Quarterly Reviews of Biophysics **36**, 2 (2003), pp. 147–227. © 2003 Cambridge University Press
DOI: 10.1017/S0033583503003871 Printed in the United Kingdom

Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution

Michel H. J. Koch[1,*], Patrice Vachette[2] and Dmitri I. Svergun[1,3]

[1] European Molecular Biology Laboratory – Hamburg Outstation, EMBL c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany
[2] LURE, Université Paris-Sud BP34, F-91898, Orsay Cédex, France
[3] Institute of Crystallography, Russian Academy of Sciences, Leninsky Prospect 59, 117333 Moscow, Russia

Abstract. A self-contained presentation of the main concepts and methods for interpretation of X-ray and neutron-scattering patterns of biological macromolecules in solution, including a reminder of the basics of X-ray and neutron scattering and a brief overview of relevant aspects of modern instrumentation, is given. For monodisperse solutions the experimental data yield the scattering intensity of the macromolecules, which depends on the contrast between the solvent and the particles as well as on their shape and internal scattering density fluctuations, and the structure factor, which is related to the interactions between macromolecules. After a brief analysis of the information content of the scattering intensity, the two main approaches for modelling the shape and/or structure of macromolecules and the global minimization schemes used in the calculations are presented. The first approach is based, in its more advanced version, on the spherical harmonics approximation and relies on few parameters, whereas the second one uses bead models with thousands of parameters. Extensions of bead modelling can be used to model domain structure and missing parts in high-resolution structures. Methods for computing the scattering patterns from atomic models including the contribution of the hydration shell are discussed and examples are given, which also illustrate that significant differences sometimes exist between crystal and solution structures. These differences are in some cases explainable in terms of rigid-body motions of parts of the structures. Results of two extensive studies – on ribosomes and on the allosteric protein aspartate transcarbamoylase – illustrate the application of the various methods. The unique bridge between equilibrium structures and thermodynamic or kinetic aspects provided by scattering techniques is illustrated by modelling of intermolecular interactions, including crystallization, based on an analysis of the structure factor and recent time-resolved work on assembly and protein folding.

* Correspondence may be addressed to any one of the authors:
Michel Koch, European Molecular Biology Laboratory – Hamburg Outstation, EMBL c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany.
Tel.: +49-40-89902-113; Fax: +49-40-89902-149; E-mail: koch@embl-hamburg.de
Patrice Vachette, LURE, Université Paris-Sud BP34, F-91898, Orsay Cédex, France.
Tel.: +33-1-64468855; Fax: +33-1-64464148; E-mail: vachette@lure.u-psud.fr
Dmitri Svergun, European Molecular Biology Laboratory – Hamburg Outstation, EMBL c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany.
Tel.: +49-40-89902-125; Fax: +49-40-89902-149; E-mail: svergun@embl-hamburg.de

1. Introduction

As pointed out by Guinier and Fournet in the introduction to their book (Guinier & Fournet, 1955) the delay between the development of crystallography starting around 1915 and that of small-angle scattering was largely due to the greater experimental difficulties and in particular to the lack of sufficiently brilliant sources. Small-angle scattering is thus one of the techniques that have gained most from the availability of neutron and synchrotron radiation sources. First X-ray applications date back to the late 1930s when it had become clear that light scattering could be

used to determine the molecular mass of proteins (Putzeys & Brosteaux, 1935) thereby also establishing a link between scattering and hydrodynamic methods.

For a long time interpretation of scattering patterns from solutions of biological samples was limited to modelling using simple bodies (e.g. ellipsoids, cylinders, etc.). Further progress was made by the introduction of spherical harmonics for data interpretation (Harrison, 1969; Stuhrmann, 1970a) and by the use of neutron contrast variation by solvent exchange (H_2O/D_2O) (Ibel & Stuhrmann, 1975) or specific deuteration (Engelman & Moore, 1972). Despite the systematic 'no crystals, no funds' policy of the last decades there has been significant progress in the field. Instruments for synchrotron radiation and position-sensitive gas proportional detectors with fast readout resulted in significant improvements in the precision of the data and in the speed of data collection. This led to progress along two avenues: strictly structural studies on equilibrium systems and time-resolved studies on macromolecular assembly and folding. In the study of equilibrium systems, full advantage of improved data-collection facilities could only be taken during the last decade when the generalized availability of computing power also allowed the development of efficient *ab initio* data interpretation methods based on spherical harmonics, global minimization algorithms and rigid-body refinement. During the same period, time-resolved solution-scattering studies contributed very significantly to the understanding of macromolecular folding processes.

After a brief reminder of the basics of scattering the present review focuses on the interpretation of X-ray and neutron-scattering patterns and applications to biological macromolecules in solution. Studies on solutions provide, besides those on fibres, membranes and gels, useful insights into the structure of non-crystalline biochemical systems. For a minimal investment in time and effort, they also provide perhaps the only direct link between high-resolution structural information obtained by protein crystallography or NMR or lower resolution imaging using various forms of microscopy, and the observations of hydrodynamic methods, light scattering and spectroscopy. Moreover, they provide unique possibilities of investigating intermolecular interactions including assembly and large-scale conformational or structural changes on which biological function often relies.

2. Basics of X-ray and neutron scattering

2.1 Elastic scattering of electromagnetic radiation by a single electron

Charged particles, like electrons, emit radiation when they are accelerated and, if the cause of acceleration is an electromagnetic wave, this emission is referred to as scattering, because the direction of the initial wave apparently changes. The geometry of the scattering problem for a single electron is illustrated in Fig. 1. For an observer located at **r**, elastic scattering by a single electron placed at the origin of a cartesian coordinate system and exposed to a monochromatic plane wave linearly polarized along X with electric field $E(t) = E_0 e^{i\omega t}$, will result in a wave of the same frequency but with an electric field $E(\mathbf{r}, t)$ given by Eq. (2.1). This corresponds to the classical doughnut representation of dipolar radiation, where Ψ is the angle between the direction of polarization and the observer's line of sight.

$$E(\mathbf{r}, t) = \frac{e^2}{mc^2} \frac{\sin \Psi}{r} \frac{\omega^2}{\omega_0^2 - \omega^2} E(t) \tag{2.1}$$

The electric field of the scattered wave is linearly related to that of the incident wave by three factors: (i) a constant, the classical electron radius, $r_0 = e^2/mc^2 = 2{\cdot}82 \times 10^{-15}$ m, where e is the

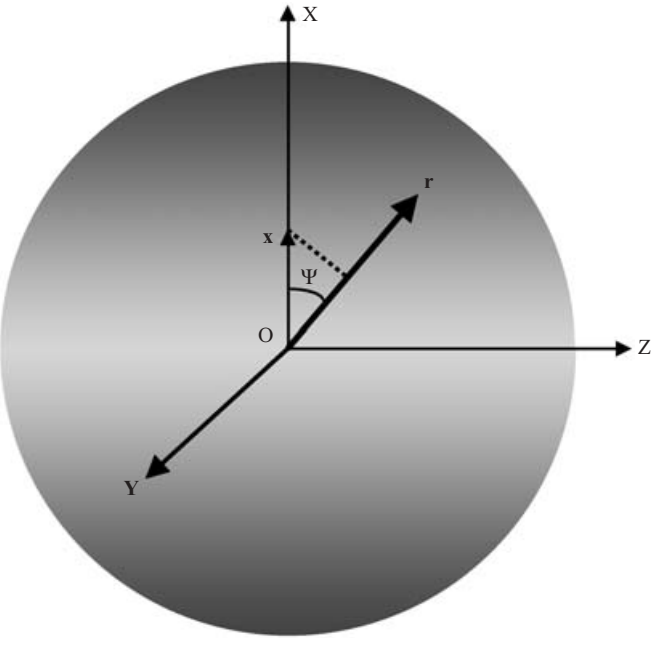

Fig. 1. Geometry of the scattering problem for an electron placed at the origin (O) of a cartesian coordinate system for an observer located at **r**. Acceleration of the electron by a monochromatic plane wave linearly polarized along X leads to scattering. The scattering intensity in any direction is proportional to $\sin^2 \Psi$ and is represented here by the grey level on a sphere. The angle Ψ defines the projection of the dipole $|e|\mathbf{x}$ on the perpendicular to the line of sight of the observer in that direction. Note that there is no scattering in the direction of acceleration of the electron.

charge of the electron, m its mass and c the velocity of light in vacuum, (ii) a geometrical factor corresponding to the projection of the electric field of the incident wave on the perpendicular to the observer's line of sight, (iii) a frequency factor, where ω_0 is the natural frequency of the undamped oscillator representing the motion of the electron. The natural frequency ω_0 corresponds to the binding strength of the electrons, which in the case of electrons in atoms lies between the ultraviolet and X-ray region of the spectrum. Consequently, if the incident radiation is visible light ($\lambda \approx 500$ nm), $\omega \ll \omega_0$ and the frequency factor reduces to ω^2/ω_0^2. The amplitude of the scattered radiation at **r** is proportional to ω^2 and in phase with the incident radiation. This is Rayleigh scattering, which is responsible for the blue sky. Light scattering (elastic or quasi-elastic) is indispensable for monitoring assembly phenomena or intermolecular interactions in solution but the wavelengths are too long to obtain much structural information, except for large structures like viruses. Small-angle laser light scattering is a useful structural technique in the study of semi-crystalline polymers where one is dealing with entities like spherulites, which have dimensions in the μm-range (see e.g. Goderis *et al.* 2002).

For X-rays, $\lambda \approx 0{\cdot}1$ nm, $\omega_0 \ll \omega$ and the frequency factor is equal to -1. The amplitude of the scattered wave thus becomes independent of the frequency and its phase is shifted by π relative to the incident radiation. This is Thomson scattering. The world of X-rays is thus colourless with only shades of grey (contrast) and Eq. (2.1) simplifies to:

$$E(\mathbf{r}, t) = -r_0 \frac{\sin \Psi}{r} E(t). \tag{2.2}$$

What the observer detects is, of course, not the amplitude but the time-averaged energy flux or intensity [i.e. energy/(unit area × unit time)] which is given by the real part of the complex Poynting vector $\mathbf{S}_p = \varepsilon_0 c^2(\mathbf{E} \times \mathbf{B})/2$. Since $|\mathbf{B}| = |\mathbf{E}|/c$ and $\mathbf{E} \cdot \mathbf{B} = 0$ (i.e. $\mathbf{E}$, the electric component of the field is perpendicular to its magnetic component $\mathbf{B}$) it follows that the time-averaged intensity is $\langle I \rangle = \langle S_p \rangle = (\varepsilon_0/\mu_0)^{\frac{1}{2}} E_0^2$, where ε_0 and μ_0 are the permittivity and permeability of vacuum, respectively (Jackson, 1998a).

When the geometrical factor for unpolarized light is expressed in terms of the scattering angle (2θ) (i.e. the angle between the wave vectors of the incident and scattered waves) the intensity of Thompson scattering for an incident beam intensity I_0 becomes:

$$I(2\vartheta) = r_0^2 \left(\frac{1 + \cos^2(2\theta)}{2} \right) \frac{1}{r^2} I_0. \tag{2.3}$$

Clearly, if the scattering angle is small ($< 5^\circ$) the polarization factor is close to 1 and can be neglected. The product of r_0^2 and the polarization factor has dimensions of an area and is called the differential scattering cross section $\mathrm{d}\sigma/\mathrm{d}\Omega$. It represents the ratio between the energy scattered/unit solid angle and unit time and the incident energy/unit area and unit time. Its value corresponds to $|b|^2$, the square of the modulus of the scattering length. The scattering length of a single electron is thus equal to $b_e(\theta) = r_0((1 + \cos^2(2\theta))/2)^{\frac{1}{2}}$. Since the scattering amplitude or scattering factor f of an object is defined as the ratio between the amplitude of the scattering of the object and that of one electron in identical conditions, the scattering amplitude of a single electron $f_e \equiv 1$. The elastic scattering cross section for X-rays thus depends on the scattering angle whereas the integral cross section for elastic scattering (i.e. the total energy elastically scattered by a single electron) is $\sigma_{el} = 8\pi r_0^2/3$.

2.2 Scattering by assemblies of electrons

To obtain the scattering from any blob of matter it suffices to add the waves originating from each electron in the volume taking into account the phase difference φ which is related to the separation between scatterers (Δ) by $\varphi = \mathbf{s} \cdot \Delta$, where $\mathbf{s}$, the momentum transfer vector ($|\mathbf{s}| = 4\pi \sin\theta/\lambda$), is the difference between the wave vectors of the incident and scattered waves, which is equal to the scattering vector multiplied by 2π. To observe interferences the scattered waves must, of course, be coherent over distances comparable to those of the objects under investigation.

The longitudinal or temporal coherence length of the beam (Λ) corresponds effectively to the distance over which two waves with wavelength λ and $\lambda + \Delta\lambda$ become out of phase. It describes the fact that the effective frequency range of monochromatic radiation is of the order of the reciprocal of the duration of a wave train and is given by $\Lambda = \lambda_0^2/\Delta\lambda$, where λ_0 is the mean wavelength and $\Delta\lambda$ the spectral full-width at half-maximum of the radiation. For a typical monochromator $\Delta\lambda/\lambda = 2 \times 10^{-4}$ and for $\lambda_0 = 0{\cdot}15$ nm, $\Lambda = 0{\cdot}75$ μm. When the difference in path length of the scattered waves becomes of the order of or exceeds Λ, the interference effects vanish.

The transverse 1σ coherence area, which corresponds to the area of the sample that is coherently illuminated by a quasi-monochromatic incoherent source is $S = \lambda_0^2 R^2/4\pi\sigma_x\sigma_y$, where σ_x and σ_y are the horizontal and vertical 1σ source sizes and R the distance between the source and the sample. Note that the transverse coherence length does not depend on the wavelength spread but only on the geometry (source size and distance source object). With microfocus beam

lines, where beams have dimensions of the order of 20 μm, the samples are (partially) coherently illuminated even if the source is not coherent. This may lead to speckle (Dierker *et al.* 1995; Sandy *et al.* 1999). From the transmission geometry of a small-angle experiment it is obvious that the transverse coherence length is the key factor that determines speckle. This must be considered in detail in the design of instruments for third- and fourth-generation X-ray sources. Note that the present designs of free electron lasers give sources with high transverse coherence where, in contrast with conventional lasers, a large longitudinal coherence length can only be achieved with a monochromator. Whether one sums the amplitudes or the intensities depends on the situation. If the separation between scatterers (Δ) can be considered fixed, as in the case of two atoms in a molecule, there is coherent scattering and one adds up the amplitudes with an appropriate phase shift and multiplies the sum by its complex conjugate to obtain the intensity.

If Δ is not fixed, as in the case of atoms in two distant molecules in solution, the scattering is incoherent and the total intensity is obtained by summing the intensities of the individual waves. Inelastic and incoherent X-ray scattering (Compton scattering) is a smaller effect than elastic scattering, which is, however, relatively more pronounced for light elements. It only contributes to the background and does not yield any structural information.

The sum of amplitudes of a distribution of N electrons located at $\mathbf{r}_i$ is the Fourier series of that distribution:

$$F(\mathbf{s}) = \sum_{i=1}^{N} f_e \exp(i\mathbf{s} \cdot \mathbf{r}_i). \tag{2.4}$$

If the scatterers can take all orientations, as in a gas or in a solution, the phase factor is spherically averaged and

$$\langle \exp(i\mathbf{s} \cdot \mathbf{r}) \rangle = \frac{\sin(sr)}{sr}. \tag{2.5}$$

It is often more practical to use a continuous function to represent the scattering density, $\rho(\mathbf{r})$, which for X-ray scattering is equivalent to the electron density since $f_e = 1$, or scattering length density in the case of neutron scattering. A simple application is provided by the expression for the scattering factor of an atom with radial electron density $\rho(r)$ given in Eq. (2.6).

$$f(s) = 4\pi \int \rho(r) r^2 \frac{\sin(sr)}{sr} \mathrm{d}r. \tag{2.6}$$

Clearly $f(0) = Z$, the number of electrons. In most strictly small-angle applications ($2\theta < 5^\circ$) involving only light elements the angular dependence of the scattering factor can be neglected, but when comparing high-angle scattering curves with those calculated from crystallographic models this is no longer the case. For oxygen the scattering factor drops from 8 at $s = 0$ to 7·93 at $s = 3{\cdot}65$ nm^{-1} (i.e. $2\theta = 5^\circ$ and $\lambda = 0{\cdot}15$ nm) and to 7·24 at $s = 12{\cdot}6$ nm^{-1} (0·5 nm resolution).

For a molecule with N spherical atoms the scattering amplitude is given by an expression similar to Eq. (2.4) but where f_e is replaced by the scattering factors of the individual atoms. For randomly oriented molecules this leads to the expression of the average scattered intensity given by Debye's formula (Debye, 1915) which involves the distances $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ between pairs of scatterers:

$$I(s) = \sum_{i=1}^{N} \sum_{j=1}^{N} f_i(s) f_j(s) \frac{\sin(sr_{ij})}{sr_{ij}}. \tag{2.7}$$

The same expression is also useful to calculate scattering patterns from low-resolution models based on assemblies of spheres. The scattering factors are then replaced by those of spheres with appropriate diameters and the s-dependence must be taken into account. Even if more sophisticated modelling tools exist, as described below, this approach retains a great didactic value. (For typical applications see, e.g. Bordas *et al.* 1983; Puigdomenech *et al.* 1989.)

2.3 Anomalous scattering and long wavelengths

When ω is close to ω_0 (i.e close to an absorption edge) the undamped oscillator approximation to Eq. (2.1) obviously breaks down and dissipative terms, due to the rearrangement of other electrons in the atom, can no longer be neglected. This leads to

$$f_{\text{anom}}(s) = f_0(s) + f'(s, \lambda) + if''(s, \lambda) \tag{2.8}$$

anomalous scattering and the scattering factor becomes a complex function.

In the X-ray range mainly the absorption edges of heavier elements ($Z > 25$ for K edges and $Z > 60$ for L edges) are used, whereas in the soft X-ray region one also has access to the sulphur, phosphorus or chlorine and calcium edges. Anomalous scattering is commonly used in the study of metallic systems but the experimental difficulties especially at longer wavelengths have limited its practical use for solutions of biological macromolecules (for a review see Stuhrmann *et al.* 1991). The phenomenon, however, plays a very important role in techniques like MAD (multiple wavelength anomalous diffraction) (Ealick, 2000) or SAD (single wavelength anomalous diffraction) (Dauter *et al.* 2002) in protein crystallography.

2.4 Neutron scattering

The wavelength (λ) of neutrons is linked to their (group) velocity (v) and mass m_{N} by de Broglie's relationship, $\lambda = h/m_{\text{N}}v$, where h is Planck's constant, or $\lambda\,(\text{nm}) = 396{\cdot}6/v\,(\text{m s}^{-1})$. The phenomena leading to neutron scattering by nuclei result from potential scattering and spin interactions and are thus fundamentally different from those described for X-rays. As nuclei are considerably smaller than the wavelength of neutrons there are usually no polarization effects, except in magnetic materials where there is also a non-negligible contribution of scattering by electrons (see Marshall & Lovesey, 1971; Bacon, 1975).

The differential scattering cross section for a nucleus located at R is linked to the probability of transition between a plane wave state Ψ_k describing the incident neutron with wavevector $\mathbf{k}$ to a state $\Psi_{k'}$ describing the scattered neutron with wavevector $\mathbf{k}'$ with the same energy $\hbar^2k^2/4\pi^2m_{\text{N}}$, where h is Planck's constant. This probability is proportional to

$$\int \Psi^*_{k'} \widehat{V} \Psi_k \mathrm{d}r. \tag{2.9}$$

The Fermi pseudo-potential $\widehat{V}(r) = h/m_{\text{N}} b\,\delta(r-R)$, where b is the scattering length of the nucleus and $\delta(r-R)$ is a Dirac delta function, is a mathematical device used to describe isotropic scattering in the first Born approximation, which does not represent the real potential. The neutron-scattering amplitude is usually expressed in terms of scattering length rather than of a dimensionless scattering factor as for X-rays. For an array of identical atoms with scattering length b it is

$$F(\mathbf{s}) = \sum_{i=1}^{N} b \exp(i\mathbf{s} \cdot \mathbf{r}_i). \tag{2.10}$$

Table 1. *X-ray and neutron bound coherent scattering length and bound coherent and incoherent and true absorption cross sections of the most important elements in biological macromolecules. The values of* σ_a *for neutrons were calculated from the inverse wavelength dependence and the value of* μ_t/ρ *corresponds to the total interaction cross section* ($\sigma_{coh}+\sigma_{inc}+\sigma_a$)

				Neutron				
	$Z=f(0)$	b_x (fm) 2·8 $f(0)$	μ_t/ρ (cm^2 g^{-1}) (CuKα) (X-rays)	b_{coh} (fm)	σ_{coh} (10^{-24} cm^2)	σ_{inc} (10^{-24} cm^2)	σ_a (10^{-24} cm^2) (0·15 nm)	μ_t/ρ (cm^2 g^{-1}) (0·15 nm)
^{1}H	1	2·8	0·39	−3·739	1·757	80·3	0·27	49·3
^{2}H	1		0·39	6·671	5·592	2·05	0·0	2·3
C	6	16·9	4·51	6·646	5·55	$<10^{-3}$	$<10^{-2}$	0·28
N	7	19·7	7·44	9·36	11·0	0·5	1·58	0·56
O	8	22·5	11·5	5·803	4·232	$<10^{-3}$	$<10^{-3}$	0·16
Na	11	30·8	29·7	3·580	1·6	1·7	0·42	0·097
P	15	42·3	75·5	5·13	3·307	$<10^{-2}$	0·14	0·067
S	16	45·0	93·3	2·847	1·02	$<10^{-2}$	0·44	0·027
Cl	17	47·6	106·0	9·577	11·53	5·3	27·9	0·757
K	19	53·2	145·0	3·67	1·69	0·27	1·75	0·057

In contrast to the situation for X-rays, the neutron-scattering processes at the different centres are independent and different isotopes can have very different scattering lengths. Moreover, a neutron with spin $\frac{1}{2}$ can interact with a nucleus of spin I to give compound nuclei with spin $I+\frac{1}{2}$ or $I-\frac{1}{2}$, which have different scattering lengths b_+ and b_- respectively. Since a spin state J has $2J+1$ possible orientations, the two possible spin states of the compound nucleus will have $2I+2$ and $2I$ possible orientations respectively corresponding to fractions $w_+=(I+1)/(2I+1)$ and $w_-=I/(2I+1)$.

To take all these combinations into account, it is thus useful to describe the scattering length as the sum of two terms, one representing the average scattering length of all isotopes of that element in the sample ($\bar{b}$) and the second one representing the deviation from this average at a particular position ($b_p-\bar{b}$).

$$b_p=\bar{b}+(b_p-\bar{b}). \tag{2.11}$$

When this expression is introduced in Eq. (2.10) two terms are obtained. The first one corresponding to the average scattering length, gives rise to coherent scattering, whereas the second one corresponding to the fluctuations of the scattering length only gives rise to isotropic disorder or incoherent scattering, which is proportional to $\overline{b^2}-\bar{b}^2$. As in the case of X-rays, this incoherent scattering does not contain any structural information.

The differential coherent scattering cross section is $d\sigma_{coh}/d\Omega=|b_{coh}|^2$. A positive scattering length corresponds to a phase shift of π and a negative one (e.g. ^{1}H) to scattering without phase shift. To compare neutron b-values with the corresponding values for X-rays it must be taken into account that the X-ray scattering lengths are multiples of the classical electron radius r_0, so that $b_x=r_0 f(0)$. As polarization effects are usually negligible, the total (integral) coherent scattering cross section (σ_{coh}) defined as the ratio of the outgoing current of coherently scattered neutrons to the incident flux of neutrons is simply given by $\sigma_{coh}=4\pi|b_{coh}|^2$.

Table 1 gives the scattering lengths and cross sections expressed in barns atom^{-1} (1 barn = 10^{-24} cm^2 = 100 fm^2) of the most important elements in biological macromolecules.

The values have been adapted from those given in the International Tables for X-ray Crystallography (Prins & Wilson, 1999) to facilitate comparison between X-rays and neutrons.

Unlike the case of X-rays where the value of the coherent scattering length increases linearly with the atomic number there is no simple relationship between the composition of the nuclei and their scattering length, but in general the coherent neutron scattering lengths tend to be smaller than the corresponding values for X-rays. The most important difference is that between the coherent scattering lengths of the hydrogen isotopes, which also provides the basis for the two main applications of neutron scattering in the study of solutions of biological macromolecules: contrast variation and specific labelling. ^{1}H also gives rise to a high isotropic incoherent neutron scattering background due to spin incoherence. This effect is used for absolute scaling by reference to the background of pure water or to determine the proton concentration in the sample by transmission measurements.

The difference between the scattering length of ^{1}H in the spin-up and spin-down states, which causes the incoherent scattering from unpolarized samples, is the basis of the promising spin contrast variation method utilizing polarized neutron beams and polarized targets (Stuhrmann *et al.* 1986).

2.5 Transmission and attenuation

Attenuation originates from different phenomena and the total cross section for interaction, σ, which must be considered, is the sum of the cross sections for true absorption (σ_a), coherent (σ_{coh}), and incoherent scattering (σ_{inc}). In the case of X-rays these effects correspond respectively to photoelectric absorption, Rayleigh–Thomson scattering and Compton scattering. For carbon at 10 keV the contributions for the photoelectric, elastic and inelastic contribution to the total scattering cross section amount to approximately 60, 6 and 3 ($\times 10^{-24}$ cm^2 per atom) respectively (Prins & Wilson, 1999). The mass attenuation coefficient ($\mu_m = \mu_t/\rho$) for an element is linked to its total cross section for interaction by $\mu_t/\rho(\text{cm}^2\ \text{g}^{-1}) = \sigma(\text{cm}^2)N_A/A(g)$, where N_A is Avogadro's number and A the atomic mass.

For neutrons σ_a increases linearly with the neutron wavelength whereas the total cross section for scattering $\sigma_s = \sigma_{coh} + \sigma_{inc}$ is generally independent of the wavelength. In the case of X-rays the wavelength dependence of the mass attenuation coefficient is more complex and $\mu_t/\rho \sim C\lambda^3 - D\lambda^4$, where C and D are empirical constants. Photoelectric (true) absorption of X-rays is accompanied by fluorescence whereby an outer shell electron fills an inner shell vacancy with emission of the characteristic radiation of the absorber. This radiative process is important for core levels with energies >8 keV and high atomic numbers. The fluorescence yield (ratio of the emitted X-rays to the number of primary vacancies) monotonically increases with Z and is usually between 5 and 10% for the K lines and 1–5% for the L lines. Fluorescence is a serious problem on the high-energy side of the absorption edge in anomalous scattering and in applications involving labelling with heavy metal atoms or clusters (e.g. Au), which requires some degree of energy discrimination in the detection system.

Transmission depends mainly on the composition of the sample (for solutions of biological macromolecules mainly that of the buffer) and on the choice of wavelength. The linear attenuation coefficient of a mixture is given by the sum of the contribution of all elements $\mu_t = \rho\Sigma g_i(\mu_m)_i$, where the g_i are the mass fractions of the elements. It determines the optimal thickness ($t_{opt} = 1/\mu_t$) of the samples, corresponding to $1/e$ transmission, for a given wavelength. Thicknesses above the optimal one should be avoided. From the numbers in Table 1 it is easily

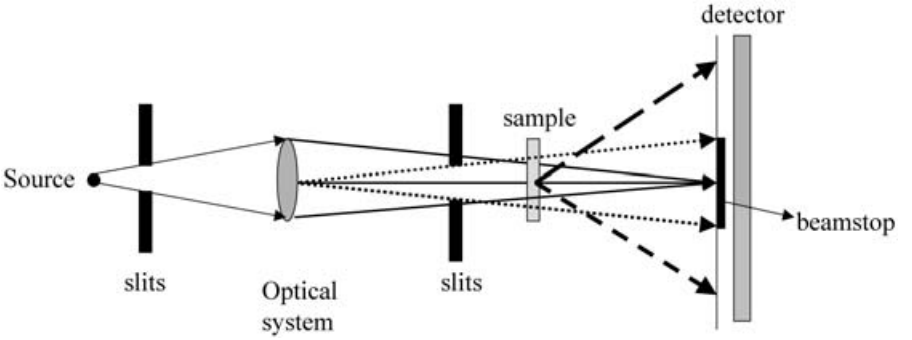

Fig. 2. Schematic representation of a small-angle scattering instrument. The optical system (monochromator/mirror for X-rays, velocity selector for neutrons) selects the appropriate wavelength range from the spectrum of the source and focuses the beam on the detector plane. The lowest angle of observation is determined not only by the focus size but also by the dimensions of the last aperture, which defines a region of high background (due e.g. to the optical system and slits). The measurements are done in transmission with samples that are about 1 mm thick. The sample-detector distances usually vary between 1 and 10 m.

calculated that the optimal thickness of water for 8·3 keV X-rays is close to 1 mm. For many applications this gives sample volumes of the order of 30–100 μl, which are still easy to handle while not requiring too much material.

3. Small-angle scattering from solutions

3.1 Instrumentation

Small-angle scattering measurements are conceptually very simple as schematically illustrated in Fig. 2, but although good engineering is certainly a prerequisite much still depends on mastering the art of reducing the background. Reviews of experimental aspects of the technique and descriptions of the different types of instruments for neutron (Ibel, 1976; Lindner *et al.* 1992; Heenan *et al.* 1997) and X-ray scattering (Koch, 1988; Boesecke & Diat, 1995; Bras & Ryan, 1998; Sandy *et al.* 1999) can be found elsewhere. Modern X-ray beam lines are generally equipped with a tunable fixed exit double monochromator and mirrors for harmonic rejection. The monochromators are usually single crystals with a very narrow bandpass ($\Delta\lambda/\lambda \sim 10^{-4}$), but as the bandpass requirements for small-angle scattering are not very high, multilayers ($\Delta\lambda/\lambda \sim 3\%$) can be used to obtain higher intensities (Tsuruta *et al.* 1998a). Most scattering experiments have hitherto been done with position sensitive gas proportional detectors with delay line readout filled with mixtures of argon or xenon and CO_2 or ethane (for an introduction see Petrascu *et al.* 1998). The most recent readout systems for these detectors are based on time to space conversion (deRaad Iseli *et al.* 2001). Solid-state CCD (charge-coupled devices) detectors are increasingly being used to cope with the high fluxes of third generation sources but special experimental procedures are required to reduce the effect of dark current and other distortions on the accuracy of the data (Pontoni *et al.* 2002).

On neutron instruments a relatively broad spectral band (FWHM $\sim 10\%$) is selected using a mechanical velocity selector. The detectors are also position sensitive gas proportional detectors filled with ^{3}He, 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 of neutron sources. The sample

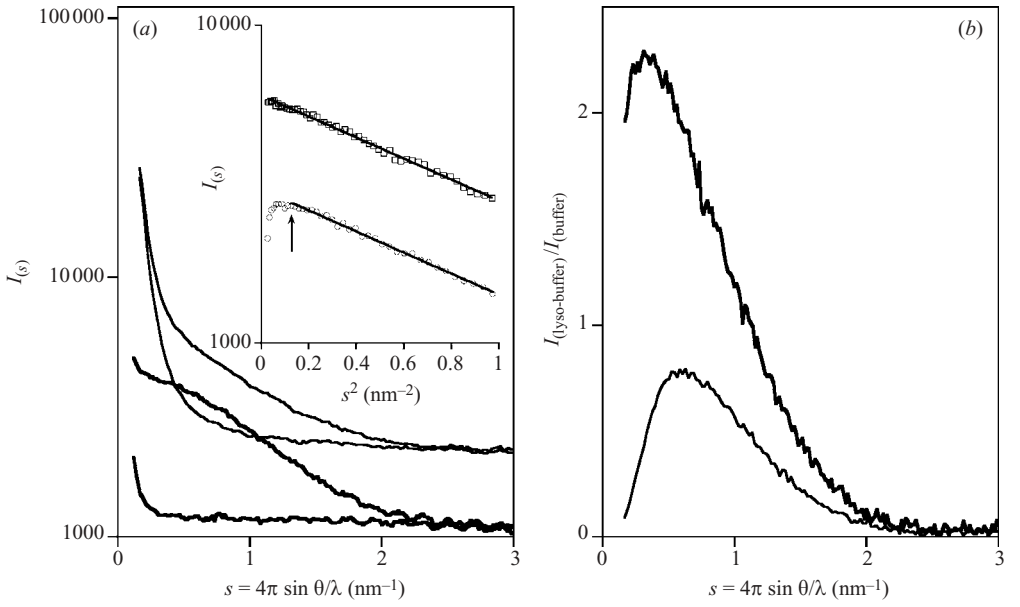

Fig. 3. (*a*) Scattering pattern of a solution of lysozyme (5 mg ml^{-1}) and its buffer recorded using a standard capillary (thin lines) and an evacuated cell (thick lines). The insets give the corresponding Guinier plots: capillary (○); evacuated cell (□). The patterns have been displaced along the ordinate for better visualization. (*b*) Signal-to-background ratio in the scattering pattern of a lysozyme solution (5 mg ml^{-1}) in a standard capillary (thin line) and an evacuated cell (thick line).

containers are usually thermostated cells with mica windows or boron glass capillaries for X-ray work and standard spectroscopic quartz cuvettes for neutron scattering. As illustrated in Fig. 3, the lowest possible background in X-ray cameras is obtained when the entire beam path including the sample container is in vacuum (Dubuisson *et al.* 1997). This is equally valid for neutron scattering.

3.2 The experimental scattering pattern

The elastic scattering of randomly oriented particles in solution and the background due to solvent scattering and other contributions result in an isotropic pattern, which is a function of s. The relevant scattering pattern of the macromolecular solute $I_{\text{exp}}(s)$, is obtained, as indicated below in Eq. (3.1), by subtracting the scattering of the buffer at dialysis equilibrium obtained in a separate measurement (buf) from that of the solution (sol). The factor x_s represents the volume fraction of the solute in the solution.

This procedure also removes any non-isotropic contributions in the background scattering, which may be due, for example, to preferential orientation in the windows. Although ideally one should also subtract the contribution of the empty sample container (*e*) this is rarely done in practice, especially if the same container is used to measure the scattering patterns of the sample and the buffer.

$$I_{\text{exp}}(s) = \frac{1}{c}[(N(n)_{\text{sol}}/I_{0\text{sol}} - (1-x_s)N(n)_{\text{buf}}/I_{0\text{buf}} - x_s N(n)_e/I_{0e})/\text{DR}(n)]. \tag{3.1}$$

$N(n)$ is the number of photons (or neutrons) counted in the nth channel of the pattern. As illustrated in Fig. 3, the subtraction in Eq. (3.1) must be done very accurately to correctly determine the difference between the experimental patterns of the solution and of the solvent, which is usually very small at high angles. On some X-ray and neutron instruments, I_0, the intensity of the transmitted direct beam, is monitored through a semi-transparent beamstop, either simultaneously with the scattering pattern or in a separate measurement. With X-rays one can also use a photodiode in the beamstop. Alternatively the intensity of the direct beam in front of the sample is measured with an ionization chamber and the transmission of the sample is obtained separately by measuring the scattering pattern of a strong scatterer (e.g. glassy carbon) placed downstream from the sample. On some installations such a procedure is part of every sample or buffer measurement and used to normalize the patterns to the intensity of the transmitted beam and put the measurements on an absolute scale. The prerequisite for this approach is, of course, to have a stable incident beam.

A number of different procedures have been described to put the measurements on an absolute scale for X-rays (Russel, 1983; Orthaber *et al.* 2000) or neutrons (Wignall & Bates, 1987) or to cross-scale X-ray and neutron data (Russel *et al.* 1988). For neutrons one uses the incoherent scattering of water as a reference.

The concentration (c in g l^{-1}) of the macromolecular solute must be accurately determined for absolute measurements, as well as for measurements relative to known standards. Experience shows that traditional methods for determining protein concentrations (e.g. the Bradford assay) are often unsatisfactory and that for proteins absorbance measurements at 280 nm are preferable.

The conversion of n to the corresponding s value is generally done using a standard sample with sufficiently large d-spacings like dry collagen ($d = 65$ nm) or organic powders [e.g. Ag behenate ($d = 5{\cdot}838$ nm) or tripalmitin ($d = 4{\cdot}06$ nm)]. The conversion factor can, of course, also be determined from the wavelength and the geometry of the instrument. The procedure may be more attractive to purists but relies on the assumption, which should be independently verified, that the position on the detector is linearly related to n (i.e. that there is no parallax or similar effect).

