruta H, Vachette P and Kantrowitz E R 1998 *Proteins* **31** 383–90
- [218] Inobe T *et al* 2003 *J. Mol. Biol.* **327** 183–91
- [219] Tauc P *et al* 1982 *J. Mol. Biol.* **155** 155–68
- [220] Rochet J-C and Lansbury P T Jr 2000 *Curr. Opin. Struct. Biol.* **10** 60–8
- [221] Levinthal C 1969 *Mossbauer Spectroscopy in Biological Systems* ed P DeBrunner *et al* (Urbana, IL: University of Illinois Press) pp 22–4
- [222] Trifonov E N and Berezovsky I N 2003 *Curr. Opin. Struct. Biol.* **13** 110–4
- [223] Doniach S 2001 *Chem. Rev.* **101** 1763–78
- [224] Akiyama S *et al* 2002 *Proc. Natl Acad. Sci. USA* **99** 1329–34
- [225] Pollack L *et al* 2001 *Phys. Rev. Lett.* **86** 4962–5
- [226] Tsuruta H *et al* 1989 *Rev. Sci. Instrum.* **60** 2356–8
- [227] Pollack L *et al* 1999 *Proc. Natl Acad. Sci. USA* **96** 10115–7
- [228] Arai M *et al* 2002 *J. Mol. Biol.* **321** 121–32
- [229] Plotkin S S and Onuchic J N 2002 *Quart. Rev. Biophys.* **35** 111–67
- [230] Plotkin S S and Wolynes P G 2003 *Proc. Natl Acad. Sci. USA* **100** 4417–22
- [231] Plaxco K W *et al* 1999 *Nat. Struct. Biol.* **6** 554–6
- [232] Jackson S E 1998 *Folding Des.* **3** R81–91
- [233] Makarov D E and Plaxco K W 2003 *Protein Sci.* **12** 17–26
- [234] Jewett A I, Pande V S and Plotkin S S 2003 *J. Mol. Biol.* **326** 247–53
- [235] Millett I S, Doniach S and Plaxco K W 2002 *Adv. Protein Chem.* **62** 241–62
- [236] Royer C A 2002 *BBA-Protein Struct. M.* **1595** 201–9
- [237] Mayor U *et al* 2003 *Nature* **421** 863–7
- [238] Fang X *et al* 2000 *Biochemistry* **39** 11107–13
- [239] Treiber D K and Williamson J R 2001 *Curr. Opin. Struct. Biol.* **11** 309–14
- [240] Russell R *et al* 2002 *Proc. Natl Acad. Sci. USA* **99** 4266–71
- [241] Pusey P N 2002 *Neutrons, X-Rays and Light: Scattering Methods Applied to Soft Condensed Matter* ed P Lindner and T Zemb (Amsterdam: North-Holland Delta Series)
- [242] Dierker S B *et al* 1995 *Phys. Rev. Lett.* **75** 449–52
- [243] Sandy A R *et al* 1999 *J. Synchrotron Radiat.* **6** 1174–84
- [244] Livet F *et al* 1998 *J. Synchrotron Radiat.* **5** 1337–45
- [245] Abernathy D L *et al* 1998 *J. Synchrotron Radiat.* **5** 37–47
- [246] Suzuki Y, Momose A and Sugiyama H 1998 *J. Synchrotron Radiat.* **5** 596–9
- [247] Lumma D *et al* 2000 *Phys. Rev. E* **62** 8258–69
- [248] Hirai M *et al* 1997 *Phys. Rev. B* **55** 3490–6
- [249] Leforestier A, Fudaley S and Livolant F 1999 *J. Mol. Biol.* **290** 481–94
- [250] Doniach S 2000 *J. Synchrotron Radiat.* **7** 116–20
- [251] Neutze R *et al* 2002 *Nature* **406** 752–7
- [252] Miao J W, Hodgson K O and Sayre D 2001 *Proc. Natl Acad. Sci. USA* **98** 6641–5
- [253] Miao J W *et al* 2003 *Proc. Natl Acad. Sci. USA* **100** 110–2
- [254] Sinha S K, Tolan M and Gibaud A 1998 *Phys. Rev. B* **57** 2740–58
- [255] Vartanyants I A and Robinson I K 2001 *J. Phys: Condens. Matter* **13** 10593–611
- [256] Kornberg M A *et al* 2002 *J. Synchrotron Radiat.* **9** 298–303
- [257] Bentson L *et al* 2002 *SLAC-PUB-9530* Stanford Linear Accelerator Center, Stanford University
- [258] Wabnitz H *et al* 2002 *Nature* **420** 482–5
- [259] Thuman-Commike P A *et al* 1999 *Biophys. J.* **76** 2249–61
- [260] Gabashvili I S *et al* 2000 *Cell* **100** 537–49
- [261] Saad A *et al* 2001 *J. Struct. Biol.* **133** 32–42

- [262] Zheng W and Doniach S 2002 *J. Mol. Biol.* **316** 173–87
- [263] Hao Q *et al* 1999 *Acta Crystallogr. D: Biol. Crystallogr.* **55** 243–6
- [264] Hao Q 2001 *Acta Crystallogr. D: Biol. Crystallogr.* **57** 1410–4
- [265] Liu Q *et al* 2003 *Acta Crystallogr. D: Biol. Crystallogr.* **59** 1016–9
- [266] Martin C S *et al* 2001 *Structure (Camb)* **9** 917–30