The value of the weighting factor x_s is equal to 0 if the patterns are to be interpreted in the frame of the formalism of contrast variation. When solutions with high concentrations (> 10 mg ml^{-1}) are measured for comparison with scattering curves of proteins obtained, for example, from crystallographic models at resolutions above 1 nm, $1 - x_s$ corresponds to the volume fraction of the bulk solvent in the solution.

Because inelastic X-ray scattering by light elements in the 8–10 keV range is not entirely negligible at higher angles, one should ideally subtract a background with the same elemental composition as the sample, especially for concentrated systems. In practice, x_s is sometimes used as a fudge factor but if its value has to be higher than 0·02 without good justification it should be taken as an indication that something is amiss with the measurements. It is therefore also preferable to measure the sample and buffer in the same cell.

DR(n) represents the detector response which, for X-rays, is measured by homogeneous irradiation with a radioactive (^{55}Fe) source or using the fluorescence of a thin metal (e.g. iron) foil. For neutrons one uses the incoherent scattering of water. Solutions of macromolecules are rarely ideal and, as explained below, attractive or repulsive interactions lead to distortions of the scattering pattern. For non-ideal (i.e. real) solutions the scattering is given by:

$$I_{\exp}(c, s) = I(s)\,\mathrm{SF}(c, s). \tag{3.2}$$

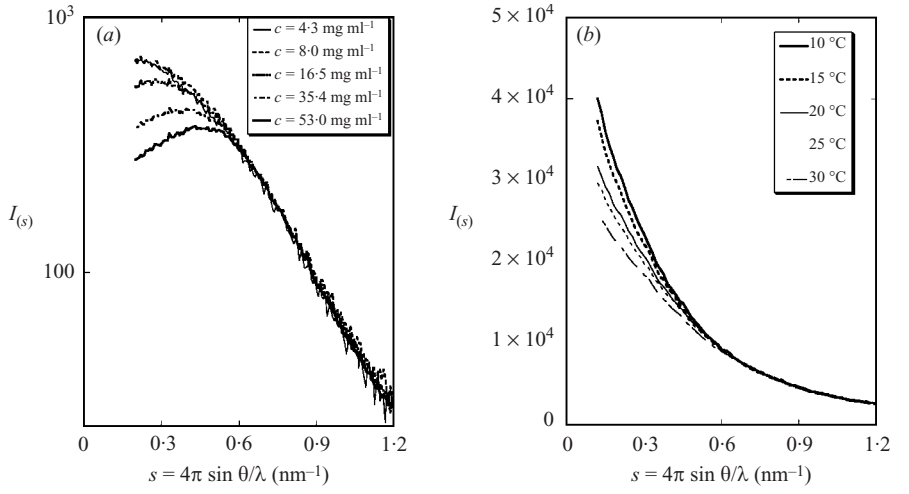

Fig. 4. (*a*) Scattering patterns of bovine serum albumin solutions in 50 mM Hepes pH 7·5, scaled for concentration at 4·3 mg ml^{-1} (top), 8·0, 16·5, 35·4 and 53·0 mg ml^{-1} (bottom). At higher concentrations the effect of the structure factor of the solution due to repulsive interactions becomes very noticeable. (*b*) Scattering patterns of a 160 mg ml^{-1} solution of γ-crystallins in 50 mM phosphate buffer, pH 7·0, as a function of temperature (after Malfois *et al.* 1996). The increase in scattering at low s values reflects the increase in attractive interactions with decreasing temperature.

SF(c, s) is the concentration dependent structure factor of the solution that takes into account the interactions, attractive or repulsive, between solute particles. At infinite dilution SF(0, s) = 1 and it is thus advisable to extrapolate scattering patterns collected at several concentrations (e.g. 3, 5, 7, 10 mg ml^{-1}) to zero concentration to obtain an undistorted pattern in the low angle region. For the high-angle region ($s > 2$ nm^{-1}) concentrated solutions ($c > 10$ mg ml^{-1}) are generally used to obtain a sufficiently high signal-to-background ratio, especially for larger proteins, where the scattering decays rapidly at increasing s values. This is a valid procedure because, as illustrated in Fig. 4 for solutions of bovine serum albumin, the effect of repulsive interactions is, in contrast with that of aggregation, negligible at higher angles. Attractive interactions, which often lead to unspecific aggregation are usually characterized by a steep increase of the scattering curve at low angles. Although visual inspection of the scattering pattern usually gives immediate information about the mono- or polydispersity of the solution this should not be taken for certain. A typical example is that of solutions of α-crystallins, where there are strong repulsive interactions between unspecific aggregates, which can better be detected by other methods (Vérétout *et al.* 1989; Vanhoudt *et al.* 2000).

Experimental solution-scattering patterns can be used independently of any further interpretation to estimate imaging parameters in cryoelectron microscopic reconstructions of large biological objects (Thuman-Commike *et al.* 1999).

3.3 Basic scattering functions

Interpretation of the scattering pattern of a dilute monodisperse solution of macromolecules usually relies, especially for neutron scattering, on the formalism of contrast variation (Stuhrmann & Kirste, 1965). By virtue of the properties of Fourier transforms (FTs), the scattering length density of the buffer, which can be considered to be uniform at low resolution, only

contributes to the forward scattering and can thus not be directly measured. The experimental scattering curve thus results from the excess scattering density of the macromolecules, which is represented at any point **r** inside the solute particles by:

$$\rho(\mathbf{r}) = (\rho_p - \rho_b)\rho_c(\mathbf{r}) + \rho_s(\mathbf{r}) = \bar{\rho}\rho_c(\mathbf{r}) + \rho_s(\mathbf{r}), \tag{3.3}$$

$\rho_c(\mathbf{r})$ represents the shape of the particle and has a value of 0 outside and 1 inside the particle. In neutron scattering a value slightly below 1 is used to take H/D exchange effects into account. $\rho_s(\mathbf{r})$ represents the fluctuation of the scattering density inside the particle around its average value (ρ_p) and ρ_b corresponds to the uniform scattering density of the solvent. The difference between ρ_p and ρ_b is called the contrast. The assumption of uniform scattering density of the solvent is valid in most practical cases but sometimes breaks down even at low angles, for example when the buffers contain fluorinated alcohols, as often used in NMR (Kuprin *et al.* 1995). The volume elements inside the particle scatter coherently and the scattering amplitude is thus the FT of $\rho(\mathbf{r})$ where the contributions of the shape and the internal structure are independent:

$$A(\mathbf{s}) = \int_V \rho(\mathbf{r}) \exp(-i\mathbf{s}\cdot\mathbf{r})\mathrm{d}\mathbf{r} = \bar{\rho}A_c(\mathbf{s}) + A_s(\mathbf{s}). \tag{3.4}$$

As the different solute particles are randomly located and oriented they scatter incoherently and the observed intensity is the (spherical) average of the intensities due to the individual particles in all possible positions and orientations. The coordinates of the momentum transfer vector in spherical coordinates are $\mathbf{s} = (s, \Omega)$ ($\Omega = (\theta, \varphi)$) and, after averaging, only the modulus (s) appears in the general expression for the scattering intensity of a solution in Eq. (3.5).

$$\langle A(\mathbf{s}) \cdot A^*(\mathbf{s})\rangle_\Omega = I(s) = \overline{\rho^2} I_c(s) + \overline{\rho} I_{cs}(s) + I_s(s). \tag{3.5}$$

The two contrast-dependent terms are functions of the shape of the solute and thus only play a role at small angles. In the diffraction patterns of crystals, they correspond to the low order reflections, which contain information on the molecular envelope and solvent level (Bragg & Perutz, 1952) but are usually not recorded. Low-angle crystallography is, however, very useful for the elucidation of large structures like those of viruses (Tsuruta *et al.* 1998b).

The contrast-independent term, which dominates the outer part of the scattering curve ($s > 2$ nm^{-1}), is due to the internal structure and corresponds to the diffraction pattern of crystals. Whereas crystal diffraction can extend to high resolution (0·1 nm) due to the local amplification of the signal by the constructive interferences due to the coherent scattering of the contents of the unit cells, the scattering from macromolecules in solutions becomes undetectable at much lower resolution (0·5–1 nm).

The three terms in Eq. (3.5) can be separated by neutron contrast variation in H_2O/D_2O mixtures. This is particularly useful for systems containing components with very different scattering density (protein, nucleic acids, lipids) as indicated in Table 2. The apparent contrast advantage of deuterated material in H_2O is, however, partly annihilated by the high incoherent background due to protons. Note also that whereas the effects of H/D exchange occurring during H_2O/D_2O contrast variation are small, specific deuteration drastically changes the internal structure of multicomponent particles.

If an H_2O/D_2O contrast variation series is available, the shape scattering function used in the *ab initio* methods below is obtained in a straightforward manner by extrapolation to infinite

Table 2. *Average contrast* ($\times 10^{10}$ *cm*$^{-2}$) *of protonated and deuterated components of biological macromolecular assemblies in water and heavy water*

Substance	X-rays	Protonated in H_2O	Protonated in D_2O	Deuterated in H_2O	Deuterated in D_2O
Proteins	2·5	2·3	–3·2	7·1	1·6
Nucleic acids	6·7	4·5	–1·7	7·0	0·9
Fatty acids	–1·1	0·3	–6·1	6·5	0·5
Carbohydrates	4·5	3·5	–2·5	7·0	1·4

contrast. If not, the scattering curve of proteins in D_2O or of nucleic acids in H_2O are usually good approximations. As illustrated below in the case of the ribosome structure, much more information can be extracted from scattering curves by combining H_2O/D_2O contrast variation and specific deuteration (Svergun & Nierhaus, 2000).

With X-rays the range of contrast variation is much more limited and there are greater experimental difficulties. Organic contrasting agents like sucrose and glycerol increase the viscosity of the solution, whereas inorganic ones like NaI, KI and CsI alter the ionic strength and lead to high absorption and organometallic compounds like Au-thioglucose combine both difficulties and give a high fluorescence. All these contribute to make the technique much less popular than its neutron counterpart. In practice, even in neutron scattering, contrast variation measurements are only rarely performed with the aim of extracting the shape scattering intensity.

For single-component macromolecules with a large molecular mass (>40 kDa), one can usually reasonably approximate the shape scattering by simply subtracting a constant from the experimental scattering curve at sufficiently high contrast. Indeed, at larger angles the scattering of a particle with $\rho_s(\mathbf{r}) \neq 0$ oscillates around a straight line given by Porod's law (Porod, 1951):

$$s^4 I(s) = Bs^4 + A. \tag{3.6}$$

Subtraction of the constant B given by the slope of a plot of $s^4 I(s)$ against s^4 from the experimental intensities thus yields an approximation to the scattering of the corresponding homogeneous body.

3.4 Global structural parameters

3.4.1 Monodisperse systems

Traditionally, the measured intensities in Eq. (3.1) are used to determine the two most popular global structural parameters – intensity at the origin ($I(0)$) and radius of gyration (R_g) – using the Guinier relation in Eq. (3.7). This relation (Guinier, 1939) is based on the fact that $I(s)$ is an even function ($I(s) = I(-s)$), which can be expanded as a series of the even powers of s. By restricting the range of momentum transfer to values such that $R_g^2 s^2 < 1$, it is found that:

$$I(s) = I(0)[1 - \tfrac{1}{3} R_g^2 s^2 + O(s^4)] \cong I(0) \exp(-\tfrac{1}{3} R_g^2 s^2). \tag{3.7}$$

The intercept of the Guinier plot ($\ln(I(s))$ *versus* s^2) gives $I(0)$ and its slope yields the radius of gyration R_g, as illustrated in the inset of Fig. 3*a*.

The value of $I(0)$ obtained after scaling for concentration corresponds to the scattering of a single particle and, since $I_s(0)$ and $I_{cs}(0)$ in Eq. (3.1) are equal to zero, it is proportional to the square of the total excess scattering length in the particle:

$$I(0) = \int_V \int_{V'} \rho(\mathbf{r})\, \rho(\mathbf{r}')\mathrm{d}V_r\, \mathrm{d}V_{r'} = \overline{\rho}^2 V^2. \tag{3.8}$$

If the measurements are made on an absolute scale, $I(0)$ can be directly related to the molecular mass of the solute. For X-rays the relation is given by:

$$M = I(0)\frac{\mu^2}{(1-\rho_0\psi)^2 N_A}. \tag{3.9}$$

For proteins the ratio between the molecular mass and the number of electrons in the particle, μ, has a value close to 1·87. ρ_0 is the average electron density of the solvent (in $e\ \mathrm{nm}^{-3}$), ψ the ratio of the volume of the particle to its number of electrons and N_A is Avogadro's number. An equivalent expression exists for neutron scattering (see Prins & Wilson, 1999).

For X-ray scattering with solutions of biological macromolecules, the simplest is often to measure a fresh solution of a well-characterized protein (e.g. bovine serum albumin) with an accurately determined concentration in the same conditions as the samples as a reference. Procedures for obtaining molecular masses of polymers without absolute intensities have also been proposed (Plestil *et al.* 1991). The radius of gyration is the second moment of the distance distribution of the particle around the centre of its scattering length density distribution. Its value depends on the contrast (Ibel & Stuhrmann, 1975):

$$R^2 = R_c^2 + \frac{\alpha}{\bar{\rho}} - \frac{\beta}{\bar{\rho}^2}, \tag{3.10}$$

$$\alpha = \frac{1}{V}\int \rho_s(\mathbf{r}) r^2 \mathrm{d}^3 r, \quad \beta = \frac{1}{V^2}\iint \rho_s(\mathbf{r}_1)\rho_s(\mathbf{r}_2)\mathbf{r}_1\mathbf{r}_2\, \mathrm{d}\mathbf{r}_1\, \mathrm{d}\mathbf{r}_2,$$

α is the second moment of the internal structure and the non-negative β describes the displacement of the centre of the scattering length density distribution with the contrast. 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.

For rod-like particles the intercept of a plot of $(\ln(sI(s))$ *versus* $s^2)$ is proportional to the mass per unit length and the slope gives the radius of gyration of the cross section R_c, defined in Eq. (3.11), whereas for lamellar particles the plot of $(\ln(s^2 I(s))$ *versus* $s^2)$ gives the intensity of thickness and the radius of gyration of the thickness, R_t (see e.g. Glatter & Kratky, 1982; Feigin & Svergun, 1987).

$$sI(s) = I_c(0)\exp\left(-\tfrac{1}{2}s^2R_c^2\right) \quad \text{and} \quad s^2I(s) = I_t(0)\exp\left(-s^2R_t^2\right). \tag{3.11}$$

The separation (Δ) between the centres of the scattering length distributions of two components with radii of gyration R_1 and R_2 and representing a fraction w_1 and $1-w_1$ of the total scattering length of a complex with radius of gyration R_{gc}, can easily be estimated using the parallel axis theorem:

$$R_{gc}^2 = w_1R_1^2 + (1-w_1)R_2^2 + w_1(1-w_1)\Delta^2. \tag{3.12}$$

This approach was used to estimate the separation between components in two-component systems like ribosomes.

In neutron scattering the distances between pairs of deuterated subunits in large complexes can be determined in this way. If all distances between pairs of subunits are known, a three-dimensional (3D) low-resolution model of the complex can be built by triangulation (Moore & Weinstein, 1979). The difficulties in obtaining accurate values of the radii of gyration of the subunits *in situ* have been discussed (Moore, 1981) and it was shown that the triple isotopic substitution method (Pavlov & Serdyuk, 1987) gives more reliable results than straightforward labelling (Harrison *et al.* 1993). The application of triangulation methods culminated in the complete mapping of the 30S subunit of the ribosome of *Escherichia coli* (Capel *et al.* 1987).

For unstructured chains like nucleic acids or the unfolded state of proteins, the Guinier approximation is no longer valid as a large number of long distances between scattering elements contributes to the s^4 term in Eq. (3.7), which can therefore no longer be neglected even at small s values. In this case a thick wormlike chain (Kratky & Porod, 1949) with finite contour length (L) and persistence length $a = b/2$, where b is the statistical chain element is a good model. The average squared end-to-end distance of such a chain is $\langle d_{ee}^2 \rangle = bL$, with $L = bn$ (see Kirste & Oberthuer, 1982). Its scattering is described by Eq. (3.13) (Rawiso *et al.* 1987), where the first term corresponds to Debye's formula for an infinitely thin gaussian chain (i.e. a chain with a gaussian distribution of intersegmental distances), with negligible persistence length and excluded volume effects.

$$\frac{I(s)}{I(0)} = \left\{ \frac{2}{x^2}(x - 1 + e^{-x}) + \frac{b}{L}\left[\frac{4}{15} + \frac{7}{15x} - \left(\frac{11}{15} + \frac{7}{15x}\right)\exp(-x)\right]\right\} \exp\left(-\frac{s^2 R_c^2}{2}\right) \tag{3.13}$$

with $x = s^2 Lb/6$.

The first two terms correspond to an infinitely thin persistence chain and the exponential factor is the correction for the finite thickness of the chain, where R_c is the radius of gyration of the cross section. The radius of gyration of the thick persistence length or wormlike chain is given by:

$$R_g^2 = (R_g^{SB})^2 + \tfrac{3}{2}R_c^2, \tag{3.14}$$

where R_g^{SB} is the radius of gyration of the thin wormlike chain (Sharp & Bloomfield, 1968)

$$(R_g^{SB})^2 = b^2\left[\frac{L}{6b} - \frac{1}{4} + \frac{4b}{L} - \frac{b^2}{8L^2}(1 - \exp(-2L/b))\right]. \tag{3.15}$$

Models taking excluded volume effects into account are also available (Pedersen & Schurtenberger, 1996).

3.4.2 Polydisperse systems and mixtures

In practice, the experimental scattering curves often correspond to mixtures, as a result of attractive interparticle interactions, which may be wanted like in titrations or time-resolved experiments, or unwanted (e.g. due to aggregation). In this case, and in absence of other phenomena (e.g. repulsive interactions) than structural modifications, complex formation or oligomerization, the forward scattering and radius of gyration are given by Eqs. (3.16) and (3.17), respectively where n_i represents the number concentration of the ith species.

$$I(0) = \sum_i n_i I_i(0), \tag{3.16}$$

$$I(0) \cdot R_g^2 = \left\{ \sum_i n_i I_i(0) R_{gi}^2 \right\}. \tag{3.17}$$

Scattering patterns from mixtures are obviously less suitable to extract geometrical structural information and the methods used in their interpretation will therefore not be discussed in detail here. Beside the classical approach (see e.g. Feigin & Svergun, 1987; Koenig *et al.* 1992) using distributions of simple bodies like spheres or cylinders, singular value decomposition (SVD) initially introduced in the analysis of small-angle X-ray scattering (SAXS) in the early 1980s (Fowler *et al.* 1983), has become popular in the analysis of scattering curves of mixtures arising in titrations and time-resolved experiments (see e.g. Chen *et al.* 1996; Bilgin *et al.* 1998; Perez *et al.* 2001). This method (see e.g. Press *et al.* 1992) gives the minimum number of eigenvectors which can be linearly combined to account for the complete set of individual experimental patterns. This number may of course be smaller than the actual number of components.

Many solution-scattering studies stop after the analysis of the forward scattering and radius of gyration data and very useful information can often be obtained in this way. It is, however, possible to obtain substantially more – and more reliable – information from a scattering curve and with modern computer programs this does not require much additional effort.

3.5 Characteristic functions

Useful information about the particle structure is provided by the characteristic function $\gamma(r)$, which is the real space counterpart of $I(s)$. For a homogeneous particle, $\gamma(r)$ gives the probability of finding a distance between r and $r + \mathrm{d}r$ inside the volume V

$$V\gamma(r) = \frac{1}{2\pi^2} \int_0^\infty s^2 I(s) \frac{\sin(sr)}{sr} \mathrm{d}s \tag{3.18}$$

of the particle.

The function $p(r)$ is the spherically averaged autocorrelation function of the excess scattering density $\langle \rho(\mathbf{r}) * \rho(-\mathbf{r}) \rangle$ and the scattering pattern is related to it by the inverse transformation in Eq. (3.19).

$$p(r) = r^2 V\gamma(r) = \frac{1}{2\pi^2} \int_0^\infty srI(s) \sin(sr) \mathrm{d}s, \tag{3.19}$$

$$I(s) = 4\pi \int_0^\infty p(r) \frac{\sin sr}{sr} \mathrm{d}r. \tag{3.20}$$

For homogeneous particles, like envelope functions, $p(r)$ represents the histogram of distances between pairs of points within the particle. Obviously, its value is uniformly zero when r exceeds $D_{\max}$, the maximum dimension of the particle. As illustrated in Fig. 5, the shape of $p(r)$ gives information about the main features of the shape of the solute particles.

It is clear from Eq. (3.20) that

$$I(0) = 4\pi \int_0^{D_{\max}} p(r) \mathrm{d}r \tag{3.21}$$

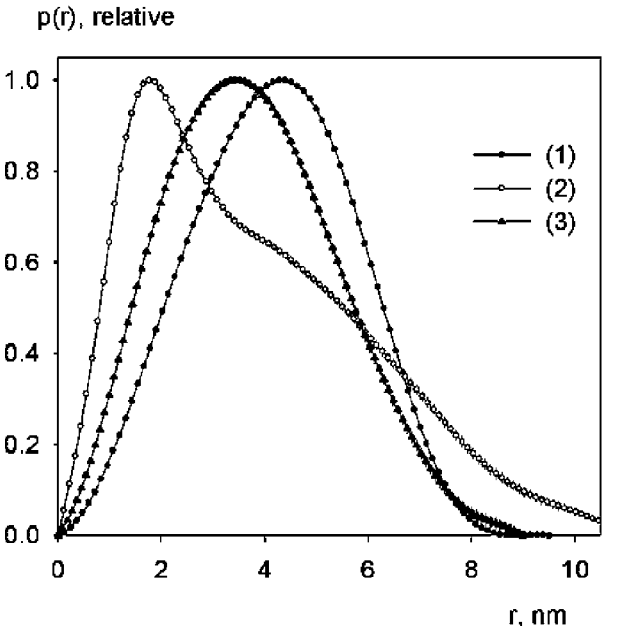

Fig. 5. Distance distribution functions of proteins with different shapes computed from experimental data: (1) hollow globular (urate oxidase); (2) elongated (Z1Z2 domain at the N terminus of titin); (3) flattened (bovine serum albumin).

and that

$$V\gamma(0) = \frac{1}{2\pi^2}\int_0^\infty s^2 I(s)\mathrm{d}s = V\langle\Delta\rho\rangle^2 = \frac{QV}{2\pi^2}. \tag{3.22}$$

The quantity Q is called the Porod invariant (Porod, 1951) because its value, which is equal to the mean square value of the excess scattering density of the sample, is independent of the structure. For a homogeneous particle $\langle\Delta\rho^2\rangle = \overline{\rho}^2$, where $\overline{\rho}$ is the contrast, and the volume of the hydrated solute particle, or Porod volume, is given by the ratio $2\pi^2 I(0)/Q$.

The characteristic function and the distance distribution function could, in principle, be calculated by Fourier transformation of the experimental data but this is not reliable because $I(s)$ is only measured at a finite number of points (s_i) in the interval [$s_{\min}$, $s_{\max}$] rather than [0, ∞]. The precision of these measurements is determined by the corresponding statistical errors (σ_i) but there are also always some systematic errors.

It is thus preferable to compute $p(r)$ indirectly by inverse transformation, over the interval [0, $D_{\max}$]:

$$I(s) = \int_0^{D_{\max}} p(r)\frac{\sin(sr)}{sr}\mathrm{d}r. \tag{3.23}$$

For this purpose, $p(r)$ is represented by a linear combination of orthogonal functions $\varphi_k(r)$ on the interval [0, $D_{\max}$]:

$$p(r) = \sum_{k=1}^{K} c_k\varphi_k(r). \tag{3.24}$$

The *a priori* estimate of $D_{\max}$, which is usually available, can be refined at a later stage by iterative calculations of $p(r)$ with different values of $D_{\max}$.

The coefficients c_k in Eq. (3.24) are determined by fitting the experimental data and minimizing the functional

$$\Phi_\alpha = \sum_{i=1}^{N} \left[\frac{I_{\exp}(s_i) - \sum_{k=1}^{K} c_k \psi_k(s_i)}{\sigma(s_i)} \right]^2 + \alpha \int_0^{D_{\max}} \left[\frac{\mathrm{d}p}{\mathrm{d}r} \right]^2 \mathrm{d}r, \tag{3.25}$$

where $\psi_k(s)$ are the Fourier transformed and smeared functions $\varphi_k(r)$. The regularizing multiplier $\alpha \geqslant 0$ is used to balance between goodness of fit to the data (first summand) and the smoothness of the $p(r)$ function (second summand). This so-called indirect transform method (Glatter, 1977) imposes strong constraints of boundedness and smoothness on $p(r)$ and is preferable to other techniques.

The main problem when using the indirect transform technique is to select the proper value of the regularizing multiplier α. With too small values the solutions are unstable to experimental errors, whereas with too large values the solutions display systematic deviations from the experimental data. In the GNOM program (Semenyuk & Svergun, 1991; Svergun, 1992) a set of perceptual criteria describes the quality of the solution to guide the user in the choice of α. The program either finds the optimal solution automatically or detects dubious assumptions about the system (e.g. the value of $D_{\max}$). The calculated $p(r)$ is used to obtain $I(0)$ from Eq. (3.21) and the radius of gyration of the particle R_g from Eq. (3.26).

$$R_g^2 = \frac{\int r^2 p(r) \mathrm{d}r}{2 \int p(r) \mathrm{d}r}. \tag{3.26}$$

As the entire scattering curve is used in the evaluation of these two parameters, rather than a limited angular range at low angles, the values are more reliable and less sensitive to residual interaction effects or low levels of aggregation than those obtained from the Guinier approximation.

The use of Eq. (3.7) requires a sufficient number of experimental points in the interval $s_{\min} < s < 1/R_g$, with $s_{\min} < \pi/D_{\max}$, whereas indirect transform methods allow to reliably compute the $p(r)$ functions under much less demanding conditions. Once the $p(r)$ function is available, everything is in place for the transmutation of the rather uninspiring 1D solution-scattering data into 3D models, as explained in the following section.

4. Modelling

4.1 Spherical harmonics

The formalism of Fourier series and transformations introduced in Eqs. (2.4) and (3.4) is ideally suited for single crystal structure analysis, where diffraction gives rise to discrete reflections with intensities $I(\mathbf{s}_{hkl})$ along specific directions in 3D reciprocal space. The spherical averaging which takes place in solution scattering makes it more convenient to use the mathematical apparatus of spherical harmonics (Harrison, 1969; Stuhrmann, 1970a, b). These form a complete set of orthogonal angular functions $Y_{lm}(\Omega)$ on the surface of the unit sphere. The scattering

density is expressed as

$$\rho(\mathbf{r}) \approx \rho_L(\mathbf{r}) = \sum_{l=0}^{L} \sum_{m=-l}^{l} \rho_{lm}(r) Y_{lm}(\omega), \tag{4.1}$$

where $(r, \omega) = (r, \theta, \varphi)$ are spherical coordinates and the

$$\rho_{lm}(r) = \int_{\omega} \rho(\mathbf{r}) Y^{*}_{lm}(\omega) \mathrm{d}\omega \tag{4.2}$$

are radial functions. The truncation value L defines the accuracy of the expansion (due to completeness of the spherical harmonics, $\rho_L(\mathbf{r}) \rightarrow \rho(\mathbf{r})$ when $L \rightarrow \infty$). The amplitudes can similarly be represented in reciprocal space,

$$A(\mathbf{s}) = \sum_{l=0}^{L} \sum_{m=-l}^{l} A_{lm}(s) Y_{lm}(\Omega). \tag{4.3}$$

The partial amplitudes ($A_{lm}(s)$) are related to the radial functions by the Hankel transformation

$$A_{lm}(s) = i^{l} \sqrt{\frac{2}{\pi}} \int_{0}^{\infty} j_l(sr) \rho_{lm}(r) r^2 \mathrm{d}r, \tag{4.4}$$

where the $j_l(sr)$ are spherical Bessel functions (Stuhrmann, 1970b). Substituting Eq. (4.3) into Eq. (3.5), all cross-terms in the average vanish due to the orthogonality of the spherical harmonics yielding a simple expression for the intensity:

$$I(s) = \sum_{l=0}^{L} I_l(s) = 2\pi^2 \sum_{l=0}^{L} \sum_{m=-l}^{l} |A_{lm}(s)|^2. \tag{4.5}$$

The scattering intensity of a particle is thus a sum of independent contributions from the substructures corresponding to different spherical harmonics $Y_{lm}(\omega)$. This property of the multipole expansion allows not only to rapidly compute scattering patterns from known structures using Eqs. (4.1)–(4.5) but also to meaningfully approach the inverse problem (i.e. that of getting information about the structure from a solution-scattering pattern).

The spherical harmonics are combinations of trigonometric functions of orders l and m, where the lower order harmonics define the gross structural features of the particle and the higher harmonics describe finer details. The correlation between the truncation value L and the accuracy of the structure representation can be illustrated by considering the evolution as a function of L of the angular envelope ($F(\omega)$) describing the shape $\rho(\mathbf{r})$ of the particle such that

$$\rho(\mathbf{r}) = \begin{cases} 1, & 0 \leqslant r < F(\omega) \\ 0, & r \geqslant F(\omega) \end{cases} \tag{4.6}$$

$$F(\omega) = \sum_{l=0}^{L} \sum_{m=-l}^{l} f_{lm} Y_{lm}(\omega) \tag{4.7}$$

with multipole coefficients

$$f_{lm} = \int_{\omega} F(\omega) Y^{*}_{lm}(\omega) \mathrm{d}\omega. \tag{4.8}$$

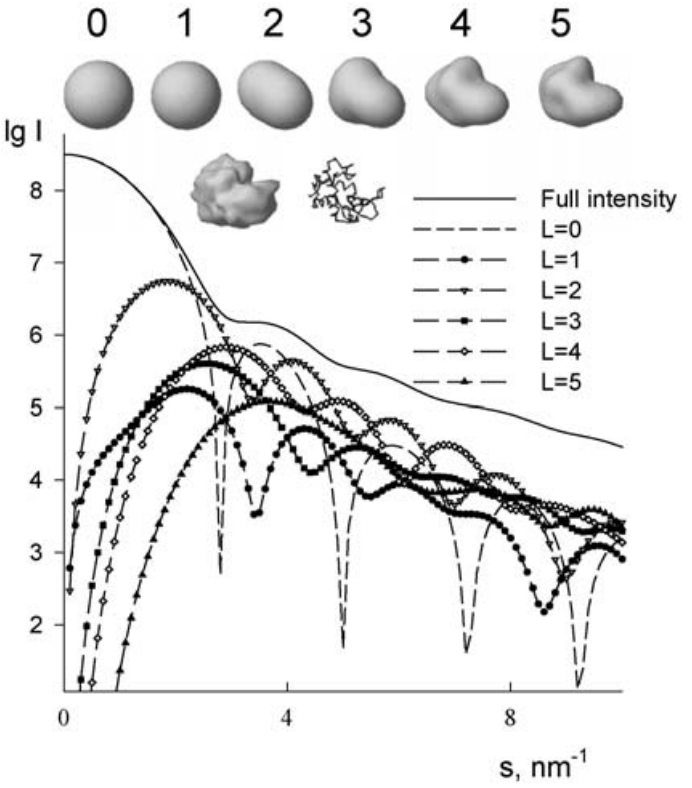

Fig. 6. Accuracy of shape representation using spherical harmonics. Top row: surface representations of truncated envelope functions of lysozyme. Second row: high-resolution envelope function and C_α trace of the protein. The shape scattering intensity from lysozyme is shown along with the contributions from different multipoles.

These equations describe a homogeneous particle with unit density inside the envelope $F(\omega)$ and zero outside as defined in Eq. (3.3). As an example, the envelope functions of lysozyme computed for different values of L are shown in Fig. 6 along with their partial intensities $I_l(s)$. In the monopole approximation ($L=0$), the particle is described by a single parameter, the radius of the equivalent sphere R_0, and $I_0(s)$ is the scattering intensity from this sphere. For $L>0$, all partial intensities are equal to zero at $s=0$ and grow as s^{2l}, so that the contribution of higher harmonics increases with scattering angle. The resolution of the structure description in Eq. (4.7) is approximately $2\pi R_0/(L+1)$, and the number of parameters $(L+1)^2$. Note that when restoring the shape from the scattering data, the number of independent parameters is reduced by 6, as the entire shape can be rotated and displaced without altering the scattering pattern. Thus, a quadrupole approximation ($L=2$) involves three independent parameters and yields a fair description of the particle anisometry. At $L=4$ (19 parameters), the main features of the shape are adequately represented and a further increase of L provides an increasingly detailed description of the envelope at the cost of a quadratically growing number of parameters. This example suggests that, if it were possible to extract, say, 19 independent parameters describing the particle shape from the scattering data, one could construct models described by low-resolution envelopes *ab initio*.

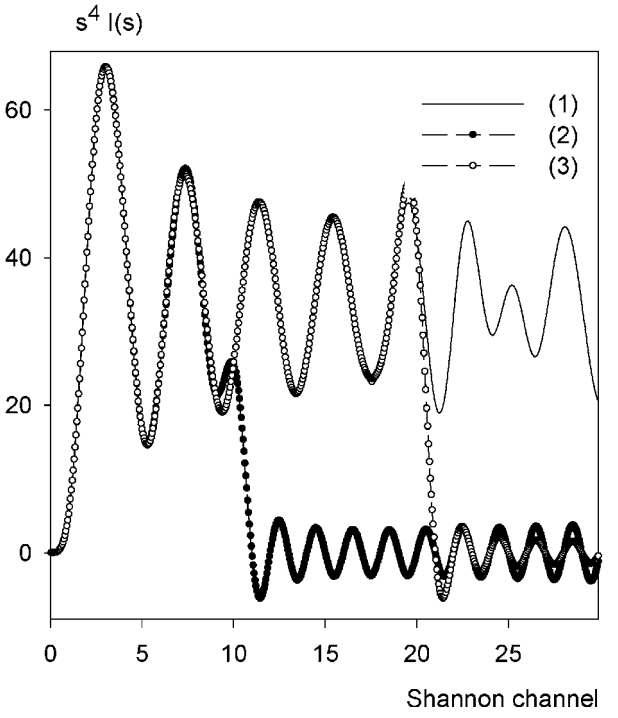

Fig. 7. Shannon approximation of the scattering intensity from an ellipsoidal particle (1) using ten, (2) twenty, and (3) Shannon channels.

4.2 Shannon sampling

The information content of solution-scattering data is usually estimated on the basis of Shannon's sampling theorem (Shannon & Weaver, 1949). As indicated in Eq. (3.20), the scattering curve $I(s)$ is the Fourier image of the characteristic function, which is uniformly equal to zero for intraparticle distances exceeding $D_{\max}$. After double integration by parts Eq. (3.20) yields:

$$I(s) = \sum_{k=1}^{\infty} s_k I(s_k) \left[\frac{\sin\left(D_{\max}(s - s_k)\right)}{D_{\max}(s - s_k)} - \frac{\sin\left(D_{\max}(s + s_k)\right)}{D_{\max}(s + s_k)} \right]. \tag{4.9}$$

This means that the continuous function $I(s)$ can be represented by its values in a discrete set of points (Shannon channels) where $s_k = k\pi/D_{\max}$ [$I(s)$ is therefore a so-called analytical function (Frieden, 1971)]. The number of parameters (or degrees of freedom) required to represent an analytical function on an interval $[s_{\min}, s_{\max}]$ is thus given by the number of Shannon channels ($N_s = D_{\max}(s_{\max} - s_{\min})/\pi$) in this interval. 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.

The question as to whether N_s represents the maximum number of independent parameters, which can be extracted from a scattering pattern has been discussed at length during the last decades (Damaschun *et al.* 1968; Moore, 1980; Taupin & Luzzati, 1982). Larger values of N_s clearly correspond to a higher information content in the data but at the same time the series in Eq. (4.9) must contain an infinite number of terms for the scattering intensity to have physical meaning. Figure 7 displays Shannon representations of the scattering intensity from an ellipsoid

model, truncated at $N_s = 10$ and 20. Inside the Shannon interval, the approximation in Eq. (4.9) provides a fair description of $I(s)$, although some systematic deviations are observed when approaching $s = s_{max}$. However, at higher angles the truncated Shannon representation oscillates around zero, which is unphysical as the scattering intensity cannot be negative. Scattering computed from any, even incorrect, physical model, would do better in this sense. Further, the experimental small-angle scattering 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 (Frieden, 1971), this oversampling allows in principle to extend the data beyond the measured range (so-called 'super-resolution') and thus to increase the effective number of Shannon channels. It is interesting to recall here that, in contrast to solution-scattering curves, crystallographic data are undersampled since the separation between reflections is twice the sampling distance required to describe the 3D scattering intensity as the Fourier image of the density in the unit cell (e.g. Baker *et al.* 1993). It is because of this undersampling that the phase problem (obtaining the missing phases of the complex amplitudes of the reflections) in crystallography cannot be unambiguously solved without additional information.

Extraction of structural parameters from solution-scattering data is in general a nonlinear problem, and the number of parameters cannot really be predicted by considering the experimental angular range only. 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 ($s_{min} < \pi/D_{max}$). It will, however, be demonstrated below that the level of detail of models, which can be deduced from solution-scattering patterns also depends on several other factors, like the accuracy of the data or the *a priori* information available.

4.3 Shape determination

Scattering curves from monodisperse solutions of randomly oriented particles contain information only about the spherically averaged 3D structure of a particle [excess scattering density distribution $\rho(\mathbf{r})$]. As the partial densities in the multipole expansion in Eq. (4.1) can be arbitrarily rotated without changing the scattering intensity (Stuhrmann, 1970b), it is clear that an infinite number of distributions $\rho(\mathbf{r})$ exist providing one and the same intensity $I(s)$. Simplifying assumptions about $\rho(\mathbf{r})$ thus have to be made to reduce the uncertainty of *ab initio* structure reconstruction from solution-scattering data. At low resolution (say, about 2–3 nm), the search is usually limited to homogeneous models, i.e. one assumes that $I(s) = I_c(s)$ and discards the second and third terms in Eq. (3.5) [Porod's law (Eq. (3.6)] is often used to obtain a homogeneous approximation).

In the past, shape modelling was done on a trial-and-error basis by computing scattering patterns from different shapes and comparing them with the experimental data. The two main strategies, which can be distinguished, also reflect the computing power available at the time of their introduction. The first one, based on Occam's razor, was to keep the number of model parameters as low as possible and usually relied on the comparison of the experimental scattering curves with charts containing $\log(I(s))$ *versus* $\log(sR_g)$ plots of the scattering patterns of three-parameter bodies like prisms, triaxial ellipsoids, elliptical or hollow circular cylinders (see e.g. Kratky & Pilz, 1978). The second strategy, which was developed somewhat later, was to use assemblies of spheres described by many parameters to construct more complex bodies constrained by additional information (e.g. from electron microscopy or hydrodynamic data). The scattering intensity from regularly packed spheres models was computed using Debye's formula

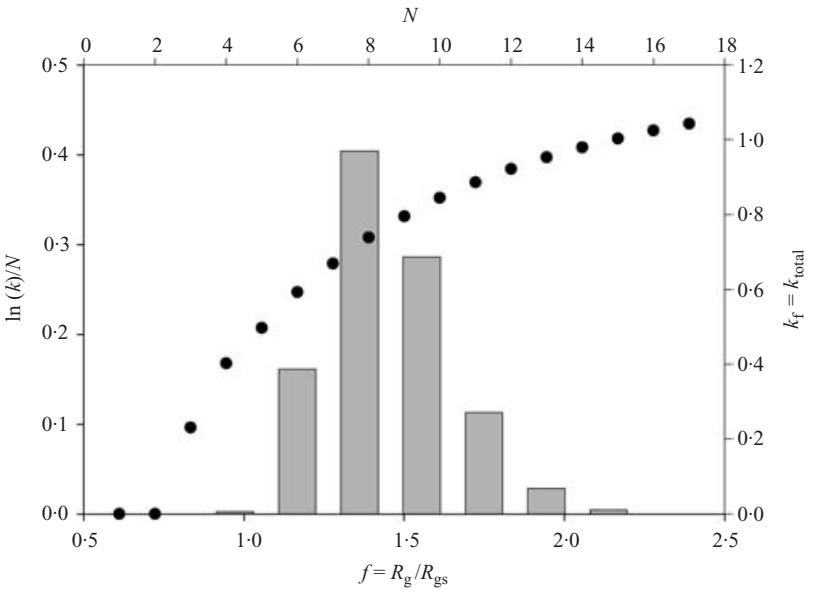

Fig. 8. Ratio between the logarithm of the number of unique configurations (k) of N adjacent squares on a lattice, excluding translations and rotations as a function of N (●). The histogram represents the fraction (k_f) of the total number of configurations (k_{total}) in each bin corresponding to the asymmetry factor f for arrangements of 16 squares ($N=16$).

(Glatter, 1972; Rolbin *et al.* 1973). Evolution of the two strategies led to the modern *ab initio* shape determination methods described below.

Modelling using packed spheres or cubes does not allow an exhaustive search. This can be illustrated by modestly considering the number of different arrangements of connected cells on a square lattice excluding those related by rotation or translation as illustrated in Fig. 8. This number increases very rapidly, but the distribution of radii of gyration of these arrangements indicates that for a given volume (number of squares) and radius of gyration, the number of possible models is related to the asymmetry factor ($f=R_g/R_{gs}$), where R_{gs} is the radius of gyration of the equivalent sphere), the most compact (sphere) and most extended arrangement (cylinder) being unique. The combinatorial increase in the number of arrangements makes it necessary to use Monte Carlo searches, as explained below.

4.3.1 Modelling with few parameters: molecular envelopes

An elegant *ab initio* procedure for determining the shape or angular envelope function $F(\omega)$, defined in Eq. (4.6), using a few parameters only was proposed by (Stuhrmann, 1970a). With the definition of the shape introduced in Eqs. (4.6)–(4.8), $\rho(\mathbf{r})=1$ in the range of integration of Eq. (4.2) [i.e. from 0 to $r=F(\omega)$].

Inserting Eq. (4.2) into Eq. (4.4) and expanding $j_l(sr)$ into a power series, the partial scattering amplitudes of the particle are

$$\mathcal{A}_{lm}(s) = (i \cdot s)^l \cdot \sqrt{2/\pi} \cdot \sum_{p=0}^{\infty} ((-1)^p \cdot f_{lm}^{(l+2p+3)} \cdot \{2^p \cdot p! \cdot (l+2p+3) \cdot [2(l+p)+1]!!\}^{-1} \cdot s^{2p}), \tag{4.10}$$

where the coefficients of the qth power of the shape function

$$f_{lm}^{(q)} = \int [F(\omega)]^q \cdot Y_{lm}^*(\omega) \mathrm{d}\omega \tag{4.11}$$

are readily computed using a recurrence formula based on Wigner coefficients (Svergun & Stuhrmann, 1991). Equations (4.10) and (4.4) allow fast computation of the partial amplitudes and beyond that of the scattering intensity from the f_{lm} coefficients, i.e. from the given envelope function. The algorithm for *ab initio* envelope determination starts from a spherical shape for which all coefficients but f_{00} are equal to zero. Subsequently, the f_{lm} coefficients are obtained which minimize the discrepancy between the experimental $\{I_{\mathrm{exp}}(s_k), k=1, \ldots, N\}$ and calculated curves

$$R^2 = \sum_{k=1}^{N} \{[I_{\mathrm{exp}}(s_k) - \mu I_{\mathrm{calc}}(s_k)] W(s_k)\}^2 \Big/ \sum_{k=1}^{N} [I_{\mathrm{exp}}(s_k) W(s_k)]^2 \tag{4.12}$$

with weighting factor $W(s_k) = s_k^2(\sigma(s_k)/I_{\mathrm{exp}}(s_k))$, where $\sigma(s_k)$ is the standard deviation in the kth point, and μ is an overall scaling factor

$$\mu = \sum_{k=1}^{N} I_{\mathrm{exp}}(s_k) I_{\mathrm{calc}}(s_k) W^2(s_k) \Big/ \sum_{k=1}^{N} [I_{\mathrm{exp}}(s_k) W(s_k)]^2. \tag{4.13}$$

This approach was further developed (Svergun *et al.* 1996, 1997c) and improvements of the method allowing to take the finite width of the particle–solvent interface into account were implemented in the shape determination program SASHA (Svergun, 1997). The main advantage of this method is that the number of parameters used in the description of the shape is comparable to the number of Shannon channels in the experimental data (the number of independent parameters in the series in Eq. (4.7) is $N_{\mathrm{p}} = (L+1)^2 - 6$).

The question arises, of course, whether this envelope determination is unique, or in other words, whether, in addition to the trivial case of an enantiomorphic envelope, different shapes exist at the same level of resolution (i.e. for the same L) yielding identical scattering curves. This problem was considered by using computer simulations on model bodies described by envelope functions exactly represented by a finite series in Eq. (4.7) of spherical harmonics up to $L=4$ (Svergun *et al.* 1996). Given the scattering intensity calculated from a model envelope, the particle shape was restored from this intensity with the above algorithm. Both error-free curves and curves containing statistical noise were simulated in different angular intervals. Shape restoration for error-free data was unique, even when using very limited ranges in the simulated curves. In the presence of errors, the uncertainty in the shape determination depended on the ratio between the number of model parameters N_{p} and of Shannon channels N_{s}. Shape restoration was found to be practically independent of the initial approximation and stable with respect to random errors when $N_{\mathrm{p}} \approx 1{\cdot}5 N_{\mathrm{s}}$. As experimental solution-scattering curves usually cover about 10–15 Shannon channels this result suggests that 15–20 variables, corresponding to a multipole resolution of $L=4$, can legitimately be used in the shape description.

Particle symmetry imposes restrictions on the multipole coefficients f_{lm} in Eq. (4.7) thereby reducing the number of independent parameters. Information about symmetry, if available, thus improves the reliability of the *ab initio* shape restoration. Consider, for example, a homodimeric particle with a twofold symmetry axis along z. In this case, all f_{lm} coefficients with odd m are zero, and the particle shape at $L=4$ is described by 12 independent parameters instead of 19 for

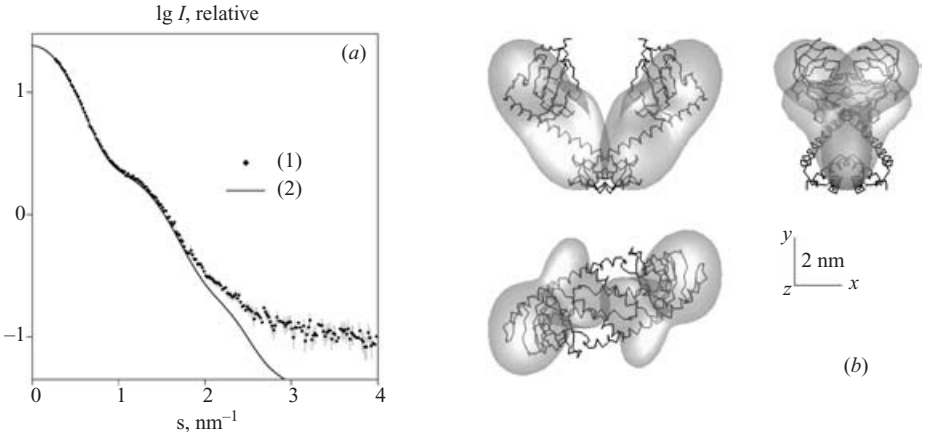

Fig. 9. (*a*) X-ray scattering from MIP (1) and scattering from the *ab initio* envelope model (2). (*b*) Atomic model of homodimeric MIP (C_α chain, Protein Data Bank entry 1fd9) superimposed to the *ab initio* model obtained by the program SASHA assuming two monomers (semi-transparent envelope) and twofold symmetry. The models are displayed on an SGI Workstation using the program ASSA. The right and bottom views are rotated counterclockwise by 90° around *y* and *x*, respectively.

a non-symmetric case. The higher the symmetry, the more multipole coefficients vanish, and the higher the resolution that can be achieved in the restoration. The most frequently occurring symmetries up to point group P62 are built into the program SASHA (for an extensive list of symmetry selection rules for spherical harmonics see Spinozzi *et al.* 1998). The quaternary structure of symmetric particles can also be restored in terms of the envelope function of their asymmetric unit. Thus, scattering from a symmetric homodimer is readily expressed via the shape of a monomer and the distance Δd between the monomers. The shape is determined as described above with a single additional parameter Δd (Svergun *et al.* 1997c).

Shape determination using envelope functions has successfully been used by several groups to generate low-resolution models of proteins (Grossmann *et al.* 2000; Krueger *et al.* 2000; Bernocco *et al.* 2001; Aparicio *et al.* 2002). The cases when high-resolution crystallographic models have become available *a posteriori* provide a good test for the validity of models generated *ab initio* from solution-scattering data. This is illustrated in Fig. 9, which displays a low-resolution model of the dimeric macrophage infectivity potentiator from *Legionella pneumophila* (Schmidt *et al.* 1995), and the crystal structure of the enzyme, determined 6 years later (Riboldi-Tunnicliffe *et al.* 2001). As illustrated below, in some cases the structures in the crystal and in solution can differ significantly.

4.3.2 Modelling with many parameters: bead models

Modelling using angular envelope functions has limitations in the description of complicated shapes like those having internal cavities. A more comprehensive description is achieved with bead models which extend the strategy of trial-and-error Debye modelling by making use of the number crunching capacity of modern computers. The concept of *ab initio* bead modelling in a confined volume (Chacon *et al.* 1998) relies on the fact that the maximum dimension $D_{\max}$ of a particle is readily obtained from its scattering pattern and that the particle must obviously fit inside a sphere of this diameter. If the sphere is filled with M densely packed spheres of radius

$r_0 \ll D_{\max}$, each of these beads may belong either to the particle (index = 1) or to the solvent (index = 0), and the particle shape is completely described by a string, X, of M bits. Starting from a random distribution of ones and zeros, the model is modified to find the binary string (i.e. the shape) that fits the experimental data. As the models may contain thousands of beads, conventional minimization techniques like gradient searches are not usable and different types of Monte Carlo searches are employed. The original method (Chacon *et al.* 1998), implemented in the program DALAI_GA, is based on a genetic algorithm. The more general 'dummy atoms' procedure (Svergun, 1999), which allows to generate models of inhomogeneous particles based on contrast variation data is described here; *ab initio* shape determination is a particular case of this procedure.

Assume a particle consisting of K components with distinctly different scattering length densities. In a nucleoprotein complex, for example, these components are the protein and RNA moieties, and $K = 2$. An arbitrary volume sufficiently large to enclose this particle (e.g. a sphere of radius $R = D_{\max}/2$) is filled with N 'dummy atoms' or (hexagonally) close packed spheres of radius $r_0 \ll R$. Each dummy atom is assigned an index X_j indicating the phase to which it belongs [X_j ranges from 0 (solvent) to K]. Given the fixed atomic positions, the shape and structure of the dummy atom model (DAM) are completely described by a phase assignment (configuration) vector X with $N \approx (R/r_0)^3$ components.

If the dummy atoms of the kth phase have contrast $\overline{\rho_k}$, the scattering intensity from the DAM is

$$I(s) = \left\langle \sum_{k=1}^{K} \overline{\rho_k} A_k^2(\mathbf{s}) \right\rangle_{\Omega}, \tag{4.14}$$

where $A_k(\mathbf{s})$ is the scattering amplitude from the volume occupied by the kth phase. Representing the amplitudes with spherical harmonics as in Eq. (4.3) one obtains

$$I(s) = 2\pi^2 \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \left\{ \sum_{k=1}^{K} \left[\overline{\rho_k} A_{lm}^{(k)}(s)\right]^2 + 2 \sum_{n>k} \overline{\rho_k} A_{lm}^{(k)}(s) \overline{\rho_n} [A_{lm}^{(n)}(s)]^* \right\}. \tag{4.15}$$

Following Eq. (4.7), the partial amplitudes from the volume occupied by the kth phase in a DAM are

$$A_{lm}^{(k)}(s) = i^l \sqrt{2/\pi} v_a \sum_j j_l(sr_j) Y_{lm}^*(\omega_j), \tag{4.16}$$

where the sum runs over all atoms of that phase $(r_j \omega_j) = \mathbf{r}_j$ are their polar coordinates, $v_a = (4\pi r_0^3/3)/0{\cdot}74$ is the displaced volume per dummy atom. The scattering curves from a multiphase DAM for an arbitrary configuration X and arbitrary contrasts ρ_k can be readily computed using Eqs. (4.14)–(4.16).

Given a set of $M \geqslant 1$ contrast variation curves $I_{\exp}^{(i)}(s)$, $i = 1, \ldots, M$, one searches for a configuration X minimizing the overall R factor between the experimental and calculated data

$$R^2(X) = \sum_{j=1}^{M} R_j^2(X), \tag{4.17}$$

where $R_j(X)$ is defined in Eq. (4.12). For an adequate description of a structure the number of dummy atoms must be large ($N \approx 10^3$), and thus significantly exceed the number of Shannon

channels. Even if the data are neatly fitted, uniqueness of such a model cannot be meaningfully discussed.

Given the resolution of the solution-scattering data, the model must be constrained to have low resolution with respect to r_0. For this, a list of neighbours (i.e. dummy atoms at a distance $2r_0$) is generated for each dummy atom. The looseness or degree of isolation of each non-solvent atom is calculated as $P(N_e) = \exp(-0{\cdot}5N_e) - \exp(-0{\cdot}5N_c)$, where N_e is the number of neighbours having the same index and $N_c = 12$ is the coordination number for hexagonal packing. The looseness of the configuration X is characterized by the average value $P(X) = \langle P(N_e) \rangle$ over all non-solvent atoms. Another condition imposes connectivity, i.e. the possibility to connect two arbitrarily selected atoms belonging to a phase by successively connecting neighbouring atoms belonging to the same phase. The measure of connectivity of the kth phase is computed as $G_k(X) = \ln(N_k/M_k) \geqslant 0$, where N_k and M_k are the numbers of dummy atoms in the entire phase and in the largest connected fragment, respectively.

The task of retrieving a low-resolution model from the scattering data can be formulated as follows: given a DAM, find a configuration X minimizing the goal function

$$f(X) = R^2(X) + \alpha \sum_K \{P_k(X) + G_k(X)\}, \tag{4.18}$$

where $\alpha > 0$ is the weight of the looseness penalty. This weight must as usually for penalties be selected in such a way that the second term yields a significant (say, 10–50%) contribution to the function at the end of the minimization.

Simulated annealing (SA; Kirkpatrick *et al.* 1983) is the method of choice for global minimization given the large number of variables and the combinatorial nature of the problem. The underlying idea in this method is to perform random modifications of the system (i.e. of the vector X) and move most often to configurations that decrease $f(X)$ but sometimes also to those increasing $f(X)$. The probability of accepting this second type of move decreases in the course of the minimization (the system is 'cooled'). Initially, the temperature is high and the changes almost random whereas towards the end a configuration corresponding (nearly) to the minimum of the goal function is reached. The algorithm was implemented in its faster 'simulated quenching' (Press *et al.* 1992; Ingber, 1993) version:

(1) Start from a random configuration X_0 at a 'high' temperature T_0 (e.g. $T_0 = f(X_0)$).
(2) Select an atom at random, randomly change its index (i.e. the phase to which it belongs) to obtain configuration X' and compute $\Delta = f(X') - f(X)$.
(3) If $\Delta < 0$, move to X'; if $\Delta > 0$, do this with probability $\exp(-\Delta/T)$. Repeat step 2 from X' (if accepted) or from X.
(4) Hold T constant for 100 N reconfigurations or $10N$ successful reconfigurations, whichever comes first, then cool the system ($T' = 0{\cdot}9T$). Continue cooling until no further improvement in $f(X)$ is observed.

The spherical harmonics expansion using Eq. (4.16) is much faster than Debye's formula in Eq. (2.7). Further, only a single dummy atom is changed at each move and hence only a single summand in Eq. (4.16) must be updated to calculate the partial amplitudes. As this is the most time consuming operation, the evaluation of $f(X)$ is accelerated about N times. This acceleration makes it possible to use robust SA (Ingber, 1993), which would otherwise be prohibitively slow as millions of function evaluations are required for a typical refinement.

This method was developed to analyse the contrast variation data from multi-component particles, and an example of its full-scale application to ribosomes is presented in Section 5.1. In the particular case of a single-component particle ($K=1$), the 'dummy atoms' approach reduces to the *ab initio* shape determination procedure implemented in the program DAMMIN (Svergun, 1999). Its ability to satisfactorily restore low-resolution shapes of macromolecules from solution-scattering data was demonstrated in test examples and in numerous applications by different groups (Funari *et al.* 2000; Svergun *et al.* 2000a; Egea *et al.* 2001; Fujisawa *et al.* 2001; Sokolova *et al.* 2001; Aparicio *et al.* 2002; Scott *et al.* 2002).

The program DAMMIN, similar to the envelope reconstruction algorithm SASHA, can take the symmetry of the particle (up to point group P62) as well as *a priori* information about its anisometry into account. The example of the hydrophilic (V_1) portion of *Manduca sexta* ATPase illustrates the importance of symmetry restrictions as well as the need to apply them cautiously. The X-ray scattering pattern of this large protein complex with a molecular mass of about 550 kDa, is presented in Fig. 10*a*. Its low-resolution model restored *ab initio* by SA without symmetry restrictions in Fig. 10*b* (left) displays an elongated mushroom-like shape, which agrees well with the results of cryo-EM (Grueber *et al.* 2000). As the major portion of the enzyme was expected to have quasi-threefold symmetry, this restriction was imposed during shape reconstruction. Surprisingly, the shape restored assuming P3 symmetry was a flat particle (Fig. 10*b*, second from the left), with both shapes yielding practically the same fit to the experimental data. The cause of this incorrect anisometry lies in the anisotropy of the search space due to the symmetry axis. By imposing an additional condition of particle prolateness, a rather detailed model of the enzyme is obtained (Fig. 10*b*, second from the right) and further a major structural transition due to redox modulation can be detected. The independently restored shape of the reduced form of the enzyme (Fig. 10*b*, right) indicates that the main structural alteration occurs in the headpiece, where the major subunits A and B are located, and at the bottom of the stalk. This structural change due to redox modulation is also corroborated by a lower susceptibility to tryptic digestion and changes in the tryptophan fluorescence of the reduced V_1 ATPase.

How can the predictions of the sampling theorem be reconciled with the restoration of the 'dummy atom' or 'bead' models described by $N \gg N_s$ atoms? First, as noted above, N_s alone does not define the number of degrees of freedom for a data set. Oversampling of the data in general increases the information content and the effective number of degrees of freedom was shown to range from zero for a signal-to-noise ratio of 1 to $15N_s$ for a signal-to-noise ratio of 10^3 (Frieden, 1971). This should not be taken as a proof that it is legitimate to build models described by $15N_s$ independent parameters, but rather as an indication that the number of degrees of freedom strongly depends on data accuracy. Secondly, the number of independent parameters in a DAM is much lower than N due to the looseness and connectivity penalty. At the later stages of annealing the program searches for a compact solution with the smallest interfacial area, whereas the fit acts as a constraint in Eq. (4.18) (the penalty is decreased rather than R^2). The higher the information content of the data, the more stringent the constraint and the more detail the DAM should retain. The compactness and connectivity conditions are thus crucial for a reliable shape reconstruction using bead modelling techniques. In the original genetic algorithm based DALAI_GA (Chacon *et al.* 1998) the solution was implicitly constrained by gradually decreasing r_0 during minimization, whereas the latest version of this program (Chacon *et al.* 2000) also contains an explicit compactness constraint.

It is clear that, in contrast to the envelope determination technique which gives only one solution, bead modelling methods may yield multiple solutions (spatial distributions of beads).

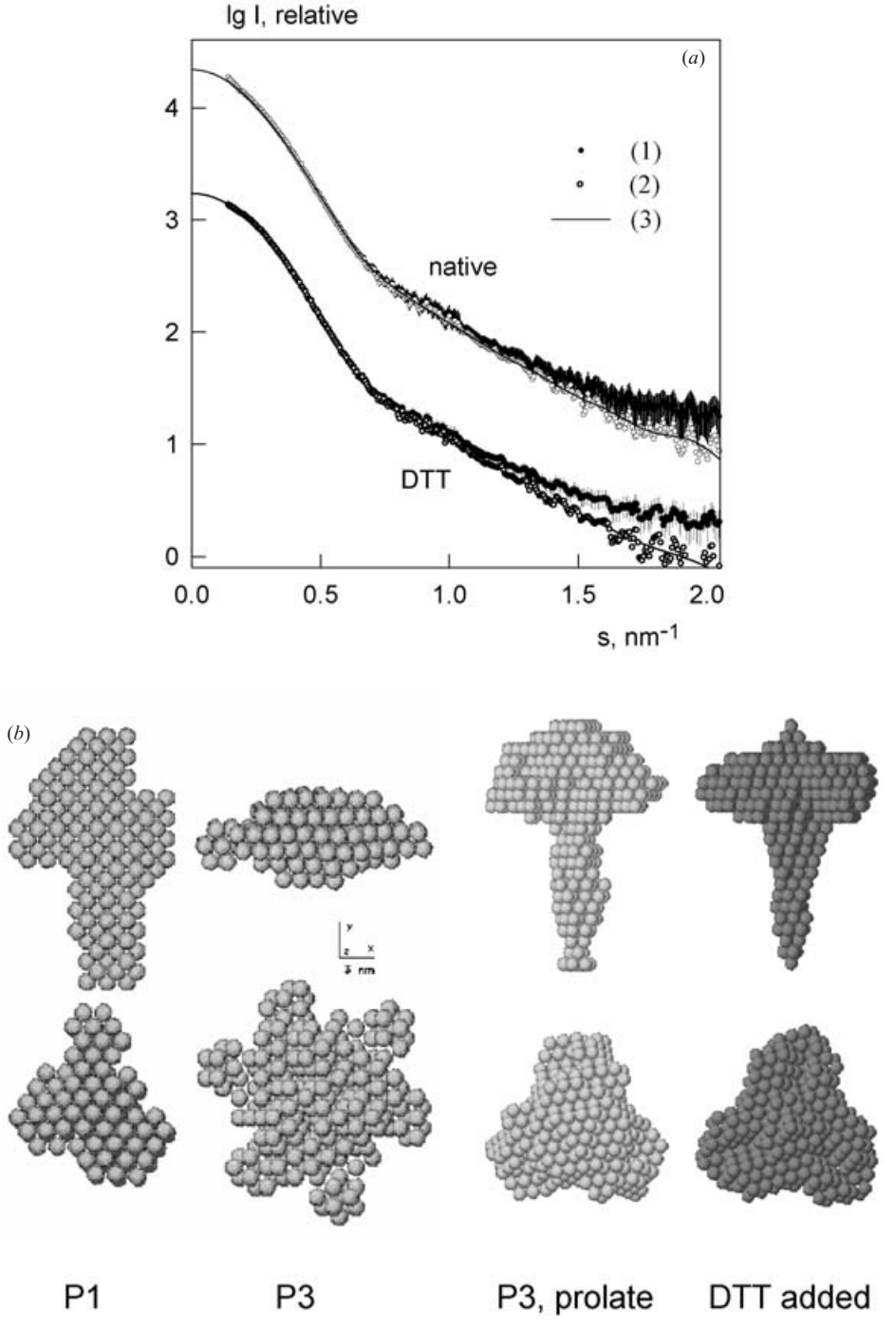

Fig. 10. (*a*) X-ray scattering patterns from the oxidized and reduced forms of V_1 ATPase. Experimental scattering before (1) and after (2) subtracting a constant and scattering from the *ab initio* models (3). The data and fit for the reduced form are shifted downwards by one logarithmic unit for better visualization. (*b*) The models obtained by DAMMIN with different symmetry and anisometry restrictions (see text for explanations). The bottom view is rotated counterclockwise by 90° around *x*.

Running any of the above shape determination programs several times with different seeds for random number generators produces a manifold of models corresponding to nearly identical scattering patterns. The variations between these models can serve as an indicator of the stability of the solution. The models obtained in independent runs can be superimposed and averaged to obtain a most probable model. This procedure is automated in the program SUPCOMB (Kozin & Svergun, 2001), which can align and superimpose two arbitrary low- or high-resolution models represented by ensembles of points without information about correspondence between these points. This is done by minimizing a dissimilarity measure between two models as a normalized spatial discrepancy (NSD). For every point (bead or atom) in the first model, the minimum value among the distances between this point and all points in the second model is found, and the same is done for the points in the second model. These distances are added and normalized against the average distances between the neighbouring points for the two models. As a rule of thumb, NSD values close to one indicate that the two models are similar. In the automated averaging procedure DAMAVER (V. V. Volkov & D. I. Svergun, unpublished observations), a reference model, with lowest average NSD, is selected by pairwise comparison within a set of models provided by a shape determination algorithm, and possible outliers with high NSD compared to the rest of the set are marked. After superposition of all models except outliers, the entire assembly of beads is remapped onto a densely packed grid of beads where each grid point is characterized by its occupancy factor (the number of beads in the set of models that map in its vicinity). An average (most probable) model is then constructed by filtering out low occupancy grid points. Alternatively, grid points with a non-zero occupancy can be used as a new search volume to construct a refined model. The diversity of the *ab initio* models and the results of the averaging procedure are illustrated in Fig. 11 for the shape determination of HIV reverse transcriptase.

Recently, other Monte Carlo based *ab initio* approaches have been proposed, which do not restrain the search space. A 'give-and-take' procedure (Bada *et al.* 2000) implemented in the program SAXS3D places beads on a hexagonal lattice, similar to the above close packed lattice of beads but unlimited in space. At each step, a new bead is added, removed or relocated to improve the agreement with the data. Although the new beads are always adjacent to already existing ones, connectivity is not explicitly imposed and this may result in unconnected models. The SASMODEL program (Vigil *et al.* 2001) does not use a fixed grid but represents the model by a superposition of interconnected ellipsoids and employs a Monte Carlo search of their positions and dimensions to fit the experimental data. It may be argued whether or not it is better to use a pre-defined search space (DALAI_GA, DAMMIN). On one side, improper restriction of the search space may lead to boundary effects and artefacts in the models. On the other side, the value of D_{max}, which is usually reliably determined from the experimental data, provides a valuable constraint reducing the uncertainty of the shape reconstruction. For this reason, the program DAMMIN reads in the output files (the data as well as the value of D_{max}) produced by the indirect transform program GNOM (Section 3.3). This use of pre-processed data also allows to desmear the scattering patterns and – in some cases – to correct for concentration effects or aggregation.

4.4 Modelling domain structure and missing parts of high-resolution models

All shape determination methods described so far are fundamentally limited by the assumption that the particle is homogeneous, i.e. by the necessity to fit the shape scattering curve. Contrast

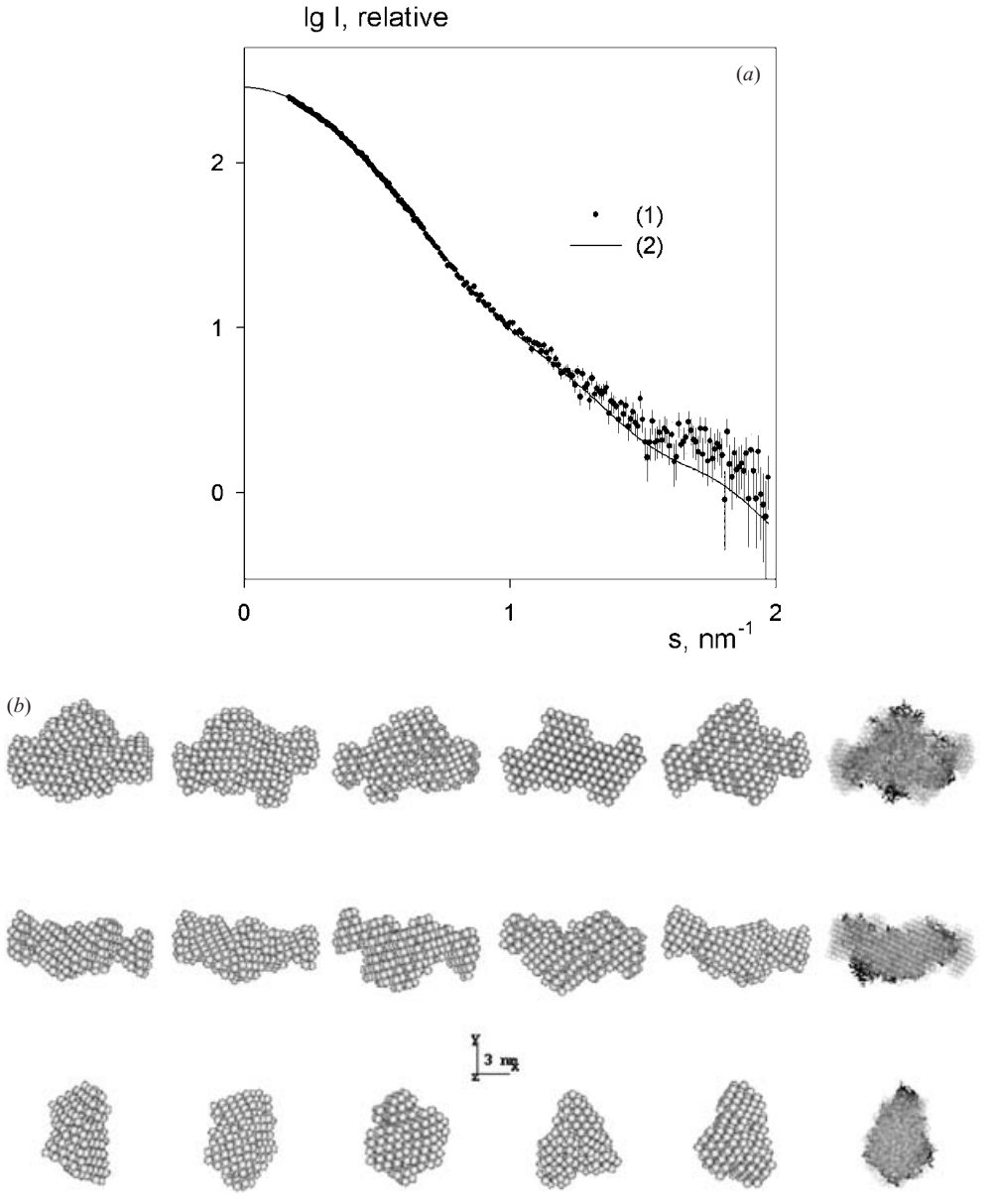

Fig. 11. Shape determination of HIV reverse transcriptase. (*a*) Synchrotron X-ray scattering patterns: experimental data (1) and the scattering from the models restored by DAMMIN (2). (*b*) Five models restored by DAMMIN and (rightmost column) the average model superimposed with the atomic model of the enzyme (PDB entry 3hvt) (Wang *et al.* 1994). The middle and bottom views are rotated counterclockwise by 90° around *x* and *y* respectively.

variation experiments required to determine this curve from Eq. (3.5) are usually impractical because of the larger amounts of material and more difficult preparations needed and of the low accuracy of this decomposition at higher angles. In most cases, the scattering data measured at higher contrasts are used 'as is', and for X-ray scattering on sufficiently large (more than 50 kDa) proteins, a reasonable approximation to the 'shape scattering' curve is obtained by

subtracting an appropriate constant (see Section 3.3). However, the increasing relative contribution from the internal structure at higher resolution severly restricts the usable part of the scattering patterns. This not only limits the resolution (to about 2–3 nm) but also the reliability of the models.

A new *ab initio* approach for building structural models of proteins from X-ray solution scattering accounting for higher resolution data has been proposed (Svergun *et al*. 2001). Proteins typically consist of folded polypeptide chains where the C_α atoms of adjacent amino-acid residues in the primary sequence are separated by approximately 0·38 nm. At a resolution of 0·5 nm, a protein can thus be considered as an assembly of dummy residues (DR) centred at the C_α positions. The scattering from a protein with K residues at positions $\mathbf{r}_i$ is computed by the Debye formula

$$I_{\mathrm{DR}}(s) = \sum_{i=1}^{K}\sum_{j=1}^{K} g_i(s)g_j(s)\frac{\sin sr_{ij}}{sr_{ij}}, \tag{4.19}$$

where $g_i(s)$ is the spherically averaged form factor of the jth residue and $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ is the distance between the ith and jth residues. The averaged residue form factor is computed from the spherically averaged scattering amplitudes of the amino-acid residues after solvent subtraction weighted according to their abundance. To account for bound solvent, the model is surrounded by dummy solvent atoms representing the first hydration shell. Simulations with known protein structures indicated that such a model adequately represents the scattering patterns up to 0·5 nm resolution (see Section 4.6 for a more detailed discussion of the computation of the scattering patterns from atomic models and bound solvent effects). A 3D model of the protein can be constructed by finding a spatial arrangement of the DRs that fits the experimental solution-scattering pattern. The reconstruction procedure resembles the 'dummy atoms' modelling in the previous section: SA is used to randomly modify the search model inside a sphere of diameter $D_{\max}$. Important differences are that (i) there is no fixed grid and a random move consists in relocating a DR taken at random to an arbitrary point at a distance of 0·38 nm from another randomly selected DR within the search volume; (ii) instead of being compact, the distribution of the DRs is required to be 'chain-compatible'. This restriction can be formulated by considering the spatial arrangement of C_α atoms in a real protein. In addition to the 0·38 nm separation along the chain, excluded volume effects and local interactions lead to a characteristic distribution of nearest neighbours. A histogram of the average number of C_α atoms in a 0·1 nm thick spherical shell surrounding a given C_α atom as a function of the shell radius $\langle N(R_k)\rangle$ for $0 < R_k < 1$ nm is presented in Fig. 12. In order to obtain a plausible chain-compatible DR model the histogram $N_{\mathrm{DR}}(R_k)$ should be similar to $\langle N(R_k)\rangle$.

The goal function to be minimized is $f(\mathbf{r}) = R^2(\mathbf{r}) + \alpha P(\mathbf{r})$, where the R factor is similar to that in Eq. (4.12) and the penalty $P(\mathbf{r})$ has the form

$$P(\mathbf{r}) = \sum_k [W(R_k)(N_{\mathrm{DR}}(R_k) - \langle N(R_k)\rangle)]^2 + G(\mathbf{r}) + R_{g0}^2. \tag{4.20}$$

The first term in Eq. (4.20) imposes a protein-like nearest neighbour distribution (the weights $W(R_k)$ are inversely proportional to the variations of $\langle N(R_k)\rangle$ in Fig. 12). The second term $G(\mathbf{r})$ ensures that the model is connected, i.e. each DR has at least one neighbour at a distance of 0·38 nm [cf. Eq. (4.19)]. The third term, proportional to the radius of gyration with respect to the origin, keeps the centre of mass of the DR model close to the origin and is gradually decreased

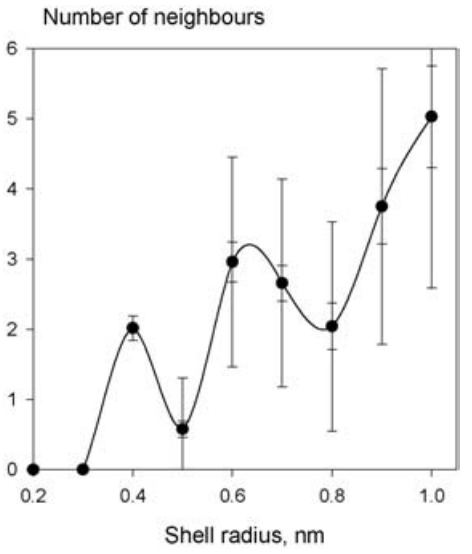

Fig. 12. Histogram of the average number of C_α atoms in 0·1 nm thick spherical shells around a given C_α atom. Smaller error bars, variation of the averaged values over all proteins; larger error bars, averaged variation within one protein.

during the SA procedure. The algorithm, implemented in the program GASBOR, can also take particle symmetry into account by generating symmetry mates for the DRs in the asymmetric unit (point groups P2 to P6 and P222 to P62 are supported).

As usual with SA, millions of function evaluations are required and it would take a prohibitively long time to fully re-compute $f(\mathbf{r})$ at each step. Fortunately, both Debye's formula [Eq. (4.19)] and the penalty [Eq. (4.20)] are computed from the distances r_{ij}. The table of off-diagonal distances $\{r_{ij}, i>j\}$ is computed only once and updated when moving one DR at a time. The performance of the Debye formula nevertheless significantly depends on the size of the protein. For smaller proteins (less than 50 kDa) the algorithm may be even faster than the multipole expansion implemented in the *ab initio* shape determination program DAMMIN, whereas for larger (more than 100 kDa) proteins, the Debye formula is much slower. Building a DR model of lysozyme (14·5 kDa) using GASBOR takes about 1 h on a 1 GHz Pentium III PC, whereas the reconstruction of a 150 kDa protein takes several days and would thus justify migrating to a more powerful computer. In the real-space version of GASBOR (Petoukhov & Svergun, 2003) a gain in speed of a factor of 5 is obtained by fitting the distance distribution function $p(r)$ rather than the intensity $I(s)$.

Compared to other shape determination methods, DR modelling employs fewer free parameters while accounting for more experimental information. Figure 13 presents models of the chitin-binding protein CHB1 (Svergun *et al.* 2000a) reconstructed using different *ab initio* methods. All models are similar at low resolution, but the less detailed shape models only fit the scattering at very low angles whereas the more detailed DR model neatly fits the entire experimental scattering pattern.

Like the bead modelling methods discussed above, the DR method produces a manifold of spatial distributions of DRs, rather than a single solution. Calculations on simulated and experimental scattering patterns indicate that the differences between the DR models are substantially

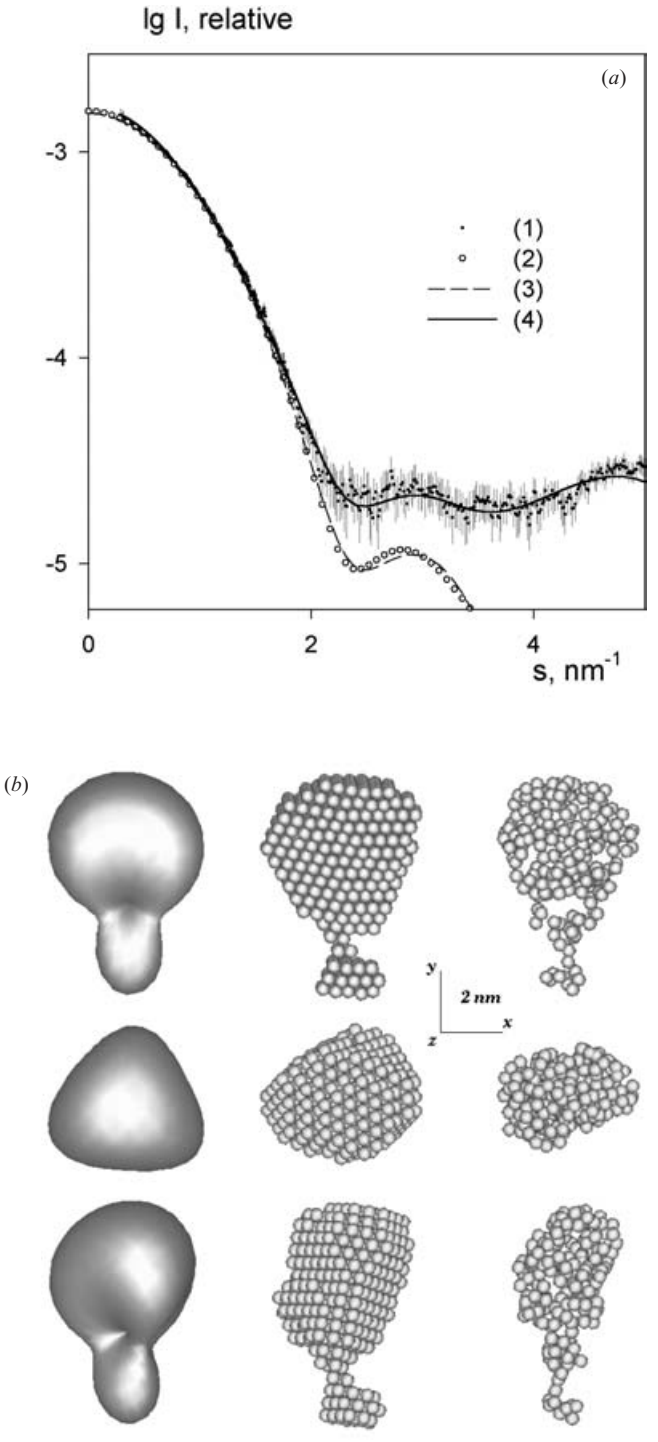

Fig. 13. (*a*) Experimental X-ray scattering from CHB1 (1) and scattering from the *ab initio* models: (2) envelope model (SASHA); (3) bead model (DAMMIN); (4) dummy residue model (GASBOR). (*b*) *Ab initio* models of CHB1 obtained by SASHA (left column), DAMMIN (middle column) and GASBOR (right column). The middle and bottom rows are rotated counterclockwise by 90° around *x* and *y* respectively.

smaller than those observed in low-resolution shape determination. The variations between the DR models preserve the domain structure of the protein, and the average (most probable) model can be generated by averaging the results of independent SA runs as described in the previous section.

DR modelling has potential for further development. One of the challenging tasks in this context is the direct determination of the protein fold from solution-scattering data using reduced complexity models and residue-specific constraints commonly employed in protein structure prediction. Scattering patterns from proteins can be fitted by native-like models consisting of a representation of the C_α trace with constraints due to secondary structure, hydrophobic contacts, bond and dihedral angle distributions, knowledge-based potentials, etc. It was shown (Zheng & Doniach, 2002) that the chances of identifying the correct topology are increased by combining constraints based on scattering data in a coarse-grained approach to *ab initio* fold prediction (Hinds & Levitt, 1994) with a homology search against the Dali domain library (Holm & Sander, 1998). Another way to reduce the uncertainty is to further simplify the search model, e.g. by using fragments with similar sequences (Simons *et al.* 1997). Alternatively, the information content of the scattering data can be increased by selective labelling of structural fragments. Selective deuteration of specific residues combined with contrast variation in H_2O/D_2O might provide sufficient constraints to establish the protein fold *ab initio* from neutron solution-scattering measurements.

Another potential application of *ab initio* models obtained from small-angle scattering is the phasing of low-resolution reflections in protein crystallography. The first successful attempts using envelopes reconstructed by the program SASHA were recently reported (Hao *et al.* 1999; Hao, 2001). The more detailed DR models should provide even better search models for molecular replacement (Aparicio *et al.* 2002).

An important application of DR-type modelling is the addition of missing fragments to incomplete models of proteins. 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 during cloning to facilitate crystallization. Domains of large proteins are also often isolated and studied separately. If high- or low-resolution models of all domains exist it may be possible to reconstruct the complete structure by rigid-body refinement as explained below. If, however, only some of the models are available, alternative approaches are required. 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 fit 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. In the case of 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. In this case near-native folds of missing loops or domains can be obtained by imposing residue-specific constraints (see above). The scoring function containing the discrepancy between the experimental and calculated patterns and relevant penalty terms is then minimized by SA. With this approach known structures can be completed with the degree of detail justified by the experimental data and by available a priori information. The algorithms for adding lacking loops or domains are implemented in the program suite CREDO (Petoukhov *et al.* 2002).

4.5 Computing scattering patterns from atomic models

As described in the previous section, solution-scattering data provide useful additional information for modelling when the high-resolution structure of a macromolecule is partially known. If a complete atomic model of the macromolecule is available its calculated scattering can be compared with the experimental patterns to determine the degree of similarity between the crystal and solution structures (Langridge *et al.* 1960; Ninio *et al.* 1972; Mueller, 1983; Pavlov *et al.* 1986; Grossmann *et al.* 1993). The solution-scattering data are also useful for validating predicted models obtained by homology modelling or other approaches. Further, if high-resolution models of individual domains are available, the quaternary structure of macromolecular complexes can be constructed by rigid-body modelling against solution-scattering data (Chamberlain *et al.* 1998; Svergun *et al.* 1998a; Boehm *et al.* 1999; Krueger *et al.* 1999).

A necessary prerequisite for high-resolution modelling in solution-scattering studies is an accurate computation of scattering patterns from atomic structures. This computation is by no means trivial, as the solvent scattering must be taken into account. Below, methods for computing X-ray and neutron scattering from atomic models and rigid-body refinement are considered in detail.

The scattering intensity from a particle in solution can be expressed as

$$I(s) = \langle |A_a(\mathbf{s}) - \rho_b A_{ex}(\mathbf{s}) + \delta\rho_h A_h(\mathbf{s})|^2 \rangle_\Omega, \tag{4.21}$$

where $A_a(\mathbf{s})$, $A_{ex}(\mathbf{s})$ and $A_h(\mathbf{s})$ are, respectively, the scattering amplitudes from the particle *in vacuo*, from the excluded volume, and from the hydration shell. The scattering density of the bulk solvent, ρ_b, may differ from that of the hydration shell, ρ_h, resulting in a non-zero contrast for the shell $\delta\rho_h = \rho_h - \rho_b$ (Svergun *et al.* 1995).

The solvent contribution in Eq. (4.21) consists of two terms, one due to the excluded volume (i.e. the volume inaccessible to the solvent) and the second to the difference between the density of the solvent in the hydration shell and in the bulk. Several methods have been proposed to compute the scattering from the excluded volume. In the effective atomic scattering method, the excluded volume is built by dummy solvent atoms located at the positions of the atoms in the macromolecule (Langridge *et al.* 1960; Fraser *et al.* 1978; Lattman, 1989; Koch *et al.* 1995; Svergun *et al.* 1995). It has been argued that this approach non-uniformly fills the excluded volume and thus does not adequately represent the scattering at resolutions above 1–2 nm. The cube method was developed to avoid inhomogeneous filling (Ninio *et al.* 1972; Mueller, 1983; Pavlov & Fedorov, 1983b). Following the approach of Lee & Richards (1971), the particle surface accessible to the solvent is defined by rolling a sphere simulating a water molecule on the van der Waals surface of the particle. The excluded volume is represented by cubes with a small edge (down to 0·05 nm) to ensure its precise and uniform filling, and it was claimed that this procedure is superior to the effective scattering factor methods at higher resolution (Mueller, 1983; Pavlov & Fedorov, 1983b). Different approaches have also been proposed to account for the scattering from the hydration shell, either by placing water molecules on the surface (Hubbard *et al.* 1988; Grossmann *et al.* 1993; Fujisawa *et al.* 1994) or by surrounding the particle by a continuous outer envelope simulating the first solvation shell (Svergun *et al.* 1995). Some of the procedures for adding hydration shells to bead models have recently been reviewed (Perkins, 2001).

There are several publicly available programs for computing solution-scattering patterns from biological macromolecules. The first one (Lattman, 1989) is based on the effective atoms method

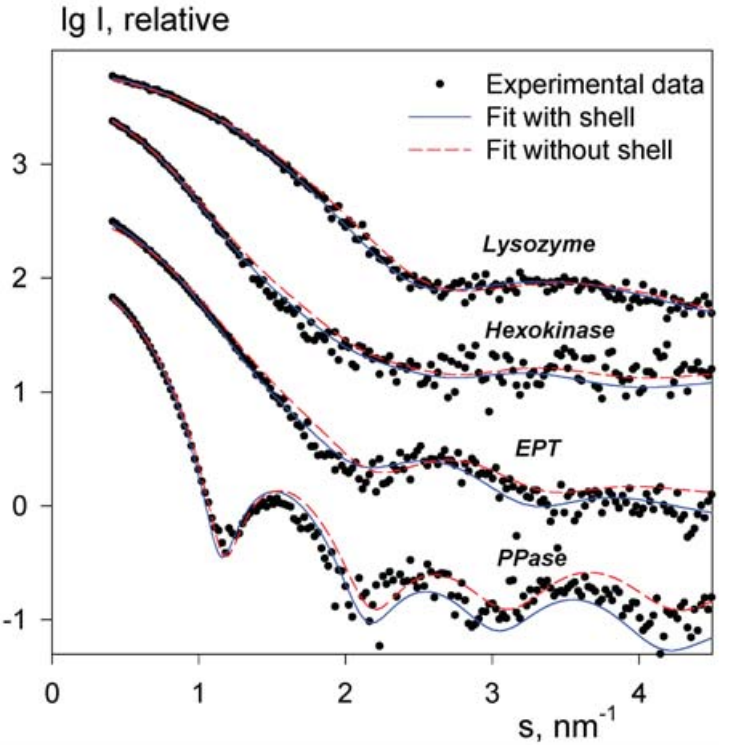

Fig. 14. X-ray scattering patterns from several proteins and scattering computed from their crystallographic model with and without accounting for the solvation shell. EPT, UDP-*N*-acetylglucosamine enolpyruvyltransferase [PDB entry 1NAW (Schonbrunn *et al.* 1996)]. PPase, pyrophosphatase [PDB entry 1JFD (Avaeva *et al.* 1997)].

and computes X-ray curves and the second one (Pavlov & Fedorov, 1983a) employs the cube method and can evaluate both X-ray and neutron profiles. Neither of these programs accounts for the hydration shell. The more recent program, CRYSOL (Svergun *et al.* 1995), employs atomic scattering factors to evaluate the X-ray scattering from the excluded volume. To account for the hydration shell, the angular envelope function of the particle $F(\omega)$ [see Eq. (4.6)] is computed and surrounded by a 0·3 nm thick homogeneous layer with variable scattering density. Given the atomic coordinates, the program predicts the scattering profile or fits the experimental data by adjusting the excluded volume of the particle and the density of the hydration layer. The terms in Eq. (4.21) are computed using the multipole expansion with equations similar to Eq. (4.16) and Eq. (4.11). This speeds up the calculations and, additionally, the partial amplitudes computed by CRYSOL can further be used for rapid computation of scattering from complexes (see next section).

Analysis of the SAXS data from numerous known protein crystal structures indicated that inclusion of the hydration shell significantly improved the agreement between the experimental and calculated X-ray scattering curves. As illustrated in Fig. 14, patterns calculated from atomic models without hydration shell systematically deviate from the experimental ones suggesting that the particle in solution is larger than in the crystal. The fit is much improved by assuming a higher density in the border layer (typically 1·05–1·25 times that of the bulk), suggesting that the hydration shell around proteins is denser than the bulk solvent. SAXS solution studies alone do not

provide an unequivocal proof of physical origin of this higher density. A similar effect on the scattering curves could indeed result from higher mobility or disorder of the surface side chains in solution compared to their average structure in the crystal, which would also increase the apparent particle size. This ambiguity was resolved in a combined X-ray and neutron scattering study of several proteins (Svergun *et al.* 1998b). The results obtained in different contrast conditions provided by X-rays and neutrons were shown to be compatible with a denser hydration shell rather than with a higher mobility of the side-chains on the protein surface. The program CRYSON (Svergun *et al.* 1998b) is an analogue of CRYSOL for computing neutron scattering in H_2O/D_2O mixtures taking H/D exchange effects into account.

The existence of a hydration shell around macromolecules in solution had long been inferred from various observations but relatively little attention was paid to it in solution scattering. The higher density of this shell compared to bulk solvent is due to electrostriction (see e.g. Bockris & Reddy, 1977), i.e. the compression of a material medium under the influence of electric fields due, in this case, to partial charges and induced dipoles on the protein surface. The solution-scattering experiments confirm the higher density in the hydration shell predicted from molecular dynamic simulations (Levitt & Sharon, 1988) and more recently from explicit all-atom computations of scattering patterns from proteins accounting for the solvent dynamics (Merzel & Smith, 2002a, b), and reported in several crystallographic studies (Badger, 1993; Burling *et al.* 1996). Such a denser water layer, where the movements of water molecules are reduced compared to bulk water (Teixeira, 1994), has also been observed in the vicinity of membranes. Proton migration along such a surface is much faster than exchange with bulk water (see e.g. Heberle *et al.* 1994). This and the clustering of proton-binding sites on the surface of proteins can lead to considerably higher proton transport rates to specific sites (Gutman & Nachliel, 1997).

The dynamic properties of the hydration layer and the resulting forces – attractive or repulsive – have, however, remained elusive (Israelachvili & Wennerstroem, 1996). There seems to be substantial agreement between the results of different methods that the thickness of the hydration layer does not exceed 1–2 water molecules. This was shown early on in dielectric studies (see e.g. Bone & Zaba, 1992), which suggested the existence of a one or two molecule thick layer of water with two classes of interactions (bulk-type and bound). More recent spectroscopic experiments (Pal *et al.* 2002) have confirmed the existence of two classes of solvation times, the first one corresponding to bulk-type solvation ($\sim$1 ps) and the second one to the formation of a more rigid water structure ($\sim$38 ps). Beyond about 0·7 nm from the surface there is essentially only bulk-type solvation.

Analysis of high angle scattering patterns at resolutions above 1 nm is an experimentally and computationally demanding task, which was attempted several times during the last decades (Pickover & Engelman, 1982; Mueller *et al.* 1990). High-resolution patterns can only be obtained with concentrated solutions ($>$2% w/v). This requires beside very accurate scattering and transmission measurements an accurate knowledge of the volume fraction of the solute in Eq. (3.1). The effects of the dissolved macromolecule on the structure of water can also be taken into account (Hirai *et al.* 2002) but the procedure is still somewhat empirical. Accounting for the excluded and bound solvent [Eq. (4.21)] is a difficult calculation as the approximations of continuous excluded volume and hydration shell break down and the atomicity of the solvent must be taken into account. Paradoxically, the cube method, which uniformly fills the excluded volume (Mueller, 1983; Pavlov & Fedorov, 1983b), inadequately represents solvent scattering at higher angles, as illustrated by comparison with the results of molecular dynamics simulations

(M. Malfois & D. I. Svergun, unpublished observations). Using CRYSOL experimental high angle X-ray scattering patterns can be reasonably well fitted up to about 0·5 nm resolution (Hirai *et al.* 2002; Svergun *et al.* 2001). How hydration can best be taken into account in modelling scattering patterns at higher resolution is still not clear as the question has hitherto only been investigated in the case of myoglobin (Seki *et al.* 2002).

The problem of the hydration layer is inseparable of the effects of small solutes (e.g. glycerol, sucrose, salts) on the structure of water which were already observed by X-ray scattering in the 1920s (see e.g. Krishnamurti, 1929) and are still actively studied especially by neutron scattering (Dixit *et al.* 2002). Much less is known about the effect of these substances on the excluded volume and hydration layer although there are some sporadic observations in the literature (e.g. Stuhrmann *et al.* 1976). In particular, the effect of polyethylene glycol (PEG) on the hydration and interactions between macromolecules in solution has received more attention than that of other substances because of its practical importance in crystallization (see Section 6).

The combination of scattering and hydrodynamic methods in the study of hydration of biological macromolecules has been reviewed (Eisenberg, 1994). Bead models provide the simplest and most direct link between the structural data and hydrodynamic parameters (diffusion coefficients, intrinsic viscosity, relaxation time, etc.), which can be calculated using the HYDRO program suite (Garcia de la Torre *et al.* 2000) and this provides an additional consistency check.

Accounting for the hydration shell is indispensable for an adequate computation of solution-scattering patterns from dissolved macromolecules. The combination of accurate neutron and X-ray scattering, including anomalous scattering (Plestil & Hlavata, 1988), measurements should allow in future to obtain more information about important phenomena like hydration and counterion condensation.

4.6 Rigid-body refinement

During the last decade X-ray crystallography and NMR have generated many protein structures, and it is expected that even larger numbers will be produced in the near future (Burley, 2000; Edwards *et al.* 2000). Most cellular functions are, however, accomplished by macromolecular complexes, which are too large for NMR studies and often possess inherent structural flexibility making them difficult to crystallize. As solution scattering is sensitive to changes in the quaternary structure of macromolecules it should be particularly useful for the analysis of such complexes.

One approach already mentioned above (Section 3.4), but which is quite heavy and expensive, is that of label triangulation of deuterated subunits in neutron scattering. If this is impractical one can often obtain similar information in an alternative way. Indeed, if high-resolution structures of individual domains or subunits composing the complex are available, detailed models can be constructed by rigid-body modelling if it can be assumed that the tertiary structure of the domains is essentially preserved upon complexation. In fact, this approach could also significantly improve the interpretation of label triangulation data.

Before describing the procedure and its application to models at atomic resolution a caveat should be entered against overinterpretation of the final models obtained by this procedure. Although they apparently retain the high resolution of the individual domains, these models are obtained by an effectively low-resolution refinement in terms of rotations and translations of the subunits or domains. In particular when the crystal structure of the entire complex

is also available, these models should not be used for detailed comparisons, but rather as a measure of the differences between the structures of the macromolecules in the crystal and in solution.

To illustrate the application of rigid-body modelling, consider a complex consisting of two subunits (A and B). The scattering amplitudes from the subunits centred at the origin in reference orientations are denoted $A(\mathbf{s})$ and $B(\mathbf{s})$, respectively. An arbitrary complex can be constructed by fixing the first subunit and rotating and translating the second one. The rotation is described by the Euler angles α, β and γ (Edmonds, 1957) and the shift by a vector $\mathbf{u} = (u_x, u_y, u_z)$, so that the entire operation is defined by six parameters. If $C(\mathbf{s})$ is the scattering amplitude from the displaced second subunit, the scattering from the complex can be expressed as (Svergun, 1991, 1994)

$$I(s) = I_A(s) + I_B(s) + 2\langle A(\mathbf{s})C^*(\mathbf{s})\rangle_{\Omega}. \tag{4.22}$$

Representing the scattering amplitudes using spherical harmonics $Y_{lm}(\Omega)$ as in Eq. (4.5) yields the closed expression

$$I(s) = 2\pi^2 \sum_{l=0}^{\infty} \sum_{m=-l}^{l} (|A_{lm}(s)|^2 + |B_{lm}(s)|^2 + 2\mathrm{Re}[A_{lm}(s)C^*_{lm}(s)]). \tag{4.23}$$

If the structures of the subunits are known, the scattering amplitudes and the partial functions $A_{lm}(s)$ and $B_{lm}(s)$ can be computed from the atomic models as described in Section 4.5. The partial functions $C_{lm}(s)$ of the rotated and translated second subunit can be expressed analytically via the $B_{lm}(s)$ functions, the elements of the finite rotation matrix and the Wigner 3*j* coefficients (Edmonds, 1957). (For explicit equations allowing rapid computation of $I(s)$ see Svergun, 1991, 1994.)

Rigid-body modelling of solution-scattering data requires determination of the six positional parameters of the second subunit minimizing the discrepancy between the experimental and calculated intensities. This search can be further restricted for the practically important case of the structures of homodimers with a twofold symmetry axis, which are entirely defined by the structure and orientation of one monomer and the distance between its centre and the twofold axis. Assuming this axis to coincide with Y and the monomers to be separated by $2u_z$ along Z, the intensity is a function of four parameters $(\alpha, \beta, \gamma, u_z)$ (see explicit equations in Svergun *et al.* 1997c).

Rapid computation of $I(s)$ using the spherical harmonics approach allows, in principle, to perform an exhaustive search of positional parameters to fit the experimental scattering from the complex. 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 based on visual criteria. Rigid-body modelling programs are available for major UNIX platforms (program ASSA; Kozin *et al.* 1997; Kozin & Svergun, 2000) as well as for Wintel-based PCs (program MASSHA; Konarev *et al.* 2001). They allow 3D display and manipulation of high-resolution atomic structures and low-resolution models represented as smooth envelopes or ensembles of beads and are coupled to computational modules for interactive or automated refinement. In the first mode, the user may shift and rotate the subunits while observing corresponding changes in the fit to the experimental data; in the automatic mode, the program performs an exhaustive search in the vicinity of the current configuration. A single computation of the scattering from

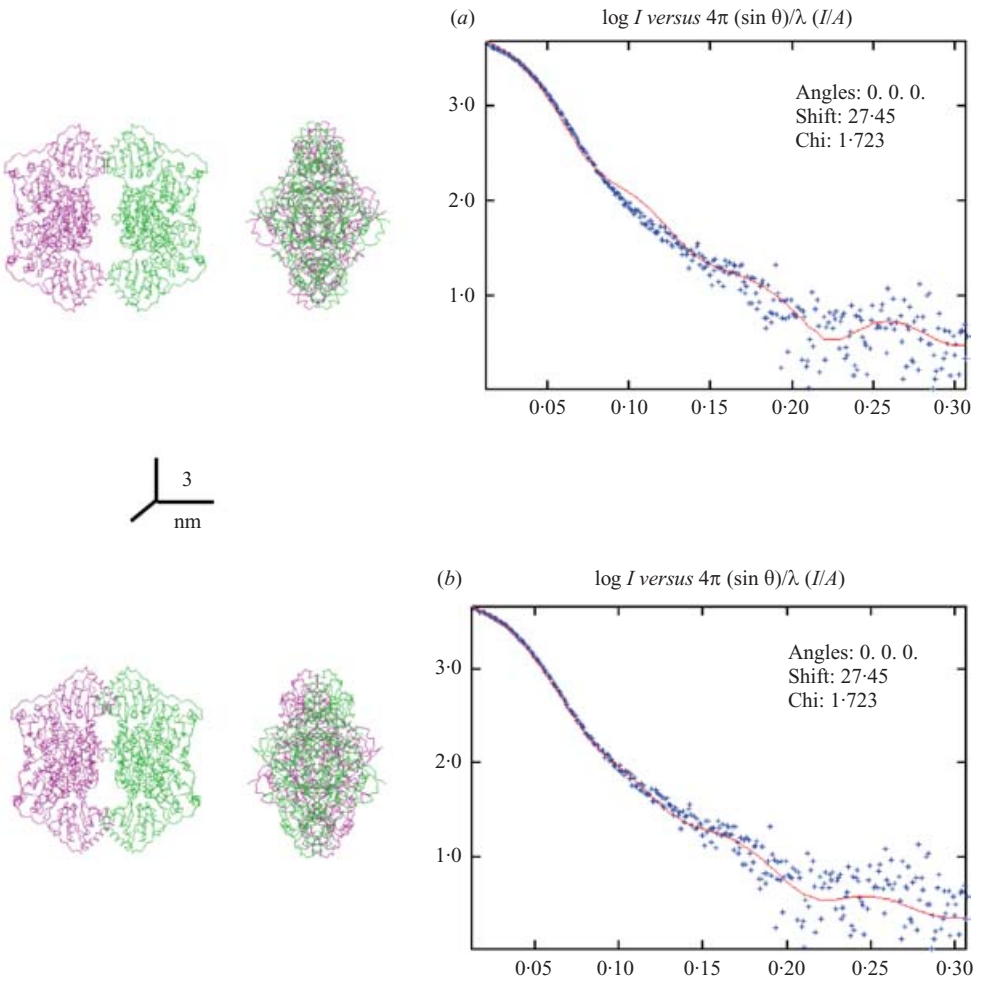

Fig. 15. Rigid-body modelling of the structure of tetrameric yeast PDC in terms of two dimers. (*a*) Original crystallographic structure; (*b*) model obtained by rigid-body refinement using the program MASSHA. Left, the models as C_α traces (side views in the middle column are rotated by 90° around *y*); right, fits to the experimental X-ray scattering data.

a complex takes less than a second on a 1 GHz Pentium-III machine, and a default refinement cycle requires about a minute. The program MASSHA has an option for rigid-body refinement of homodimeric complexes with a twofold symmetry axis. Another useful option allows to automatically superpose two arbitrary low- or high-resolution models by invoking the alignment program SUPCOMB (Kozin & Svergun, 2001) described in Section 4.3.

Rigid-body refinement of the structure of tetrameric yeast pyruvate decarboxylase (PDC) in Fig. 15 illustrates the modelling of a homodimer. Yeast PDC is a thiamin diphosphate-dependent enzyme involved in some steps of alcoholic fermentation. The enzyme is a tetramer of molecular mass 236 kDa at low pH which dissociates into dimers at high pH (Huebner *et al.* 1990; Koenig *et al.* 1993). The solution-scattering pattern computed from the crystallographic model [PDB entry 1 pvd (Arjunan *et al.* 1996), Fig. 15*a*] deviates significantly from the experimental data (Fig. 15*a*, right panel, $\chi = 1{\cdot}723$). Using the automatic refinement in MASSHA and starting from

the crystallographic model, nearly the same model was obtained in a matter of minutes as in the original time-consuming interactive rigid-body refinement of the structure in terms of two dimers (Svergun *et al.* 2000b) (Fig. 14*b*). The dimers in the final model are tilted by 11° around the *x*-axis (left view) and the distance between them is reduced by 0·4 nm. The model provides a good fit to the experimental data with $\chi = 0{\cdot}66$. The root mean-square displacement of atoms in the refined model compared to the crystal structure is 0·57 nm. This large conformational change can be explained by weak contacts between the dimers in the crystallographic model (Fig. 15*b*, left panel) so that the crystal environment significantly influences the quaternary structure of the enzyme.

Other approaches to rigid-body modelling include an 'automated constrained fit' procedure (Boehm *et al.* 1999), which generates thousands of possible bead models in an exhaustive search for the best fit. This procedure was applied to a number of proteins, e.g. a folded-back model of monomeric factor H of human complement was deduced from X-ray and neutron small-angle scattering and ultracentrifugation (Aslam & Perkins, 2001). In another approach (Krueger *et al.* 1998; Zhao *et al.* 1998; Tung *et al.* 2000, 2002), a representation of the domains in terms of triaxial ellipsoids is used to find their approximate arrangement in the complex. The atomic models of the domains are subsequently positioned within the ellipsoids utilizing information from other methods including NMR, homology modelling and energy minimization, and its applications have been comprehensively reviewed (Wall *et al.* 2000).

5. Applications

The two examples below illustrate the potential of recent data analysis and modelling techniques on two very different systems requiring also different methodological approaches. They illustrate the variety of problems where solution-scattering can be of use either in itself or as a complementary source of information. The second example, which gives an overview of the SAXS studies on the allosteric enzyme aspartate transcarbamoylase from *Escherichia coli* (ATCase), is treated in detail to illustrate the biochemical relevance of such investigations on a well-known textbook case.

5.1 Contrast variation studies of ribosomes

The various aspects of the application of the contrast variation method are best illustrated by the studies on the structure of the ribosome. This supramolecular assembly, which is responsible for protein synthesis in all organisms, consists of two different complex subunits. The most extensively characterized form – the 70S ribosome from *Escherichia coli* – has a total molecular mass around $2{\cdot}3 \times 10^6$ Da and consists of a 30S subunit containing 21 individual proteins and a single 16S rRNA molecule, and a 50S subunit with 33 different proteins and two rRNA molecules (5S rRNA + 23S rRNA) (Wittmann, 1982). In each of the subunits the rRNA moieties account for about two-thirds of the mass. The difference in scattering length between the protein and RNA moieties make the ribosome a very suitable object for contrast variation by H_2O/D_2O substitution (see Table 2). The early studies (for a review see Koch & Stuhrmann, 1979) provided valuable information about the structure of the ribosome and its subunits in solution in terms of integral parameters of the protein and rRNA moieties. They also provided the first models of the shape of the ribosome and its subunits based on spherical harmonics.

More detailed information was obtained by selective deuteration of the ribosomal components – individual proteins and RNA. Triangulation of labelled protein pairs in the small ribosomal subunit is certainly the largest small-angle scattering project ever undertaken. In an effort lasting fifteen years it led to a map of the 21 individual protein positions in the small ribosomal subunit from *E. coli* (Capel *et al.* 1987).

In another comprehensive study of the ribosome (Svergun *et al.* 1997b; Svergun & Nierhaus, 2000), a set of 42 synchrotron X-ray and neutron solution-scattering curves from hybrid *E. coli* ribosomes was obtained (Fig. 16), where the protein and rRNA moieties in the subunits were either protonated or deuterated in all possible combinations. This is probably the most extensive set of consistent X-ray and neutron contrast-variation data collected on a single object. This data set was analysed using the dummy atoms technique described in Section 4.3.2. The search volume defined by the cryo-electron microscopic model (Frank *et al.* 1995) was divided into 7890 densely packed spheres of radius 0·5 nm. SA was employed to assign each sphere to solvent, protein or RNA to simultaneously fit all scattering curves (Fig. 16). Twelve independent reconstructions starting from random approximations yielded reproducible results and were averaged to yield the model presented in Fig. 17 (top) displaying 15 and 20 protein sub-volumes in the 30S and 50S subunit, respectively, connected by RNA. The models in the middle and bottom row in Fig. 17 illustrate the comparison of the map obtained from solution scattering with the high-resolution X-ray crystallographic maps of the ribosomal subunits, which became available soon after (Nissen *et al.* 2000; Schluenzen *et al.* 2000). The positions of protein globules predicted from solution scattering agree astonishingly well with the crystallographic results, and the agreement is even more striking given that the resolution of the neutron-derived map was only 3 nm and that the crystal stuctures belong to different species (*Th. thermophilus* and *H. marismortui*). This *a posteriori* comparison underlines the potential of contrast variation and of the joint use of X-ray and neutron scattering in the study of large macromolecular complexes.

5.2 Structural changes and catalytic activity of the allosteric enzyme ATCase

Aspartate transcarbamoylase (or aspartate transcarbamylase) from *E. coli* (ATCase) catalyses the first committed step of the biosynthetic pathway of pyrimidines, namely the carbamylation of aspartate by carbamyl phosphate (CP) yielding carbamyl aspartate and phosphate. The two substrates display preferred order binding, with CP binding first. This highly regulated enzyme with positive homotropic cooperativity for the binding of the substrate L-aspartate, heterotropic activation by ATP and inhibition by CTP and by UTP in synergy with CTP (for reviews see Hervé, 1989; Lipscomb, 1994) is a paradigm of allostery. The enzyme is a heterododecamer comprising two trimers of catalytic chains (MM = 34 kDa) and three dimers of regulatory (r) chains (MM = 17 kDa) forming the 306 kDa holoenzyme with quasi-D_3 symmetry. All regulatory nucleotides bind to the same site on the regulatory chain, about 6 nm away from the nearest active site.

The regulatory properties of ATCase have been interpreted in terms of various states of the enzyme with different catalytic activity level and different quaternary structures. The seminal MWC model for allostery (Monod *et al.* 1965) proposes that the enzyme is in equilibrium between two symmetrical quaternary structures with low and high affinity for the substrates referred to as the T (tense) and R (relaxed) structures, and that homotropic or heterotropic effects are due to displacements of the equilibrium between these two forms of the enzyme. The affinity only depends on the quaternary structure of the enzyme and not on the fractional

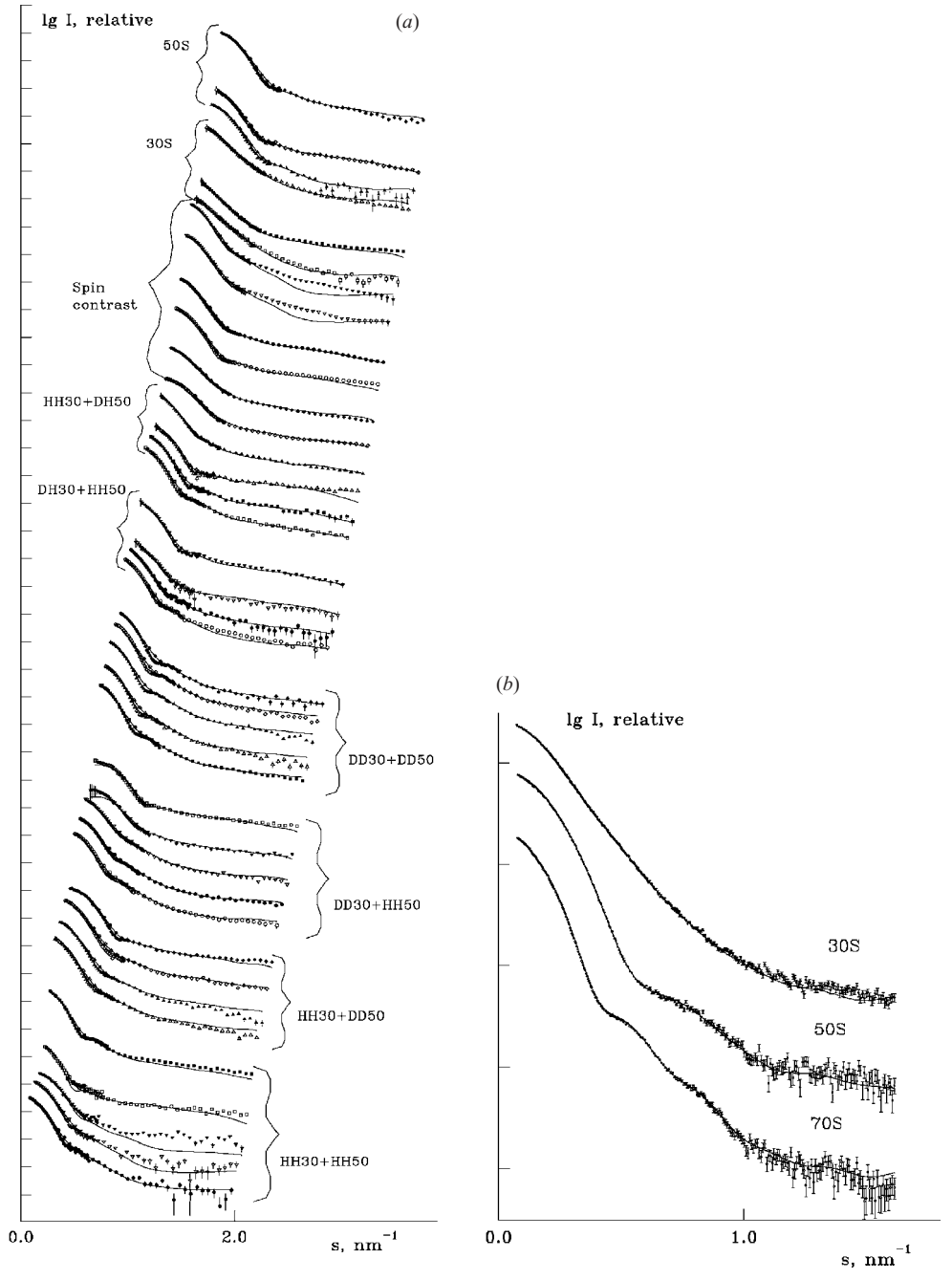

Fig. 16. Neutron (*a*) and X-ray (*b*) scattering data from hybrid 70S ribosomes and free subunits fitted by a four-component dummy atoms model. In (*a*), H and D denote protonated and deuterated components, respectively, whereby the first letter is related to proteins, the second to RNA (e.g. HH30 + DH50 describes a particle with fully protonated 30S subunit and the 50S subunit with proteins deuterated, proteins protonated). 'Spin contrast' denotes the data obtained by spin-dependent contrast variation (Svergun *et al.* 1996), the upper six curves were collected on free ribosomal subunits. The experimental data are presented as dots with error bars, the fits as solid lines. Successive curves are displaced up by one logarithmic unit corresponding to the distance between the ordinate tick marks [in panel (*a*), also by $\Delta s = 0{\cdot}05$ nm^{-1} along the abscissa] for better visualization.

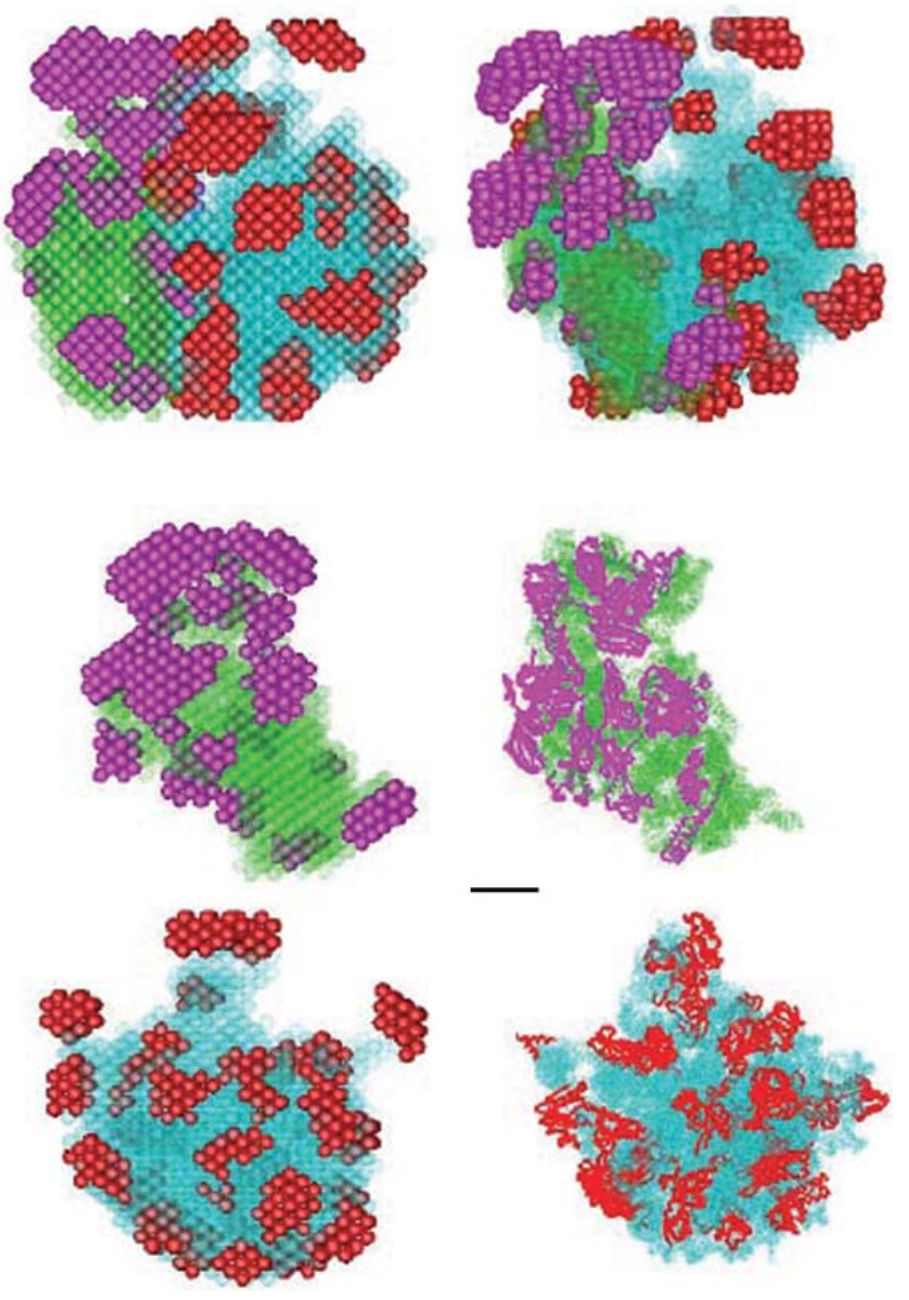

Fig. 17. The dummy atoms model of the 70S ribosome. Magenta and red, proteins in the 30S and 50S respectively; green and cyan, in the 30S and 50S respectively. The sphere radius is $r_0 = 0{\cdot}5$ nm. The spheres representing proteins are displayed as solids, those representing RNA as semi-transparent bodies. Top row, entire 70S (the left view is rotated counterclockwise by 45° around the vertical axis). Middle and bottom rows provide comparison of the protein–RNA distribution in the ribosomal subunits (right) with the later high-resolution crystallographic models (right). Middle row, crystallographic model of *Th. thermophilus* 30S subunit, resolution 0·33 nm (Schluenzen *et al.* 2000). Bottom row, 50S subunit from *H. marismortui* (Nissen *et al.* 2000), resolution 0·24 nm (note that the peripheral proteins L1 and L7/L12 are not seen in the crystal).The solvent view is displayed for both subunits (bar length, 5 nm).

occupancy of active sites by the substrates. The model is entirely determined by two parameters, the quaternary structure equilibrium constant $L = [T]_0/[R]_0$ of the unliganded enzyme and the partition coefficient $c = k_R/k_T$. It is most often contrasted with the KNF model (Koshland *et al.* 1966), which was proposed almost simultaneously and postulates that the tertiary structure of a protomer is modified upon ligand binding, altering the interaction with adjacent protomers, thereby changing the affinity of their active site for substrates. More recent models combine some of the features of the MWC model for the transition with local KNF-type interactions (Ackers *et al.* 1992). In terms of structure, the MWC model with its postulated symmetry proposes a concerted transition without intermediates, while a whole spectrum of intermediates is predicted in the incremental site by site KNF model. The regulatory properties of ATCase have been interpreted by some groups in terms of the MWC model (Howlett *et al.* 1977) while some other groups have reported observations incompatible with this model and put forward an alternative model combining a two-state equilibrium for the homotropic cooperativity and a local, incremental effect of ATP on the affinity of the nearest active site (Tauc *et al.* 1982).

Crystal structures of the unliganded enzyme (T state) and of the enzyme complexed with *N*-phosphonacetyl-L-aspartate (PALA) (R state) have been determined by Lipscomb's group (Ke *et al.* 1988; Stevens *et al.* 1990). PALA is a transition state analogue with a very high affinity for the active site ($K_D = 2 \times 10^{-7}$ M for the isolated catalytic trimer). The two structures are shown in Fig. 18 and display very conspicuous differences, which, beyond numerous local changes, can be described in terms of rigid-body movements of the catalytic and regulatory subunits. Thus, the two catalytic trimers move apart by 0·54 nm each (total 1·08 nm) along the threefold axis and rotate in opposite directions by 6° each around this axis, while the regulatory dimers rotate by 15° around the quasi-twofold axis to span the larger distance between the two trimers (Stevens *et al.* 1991). Each chain comprises two domains, the CP domain and the aspartate domain for the c chain, the Zn domain, which contains one structural Zn atom, interacting with the c chain and the allosteric domain with the nucleotide binding site for the r chain. The active site is located in the cleft between the two domains of the c chain which closes in the R state. In contrast, the angle between the two regulatory domains opens up in the R state. Finally, a loop comprising amino acids 230–245 from the c chain, the 240s loop, involved in inter trimer c–c and in r–c contacts and indicated by arrows in Fig. 18, changes both its conformation and its position in the R state, and has been proposed to play a key role in the T to R transition (Ladjimi & Kantrowitz, 1988).

Such a large-amplitude conformational change could be expected to translate into changes in the SAXS pattern. The scattering patterns of the two forms are shown in Fig. 19. The radius of gyration increases by about 5% from T to R while the largest dimension $D_{max} = 14$ nm is essentially unchanged, in agreement with the crystal structures (Moody *et al.* 1979). But, much more interesting are the conspicuous minima and maxima seen in the T pattern in the s range $0{\cdot}6\ \text{nm}^{-1} < s < 3\ \text{nm}^{-1}$: the first one undergoes changes in intensity and position while the second one near $1{\cdot}8\ \text{nm}^{-1}$ is almost absent from the R pattern. The scattering pattern therefore provides a sensitive and specific probe of the quaternary structure of ATCase (Hervé *et al.* 1985), which has been used to study several aspects of the structural basis of the regulatory properties of ATCase, some of which will be presented below.

In view of some critical issues regarding conformational transitions of allosteric enzymes addressed below it should be stressed at this point that, even going strictly by the Shannon theorem (see Section 4.2 above), SAXS patterns typically recorded out to $s_{max} = 3\ \text{nm}^{-1}$ are described by about 14 independent Shannon channels, i.e. at least 14 independent structural

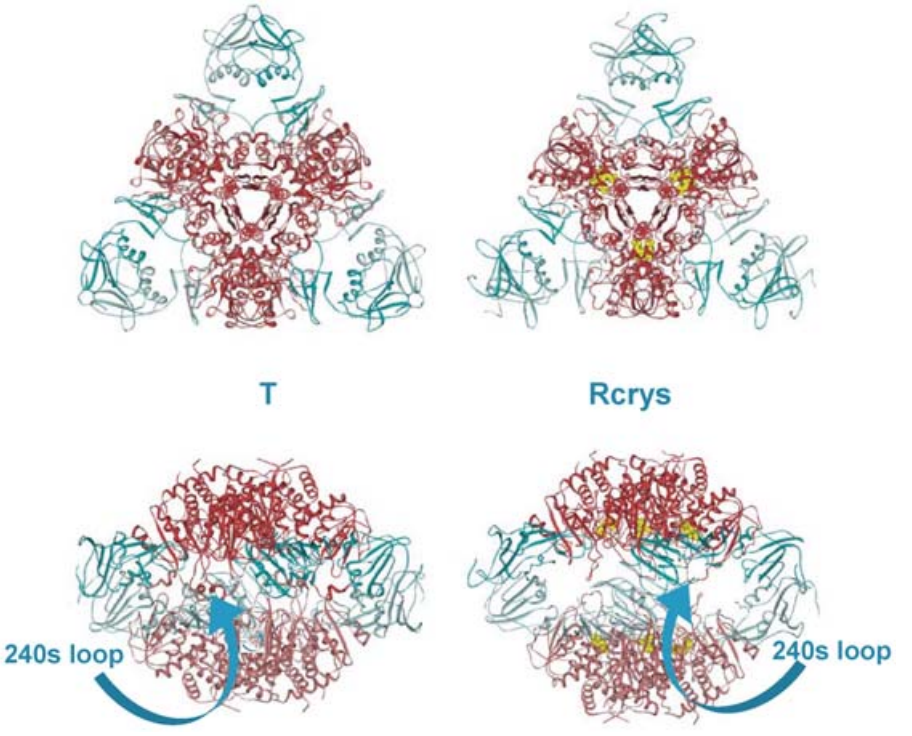

Fig. 18. Ribbon drawings of ATCase viewed down the threefold axis (top row) and along one of the twofold axes (bottom row). The T state (PDB entry 6at1) is shown in the left column, the crystal structure of the R state (PDB entry 1d09) in the right column. The arrows point to one of the 240 s loops of the catalytic chain (see text for details). The catalytic trimers are in red, the regulatory dimers in blue; PALA molecules are in yellow space-filling representation. The figure has been prepared using the freeware WebLab ViewerLite™ v. 4.0 (Molecular Simulations Inc.) (After Fetler & Vachette, 2001.)

parameters can be extracted from such patterns. This information content, though limited in comparison with that of high-resolution methods, is much higher than that of hydrodynamic methods, such as sedimentation velocity measurements, widely used to detect conformational changes although they yield only a single parameter.

Since crystal structures and scattering patterns in solution are available, it comes immediately to mind to compare the quaternary structures in both environments by calculating the scattering pattern from the crystal coordinates. Early attempts using rudimentary calculations concluded from the similarity of the scattering patterns to the similarity of the quaternary structures, thereby validating the changes observed in the crystal (Altman *et al.* 1982; Rey & Dumas, 1984). Indeed, crystals were grown at a pH at which the enzyme was not active. The availability of the CRYSOL program (Svergun *et al.* 1995) and of much improved experimental patterns prompted a fresh attack on the question. The calculated and experimental patterns of the unliganded enzyme are in excellent agreement (Fig. 19, bottom curves), suggesting that the quaternary structure is the same in both crystal and solution. In contrast, large deviations, in excess of twenty times the experimental error, are observed between the calculated and experimental R patterns (Fig. 19, top curves), reflecting significant differences between the two quaternary structures in the crystal and in solution (Svergun *et al.* 1997a). The experimental curve of the R state was fitted using rigid-body movements of the c_3 and r_2 subunits in keeping with the crystallographic description of the

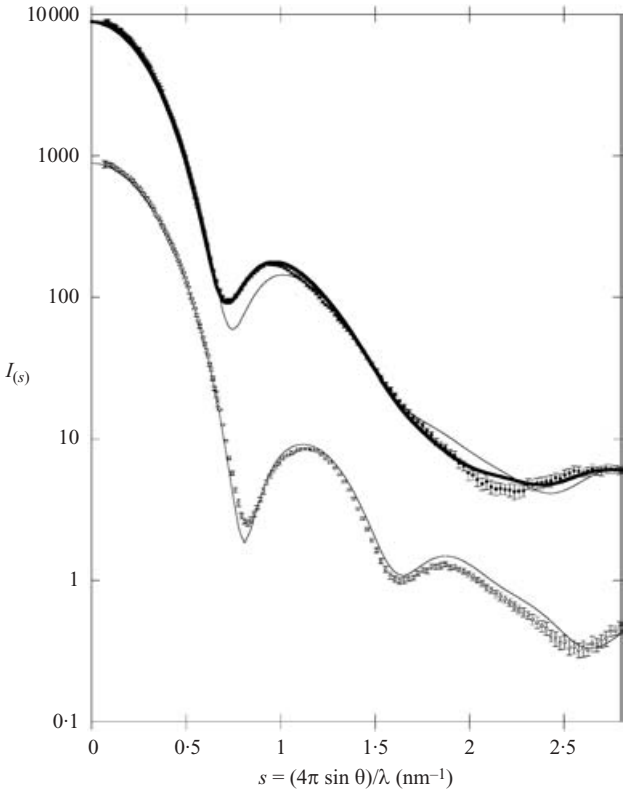

Fig. 19. Comparison of the experimental X-ray scattering patterns of ATCase (dots with associated error bars) with the scattering curve calculated from the atomic structures of the T and R states (thin lines). The curve from the solution model of the R state is shown in thick line. The curves for the T state have been divided by 10 for the sake of clarity.

transition (Fig. 19, top curves). Taking the crystal R structure as a reference, the distance between the catalytic trimers along the threefold axis increases by 0·34 nm (from 1·08 to 1·42 nm with respect to the T structure) and the trimers rotate by 8° around the same axis, while the regulatory dimers rotate by 9° around the corresponding twofold axis (Fig. 20). The crystal structures used in this work were lacking the last seven N-terminal residues from the r chain. Later calculations using higher resolution structures with complete r chains (Jin *et al.* 1999; Kosman *et al.* 1993) essentially confirmed the initial findings with a slightly smaller increase in the distance between the two catalytic trimers (0·28 *versus* 0·34 nm) (Fetler & Vachette, 2001). As allosteric enzymes have been selected for easy reorganization upon ligand binding, involving only low-energy non-covalent interactions, it is not surprising that the crystal packing forces, which also originate from non-covalent interactions between neighbouring molecules, could distort these subtle architectures.

Similar differences, though of smaller amplitude, have been reported over the last years on a variety of systems, mostly large multimeric proteins (Svergun *et al.* 2000b; Nakasako *et al.* 2001). The deformability of these complexes seems to be related to the area of the interfaces between subunits (Svergun *et al.* 2000b). A different and interesting case is that of the small adaptor protein Grb2 comprising an SH2 domain flanked by two SH3 domains which has been shown,

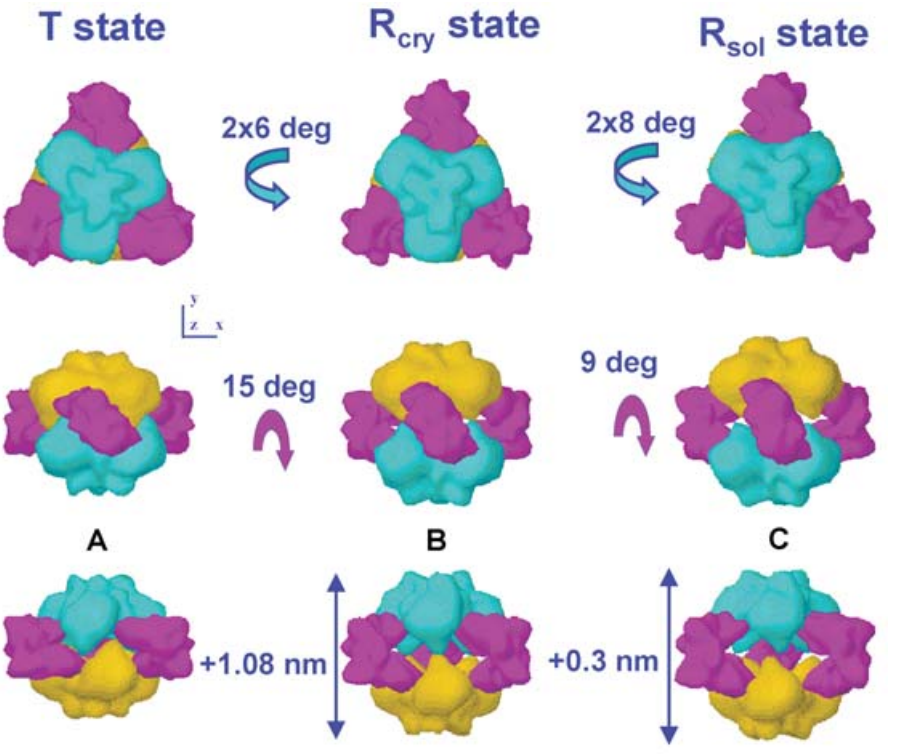

Fig. 20. Models of ATCase quaternary structure 'as seen' by X-rays at low resolution in the T state (left), and in the R state both in the crystal (centre) and in solution (right). The two c trimers are coloured in yellow and cyan, the three r dimers in magenta. Top, view along the threefold axis; middle and bottom, views along one of the twofold axes (top view rotated along x axis by 90° counterclockwise and clockwise respectively). The domains are represented by the envelope functions evaluated by the program CRYSOL. Axial unit, 2 nm.

using a combination of NMR and SAXS, to explore in solution an ensemble of conformations with relatively open structures (Yuzawa *et al.* 2001) as compared to the compact crystal structure (Maignan *et al.* 1995). Structural differences between solution and crystal conformations are likely to be observed with increasing frequency in the case of allosteric enzymes, complexes or larger assemblies. In such cases, the high-resolution crystal structure will best be combined with solution work to validate the gross conformation of the molecule.

Beyond the actual quaternary structure, the main objective of SAXS studies is the conformational transition. Indeed, the scattering from a mixture of conformations is the sum of the various contributions weighted by their respective fractional concentration. Titration experiments can be performed, in which a series of patterns (of the order of 20 curves) are recorded at different substrate concentrations. The whole data set is then analysed (Fetler *et al.* 1995a) using SVD (see Section 3.4.2). The number of eigenvectors, which also represents the minimum number of conformations involved in the transition, is a key parameter in trying to discriminate between a concerted and a progressive transition. Some representative curves from such an experiment performed with PALA are shown in Fig. 21. The arrows point to three crossing points common to all curves, a feature very suggestive of an all-or-none transition. This is confirmed by the SVD analysis showing that all experimental curves can be approximated within

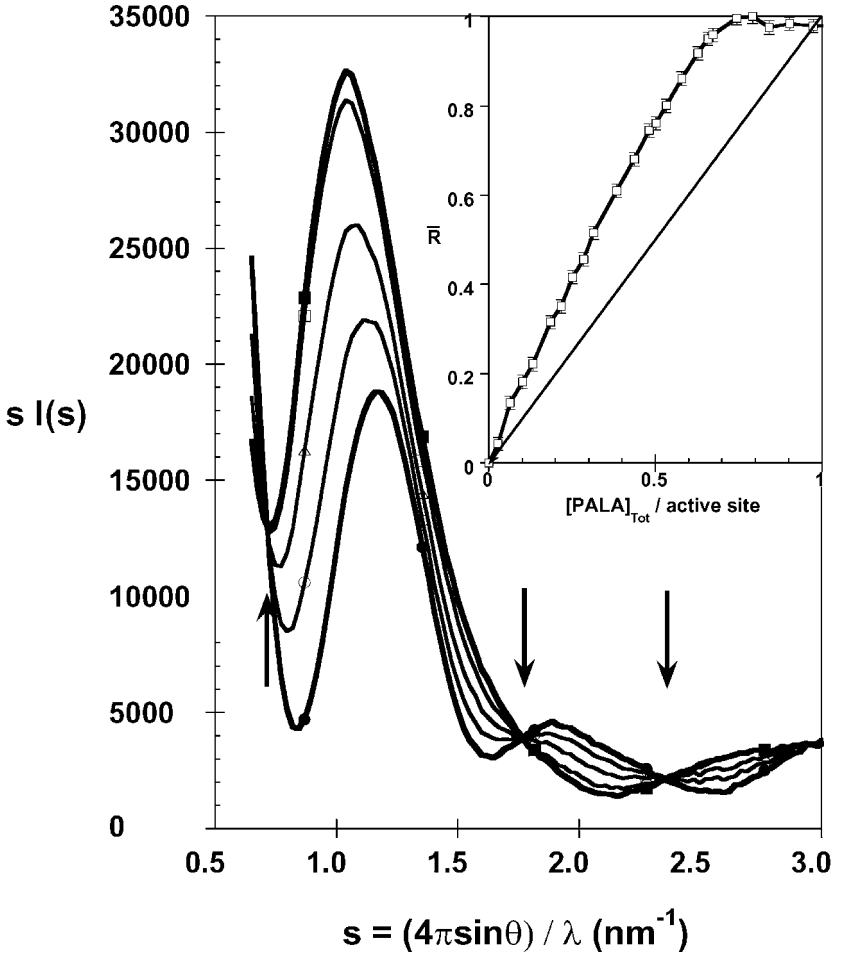

Fig. 21. X-ray scattering patterns of ATCase in the presence of increasing PALA concentrations expressed in moles of PALA mole^{-1} of active site [(●) 0; (○) 0·20; (△) 0·40; (□) 0·60; (■) 2]. The arrows point towards three crossing points common to all curves (after Fetler *et al.* 1997). Inset: titration curve of the structural transition (see text for details).

experimental error by a linear combination of the first two eigenvectors. The transition triggered by PALA only involves two quaternary structures, T and R, with no detectable intermediate species. This supports the view of an equilibrium between the T and R forms of the MWC model. In the simple case of a two-state transition, the SVD analysis directly yields the concentration fraction of each form. The ensuing titration curve of the transition, i.e. the fraction of molecules in the R state *versus* active site occupancy, is represented by the thin line in the inset to Fig. 21. All molecules appear to be in the R state when only two-thirds of the active sites are occupied (Fetler *et al.* 1995a). Similar experiments were performed in the presence of saturating amount of CP using succinate, a non-reactive analogue of aspartate, known to induce the co-operative transition (Fetler *et al.* 1997). Here also, only two quaternary structures are detected. This experiment is, however, best viewed as a titration of the complex of ATCase with CP by succinate. Small but significant differences were seen when comparing the pattern of the unliganded enzyme (T state) with that of the ATCase–CP complex (Fetler *et al.* 1997). An SVD

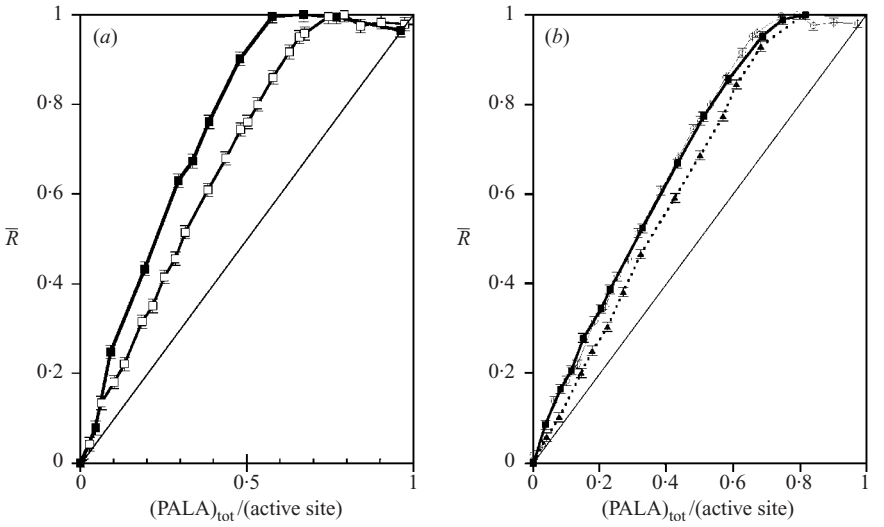

Fig. 22. Titration curves ($\overline{R}$ *versus* total PALA concentration $[PALA]_{tot}$) per active site. (*a*) PALA alone (○ and thin line); PALA + 5 mM CP (■ and thick line). The lines simply join the successive data-points. The diagonal corresponds to the fractional occupancy of active sites by PALA. (After Fetler *et al.* 1997.) (*b*) PALA alone (○ and thin line); PALA + 5 mM ATP (■ and thick line); PALA + 5 mM CTP (▲ and dashed thick line). The lines simply join the successive data-points. The diagonal corresponds to the fractional occupancy of active sites by PALA. (After Fetler *et al.* 1995.)

analysis established that this change was not due to a shift in the T ↔ R equilibrium but that a third quaternary structure was involved. Therefore, CP modifies the quaternary structure of ATCase to a T′ state, close to, but different, from T. A titration experiment by PALA in the presence of saturating amounts of CP also follows a two-state process, but the titration curve is systematically shifted by 10–15% towards R (thick line in Fig. 22*a*). In conclusion, CP changes the quaternary structure of ATCase to T′, a structure more readily converted to R by PALA (Fetler *et al.* 1997). The facilitating effect of CP had already been detected using sedimentation velocity measurements, but this was interpreted as an indication that CP was shifting the T ↔ R equilibrium towards R, with L changing from 250 to 7 (Howlett *et al.* 1977; Howlett & Schachman, 1977). This illustrates the advantage of the higher information content of SAXS compared to one-parameter hydrodynamic methods.

The study of the mechanism of action of nucleotide allosteric effectors such as CTP and ATP is another case in point. On the basis of sedimentation velocity measurements, it had been proposed that CTP (resp. ATP) inhibits ATCase by shifting the T ↔ R equilibrium towards T (resp. R), in agreement with the MWC model (Howlett *et al.* 1977; Howlett & Schachman, 1977). Titration experiments with PALA were performed in the presence of saturating amounts of CTP and ATP. The transition appears to be concerted in both cases, and the titration curve in the presence of CTP indicates that, over most of the transition, the fraction of molecules in the R state observed at a given occupancy of the active sites by PALA is about 6% lower in the presence than in the absence of CTP (Fig. 22*b*). This confirms the result of the earlier hydrodynamic study. In contrast, no corresponding effect is observed with ATP: the two titration curves with and without ATP are strictly superimposable, proving that ATP does not modify the T ↔ R equilibrium (Fig. 22*b*) (Fetler *et al.* 1995a). Therefore, the MWC model does not account for the activation by ATP. More recently, this question has been reinvestigated to

understand the origin of the differences between the modification in sedimentation velocity coefficient and the absence of effect in SAXS (Fetler & Vachette, 2001). The differences in pH (7.0 *versus* 8.3) and buffer (phosphate *versus* Tris) were shown to have no effect on the scattering pattern. It was reported, however, that the effects of ATP in the sedimentation velocity experiments were clearer when using the Mg salt (Foote & Schachman, 1985). The effects of Na- and Mg-ATP on the scattering pattern were thus compared, and Mg-ATP was found to significantly modify the scattering pattern of ATCase, even at saturating concentration levels of PALA (R state), while the only effect of Na-ATP is to slightly fill up the first minimum of the scattering pattern. Interestingly, while the pattern recorded at subsaturating PALA concentration is a perfect combination of the T and R pattern (concerted transition), neither of the patterns recorded in the presence of Na-ATP or Mg-ATP can be approximated in this way. In contrast, the two patterns can be very well approximated by a combination of the two extreme patterns recorded in the presence of the same form of the nucleotide. Furthermore, the coefficients of the linear combination are the same and are equal to those for the corresponding curve without nucleotide. Several important conclusions can therefore be drawn from these experiments: (i) the T to R transition is always concerted; (ii) neither Na-ATP nor Mg-ATP alters the T↔R equilibrium; (iii) the minor modifications of the SAXS pattern observed with Na-ATP are due to the contribution of ATP to the scattering and do not imply any conformational change; (iv) in contrast, Mg-ATP modifies the quaternary structure of the R state. The sedimentation velocity results can now be reconciled with those of SAXS: there is indeed a change in the sedimentation velocity coefficient (scattering pattern) in the presence of Mg-ATP, which is, however, not due to a shift of the T↔R equilibrium towards R, but to a change in the quaternary structure of the R state (Fetler & Vachette, 2001). The hydrodynamic measurement, yielding only one parameter, could only be interpreted as a variation along the T↔R coordinate, while the higher content in structural information of the scattering pattern allows to discriminate unambiguously between this and a modification of the R state quaternary structure.

Using a similar approach to that used for modelling the quaternary structure of the R state in solution, a model has been proposed for this new quaternary structure in the presence of both PALA and Mg-ATP. The structural units considered as rigid bodies here are the dimer of allosteric domains and the catalytic trimers together with the Zn domain of the regulatory chain. This is based both on the observation in the T to R transition in the crystal of an opening of the angle between the two regulatory domains (Ke *et al.* 1988) and on a TLS analysis of the *B* factors of the R structure indicating that the dimers of allosteric domains are moving as a rigid body, with the catalytic trimer with the Zn domain forming another dynamic unit (Tanner *et al.* 1993). The distance between catalytic trimers in the resulting model is 0·16 nm larger than in the structure of the R state in solution with a further 5° rotation around the threefold axis, while the allosteric domains rotate by an additional 6° around the twofold axis, differences of significant amplitude. On the basis of the crystal structure of ATCase in the presence of ATP (likely to be the Na form), a proposal has been put forward to account for the different effects of the two forms of the nucleotide (Fetler & Vachette, 2001).

A number of point mutants suggested by the analysis of the crystal structure of the T and the R states and further selected on the basis of their actual functional consequences were also studied by SAXS. Without entering into a description of all the observations made on these mutant enzymes, it suffices to mention that many different cases were encountered. Some mutants only displayed very specific and minor differences with the wild-type enzyme

(e.g. Cherfils *et al.* 1989), while some others like E239Q–ATCase were shown to adopt a different structure than the wild type when unliganded (Tauc *et al.* 1990). The mere addition of CP, which has the small effect reported above on the wild type (Fetler *et al.* 1997), was shown to convert practically the entire population of E239Q–ATCase to the R state, thereby explaining the absence of cooperativity for aspartate of this mutant. One mutant enzyme with reduced but significant cooperativity for aspartate was shown to adopt the R conformation in the presence of saturating concentrations of PALA, but, in contrast to the wild-type enzyme (Tauc *et al.* 1994), not in the combined presence of CP and succinate, an analogue of aspartate. Time-resolved SAXS measurements made it possible to study the transient steady state during catalysis using the physiological substrates in saturating amounts. The steady state appears to be a mixture of 60% T and 40% R form, which further converts entirely to R in the additional presence of ATP, thereby explaining the cooperativity for aspartate binding and the stimulation by ATP at saturating concentrations of substrates (Tsuruta *et al.* 1998c). One (and up to now unique) mutant, D162A–ATCase, has been shown to be unable to convert to the R state in the presence of very high PALA concentration while displaying a strong response to free ATP together with a synergistic effect of the two ligands (Fetler *et al.* 2002). Finally, the unliganded form of a particularly interesting mutant, D236A–ATCase, has been shown to be in equilibrium between two structures as revealed by the variations of the scattering pattern with temperature (L. Fetler & P. Vachette, unpublished observations).

A particular feature of the catalytic chain, the 240s loop has been proposed to play a key role in the T to R transition (Ladjimi & Kantrowitz, 1988). In order to test this proposal, a triple mutant was constructed. First, the two tryptophans of the catalytic chain were replaced by phenylalanines and in this tryptophan-free context Tyr240 was substituted by tryptophan. The resulting enzyme displayed enzymic characteristics very close to that of wild-type ATCase. A SAXS titration experiment with PALA was performed, which indicated that the transition was concerted and yielded a titration curve identical to that of wild-type ATCase, with a transition already complete when four out of six active sites are occupied by PALA. In parallel, the fluorescence emission spectrum displays a marked red shift upon PALA binding, indicative of an increased exposition of tryptophan to the surrounding solvent. Titration of this red shift reveals a linear variation with the active site occupancy, the shift being complete only when all sites are occupied (Fig. 23). Comparison of these two results allowed to conclude that, far from being a key step of the T to R quaternary structure transition, the loop changes conformation upon ligand binding at the nearest active site (Fetler *et al.* 1995b).

Taking advantage of the high flux from synchrotron radiation X-ray beams, time-resolved measurements were performed using the SAXS beamline BL15A at the Photon Factory, a second-generation machine (Wakabayashi & Amemiya, 1991), and a stopped-flow apparatus specifically designed for SAXS measurements and fast mixing of viscous solutions (Tsuruta *et al.* 1989). Indeed, the expected rate of the conformational change [typical times of the order of 10 ms at 4 °C (Kihara *et al.* 1984)], left no hope of following the transition at room temperature. Experiments were thus performed in 20–30% ethylene glycol at -5 to -10 °C, conditions in which the enzyme was shown to be still cooperative and sensitive to allosteric effectors. Using succinate, an aspartate analogue, series of time-resolved scattering patterns were recorded at various succinate concentrations which, like the titration experiments, were indicative of a concerted T to R transition, no intermediate species being detected during the course of the conformational change (Tsuruta *et al.* 1998c). The apparent rate constant of

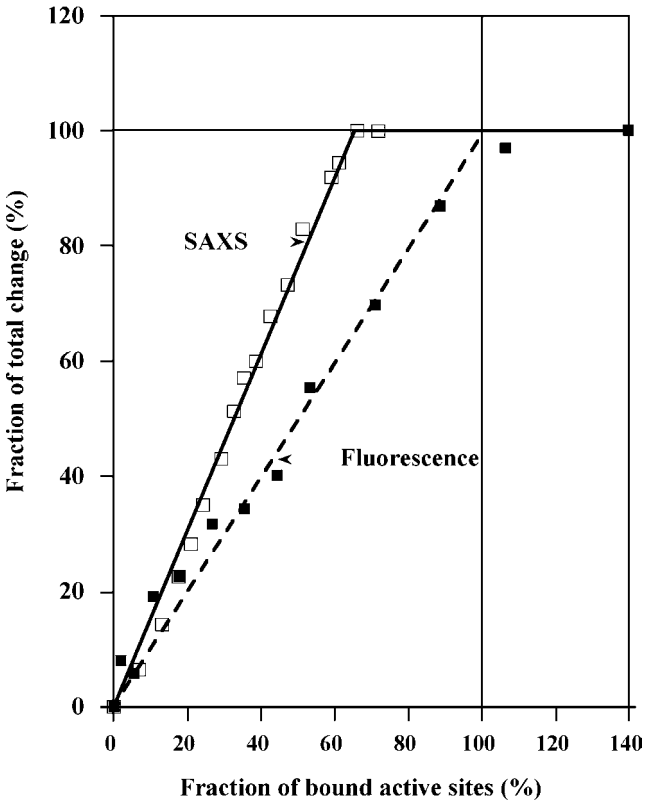

Fig. 23. Titration curves of F209F284W240-ATCase. X-ray titration ($\bar{R}$ *versus* total PALA concentration $[PALA]_{tot}$) (□ and continuous line); fluorescence spectroscopy titration (■ and dashed line).

the structural change from T to R varied between 0·05 and 3 s^{-1} and its dependence on succinate concentration could be accounted for by a simplified kinetic MWC-type model, yielding equilibrium and kinetic constants in reasonable agreement with previous work performed using spectroscopic probes (Hammes & Wu, 1971). Using equal amounts of the physiological substrates CP and aspartate, the enzyme is seen to convert from T towards R, to reach a steady-state plateau during which substrates are consumed before reverting to the T form (Tsuruta *et al.* 1990). Above a certain concentration, the enzyme entirely converted to the R state. The effect of nucleotides in the presence of the physiological substrates CP and aspartate was studied. While CTP does not seem to significantly modify the apparent rate constant of the T to R transition, ATP increases it, a result which could not be obtained directly by any other method.

This SAXS study of ATCase illustrates the various applications of the method to a biochemical problem. Beyond the elucidation of several conformations, the method has allowed a more complete characterization of the structural transition. It has also made it possible to settle some pending issues of direct functional relevance by providing a direct monitoring of the enzyme conformational behaviour, albeit at low resolution. Furthermore, very high flux instruments at third-generation sources and faster area detectors have considerably increased the accessible time resolution. The kinetics of the structural transition of ATCase is currently being revisited (H. Tsuruta & E. R. Kantrowitz, personal communication).

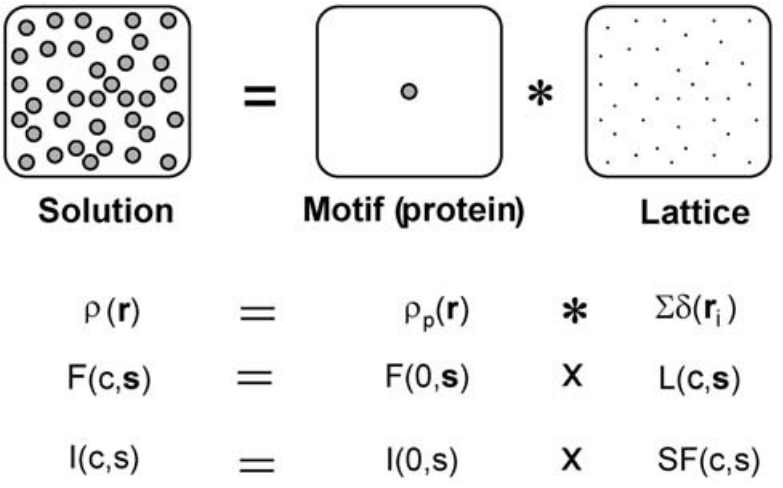

Fig. 24. Representation of a solution of macromolecules as the convolution of the excess scattering density of the particles and a distribution of delta functions.

6. Interactions between molecules in solution

The previous discussion has essentially been restricted to scattering by ideal solutions of macromolecules where the experimental scattering from the solute is proportional to the intensity scattered by a single particle. This is also what justified the direct retrieval of structural information from small-angle scattering data. In real solutions, interaction forces of various origins influence the time-averaged spatial distribution of the particles and hence also the scattering. This can be clarified by representing a solution of macromolecules as the convolution of a motif, the excess scattering density $\rho(\mathbf{r})$ associated with the particle, and of a distribution of delta functions $\delta(\mathbf{r}_i)$ corresponding to the centers of the individual particles as illustrated in Fig. 24. If the particles can be considered to be spherical on the scale of their average separation, the general expression of the intensity scattered by the solution can be factored in two parts, the first one corresponding to the shape of the particles and the second one reflecting their spatial distribution, as expressed in Eq. (3.2). This equation has been shown to be valid for proteins albeit in a restricted s range, in the case of quasi-spherical particles and weak or moderate interactions (Vérétout *et al.* 1989; Tardieu, 1994). In conventional structural studies the structure factor is regarded as a nuisance to be minimized and ways of detecting and accounting for weak or moderate interactions are presented below. Considered in a different light, the structure factor obtained in experiments where interactions make a significant and occasionally even major contribution to the scattered intensity provides a unique source of information about the forces determining the spatial distribution of macromolecules in solution. For a systematic study of the influence of physicochemical parameters on the structure factor, the form factor must, however, either be known beforehand or experimentally determined in preliminary measurements.

The systematic analysis of the interactions of biological macromolecules in solution, initiated about 20 years ago with initially only qualitative conclusions has now reached a semi-quantitative level where parameterized potentials are used to account for an increasing proportion of actual experimental situations. The main results from recent work presented below are illustrated by examples covering the complete range from the benchmark protein lysozyme to viruses.

6.1 Linearizing the problem for moderate interactions: the second virial coefficient

The X-ray structure factor at the origin is related to the osmotic pressure of the solution by the following relationship:

$$\mathrm{SF}(c, 0) = (RT/M)(\partial\Pi/\partial c)^{-1}, \tag{6.1}$$

where $R = 8{\cdot}31$ J mol^{-1} K^{-1} is the gas constant and M the molecular mass of the solute in Daltons. The osmotic pressure can be expanded as a power series of the concentration c:

$$\Pi/cRT = 1/M + A_2 c + A_3 c^2 + \cdots. \tag{6.2}$$

When the interactions are weak and the protein concentration low, the linear approximation is valid and hence

$$1/\mathrm{SF}(c, 0) = 1 + 2MA_2 c. \tag{6.3}$$

In the case of net repulsive interactions, the particles tend to be evenly distributed while attractive interactions lead to large fluctuations in their distribution with particle-rich regions surrounded by depleted regions. Accordingly, the osmotic pressure is higher (repulsion) or lower (attraction) than in the ideal case, the structure factor at the origin has a value above or below 1 and A_2, the second virial coefficient, will be positive or negative, as illustrated in Fig. 4. The virial coefficient depends on temperature and other physico-chemical conditions and changes sign from negative to positive under the so-called 'theta conditions' in which all attractive and repulsive interactions cancel out.

In order to eliminate the effect of interactions on the scattering patterns the measurements are usually performed at several concentrations, and extrapolation of the scattering curve to zero concentration yields the 'ideal' value of the intensity at the origin $I_{\mathrm{ideal}}(0)$. In practice, it is usually possible to closely approach ideality by working at sufficiently low concentrations and adjusting the physico-chemical parameters like pH and ionic strength. Indeed, provided the necessary precautions are taken to minimize the background with small-angle scattering instruments on storage rings, the scattering from solutions with a few mg ml^{-1} of 10 kDa proteins and of a fraction of mg ml^{-1} of 200 kDa proteins can be satisfactorily measured. With neutrons, the absence of instrumental scattering at low s values and the use of longer wavelengths more than compensate for the lower flux. As an example, the scattering patterns of solutions of nucleosome core particles in a 10 mM Tris buffer (pH 7·6), with 15 mM NaCl are shown in Fig. 25*a*. Under these low ionic strength conditions, the interactions between particles are repulsive as indicated by the decrease in intensity with increasing concentration. Since the aim of the study was to investigate shape changes as a function of ionic strength, the particle concentration had to be sufficiently low to allow extrapolation to zero angle ($I(0, c)$ and subsequently, to infinite dilution to obtain $I_{\mathrm{ideal}}(0)$ (Fig. 25*b*) (Mangenot *et al.* 2002).

In the more general case where there are attractive interactions and the polydispersity of the solute depends on its concentration the generalized indirect Fourier transformation (GIFT) can be used to obtain the structure factor for spherical particles, whereas for non-spherical particles (e.g. rods) an effective structure factor is obtained (Brunner-Popela & Glatter, 1997; Weyerich *et al.* 1999).

For thin rods like DNA at low ionic strength, the length distribution has little influence on the structure factor (Koch *et al.* 1995). In the dilute regime the position of its first maximum, which essentially depends on the centre to centre separation between rigid fragments, varies like the

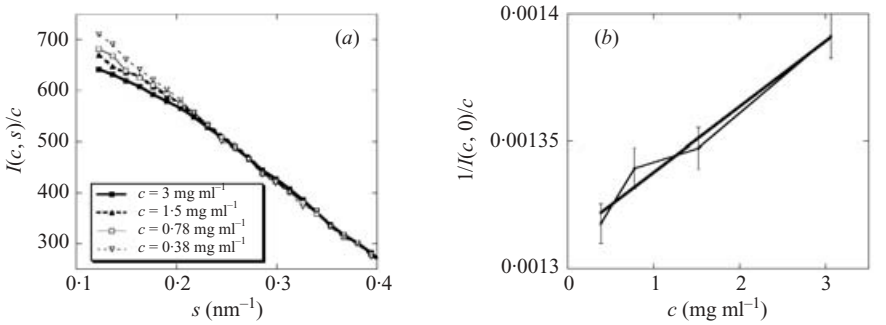

Fig. 25. (*a*) Scattering curves of nucleosome core particles in a 10 mM Tris buffer (pH 7·6) with 15 mM NaCl, as a function of concentration; (*b*) plot of $1/I(0, c)$ *versus* c derived from the curves in (*a*). The linear regression yields a value of $4{\cdot}78 \times 10^{-5}$ mol ml g^{-2} for A_2. (Courtesy of D. Durand.)

square root of the concentration. At higher concentrations other effects also play a role as shown by extensive theoretical calculations (Weyerich *et al.* 1990). The length distribution has, however, a strong influence on the relaxation times observed in electric field scattering (Koch *et al.* 1988, 1995) and on the slow mode observed in dynamic light scattering (Skibinska *et al.* 1999).

6.2 Determination of the structure factor

As mentioned above, small-angle scattering studies of solutions can provide precious information on the interaction forces between macromolecules which can be of biological or physicochemical relevance. As an example, the SAXS study of interactions between eye lens proteins has shed light on the molecular basis for fundamental physiological properties of the eye lens like its transparency and the correction for some chromatic aberrations that had been initially investigated by light scattering (Benedek, 1971; Delaye & Tardieu, 1983; Vérétout *et al.* 1989). Nowadays, the main incentive behind the study of interactions between proteins is certainly crystallization, which remains the bottleneck in many structural studies, and any improvement of the yield of crystallization trials brought about by a better understanding of the process would have a major impact.

Previous studies of protein crystal growth essentially determined 'solubility curves', i.e. the line in a phase diagram separating the upper supersaturation region where crystals may grow from the lower, undersaturated region where macromolecules are soluble. The first SAXS studies of protein crystallization undertaken on lysozyme (Guilloteau, 1991; Ducruix *et al.* 1996) showed that salts known to induce crystallization of lysozyme turned interactions between macromolecules from repulsive to attractive, a conclusion also reached at about the same time by light-scattering studies on lysozyme (Muschol & Rosenberger, 1995). A series of subsequent SAXS studies (e.g. Gripon *et al.* 1997; Velev *et al.* 1998; Bonneté *et al.* 1999) led to the general conclusion that solubility increases when interactions are more repulsive and that, for crystallization purposes, it may be equivalent to determine solubility by lengthy experiments or directly measure interactions in solution using SAXS or light scattering. Based on light-scattering results, it had been proposed that proteins crystallize under conditions where the second virial coefficient is slightly negative, defining what was called the 'crystallization slot', corresponding to weak attractive forces between molecules (George & Wilson, 1994). This initial success contributed to turn light scattering into a routine tool for rapid screening of crystallization conditions. However,

SAXS, with its short wavelength, is not restricted to the origin of the scattering pattern and analysis of the s dependence of the structure factor can yield information, which cannot otherwise be unambiguously retrieved in a direct way. Numerical simulations, initially developed in liquid-state physics (Hansen & McDonald, 1986; Belloni, 1991) were applied to solutions of macromolecules. In these simulations only interactions between pairs of macromolecules are taken into account, although all interactions are mediated via the surrounding solvent and ions. Each interaction is represented by a Yukawa potential of the form

$$u(r)/k_{\mathrm{B}}T = J(\sigma/r)\exp(-(r-\sigma)/d), \tag{6.4}$$

which depends on the hard-sphere diameter σ, the potential depth (J in units of kT) and its range d (k_{B} is the Boltzmann constant). These parameters are determined by a trial-and-error procedure in which the structure factor is calculated for various combinations of values.

In brief, statistical mechanics models based on the Ornstein–Zernicke (OZ) and the hypernetted chain (HNC) integral equations are used to calculate SF(c, s) which is related by FT to the pair distribution function $g(r)$:

$$S(c,\ s) = 1 + \rho \int_0^\infty 4\pi^2 (g(r)-1)(\sin(rs)/rs)\mathrm{d}r, \tag{6.5}$$

where ρ is the number density of particles in the solution. The OZ relationship between the total and direct correlation functions $h(r) = g(r) - 1$ and $c(r)$ can be written using their FTs:

$$(1 + \mathrm{FT}(h(r))(1 - \mathrm{FT}(c(r)) = 1, \tag{6.6}$$

while the HNC equation is

$$g(r) = \exp[-u(r)/k_{\mathrm{B}}T + h(r) - c(r)], \tag{6.7}$$

where $u(r)$ is the interaction potential between macromolecules. The OZ equation is solved by iteratively using the closure relationship in Eq. (6.7) followed by the calculation of the structure factor using Eq. (6.5).

This approach was first applied to the study of a liquid–liquid phase separation (Ishimoto & Tanaka, 1977; Thomson *et al.* 1987) observed when lowering the temperature of solutions of small proteins like lysozyme or γ-crystallins below a critical value (Malfois *et al.* 1996). An attractive Yukawa potential was shown to account for the structure factor and the phase separation at low temperature. In both cases, a qualitative agreement between calculated and experimental structure factors was obtained for a value of σ yielding an excluded volume close to or equal to the protein dry volume, a mean value of 0·3 nm for the potential range and a potential depth around $-2{\cdot}7$ kT as illustrated in Fig. 26 (Malfois *et al.* 1996). Similar conclusions were reached using Monte-Carlo simulations (Lomakin *et al.* 1996).

This phenomenological potential was then confronted to the actual physical forces between molecules in solution. Three contributions were considered, as originally proposed by Derjaguin, Landau, Verwey and Overbeek in their study of colloids (the so-called DLVO potential) (Verwey & Overbeek, 1948): a hard-sphere term which accounts for the fact that proteins are mutually impenetrable, coulombic repulsion and van der Waals interactions. In the cases that were investigated, the coulombic repulsion was negligible since the pH was close to the isoelectric point for γ-crystallins and a high ionic strength providing efficient screening was used with lysozyme. Finally, calculations showed that the reported values of the Yukawa potential parameters

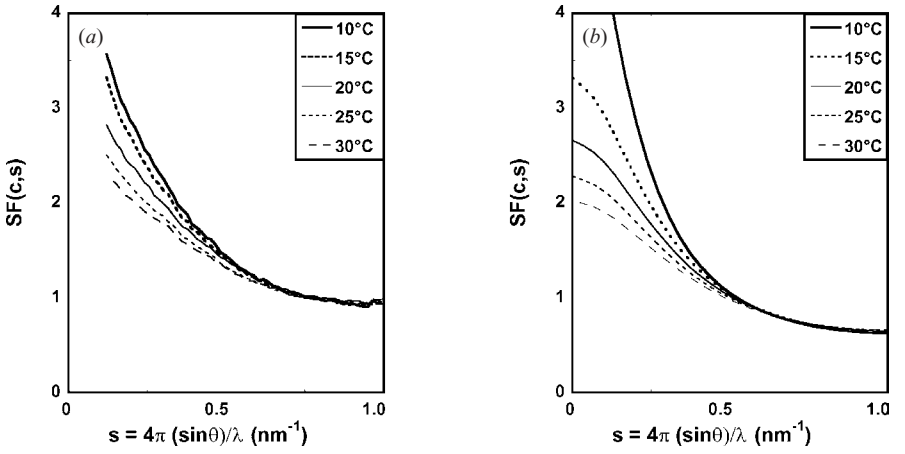

Fig. 26. (*a*) Experimental structure factors of γ-crystallins derived from the curves in Fig. 17*a*. (*b*) Theoretical structure factors calculated for a diameter of 3·6 nm, a potential range of 0·3 nm, and a potential depth of $-2{\cdot}63$ kT at 10 °C. The temperature series is obtained by modifying only the temperature. (After Malfois *et al.* 1996.)

correspond to the contribution from the van der Waals interactions. This confirms the validity of the approach, which met with similar successes in other studies of the dependence on pH, temperature, ionic strength and protein concentration of interactions between small proteins in solution (Tardieu *et al.* 1999). In the cases were coulombic repulsions could not be neglected they were also described by a Yukawa potential, with a depth J related to the charge of the protein Z_P by the following expressions:

$$J = (Z_p^2/\sigma) L_B/(1 + 0{\cdot}5\sigma/\lambda_D)^2 \tag{6.8}$$

with

$$\lambda_D = 1 \Big/ \sqrt{4\pi L_B \sum_i \rho_i Z_i^2}, \tag{6.9}$$

where the sum runs over all types of ions in the solution, ρ_i is the number concentration of each type of ion and L_B, the Bjerrum length, has a value of 0·72 in water at 300 K and $\lambda_D(\text{nm}) = 0{\cdot}3I^{-\frac{1}{2}}$, where I is the ionic strength.

At this point, one could believe that the DLVO potential would allow the *a priori* calculation of interactions between proteins at low ionic strength. The situation is unfortunately more complex. If the excluded volume is close to the dry volume in the case of small compact proteins, such is not the case for large oligomeric proteins or protein–nucleic acid (resp. protein–detergent) complexes in which it may be significantly larger. Furthermore, with large and less compact particles, the van der Waals contribution seems to vanish, for reasons which have not yet been fully elucidated, but which are associated with the presence of water in the excluded volume. Finally, the charge to be considered in the calculation of coulombic repulsions is an effective one, generally smaller than that calculated from the chemical composition of the particle and the pH of the solution, which can only be obtained experimentally (Tardieu *et al.* 1999). This suggests that interparticle interactions cannot be computed *a priori* but must be measured.

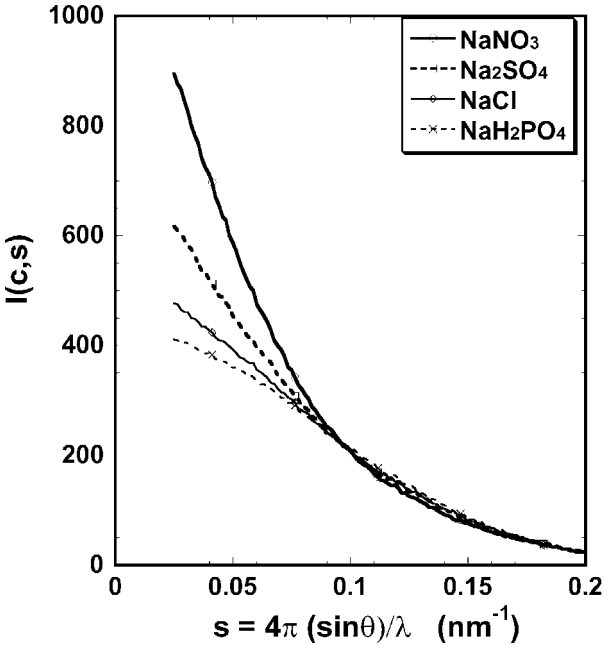

Fig. 27. Scattering patterns of 100 mg ml^{-1} solutions of lysozyme in 50 mM Na acetate buffer (pH 4·5) in the presence of various salts at a concentration of 150 mM. (Courtesy of J. P. Guilloteau & F. Bonneté.)

Other observations made during the course of these studies also underline the intrinsic limits of the approach. Indeed, neither the interaction potentials nor the calculation of the structure factors take into account departures from sphericity of the protein shape, as recently pointed out in a general model including short-range interactions, which vary over the protein surface (Lomakin *et al.* 1999). The charge distribution on the protein surface is also neglected in the expression of the potentials, while ions are represented by point charges. This might explain the failure to obtain a fully quantitative description of the structure factors and to account for specific effects associated with the nature of the added salts.

The effect of salt at concentrations below 0·2 M is essentially to screen coulombic repulsions thereby making the net interactions more attractive. However, at higher ionic strength, another attractive contribution appears which increases when lowering the temperature and depends on the nature of the anion (Fig. 27), following the order of the Hofmeister series (Hofmeister, 1888) (Fig. 28) or the reverse order depending on whether the pH of the solution is above or below the isoelectric point of the protein (Ducruix *et al.* 1996). This result obtained by X-ray as well as light scattering (Muschol & Rosenberger, 1995) is one of a number of observations displaying such a dependence (reviewed in Collins & Washabaugh, 1985; Cacace *et al.* 1997). The origin of this effect is still unclear, although it is thought to be associated with the ion distribution around the macromolecule (Ninham & Yaminsky, 1997).

Salts cannot always make the net interactions attractive, especially when dealing with large oligomeric proteins like ATCase (see Section 5.1). In those cases, crystallographers often resort to addition of polymers, mostly PEGs with molecular masses between 0·4 and 20 kDa. The addition of 20% (w/v) PEG with a molecular mass of 20 kDa to a solution of ATCase suffices to establish an attractive regime (Fig. 29*a*) (Budayova *et al.* 1999). Similar observations have been

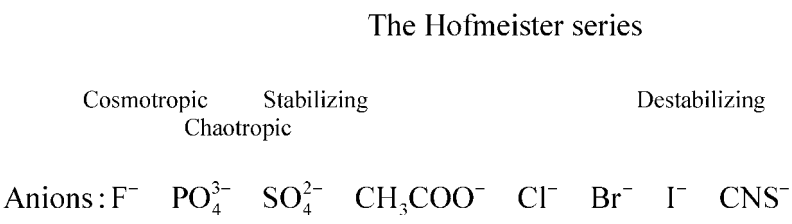

Fig. 28. Hofmeister series of anions. (After Cacace *et al.* 1997.)

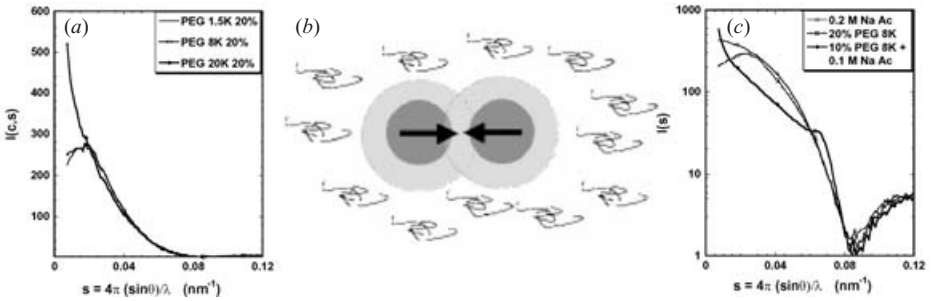

Fig. 29. (*a*) Scattering patterns of 30 mg ml^{-1} solutions of ATCase in borate buffer 10 mM (pH 8·3), $T = 20$ °C, in the presence of 20% PEG of various molecular masses (after Budayova *et al.* 1999). (*b*) Schematic representation of the origin of the depletion attraction induced by polymers. (*c*) Scattering patterns of 30 mg ml^{-1} solutions of ATCase in borate buffer 10 mM (pH 8·3), $T = 20$ °C, in the presence of 20% PEG 8K, of 0·2 M Na acetate and of the combination of 10% PEG 8K + 0·1 M Na acetate (after Budayova *et al.* 1999).

made with solutions of urate oxydase, a 128 kDa tetrameric enzyme (Bonneté *et al.* 2001), of α-crystallins, which are polydisperse oligomers with molecular masses between 800 and 900 kDa (Bernocco *et al.* 2001) and of brome mosaic virus (BMV), a 4·6 MDa plant virus with a $T = 3$ protein capsid and a mean diameter of 26·8 nm (Casselyn *et al.* 2001). Here, the molecular origin of the attraction is better understood. The effect is actually well-documented for colloids and best explained by considering a suspension of hard spheres in a solvent containing a much larger number of polymer molecules (Lekkerkerker, 1997). When the distance between the two particle surfaces is smaller than the diameter of the polymer chains, these chains are excluded from the region between macromolecules. To maximize the translational entropy of the polymer chains, the inaccessible volume between colloidal particles must be minimized. This results in an additional attraction between macromolecules, the so-called 'depletion force' (Fig. 29*b*). The range of this additional potential depends on the size of the added polymer. Numerical simulations were performed using a 'two component' approach to account for the coexistence in the solution of two types of macromolecules, the polymer and the protein, and of the corresponding three pair potentials. While temperature induced and salt induced attractions are always short range (a fraction of a nm), the depth and range of the depletion potential varies with the size and concentration of the polymer (Vivarès *et al.* 2002). Interestingly, the range of the potential can become comparable to the macromolecular dimensions, as observed in the case of urate oxydase in the presence of 8 kDa PEG (Vivarès *et al.* 2002). The existence of long-range interactions was predicted following a spectacular observation on ATCase, in the presence of 10% of 8 kDa PEG

and of 0·1 M Na acetate (Fig. 29*c*) (Budayova *et al.* 1999). The sharp upward curvature near the origin of the scattering pattern is the hallmark of attractive interactions, but the clear correlation peak around $s = 0{\cdot}6$ nm^{-1} is indicative of repulsive interactions. This was interpreted in terms of the formation in the solution of 'protein-rich, PEG-depleted' domains surrounded by 'protein-depleted, PEG-enriched' areas, which contributes to the scattering at very small angles, in prelude to phase separation. The protein concentration in the protein-enriched phase is significantly larger than the average concentration, with a correlative reduction in the average intermolecular distance, while the PEG concentration is low. Hard sphere and coulombic repulsions, which are incompletely screened at the salt concentration used, become therefore significant, giving rise to short range order and to the associated interference peak observed in the scattering pattern.

It might seem *a priori* unrealistic to follow the actual nucleation and growth of crystals from the solution using SAXS. In fact, during the first attempts crystals disappeared from the volume probed by the X-ray beam as soon as they became sufficiently large for gravity to overcome thermal agitation. Crystal growth was, however, shown to occur and was optically monitored in an agarose gel. Due to reduction of convection in the gel, larger and often less mosaic crystals were obtained (Robert & Lefaucheux, 1988). Using a specially designed sample holder, and conditions under which crystals grow over a period of hours, supersaturated solutions could be studied at various times and the appearance of crystals followed in lysozyme (Finet *et al.* 1998). Diffraction peaks are superimposed on the scattering pattern of the protein in solution and crystals grow by direct addition of lysozyme monomers, without preliminary formation of oligomers.

This is to be contrasted with a SAXS study of the growth of low pH crystals of bovine pancreatic trypsin inhibitor (BPTI) combining diffraction and SAXS in solution (Hamiaux *et al.* 2000). Crystallization of BPTI at acidic pH in the presence of thiocyanate, chloride and sulphate ions leads to three different polymorphs in $P2_1$, $P6_422$ and $P6_322$ space groups. The same decamer with 10 BPTI molecules organized through two perpendicular twofold and fivefold axes forming a well-defined compact object is found in the three polymorphs. This is at variance with the monomeric crystal forms observed at basic pH. The SAXS data recorded during crystallization of BPTI at pH = 4·5 in both undersaturated and supersaturated BPTI solutions were analysed in terms of the formation of discrete oligomers ($n = 1$–10). In addition to the monomer, a dimer, a pentamer and a decamer were identified within the crystal structure and their scattering patterns were computed using CRYSOL (Svergun *et al.* 1995). The experimental curves were then analyzed as linear combinations of these theoretical patterns using a nonlinear curve-fitting procedure. The results, confirmed by gel filtration experiments, unambiguously demonstrate the co-existence of only two types of BPTI particles in solution: monomers and decamers, without any evidence for other intermediates. The fraction of decamers increases with salt concentration (i.e. when reaching and crossing the solubility curve). This suggests that crystallization of BPTI at acidic pH is a two-step process whereby decamers first form in under- and supersaturated solutions followed by the growth of what are best described as 'BPTI decamer' crystals.

In conclusion, much information regarding interactions between macromolecules in solution can be derived from small-angle scattering. Although structure factors cannot be calculated *a priori*, phenomenological parameterized potentials can account for the experimental observations (Tardieu *et al.* 2002). A complete, quantitative agreement is not yet at hand, mostly due to our ignorance of the origin of the Hofmeister attraction. Progress is, however, being made regarding

the modelling of the depletion attraction associated with neutral polymers (Vivarès *et al.* 2002). Application of small-angle scattering methods to the study of macromolecular interactions has definitively come of age and significantly contributed to the knowledge about macromolecular interactions. Although small-angle scattering cannot compete with light scattering for routine large-scale screening of crystallization conditions, it is the only way to obtain structure factors from which interaction potentials can be derived.

7. Time-resolved measurements

Time-resolved experiments provided the initial motivation for using synchrotron radiation for X-ray diffraction and scattering. In parallel with applications to the study of muscle contraction (Huxley *et al.* 1980), the techniques were also applied to the study of assembly phenomena on time scales ranging between a few hundred milliseconds and minutes, involving mainly components of the cytoskeleton, virus capsids and multisubunit enzymes. Time-resolved experiments necessarily imply the investigation of mixtures and it was realized from the outset that this would require independent information from other sources (e.g. electron microscopy, spectroscopy, crystallography, light scattering, ultracentrifugation) and some of the early studies still provide good examples of attempts at obtaining consistent models (e.g. Bordas *et al.* 1983; Koch, 1989; Marx & Mandelkow, 1994). Experiments on the kinetics of conformational transitions in the multisubunit enzyme ATCase have been described in Section 5.2.

Recent examples of assembly studies include those on rapid large-scale protein quaternary structural changes in *Nudaurelia capensis* omega virus (Canady *et al.* 2001) and of the formation of intermediates during assembly and maturation the capsid of HK97 (Lata *et al.* 2000) or of tomato bushy stunt virus (TBSV) (Perez *et al.* 2000). Latest results obtained as part of this last study, shown in Fig. 30, illustrate the entirely new possibilities opened by third-generation sources for this type of experiments.

Time-resolved measurements can, of course, equally well be used to study transitions linked to changes in intermolecular interactions. The kinetics of two phase transitions induced by addition of PEG have been studied: the fluid–fluid phase separation of α-crystallins (S. Finet, T. Narayanan & A. Tardieu, personal communication) and the fluid–solid phase separation of BMV, which yields crystals. In this last system, microcrystals can already be detected after a few seconds under the experimental conditions chosen as illustrated in Fig. 31 (Casselyn *et al.* 2002). Interestingly, no amorphous intermediate is detected beside the isolated virus and microcrystals.

The rate of perturbation is often a limiting factor in time-resolved experiments and temperature (Bordas *et al.* 1983), pressure (Woenckhaus *et al.* 2001), concentration (Moody *et al.* 1980) and field jumps (Koch *et al.* 1988) as well as light flashes (Sasaki *et al.* 2002) have been used with varying degrees of success to trigger reactions or transitions.

Most advances in instrumentation are presently being made in the frame of investigations on protein and RNA (un)folding taking place at third-generation sources. Beyond its intrinsic relevance, the folding problem has attracted renewed interest partly due to its relationship to formation of amyloid fibrils (Lynn & Meredith, 2000) in various pathologies and the importance of misfolding in the expression of cloned proteins. A recent review (Doniach, 2001) illustrates the progress made in this area since the early feasibility studies (Phillips *et al.* 1988). Time resolutions in the 0·1–10 ms range are achieved in such experiments by using microfabricated rapid mixing devices with continuous flow and varying the distance between the fixed mixing point and the observation point (Pollack *et al.* 2001; Akiyama *et al.* 2002). For lower time

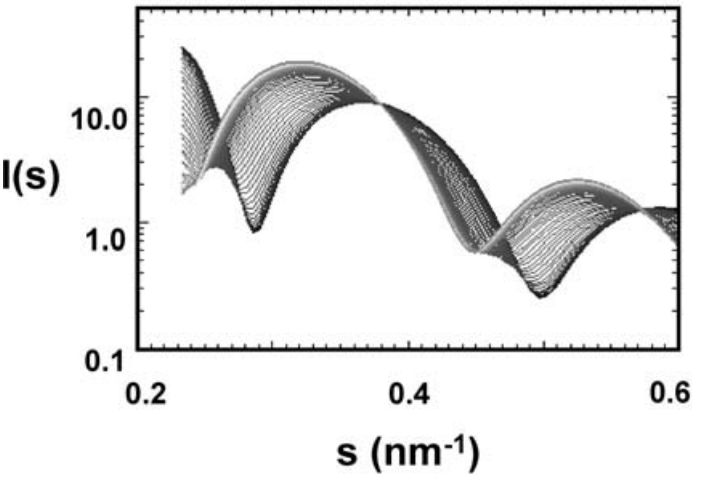

Fig. 30. Time-resolved SAXS patterns from a solution of tomato bushy stunt virus (TBSV) after mixing with a buffer containing EDTA. Each pattern corresponds to 50 ms counting time. (Courtesy of J. Pérez.)

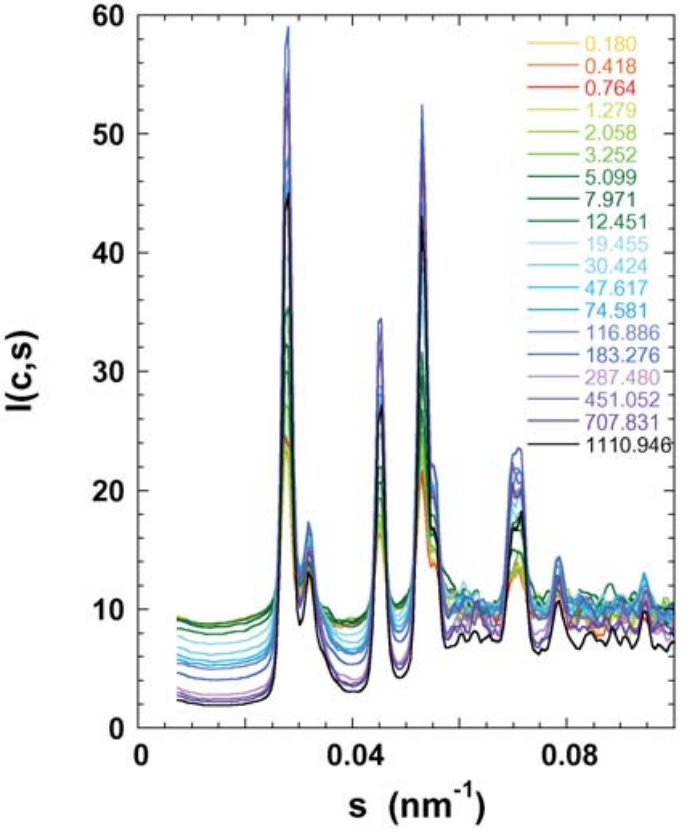

Fig. 31. Time-course (indicated in seconds) after mixing with PEG 20000 of the scattering pattern of a 10 mg ml^{-1} solution of BMV. The final PEG concentration is 5%. The data were divided by the form factor of the virus. (Courtesy of M. Casselyn and colleagues.)

resolutions (> 20 ms) stopped-flow mixers are preferred (Moody *et al.* 1980; Berthet-Colominas *et al.* 1984; Tsuruta *et al.* 1989; Casselyn *et al.* 2002).

Like assembly, folding of proteins and RNA is largely an entropy-driven process. Indeed, the folding of polypeptides is opposed by the loss of conformational entropy essentially of the backbone of the chain, which must be more than compensated by the gain in conformational entropy of the solvent. Theoretical considerations on the folding time of protein and RNA chains indicate that the possibility for folding based on an exclusively random search can be eliminated (Levinthal, 1969) and this justifies the view that at least globular proteins have been selected for 'foldability'. Whereas the classical concept of a well defined folding pathway would suggest that information can be obtained about the structure of possible intermediates, the more recent funnelled energy landscape model of unchaperoned protein folding (for a review see Plotkin & Onuchic, 2002) implies that only limited information on the distribution of intermediate structures can be extracted from such patterns. From top to bottom the funnel contains unfolded or denatured states, which may, however, still contain secondary structure, molten globule states and native states. The largest decrease in entropy is associated with formation of the tertiary structure. In this view, small-angle scattering can give information about the ensemble of states at these three levels but in order to be useful, the results of methods with the global character of small-angle scattering must be combined, as illustrated below, with independent information obtained from local, usually spectroscopic, probes.

Hitherto, the folding landscape that has been studied in most detail by SAXS and spectroscopic methods is that of the unfolding and refolding of cytochrome *c* following changes in acidity of the buffer (Akiyama *et al.* 2002). Upon a pH jump from 2·0 to 4·5, two intermediates were found beside the initial and final states by SVD of the SAXS patterns. The correlation between the radius of gyration (R_g) obtained by SAXS and the helical content (f_h) determined by circular dichroism reveals that the acid unfolded state ($R_g = 2{\cdot}43$ nm, $f_h = 20\%$) collapses in less than 160 μs into a considerably more compact structure ($R_g = 2{\cdot}05$ nm, $f_h = 17\%$), which further evolves over another intermediate ($R_g = 1{\cdot}77$ nm, $f_h = 35\%$) to the native state ($R_g = 1{\cdot}38$ nm, $f_h = 50\%$) in 10–20 ms. SAXS and circular dichroism clearly give two different but complementary pictures of the process since the radius of gyration can significantly change without this having to be the case for the helical content.

The initial unfolded acid denatured state differs from those of equilibrium unfolding with guanidine hydrochloride at neutral pH (Segel *et al.* 1998) or with methanol in acidic conditions (Kamatari *et al.* 1996). Similar observations have been made for other proteins. The refolding mechanism of cytochrome *c* in the presence of guanidine hydrochloride and imidazole, where transient dimers are formed, also differs from the one above (Segel *et al.* 1998). Transient intermediates have also been found during folding of lysozyme (Chen *et al.* 1998; Segel *et al.* 1999).

Clearly, for the same protein, the folding mechanisms and even the corresponding initial or final states may differ depending on the solvent and co-solutes. Changes in the forward scattering $I(0)$ attributed to solvation effects have been detected in several systems, where aggregation can reasonably be excluded (e.g. Chen *et al.* 1998). A large increase in effective volume of the protein upon folding has been described in pressure relaxation experiments on staphylococcal nuclease (Panick *et al.* 1999) where chain collapse, secondary and tertiary structure formation all depend on the same rate limiting step. These observations leave a rather mixed picture of unfolded states and this has prompted attempts at classifying them (Millett *et al.* 2002). Indeed, the denaturant induced unfolded state of lysozyme would be more hydrated than the native one, whereas the pressure induced counterpart in staphylococcal nuclease would be less hydrated.

These results seem to conflict with recent theoretical models suggesting that most of the structure of the protein is formed before water is expelled from the hydrophobic core (Cheung *et al.* 2002).

The effects of hydration on protein folding clearly require further investigation, especially as the forward scattering depends on both the volume of the protein and its contrast. Denaturing agents like urea or guadinium hydrochloride may significantly alter the scattering and absorption of the solvent, as is also the case for alcohols (Kuprin *et al.* 1995).

In view of this complexity, equilibrium studies remain an indispensable complement of time-resolved experiments keeping in mind, however, that equilibrium intermediates may not be representative of on-pathway(s) intermediates (i.e. the thermodynamic and kinetic landscapes may differ). A study of the thermal denaturation of the small all-β protein neocarzinostatin by SAXS, differential scanning calorimetry (DSC) and tryptophan fluorescence spectroscopy (Perez *et al.* 2001) provides a good example of the difficulties that can be encountered even in the interpretation of equilibrium folding experiments. SVD suggests the existence of at least one well defined intermediate, whereas a thermodynamic model with a single intermediate state yields estimates for the free energy changes involved that are an order of magnitude larger than expected. Such a model had previously been found to adequately describe the situtation for lysozyme (Chen *et al.* 1996) and cytochrome *c* (Segel *et al.* 1998). It is still unclear how representative the mechanisms above are, especially as protein folding without early collapse (Plaxco *et al.* 1999) as well as folding without detectable stable intermediates have been documented (Jackson, 1998b).

RNA folding has only been investigated on a very small number of examples but these suffice to illustrate some of the fundamental differences with protein folding. The difference in the balance of the various types of forces (e.g. hydrophobic *versus* electrostatic) determining the folding of proteins and RNA and specific binding of divalent cations in the case of RNA allow at most superficial analogies to be made between the two categories of phenomena. These studies usually integrate the results of hydroxyl radical protection and various forms of spectroscopy, relying on SAXS to provide the global view on the structural changes. The concept of folding landscape is also used in describing RNA folding and it is generally accepted that these landscapes are more rugged than those of proteins, with different pathways separated by large free-energy barriers (Treiber & Williamson, 2001).

Compaction of yeast $tRNA^{Phe}$ and of a ribozyme, the catalytic domain of the *B. subtilis* RNase P RNA induced by Mg^{2+} ions has been investigated at equilibrium (Fang *et al.* 2000). Starting from a completely unfolded state (U) in 4–8 M urea in absence of Mg^{2+} an intermediate (I) is formed upon dilution of the denaturant and addition of micromolar concentrations of Mg^{2+}. This intermediate transforms into the native state (N) upon cooperative binding of 3–4 Mg^{2+} ions. For tRNA, compaction occurs between U and I and I and N, for the catalytic domain mainly between U and I. Comparison of the stability of a thermophilic and a mesophilic form of the ribozyme suggests that there may be two intermediate ensembles on the path between the unfolded and native states. A switch of the relative stability of these two intermediates would increase the stability of the thermophilic form by increasing folding cooperativity.

The most extensive studies so far focused on the folding of the Tetrahymena group I ribozyme induced by Mg^{2+}, where synchrotron hydroxyl radical footprinting had revealed a rapid collapse to a partially disordered state followed by a slow search for the active structure (Sclavi *et al.* 1998). The existence of a compact intermediate, corresponding to a kinetic trap, was detected by time-resolved SAXS (Russell *et al.* 2000). The folding landscape was further explored

in subsequent fluorescence energy transfer (FRET) studies with immobilized ribozyme. Starting from different initial ensembles obtained by incubation with various Na^+ concentrations, folding was induced by addition of Mg^{2+} with dilution of Na^+ to the same final concentration to eliminate this factor as a possible determinant of the pathway (Russell *et al.* 2002b).

Different Na^+ concentrations during incubation result in three different unfolded initial ensembles (U_1, U_2, U_3) being differently populated. At initial Na^+ concentrations below 150 mM U_1 is most populated and evolves into an intermediate kinetic trap I_1 from which it escapes slowly to a second intermediate I_2 which partitions rapidly between a long-lived misfolded state (M) and the native state (N). At Na^+ concentrations of 150–250 mM folding starts mainly from U_2 which evolves rapidly to I_2 and from there to M or N. Finally at salt concentrations above 250 mM unfolding starts from U_3 which rapidly and completely converts to N. The structural features of the compact intermediates were established by SAXS at equilibrium. The effect of specific structural features of the ribozyme in determining the correct folding pathway and the overall rate was shown to depend on the exact timing of their occurrence during folding. Folding thus depends not only on the starting point in the landscape and early pathways separated by high-energy barriers may join later in the process.

In the most recent work the kinetics of refolding was investigated by time-resolved SAXS over five orders of magnitude in time (0·4 ms to1000 s) at low Na^+ concentrations were most of the molecules refold in the misfolded M state, using continuous flow, stopped flow and manual mixing (Russell *et al.* 2002a).

SVD analysis indicates that the curves can be fitted by linear combination of two components, but the presence of two kinetic phases implies the existence of at least three components. This further illustrates that some caution is required in the interpretation of SVD results. The method gives the minimum number of components necessary to describe the system as a linear combination. If any of the components in the real system can itself be described by a linear superposition of the other ones, it may pass unnoticed. The early steps in the folding of the Tetrahymena group I ribozyme are probably consecutive rather than competitive with a transition from an unfolded ensemble to a partially collapsed intermediate with a time constant of 7 ms followed by a conversion to a collapsed intermediate with a time constant of 140 ms. This compaction occurs before any stable tertiary structure is detected. The much longer time scales of these processes compared to protein folding are characteristic of RNA folding.

The short overview above suggests that the complexity and variety of folding mechanisms both for proteins and RNA will continue to be a source of challenging problems for experimentalists and theorists in the foreseeable future.

8. Conclusion

Modelling of small-angle scattering data has made very significant progress and the method has certainly borne out most of the dreams of its few practitioners. If modelling of equilibrium structures has reached a point where routine applications can be envisaged for monodisperse solutions, much remains to be done for the interpretation of mixtures and time-resolved experiments. In the current trend towards structural studies of larger integrated systems small-angle scattering and electron microscopy constitute the low-resolution pole providing the structural framework in which the models from the high-resolution pole formed by crystallography and NMR can meaningfully be fitted.

More importantly perhaps, small-angle scattering provides a much-needed bridge between many experimental methods and an opportunity to unify the images of structures with models describing the response of these structures to the (changing) forces that act upon them. Such a unifying program will require all the skills and enthusiasms available for many years to come.

9. Acknowledgements

The authors are indebted to their many collaborators and co-workers over the years and in particular to R. Kläring, (EMBL – Hamburg), A. Gabriel (EMBL – Grenoble), C. Boulin and A. Epstein (EMBL – Heidelberg), and F. Golding (F. Golding Associates, Lancaster), C. Depautex, J. M. Dubuisson and M. Lemonnier (LURE) who provided during the past 25 years, the sample environment, detectors, data acquisition hardware and software, for many of the projects described above. M.K. also thanks F. Golding for developing the computer program for the work with polyominoes and interesting discussions on the subject. P.V. is indebted to D. Durand and J. Pérez (LURE) and to L. Fetler (Curie Institute, Paris) for many stimulating discussions and for their decisive contribution to several studies. 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).

10. References

Ackers, G. K., Doyle, M. L., Myers, D. & Daugherty, M. A. (1992). Molecular code for cooperativity in hemoglobin. *Science* **255**, 54–63.

Akiyama, S., Takahashi, S., Kimura, T., Ishimori, K., Morishima, I., Nishikawa, Y. & Fujisawa, T. (2002). Conformational landscape of cytochrome c folding studied by microsecond-resolved small-angle X-ray scattering. *Proceedings of the National Academy of Sciences USA* **99**, 1329–1334.

Altman, R. B., Ladner, J. E. & Lipscomb, W. N. (1982). Quaternary structural changes in aspartate carbamoyl transferase EC-2.1.3.2 of *Escherichia coli* at pH 8·3 and pH 5·8. *Biochemical and Biophysical Research Communications* **108**, 592–595.

Aparicio, R., Fischer, H., Scott, D. J., Verschueren, K. H., Kulminskaya, A. A., Eneiskaya, E. V., Neustroev, K. N., Craievich, A. F., Golubev, A. M. & Polikarpov, I. (2002). Structural insights into the beta-mannosidase from *T. reesei* obtained by synchrotron small-angle X-ray solution scattering enhanced by X-ray crystallography. *Biochemistry* **41**, 9370–9375.

Arjunan, P., Umland, T., Dyda, F., Swaminathan, S., Furey, W., Sax, M., Farrenkopf, B., Gao, Y., Zhang, D. & Jordan, F. (1996). Crystal structure of the thiamin diphosphate-dependent enzyme pyruvate decarboxylase from the yeast *Saccharomyces cerevisiae* at 2·3 Å resolution. *Journal of Molecular Biology* **256**, 590–600.

Aslam, M. & Perkins, S. J. (2001). Folded-back solution structure of monomeric factor H of human complement by synchrotron X-ray and neutron scattering, analytical ultracentrifugation and constrained molecular modelling. *Journal of Molecular Biology* **309**, 1117–1138.

Avaeva, S. M., Kurilova, S., Nazarova, T., Rodina, E., Vorobyeva, N., Sklyankina, V., Grigorjeva, O., Harutyunyan, E. H., Oganessyan, V., Wilson, K., Dauter, Z., Huber, R. & Mather, T. (1997). Crystal structure of *Escherichia coli* inorganic pyrophosphatase complexed with SO4(2-). Ligand-induced molecular asymmetry. *FEBS Letters* **410**, 502–508.

Bacon, G. E. (1975). *Neutron Diffraction*. Oxford: Oxford University Press.

Bada, M., Walther, D., Arcangioli, B., Doniach, S. & Delarue, M. (2000). Solution structural studies and low-resolution model of the *Schizosaccharomyces pombe* sap1 protein. *Journal of Molecular Biology* **300**, 563–574.

Badger, J. (1993). Multiple hydration layers in cubic insulin crystals. *Biophysical Journal* **65**, 1656–1659.

Baker, D., Krukowski, A. E. & Agard, D. A. (1993). Uniqueness and the ab initio phase problem in macromolecular crystallography. *Acta Crystallographica* (*D*) **49**, 440–448.

Belloni, L. (1991). *Interacting Monodisperse and Polydisperse Spheres*. Amsterdam: Elsevier Science Publishers.

Benedek, G. B. (1971). Theory of transparency of the eye. *Applied Optics* **10**, 459–473.

Bernocco, S., Finet, S., Ebel, C., Eichenberger, D., Mazzorana, M., Farjanel, J. & Hulmes, D. J. (2001). Biophysical characterization of the C-propeptide trimer from human procollagen III reveals a tri-lobed structure. *Journal of Biological Chemistry* **276**, 48930–48936.

Berthet-Colominas, C., Bois, J. M., Cuillel, M., Sedita, J. & Vachette, P. (1984). An apparatus for stopped-flow X-ray scattering. *Review of Physics Applications* **19**, 769–772.

Bilgin, N., Ehrenberg, M., Ebel, C., Zaccai, G., Sayers, Z., Koch, M. H. J., Svergun, D. I., Barberato, C., Volkov, V., Nissen, P. & Nyborg, J. (1998). Solution structure of the ternary complex between aminoacyl-tRNA, elongation factor Tu, and guanosine triphosphate. *Biochemistry* **37**, 8163–8172.

Bockris, J. O. M. & Reddy, A. K. N. (1977). *Modern Electrochemistry*. New York: Plenum Press.

Boehm, M. K., Woof, J. M., Kerr, M. A. & Perkins, S. J. (1999). The Fab and Fc fragments of IgA1 exhibit a different arrangement from that in IgG: a study by X-ray and neutron solution scattering and homology modelling. *Journal of Molecular Biology* **286**, 1421–1447.

Boesecke, P. & Diat, O. (1995). High-brilliance beamline at the European Synchrotron Radiation Facility. *Review of Scientific Instruments* **66**, 1636–1638.

Bone, S. & Zaba, B. (1992). *Bioelectronics*. New York: John Wiley and Sons.

Bonneté, F., Finet, S. & Tardieu, A. (1999). Second virial coefficient: variations with lysozyme crystallization conditions. *Journal of Crystal Growth* **196**, 403–414.

Bonneté, F., Vivares, D., Robert, C. & Colloc'h, N. (2001). Interactions in solution and crystallization of *Aspergillus flavus* urate oxidase. *Journal of Crystal Growth* **232**, 330–339.

Bordas, J., Mandelkow, E. M. & Mandelkow, E. (1983). Stages of tubulin assembly and disassembly studied by time-resolved synchrotron X-ray scattering. *Journal of Molecular Biology* **164**, 89–135.

Bragg, W. L. & Perutz, M. F. (1952). The external form of the haemoglobin molecule. I. *Acta Crystallographica* **5**, 277–283.

Bras, W. & Ryan, A. J. (1998). Sample environment and techniques combined with small angle scattering. *Advances in Colloid and Interface Science* **75**, 1–43.

Brunner-Popela, J. & Glatter, O. (1997). Small-angle scattering of interacting particles. I. Basic principles of a global evaluation technique. *Journal of Applied Crystallography* **30**, 431–442.

Budayova, M., Bonneté, F., Tardieu, A. & Vachette, P. (1999). Interactions in solution of large oligomeric protein. *Journal of Crystal Growth* **196**, 210–219.

Burley, S. K. (2000). An overview of structural genomics. *Nature Structural Biology* **7** (Suppl.), 932–934.

Burling, F. T., Weis, W. I., Flaherty, K. M. & Brunger, A. T. (1996). Direct observation of protein solvation and discrete disorder with experimental crystallographic phases. *Science* **271**, 72–77.

Cacace, M. G., Landau, E. M. & Ramsden, J. J. (1997). The Hofmeister series: salt and solvent effects on interfacial phenomena. *Quarterly Reviews of Biophysics* **30**, 241–277.

Canady, M. A., Tsuruta, H. & Johnson, J. E. (2001). Analysis of rapid, large-scale protein quaternary structural changes: time-resolved X-ray solution scattering of Nudaurelia capensis omega virus (NomegaV) maturation. *Journal of Molecular Biology* **311**, 803–814.

Capel, M. S., Engelman, D. M., Freeborn, B. R., Kjelgaard, M., Langer, J. A., Ramakrishnan, V., Schindler, D. G., Schneider, D. K., Schoenborn B. P., Sillers, I. Y., Yabuki, S. & Moore, P. B. (1987). A complete mapping of the proteins in the small ribosomal subunit of *Escherichia coli*. *Science* **238**, 1403–1406.

Casselyn, M., Finet, S., Tardieu, A. & Delacroix, H. (2002). Time-resolved scattering investigations of brome mosaic virus microcrystals appearance. *Acta Crystallographica* (*D*) **58**, 1568–1570.

Casselyn, M., Perez, J., Tardieu, A., Vachette, P., Witz, J. & Delacroix, H. (2001). Spherical plant viruses: interactions in solution, phase diagrams and crystallization of brome mosaic virus. *Acta Crystallographica* (*D*) **57**, 1799–1812.

Chacon, P., Diaz, J. F., Moran, F. & Andreu, J. M. (2000). Reconstruction of protein form with X-ray solution scattering and a genetic algorithm. *Journal of Molecular Biology* **299**, 1289–1302.

Chacon, P., Moran, F., Diaz, J. F., Pantos, E. & Andreu, J. M. (1998). Low-resolution structures of proteins in solution retrieved from X-ray scattering with a genetic algorithm. *Biophysical Journal* **74**, 2760–2775.

Chamberlain, D., Ullman, C. G. & Perkins, S. J. (1998). Possible arrangement of the five domains in human complement factor I as determined by a combination of X-ray and neutron scattering and homology modeling. *Biochemistry* **37**, 13918–13929.

Chen, L., Wildegger, G., Kiefhaber, T., Hodgson, K. O. & Doniach, S. (1998). Kinetics of lysozyme refolding: structural characterization of a non-specifically collapsed state using time-resolved X-ray scattering. *Journal of Molecular Biology* **276**, 225–237.

Chen, L. L., Hodgson, K. O. & Doniach, S. (1996). A lysozyme folding intermediate revealed by X-ray solution scattering. *Journal of Molecular Biology* **261**, 658–671.

Cherfils, J., Sweet, R. M., Middleton, S. A., Kantrowitz, E. R., Tauc, P. & Vachette, P. (1989). Structural consequences of a one atom mutation on aspartate transcarbamylase from *E. coli*. *FEBS Letters* **247**, 361–366.

Cheung, M. S., Garcia, A. E. & Onuchic, J. N. (2002). Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proceedings of the National Academy of Sciences USA* **99**, 685–690.

Collins, K. D. & Washabaugh, M. W. (1985). The Hofmeister effect and the behaviour of water at interfaces. *Quarterly Reviews of Biophysics* **18**, 323–422.

Damaschun, G., Puerschel, H. V. & Mueller, J. J. (1968). Ueber die Messstrategie bei der Untersuchung der

Roentgen-Kleinwinkelstreuung von verduennten monodispersen Loesungen von Makromolekuelen. *Monatshefte fuer Chemie* **99**, 2343–2348.

Dauter, Z., Dauter, M. & Dodson, E. (2002). Jolly SAD. *Acta Crystallographica* (*D*) **58**, 494–506.

Debye, P. (1915). Zerstreuung von Roentgenstrahlen. *Annalen der Physik* **46**, 809–823.

Delaye, M. & Tardieu, A. (1983). Short-range order of crystallin proteins accounts for eye lens transparency. *Nature* **302**, 415–417.

deRaad Iseli, C., Reimann, T., Golding, F., Boulin, C., Epstein, A., Beloeuvre, E., Gabriel, A. & Koch, M. H. J. (2001). A data acquisition system for gas proportional detectors with delay line readout based on space–time–space conversion. *Nuclear Instrumentation and Methods* **A 467–468**, 1152–1155.

Dierker, S. B., Pindak, R., Fleming, R. M., Robinson, I. K. & Berman, L. (1995). X-ray photon-correlation spectroscopy study of Brownian-motion of gold colloids in glycerol. *Physical Review Letters* **75**, 449–452.

Dixit, S., Crain, J., Poon, W. C. K., Finney, J. L. & Soper, A. K. (2002). Molecular segregation observed in a concentrated alcohol–water solution. *Nature* **416**, 829–832.

Doniach, S. (2001). Changes in biomolecular conformation seen by small angle X-ray scattering. *Chemical Reviews* **101**, 1763–1778.

Dubuisson, J.-M., Decamps, T. & Vachette, P. (1997). Improved signal-to background ratio in small-angle scattering experiments with synchrotron radiation using an evacuated cell for solutions. *Journal of Applied Crystallography* **30**, 49–54.

Ducruix, A., Guilloteau, J. P., Riès-Kautt, M. & Tardieu, A. (1996). Protein interactions as seen by solution X-ray scattering prior to crystallogenesis. *Journal of Crystal Growth* **168**, 28–39.

Ealick, S. E. (2000). Advances in multiple wavelength anomalous diffraction crystallography. *Current Opinion in Chemistry and Biology* **4**, 495–499.

Edmonds, A. R. (1957). *Angular Momentum in Quantum Mechanics*. Princeton, NJ: Princeton University Press.

Edwards, A. M., Arrowsmith, C. H., Christendat, D., Dharamsi, A., Friesen, J. D., Greenblatt, J. F. & Vedadi, M. (2000). Protein production: feeding the crystallographers and NMR spectroscopists. *Nature Structural Biology* **7** (Suppl.), 970–972.

Egea, P. F., Rochel, N., Birck, C., Vachette, P., Timmins, P. A. & Moras, D. (2001). Effects of ligand binding on the association properties and conformation in solution of retinoic acid receptors RXR and RAR. *Journal of Molecular Biology* **307**, 557–576.

Eisenberg, H. (1994). Protein and nucleic acid hydration and cosolvent interactions: establishment of reliable baseline values at high cosolvent concentrations. *Biophysical Chemistry* **53**, 57–68.

Engelman, D. M. & Moore, P. B. (1972). A new method for the determination of biological quarternary structure by neutron scattering. *Proceedings of the National Academy of Sciences USA* **69**, 1997–1999.

Fang, X., Littrell, K., Yang, X.-J., Henderson, S. J., Siefert, S., Thiyagarajan, P., Pan, T. & Sosnick, T. R. (2000). Mg^{2+}-dependent compaction and folding of yeast $tRNA^{Phe}$ and the catalytic domain of the *B. subtilis* RNase P RNA determined by small-angle X-ray scattering. *Biochemistry* **39**, 11107–11113.

Feigin, L. A. & Svergun, D. I. (1987). *Structure Analysis by Small-angle X-ray and Neutron Scattering*. New York: Plenum Press.

Fetler, L., Tauc, P., Baker, D. P., Macol, C. P., Kantrowitz, E. R. & Vachette, P. (2002). Replacement of Asp-162 by Ala prevents the cooperative transition by the substrates while enhancing the effect of the allosteric activator ATP on *E. coli* aspartate transcarbamoylase. *Protein Science* **11**, 1074–1081.

Fetler, L., Tauc, P., Hervé, G., Moody, M. F. & Vachette, P. (1995a). X-ray scattering titration of the structure transition of aspartate transcarbamylase with a bisubstrate analogue. Influence of nucleotide effectors. *Journal of Molecular Biology* **251**, 243–255.

Fetler, L., Tauc, P. & Vachette, P. (1997). Carbamyl phosphate modifies the T quaternary structure of aspartate transcarbamylase, thereby facilitating the structural transition associated with cooperativity. *Journal of Applied Crystallography* **30**, 781–786.

Fetler, L. & Vachette, P. (2001). The allosteric activator Mg-ATP modifies the quaternary structure of the R-state of *E. coli* aspartate transcarbamylase without altering the T ↔ R equilibrium. *Journal of Molecular Biology* **309**, 817–832.

Fetler, L., Vachette, P., Hervé, G. & Ladjimi, M. M. (1995b). Unlike the quaternary structure transition, the tertiary structure change of the 240s loop in allosteric aspartate transcarbamylase requires active site saturation by substrate for completion. *Biochemistry* **34**, 15654–15660.

Finet, S., Bonneté, F., Frouin, J., Provost, K. & Tardieu, A. (1998). Lysozyme crystal growth, as observed by small angle X-ray scattering, proceeds without crystallization intermediates. *European Biophysical Journal* **27**, 263–271.

Foote, J. & Schachman, H. K. (1985). Homotropic effects in aspartate transcarbamoylase. What happens when the enzyme binds a single molecule of the bisubstrate analog *N*-phosphonacetyl-L-aspartate? *Journal of Molecular Biology* **186**, 175–184.

Fowler, A. G., Foote, A. M., Moody, M. F., Vachette, P., Provencher, S. W., Gabriel, A., Bordas, J. & Koch, M. H. J. (1983). Stopped-flow solution scattering using synchrotron radiation; apparatus, data collection and data analysis. *Journal of Biochemical and Biophysical Methods* **7**, 387–392.

Frank, J., Zhu, J., Penczek, P., Li, Y., Srivastava, S., Verschoor, A., Radermacher, M., Grassucci, R., Lata, R. K. & Agrawal, R. K. (1995). A model of protein synthesis based on a new cryo-electron microscopy reconstruction of the *E. coli* ribosome. *Nature* **376**, 441–444.

Fraser, R. D. B., MacRae, T. P. & Suzuki, E. (1978). An improved method for calculating the contribution of solvent to the X-ray diffraction pattern of biological molecules. *Journal of Applied Crystallography* **11**, 693–694.

Frieden, B. R. (1971). Evaluation, design and extrapolation methods for optical signals, based on the use of the prolate functions. In *Progress in Optics*, vol. 9 (ed. E. Wolf), pp. 312–407. Amsterdam: North Holland.

Fujisawa, T., Kostyukova, A. & Maeda, Y. (2001). The shapes and sizes of two domains of tropomodulin, the P-end-capping protein of actin-tropomyosin. *FEBS Letters* **498**, 67–71.

Fujisawa, T., Uruga, T., Yamaizumi, Z., Inoko, Y., Nishimura, S. & Ueki, T. (1994). The hydration of Ras p21 in solution during GTP hydrolysis based on solution X-ray scattering profile. *Journal of Biochemistry* (*Tokyo*) **115**, 875–880.

Funari, S. S., Rapp, G., Perbandt, M., Dierks, K., Vallazza, M., Betzel, C., Erdmann, V. A. & Svergun, D. I. (2000). Structure of free thermus flavus 5 S rRNA at 1·3 nm resolution from synchrotron X-ray solution scattering. *Journal of Biological Chemistry* **275**, 31283–31288.

Garcia de la Torre, J., Huertas, M. L. & Carrasco, B. (2000). Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophysical Journal* **78**, 719–730.

George, A. & Wilson, W. W. (1994). Predicting protein crystallization from a dilute solution property. *Acta Crystallographica* (*D*) **50**, 361–365.

Glatter, O. (1972). X-ray small angle scattering of molecules composed of subunits. *Acta Physica Austriaca* **36**, 307–315.

Glatter, O. (1977). A new method for the evaluation of small-angle scattering data. *Journal of Applied Crystallographica* **10**, 415–421.

Glatter, O. & Kratky, O. (1982). *Small Angle X-ray Scattering*. London: Academic Press.

Goderis, B., Reynaers, H. & Koch, M. H. J. (2002). Primary and secondary crystallization in a homogeneous ethylene-1-octene copolymer: crystallinity heterogeneity studied by SAXS. *Macromolecules* **35**, 5025–5033.

Gripon, C., Legrand, L., Rosenman, I., Vidal, O., Robert, M. C. & Boué, F. (1997). Lysozyme-lysozyme interactions in under- and super-saturated solutions: a simple relation between the second virial coefficients in H_2O and D_2O. *Journal of Crystal Growth* **178**, 575–584.

Grossmann, J. G., Abraham, Z. H., Adman, E. T., Neu, M., Eady, R. R., Smith, B. E. & Hasnain, S. S. (1993). X-ray scattering using synchrotron radiation shows nitrite reductase from *Achromobacter xylosoxidans* to be a trimer in solution. *Biochemistry* **32**, 7360–7366.

Grossmann, J. G., Ali, S. A., Abbasi, A., Zaidi, Z. H., Stoeva, S., Voelter, W. & Hasnain, S. S. (2000). Low-resolution molecular structures of isolated functional units from arthropodan and molluscan hemocyanin. *Biophysical Journal* **78**, 977–981.

Grueber, G., Svergun, D. I., Godovac-Zimmermann, J., Harvey, W. R., Wieczorek, H. & Koch, M. H. J. (2000). Evidence for major structural changes in the *Manduca sexta* midgut V1 ATPase due to redox modulation. A small-angle X-ray scattering study. *Journal of Biological Chemistry* **275**, 30082–30087.

Guilloteau, J. P. (1991). Cristallogenèse de macromolécules biologiques: influence de la nature des sels et de la température sur la solubilité du lysozyme et sur les interactions protéine–protéine. Université Louis Pasteur, Strasbourg, France.

Guinier, A. (1939). La diffraction des rayons X aux très petits angles: application à l'étude de phénomènes ultramicroscopiques. *Annales de Physique* (*Paris*) **12**, 161–237.

Guinier, A. & Fournet, G. (1955). *Small Angle Scattering of X-Rays*. New York: Wiley.

Gutman, M. & Nachliel, E. (1997). Time-resolved dynamics of proton transfer in proteinous systems. In *Annual Review of Physical Chemistry*, vol. 48, pp. 329–356. Palo Alto: Annual Reviews Inc.

Hamiaux, C., Perez, J., Prange, T., Veesler, S., Ries-Kautt, M. & Vachette, P. (2000). The BPTI decamer observed in acidic pH crystal forms pre-exists as a stable species in solution. *Journal of Molecular Biology* **297**, 697–712.

Hammes, G. G. & Wu, C. W. (1971). Relaxation spectra of aspartate transcarbamylase. Interaction of the native enzyme with aspartate analogs. *Biochemistry* **10**, 1051–1057.

Hansen, J. P. & McDonald, I. R. (1986). *Theory of Simple Liquids*. New York: Academic Press.

Hao, Q. (2001). Phasing from an envelope. *Acta Crystallographica* (*D*) **57**, 1410–1414.

Hao, Q., Dodd, F. E., Grossmann, J. G. & Hasnain, S. S. (1999). Ab initio phasing using molecular envelope from solution X-ray scattering. *Acta Crystallographica* (*D*) **55**, 243–246.

Harrison, D., May, R. P. & Moore, P. B. (1993). Measurement of the radii of gyration of ribosomal components in situ by neutron scattering. *Journal of Applied Crystallography* **26**, 198–206.

Harrison, S. C. (1969). Structure of tomato bushy stunt virus. I. The spherically averaged electron density. *Journal of Molecular Biology* **42**, 457–483.

HEBERLE, J., RIESLE, J., THIEDEMANN, G., OESTERHELT, D. & DENCHER, N. A. (1994). Proton migration along the membrane surface and retarded surface to bulk transfer. *Nature* **370**, 379–382.

HEENAN, R. K., PENFOLD, J. & KING, S. M. (1997). SANS at pulsed neutron sources: present and future prospects. *Journal of Applied Crystallography* **30**, 1140–1147.

HERVÉ, G. (1989). Aspartate transcarbamylase from *Escherichia coli*. In *Allosteric Enzymes* (ed. G. Hervé), pp. 61–79. Boca Raton, FL: CRC Press.

HERVÉ, G., MOODY, M. F., TAUC, P., VACHETTE, P. & JONES, P. T. (1985). Quaternary structure changes in aspartate transcarbamylase studied by X-ray solution scattering. Signal transmission following effector binding. *Journal of Molecular Biology* **185**, 189–199.

HINDS, D. A. & LEVITT, M. (1994). Exploring conformational space with a simple lattice model for protein structure. *Journal of Molecular Biology* **243**, 668–682.

HIRAI, M., IWASE, H., HAYAKAWA, T., MIURA, K. & INOUE, K. (2002). Structural hierarchy of several proteins observed by wide angle solution scattering. *Journal of Synchrotron Radiation* **9**, 202–205.

HOFMEISTER, F. (1888). Zur Lehre der Wirkung der Salze II. *Archiv fuer Experimentelle Pathologie und Pharmakologie* **24**, 247–260.

HOLM, L. & SANDER, C. (1998). Touring protein fold space with Dali/FSSP. *Nucleic Acids Research* **26**, 316–319.

HOWLETT, G. J., BLACKBURN, M. N., COMPTON, J. G. & SCHACHMAN, H. K. (1977). Allosteric regulation of aspartate transcarbamoylase. Analysis of the structural and functional behavior in terms of a two-state model. *Biochemistry* **16**, 5091–5099.

HOWLETT, G. J. & SCHACHMAN, H. K. (1977). Allosteric regulation of aspartate transcarbamoylase. Changes in the sedimentation coefficient promoted by the bisubstrate analogue *N*-(phosphonacetyl)-L-aspartate. *Biochemistry* **16**, 5077–5083.

HUBBARD, S. R., HODGSON, K. O. & DONIACH, S. (1988). Small-angle X-ray scattering investigation of the solution structure of Troponin C*. *Journal of Biological Chemistry* **263**, 4151–4158.

HUEBNER, G., KOENIG, S., SCHELLENBERGER, A. & KOCH, M. H. J. (1990). An X-ray solution scattering study of the cofactor and activator induced structural changes in yeast pyruvate decarboxylase (PDC). *FEBS Letters* **266**, 17–20.

HUXLEY, H. E., FARUQI, A. R., BORDAS, J., KOCH, M. H. J. & MILCH, J. (1980). The use of synchrotron radiation time resolved X-ray diffraction studies of myosin layer-line reflections during muscle contraction. *Nature* **284**, 140–143.

IBEL, K. (1976). The neutron small-angle camera D11 at the high-flux reactor, Grenoble. *Journal of Applied Crystallography* **9**, 296–309.

IBEL, K. & STUHRMANN, H. B. (1975). Comparison of neutron and X-ray scattering of dilute myoglobin solutions. *Journal of Molecular Biology* **93**, 255–265.

INGBER, L. (1993). Simulated annealing: practice versus theory. *Mathematics and Computer Modelling* **18**, 29–57.

ISHIMOTO, C. & TANAKA, T. (1977). Critical behavior of a binary mixture of protein and salt water. *Physical Review Letters* **39**, 474–477.

ISRAELACHVILI, J. & WENNERSTROEM, H. (1996). Role of hydration and water structure in biological and colloidal interactions. *Nature* **379**, 219–225.

JACKSON, J. D. (1998a). *Classical Electrodynamics*, 3rd edn. New York: John Wiley & Sons.

JACKSON, S. E. (1998b). How do small single-domain proteins fold? *Folding and Design* **3**, R81–R91.

JIN, L., STEC, B., LIPSCOMB, W. N. & KANTROWITZ, E. R. (1999). Insights into the mechanisms of catalysis and heterotropic regulation of *Escherichia coli* aspartate transcarbamoylase based upon a structure of the enzyme complexed with the bisubstrate analogue *N*-phosphonacetyl-L-aspartate at 2·1 Å. *Proteins* **37**, 729–742.

KAMATARI, Y. O., KONNO, T., KATAOKA, M. & AKASAKA, K. (1996). The methanol-induced globular and expanded denatured states of cytochrome c: a study by CD fluorescence, NMR and small-angle X-ray scattering. *Journal of Molecular Biology* **259**, 512–523.

KE, H., LIPSCOMB, W. N., CHO, Y. & HONZATKO, R. B. (1988). Complex of *N*-phosphonacetyl-L-aspartate with aspartate carbamoyltransferase. X-ray refinement, analysis of conformational changes and catalytic and allosteric mechanisms. *Journal of Molecular Biology* **204**, 725–748.

KIHARA, H., BARMAN, T. E., JONES, P. T. & MOODY, M. F. (1984). Kinetics of the allosteric transition of aspartate transcarbamylase. *Journal of Molecular Biology* **176**, 523–534.

KIRKPATRICK, S., GELATT, C. D. JR. & VECCI, M. P. (1983). Optimization by simulated annealing. *Science* **220**, 671–680.

KIRSTE, R. G. & OBERTHUER, R. C. (1982). Synthetic polymers in solution. In *Small Angle X-ray Scattering* (eds. O. Glatter & O. Kratky). New York: Academic Press.

KOCH, M. H. & STUHRMANN, H. B. (1979). Neutron-scattering studies of ribosomes. *Methods in Enzymology* **59**, 670–706.

KOCH, M. H. J. (1988). Instruments and methods for small angle scattering with synchrotron radiation. *Makromolekulare Chemie: Macromolecular Symposia* **15**, 79–90.

KOCH, M. H. J. (1989). The structure of chromatin and its condensation mechanism. In *Protein–Nucleic Acid Interaction* (eds. U. Heinemann & W. Saenger), pp. 163–204. London: Macmillan.

KOCH, M. H. J., DORRINGTON, E., KLAERING, R., MICHON, A. M., SAYERS, Z., MARQUET, R. & HOUSSIER, C. (1988). Electric field X-ray scattering measurements on tobacco mosaic virus. *Science* **240**, 194–196.

Koch, M. H. J., Sayers, Z., Sicre, P. & Svergun, D. (1995). A synchrotron radiation electric field X-ray solution scattering study of DNA at very low ionic strength. *Macromolecules* **28**, 4904–4907.

Koenig, S., Svergun, D., Koch, M. H. J., Huebner, G. & Schellenberger, A. (1993). The influence of the effectors of yeast pyruvate decarboxylase (PDC) on the conformation of the dimers and tetramers and their pH-dependent equilibrium. *European Biophysical Journal* **22**, 1894.

Koenig, S., Svergun, D., Koch, M. H. J., Huebner, G. & Schellenberger, A. (1992). Synchrotron radiation solution X-ray scattering study of the pH dependence of the quaternary structure of yeast pyruvate decarboxylase. *Biochemistry* **31**, 8726–8731.

Konarev, P. V., Petoukhov, M. V. & Svergun, D. I. (2001). MASSHA – a graphic system for rigid body modelling of macromolecular complexes against solution scattering data. *Journal of Applied Crystallography* **34**, 527–532.

Koshland, J. D. E., Nemethy, G. & Filmer, D. (1966). Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* **5**, 365–385.

Kosman, R. P., Gouaux, J. E. & Lipscomb, W. N. (1993). Crystal structure of CTP-ligated T state aspartate transcarbamoylase at 2·5 Å resolution: implications for ATCase mutants and the mechanism of negative cooperativity. *Proteins* **15**, 147–176.

Kozin, M. B. & Svergun, D. I. (2000). A software system for automated and interactive rigid body modeling of solution scattering data. *Journal of Applied Crystallography* **33**, 775–777.

Kozin, M. B. & Svergun, D. I. (2001). Automated matching of high- and low-resolution structural models. *Journal of Applied Crystallography* **34**, 33–41.

Kozin, M. B., Volkov, V. V. & Svergun, D. I. (1997). ASSA – a program for three-dimensional rendering in solution scattering from biopolymers. *Journal of Applied Crystallography* **30**, 811–815.

Kratky, O. & Pilz, I. (1978). A comparison of X-ray small-angle scattering results to crystal structure analysis and other physical techniques in the field of biological macromolecules. *Quarterly Reviews of Biophysics* **11**, 39–70.

Kratky, O. & Porod, G. (1949). Roentgenuntersuchung Geloester Fadenmolekuele. *Recueil de travaux chimiques des Pays-Bas* **68**, 1106–1122.

Krishnamurti, P. (1929). X-ray diffraction in liquid mixtures. *Proceedings of the Indian Association for the Cultivation of Science* **12**, 331–355.

Krueger, J. K., Gallagher, S. C., Wang, C. A. & Trewhella, J. (2000). Calmodulin remains extended upon binding to smooth muscle caldesmon: a combined small-angle scattering and Fourier transform infrared spectroscopy study. *Biochemistry* **39**, 3979–3987.

Krueger, J. K., McCrary, B. S., Wang, A. H., Shriver, J. W., Trewhella, J. & Edmondson, S. P. (1999). The solution structure of the Sac7d/DNA complex: a small-angle X-ray scattering study. *Biochemistry* **38**, 10247–10255.

Krueger, J. K., Zhi, G., Stull, J. T. & Trewhella, J. (1998). Neutron-scattering studies reveal further details of the Ca^{2+}/calmodulin-dependent activation mechanism of myosin light chain kinase. *Biochemistry* **37**, 13997–14004.

Kuprin, S., Graslund, A., Ehrenberg, A. & Koch, M. H. J. (1995). Nonideality of water-hexafluoropropanol mixtures as studied by X-ray small angle scattering. *Biochemical and Biophysical Research Communications* **217**, 1151–1156.

Ladjimi, M. M. & Kantrowitz, E. R. (1988). A possible model for the concerted allosteric transition in *Escherichia coli* aspartate trancarbamylase as deduced from site-directed mutagenesis studies. *Biochemistry* **27**, 276–283.

Langridge, R., Marvin, D. A., Seeds, W. E., Wilson, H. R., Hooper, C. W., Wilkins, M. H. F. & Hamilton, L. D. (1960). The molecular configuration of deoxyribonucleic acid: molecular models and their Fourier transforms. *Journal of Molecular Biology* **2**, 38–64.

Lata, R., Conway, J. F., Cheng, N., Duda, R. L., Hendrix, R. W., Wikoff, W. R., Johnson, J. E., Tsuruta, H. & Steven, A. C. (2000). Maturation dynamics of a viral capsid: visualization of transitional intermediate states. *Cell* **100**, 253–263.

Lattman, E. E. (1989). Rapid calculation of the solution scattering profile from a macromolecule of known structure. *Proteins* **5**, 149–155.

Lee, B. & Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *Journal of Molecular Biology* **55**, 379–400.

Lekkerkerker, H. N. W. (1997). Strong, weak and metastable liquids. *Physica A* **244**, 227–237.

Levinthal, C. (1969). How to fold graciously. In *Mossbauer Spectroscopy in Biological Systems* (eds. P. DeBrunner, J. Tsibris & E. Munck), pp. 22–24. Urbana, IL: University of Illinois Press.

Levitt, M. & Sharon, R. (1988). Accurate simulation of protein dynamics in solution. *Proceedings of the National Academy of Sciences USA* **85**, 7557–7561.

Lindner, P., May, R. P. & Timmins, P. A. (1992). Upgrading of the SANS Instrument D11 at the ILL. *Physica B* **180–181**, 967–972.

Lipscomb, W. N. (1994). Aspartate transcarbamylase from *Escherichia coli*: activity and regulation. *Advances in Enzymology* **68**, 67–152.

Lomakin, A., Asherie, N. & Benedek, G. B. (1996). Monte-Carlo study separation in aqueous protein solutions. *Journal of Chemical Physics* **104**, 1646–1656.

Lomakin, A., Asherie, N. & Benedek, G. B. (1999). Aleotopic interactions of globular proteins. *Proceedings of the National Academy of Sciences USA* **96**, 9465–9468.

LYNN, D. G. & MEREDITH, S. C. (2000). Review: model peptides and the physicochemical approach to beta-amyloids. *Journal of Structural Biology* **130**, 153–173.

MAIGNAN, S., GUILLOTEAU, J. P., FROMAGE, N., ARNOUX, B., BECQUART, J. & DUCRUIX, A. (1995). Crystal structure of the mammalian Grb2 adaptor. *Science* **268**, 291–293.

MALFOIS, M., BONNETÉ, F., BELLONI, L. & TARDIEU, A. (1996). A model of attractive interactions to account for liquid–liquid phase separation of protein solutions. *Journal of Chemical Physics* **105**, 3290–3300.

MANGENOT, S., LEFORESTIER, A., VACHETTE, P., DURAND, D. & LIVOLANT, F. (2002). Salt-induced conformation and interaction changes of nucleosome core particles. *Biophysical Journal* **82**, 345–356.

MARSHALL, W. & LOVESEY, S. W. (1971). *Theory of Thermal Neutron Scattering*. London: Oxford University Press.

MARX, A. & MANDELKOW, E. (1994). A model of microtubule oscillations. *European Biophysical Journal* **22**, 405–421.

MERZEL, F. & SMITH, J. C. (2002a). Is the first hydration shell of lysozyme of higher density than bulk water? *Proceedings of the National Academy of Sciences USA* **99**, 5378–5383.

MERZEL, F. & SMITH, J. C. (2002b). SASSIM: a method for calculating small-angle X-ray and neutron scattering and the associated molecular envelope from explicit-atom models of solvated proteins. *Acta Crystallographica* (*D*) **58**, 242–249.

MILLETT, I. S., DONIACH, S. & PLAXCO, K. W. (2002). Toward a taxonomy of the denatured state: small angle scattering studies on unfolded proteins. *Advances in Protein Chemistry* **62**, 241–262.

MONOD, J., WYMAN, J. & CHANGEUX, J. P. (1965). On the nature of allosteric transitions. A plausible model. *Journal of Molecular Biology* **12**, 88–118.

MOODY, M. F., VACHETTE, P. & FOOTE, A. M. (1979). Changes in the X-ray solution scattering of aspartate transcarbamylase following the allosteric transition. *Journal of Molecular Biology* **133**, 517–532.

MOODY, M. F., VACHETTE, P., FOOTE, A. M., TARDIEU, A. M., KOCH, M. H. J. & BORDAS, J. (1980). J. Stopped-flow X-ray solution scattering: the dissociation of aspartate transcarbamylase. *Proceedings of the National Academy of Sciences USA* **77**, 4040–4043.

MOORE, P. B. (1980). Small-angle scattering: information content and error analysis. *Journal of Applied Crystallography* **13**, 168–175.

MOORE, P. B. (1981). On the estimation of the radius of gyration of the subunits of macromolecular aggregates of biological origin *in situ*. *Journal of Applied Crystallography* **14**, 237–240.

MOORE, P. B. & WEINSTEIN, E. (1979). On the estimation of the locations of subunits within macromolecular aggregates from neutron interference data. *Journal of Applied Crystallography* **12**, 321–326.

MUELLER, J. J. (1983). Calculation of scattering curves for macromolecules in solution and comparaison with results of methods using effective atomic scattering factors. *Journal of Applied Crystallography* **16**, 74–82.

MUELLER, J. J., DAMASCHUN, G. & SCHRAUBER, H. (1990). The highly resolved excess electron distance distribution of biopolymers in solution – calculation from intermediate-angle X-ray scattering and interpretation. *Journal of Applied Crystallography* **23**, 26–34.

MUSCHOL, M. & ROSENBERGER, F. (1995). Interactions in under- and supersaturated lysozyme solutions. Static and dynamic light scattering results. *Journal of Chemical Physics* **103**, 10424–10432.

NAKASAKO, M., FUJISAWA, T., ADACHI, S., KUDO, T. & HIGUCHI, S. (2001). Large-scale domain movements and hydration structure changes in the active-site cleft of unligated glutamate dehydrogenase from *Thermococcus profundus* studied by cryogenic X-ray crystal structure analysis and small-angle X-ray scattering. *Biochemistry* **40**, 3069–3079.

NINHAM, B. W. & YAMINSKY, V. (1997). Ion binding and ion specificity: the Hofmeister effect and Onsager and Lifshitz theories. *Langmuir* **13**, 2097–2108.

NINIO, J., LUZZATI, V. & YANIV, M. (1972). Comparative small-angle X-ray scattering studies on unacylated, acylated and cross-linked *Escherichia coli* transfer RNA I Val. *Journal of Molecular Biology* **71**, 217–229.

NISSEN, P., HANSEN, J., BAN, N., MOORE, P. B. & STEITZ, T. A. (2000). The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920–930.

ORTHABER, D., BERGMANN, A. & GLATTER, O. (2000). SAXS experiments on absolute scale with Kratky systems using water as a secondary standard. *Journal of Applied Crystallography* **33**, 218–225.

PAL, S. K., PEON, J. & ZEWAIL, A. H. (2002). Biological water at the protein surface: Dynamical solvation probed directly with femtosecond resolution. *Proceedings of the National Academy of Sciences USA* **99**, 1763–1768.

PANICK, G., VIDUGIRIS, G. J., MALESSA, R., RAPP, G., WINTER, R. & ROYER, C. A. (1999). Exploring the temperature–pressure phase diagram of staphylococcal nuclease. *Biochemistry* **38**, 4157–4164.

PAVLOV, M. Y. & FEDOROV, B. A. (1983a). Calculation of X-ray diffuse scattering curves for DNA A- and B-forms in solution. *Biofizika* **28**, 931–936.

PAVLOV, M. Y. & FEDOROV, B. A. (1983b). Improved technique for calculating X-ray scattering intensity of biopolymers in solution: evaluation of the form, volume, and surface of a particle. *Biopolymers* **22**, 1507–1522.

PAVLOV, M. Y. & SERDYUK, I. N. (1987). Three-isotopic-substitutions method in small angle neutron scattering. *Journal of Applied Crystallography* **20**, 105–110.

PAVLOV, M. Y., SINEV, M. A., TIMCHENKO, A. A. & PTITSYN, O. B. (1986). A study of apo- and holo-forms

of horse liver alcohol dehydrogenase in solution by diffuse X-ray scattering. *Biopolymers* **25**, 1385–1397.

Pedersen, J. S. & Schurtenberger, P. (1996). Scattering functions of semiflexible polymers with and without excluded volume effects. *Macromolecules* **29**, 7602–7612.

Perez, J., Defrenne, J., Witz, J. & Vachette, P. (2000). Detection and characterization of an intermediate conformation during the divalent ion-dependent swelling of tomato bushy stunt virus. *Cellular and Molecular Biology* **46**, 937–948.

Perez, J., Vachette, P., Russo, D., Desmadril, M. & Durand, D. (2001). Heat-induced unfolding of neocarzinostatin, a small all-β protein investigated by small-angle X-ray scattering. *Journal of Molecular Biology* **308**, 721–743.

Perkins, S. J. (2001). X-ray and neutron scattering analyses of hydration shells: a molecular interpretation based on sequence predictions and model fits. *Biophysical Chemistry* **93**, 129–139.

Petoukhov, M. V., Eady, N. A. J., Brown, K. A. & Svergun, D. I. (2002). Addition of missing loops and domains to protein models using X-ray solution scattering. *Biophysical Journal* **83**, 3113–3125.

Petoukhov, M. V. & Svergun, D. I. (2003). New methods for domain structure determination of proteins from solution scattering data. *Journal of Applied Crystallography* **36**, 540–544.

Petrascu, A.-M., Koch, M. H. J. & Gabriel, A. (1998). A beginners guide to gas filled proportional detectors with delay line readout. *Journal of Macromolecular Science* (*B: Physics*) **B37**, 463–483.

Phillips, J. C., Legrand, A. D. & Lehnert, W. F. (1988). Protein-folding observed by time-resolved synchrotron X-ray scattering – a feasibility study. *Biophysical Journal* **53**, 461–464.

Pickover, C. A. & Engelman, D. M. (1982). On the interpretation and prediction of X-ray scattering profiles of biomolecules in solution. *Biopolymers* **21**, 817–831.

Plaxco, K. W., Millett, I. S., Segel, D. J., Doniach, S. & Baker, D. (1999). Chain collapse can occur concomitantly with the rate-limiting step in protein folding. *Nature Structural Biology* **6**, 554–556.

Plestil, J. & Hlavata, D. (1988). Small-angle scattering from poly-electrolyte solutions – a novel method for studying counterion condensation. *Polymer* **29**, 2216–2220.

Plestil, J., Pospisil, H., Ostanevitch, Y. M. & Degovics, G. (1991). Molecular weight determination for small-angle scattering without absolute intensities: advantages and limitations. *Journal of Applied Crystallography* **24**, 659–664.

Plotkin, S. S. & Onuchic, J. N. (2002). Understanding protein folding with energy landscapes. Part I: Basic concepts. *Quarterly Reviews of Biophysics* **35**, 111–167.

Pollack, L., Tate, M. W., Finnefrock, A. C., Kalidas, C., Trotter, S., Darnton, N. C., Lurio, L., Austin, R. H., Batt, C. A., Gruner, S. M. & Mochrie, S. G. J. (2001). Time resolved collapse of a folding protein observed with small angle X-ray scattering. *Physical Review Letters* **86**, 4962–4965.

Pontoni, D., Narayanan, T. & Rennie, A. R. (2002). High-dynamic range SAXS data acquisition with an X-ray image intensifier. *Journal of Applied Crystallography* **35**, 207–211.

Porod, G. (1951). Die Röntgenkleinwinkelstreuung von dichtgepackten kolloiden Systemen, I. *Kolloidnyi Zhurnal* **124**, 83–114.

Press, W. H., Teukolsky, S. A., Wetterling, W. T. & Flannery, B. P. (1992). *Numerical Recipes*. Cambridge: Cambridge University Press.

Prins, E. & Wilson, A. J. C. (1999). International tables for crystallography. In *Mathematical, Physical and Chemical Tables, Vol. C*. Dordrecht: Kluwer Academic Publishers.

Puigdomenech, J., Perez-Grau, L., Porta, J., Vega, M. C., Sicre, P. & Koch, M. H. J. (1989). A time-resolved synchrotron radiation X-ray solution scattering study of DNA melting. *Biopolymers* **28**, 1505–1514.

Putzeys, P. & Brosteaux, J. (1935). The scattering of light in protein solutions. *Transactions of the Faraday Society* **31**, 1314–1325.

Rawiso, M., Duplessix, R. & Picot, C. (1987). Scattering function of polystyrene. *Macromolecules* **20**, 630–648.

Rey, F. A. & Dumas, C. (1984). FFT method to compute solution X-ray scattering curves. *Biochimie* **66**, 121–126.

Riboldi-Tunnicliffe, A., Konig, B., Jessen, S., Weiss, M. S., Rahfeld, J., Hacker, J., Fischer, G. & Hilgenfeld, R. (2001). Crystal structure of Mip, a prolylisomerase from *Legionella pneumophila*. *Nature Structural Biology* **8**, 779–783.

Robert, M. C. & Lefaucheux, F. (1988). Crystal growth in gels: principle and applications. *Journal of Crystal Growth* **90**, 358–367.

Rolbin, Y. A., Kayushina, R. L., Feigin, L. A. & Schedrin, B. M. (1973). Computer calculations of the X-ray small-angle scattering by macromolecule models [in Russian]. *Kristallografia* **18**, 701–705.

Russel, T. P. (1983). An absolute intensity standard for small-angle X-ray scattering measured with position sensitive detectors. *Journal of Applied Crystallography* **16**, 473–478.

Russel, T. P., Lin, J. S., Spooner, S. & Wignall, G. D. (1988). Intercalibration of small-Angle X-ray and neutron scattering data. *Journal of Applied Crystallography* **21**, 629–638.

Russell, R., Millett, I. S., Doniach, S. & Herschlag, D. (2000). Small angle X-ray scattering reveals a compact intermediate in RNA folding. *Nature Structural Biology* **7**, 367–370.

Russell, R., Millett, I. S., Tate, M. W., Kwok, L. W., Nakatani, B., Gruner, S. M., Mochrie, S. G., Pande, V., Doniach, S., Herschlag, D. & Pollack, L. (2002a). Rapid compaction during RNA folding. *Proceedings of the National Academy of Sciences USA* **99**, 4266–4271.

Russell, R., Zhuang, X., Babcock, H. P., Millett, I. S., Doniach, S., Chu, S. & Herschlag, D. (2002b). Exploring the folding landscape of a structured RNA. *Proceedings of the National Academy of Sciences USA* **99**, 155–160.

Sandy, A. R., Lurio, L. B., Mochrie, S. G. J., Malik, A., Stephenson, G. B., Pelletier, J. F. & Sutton, M. (1999). Design and characterization of an undulator beam line optimized for small-angle coherent X-ray scattering at the advanced photon source. *Journal of Synchrotron Radiation* **6**, 1174–1184.

Sasaki, J., Kumauchi, M., Hamada, N., Oka, T. & Tokunaga, F. (2002). Light-induced unfolding of photoactive yellow protein mutant M100L. *Biochemistry* **41**, 1915–1922.

Schluenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F. & Yonath, A. (2000). Structure of functionally activated small ribosomal subunit at 3·3 Å resolution. *Cell* **105**, 615–623.

Schmidt, B., Koenig, S., Svergun, D., Volkov, V., Fischer, G. & Koch, M. H. (1995). Small-angle X-ray solution scattering study on the dimerization of the FKBP25mem from *Legionella pneumophila*. *FEBS Letters* **372**, 169–172.

Schonbrunn, E., Sack, S., Eschenburg, S., Perrakis, A., Krekel, F., Amrhein, N. & Mandelkow, E. (1996). Crystal structure of UDP-*N*-acetylglucosamine enolpyruvyltransferase, the target of the antibiotic fosfomycin. *Structure* **4**, 1065–1075.

Sclavi, B., Sullivan, M., Chance, M. R., Brenowitz, M. & Woodson, S. A. (1998). RNA folding at millisecond intervals by synchrotron hydroxyl radical footprinting. *Science* **279**, 1940–1943.

Scott, D. J., Grossmann, J. G., Tame, J. R., Byron, O., Wilson, K. S. & Otto, B. R. (2002). Low resolution solution structure of the Apo form of *Escherichia coli* haemoglobin protease Hbp. *Journal of Molecular Biology* **315**, 1179–1187.

Segel, D. J., Bachmann, A., Hofrichter, J., Hodgson, K. O., Doniach, S. & Kiefhaber, T. (1999). Characterization of transient intermediates in lysozyme folding with time-resolved small-angle X-ray scattering. *Journal of Molecular Biology* **288**, 489–499.

Segel, D. J., Fink, A. L., Hodgson, K. O. & Doniach, S. (1998). Protein denaturation: a small-angle X-ray scattering study of the ensemble of unfolded states of cytochrome *c*. *Biochemistry* **37**, 12443–12451.

Seki, Y., Tomizawa, T., Khechinashvili, N. & Soda, K. (2002). Contribution of solvent water to the solution X-ray scattering of proteins. *Biophysical Chemistry* **95**, 235–252.

Semenyuk, A. V. & Svergun, D. I. (1991). GNOM – a program package for small-angle scattering data processing. *Journal of Applied Crystallography* **24**, 537–540.

Shannon, C. E. & Weaver, W. (1949). *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press.

Sharp, P. & Bloomfield, V. A. (1968). Light scattering from wormlike chains with excluded volume effects. *Biopolymers* **6**, 1201–1211.

Simons, K. T., Kooperberg, C., Huang, E. & Baker, D. (1997). Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *Journal of Molecular Biology* **268**, 209–225.

Skibinska, L., Gapinski, J., Liu, H., Patkowski, A., Fischer, E. W. & Pecora, R. (1999). Effect of electrostatic interactions on the structure and dynamics of a model polyelectrolyte. II Intermolecular correlations. *Journal of Chemical Physics* **110**, 1794–1800.

Sokolova, A., Malfois, M., Caldentey, J., Svergun, D. I., Koch, M. H. J., Bamford, D. H. & Tuma, R. (2001). Solution structure of bacteriophage PRD1 vertex complex. *Journal of Biological Chemistry* **49**, 46187–46191.

Spinozzi, F., Carsughi, F. & Mariani, P. (1998). Particle shape reconstruction by small-angle scattering: Integration of group theory and maximum entropy to multipole expansion method. *Journal of Chemical Physics* **109**, 10148–10158.

Stevens, R. C., Chook, Y. M., Cho, C. Y., Lipscomb, W. N. & Kantrowitz, E. R. (1991). *Escherichia coli* aspartate carbamoyltransferase: the probing of crystal structure analysis via site-specific mutagenesis. *Protein Engineering* **4**, 391–408.

Stevens, R. C., Gouaux, J. E. & Lipscomb, W. N. (1990). Structural consequences of effector binding to the T state of aspartate carbamoyltransferase: crystal structures of the unligated and ATP- and CTP-complexed enzymes at 2·6-Å resolution. *Biochemistry* **29**, 7691–7701.

Stuhrmann, H. B. (1970a). Ein neues Verfahren zur Bestimmung der Oberflaechenform und der inneren Struktur von geloesten globularen Proteinen aus Roentgenkleinwinkelmessungen. *Zeitschrift fuer Physikalische Chemie* (*Neue Folge*) **72**, 177–198.

Stuhrmann, H. B. (1970b). Interpretation of small-angle scattering functions of dilute solutions and gases. A representation of the structures related to a one-particle-scattering function. *Acta Crystallographica* (*A*) **26**, 297–306.

Stuhrmann, H. B., Goerigk, G. & Munk, B. (1991). Anomalous X-ray scattering. In *Handbook on Synchrotron Radiation*, vol. 4 (eds. S. Ebashi, M. Koch & E. Rubenstein). Amsterdam: Elsevier Science Publishers.

Stuhrmann, H. B., Haas, J., Ibel, K., Koch, M. H. J. & Crichton, R. R. (1976). Low angle neutron scattering of

ferritin studied by contrast variation. *Journal of Molecular Biology* **100**, 399–413.

Stuhrmann, H. B. & Kirste, R. G. (1965). Elimination der intrapartikulaeren Untergrundstreuung bei der Roentgenkleinwinkelstreuung am kompakten Teilchen (Proteinen). *Zeitschrift fuer Physikalische Chemie* (*Neue Folge*) **46**, 247–250.

Stuhrmann, H. B., Scharpf, O., Krumpolc, M., Niinikoski, T. O., Rieubland, M. & Rijllart, A. (1986). Dynamic nuclear polarisation of biological matter. *European Biophysical Journal* **14**, 1–6.

Svergun, D. I. (1991). Mathematical methods in small-angle scattering data analysis. *Journal of Applied Crystallography* **24**, 485–492.

Svergun, D. I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography* **25**, 495–503.

Svergun, D. I. (1994). Solution scattering from biopolymers: advanced contrast variation data analysis. *Acta Crystallographica* (*A*) **50**, 391–402.

Svergun, D. I. (1997). Restoring three-dimensional structure of biopolymers from solution scattering. *Journal of Applied Crystallography* **30**, 792–797.

Svergun, D. I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophysical Journal* **76**, 2879–2886.

Svergun, D. I., Aldag, I., Sieck, T., Altendorf, K., Koch, M. H. J., Kane, D. J., Kozin, M. B. & Grueber, G. (1998a). A model of the quaternary structure of the *Escherichia coli* F1 ATPase from X-ray solution scattering and evidence for structural changes in the delta subunit during ATP hydrolysis. *Biophysical Journal* **75**, 2212–2219.

Svergun, D. I., Barberato, C., Koch, M. H. J., Fetler, L. & Vachette, P. (1997a). Large differences are observed between the crystal and solution quaternary structures of allosteric aspartate transcarbamylase in the R state. *Proteins* **27**, 110–117.

Svergun, D. I., Barberato, C. & Koch, M. H. J. (1995). CRYSOL – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of Applied Crystallography* **28**, 768–773.

Svergun, D. I., Becirevic, A., Schrempf, H., Koch, M. H. J. & Grueber, G. (2000a). Solution structure and conformational changes of the *Streptomyces* chitin-binding protein (CHB1). *Biochemistry* **39**, 10677–10683.

Svergun, D. I., Burkhardt, N., Pedersen, J. S., Koch, M. H. J., Volkov, V. V., Kozin, M. B., Meerwink, W., Stuhrmann, H. B., Diedrich, G. & Nierhaus, K. H. (1997b). Solution scattering structural analysis of the 70 S *Escherichia coli* ribosome by contrast variation. II. A model of the ribosome and its RNA at 3·5 nm resolution. *Journal of Molecular Biology* **271**, 602–618.

Svergun, D. I. & Nierhaus, K. H. (2000). A map of protein–rRNA distribution in the 70 S *Escherichia coli* ribosome. *Journal of Biological Chemistry* **275**, 14432–14439.

Svergun, D. I., Petoukhov, M. V. & Koch, M. H. J. (2001). Determination of domain structure of proteins from X-ray solution scattering. *Biophysical Journal* **80**, 2946–2953.

Svergun, D. I., Petoukhov, M. V., Koch, M. H. J. & Koenig, S. (2000b). Crystal versus solution structures of thiamine diphosphate-dependent enzymes. *Journal of Biological Chemistry* **275**, 297–302.

Svergun, D. I., Richard, S., Koch, M. H. J., Sayers, Z., Kuprin, S. & Zaccai, G. (1998b). Protein hydration in solution: experimental observation by X-ray and neutron scattering. *Proceedings of the National Academy of Sciences USA* **95**, 2267–2272.

Svergun, D. I. & Stuhrmann, H. B. (1991). New developments in direct shape determination from small-angle scattering 1. Theory and model calculations. *Acta Crystallographica* (*A*) **47**, 736–744.

Svergun, D. I., Volkov, V. V., Kozin, M. B. & Stuhrmann, H. B. (1996). New developments in direct shape determination from small-angle scattering 2. Uniqueness. *Acta Crystallographica* (*A*) **52**, 419–426.

Svergun, D. I., Volkov, V. V., Kozin, M. B., Stuhrmann, H. B., Barberato, C. & Koch, M. H. J. (1997c). Shape determination from solution scattering of biopolymers. *Journal of Applied Crystallography* **30**, 798–802.

Tanner, J. J., Smith, P. E. & Krause, K. L. (1993). Molecular dynamics simulations and rigid body (TLS) analysis of aspartate carbamoyltransferase: evidence for an uncoupled R state. *Protein Science* **2**, 927–935.

Tardieu, A. (1994). Thermodynamics and structure – concentrated solutions – structured disorder in vision. In *Neutron and Synchrotron Radiation for Condensed Matter Studies*, vol. III: *Applications to Soft Condensed Matter and Biology*, pp. 145–160. Springer-Verlag: Les éditions de Physique (France).

Tardieu, A., Bonneté, F., Finet, S. & Vivarès, D. (2002). Understanding salt or PEG induced attractive interactions to crystallize biological macromolecules. *Acta Crystallographica* (*D*) **58**, 1564–1567.

Tardieu, A., Le Verge, A., Riès-Kautt, M., Malfois, M., Bonneté, F., Finet, S. & Belloni, L. (1999). Proteins in solution: from X-ray scattering intensities to interaction potentials. *Journal of Crystal Growth* **196**, 193–203.

Tauc, P., Keiser, R. T., Kantrowitz, E. R. & Vachette, P. (1994). Glu-50 in the catalytic chain of *Escherichia coli* aspartate transcarbamoylase plays a crucial role in the stability of the R quaternary structure. *Protein Science* **3**, 1998–2004.

Tauc, P., Leconte, C., Kerbiriou, D., Thiry, L. & Hervé, G. (1982). Coupling of homotropic and heterotropic interactions in *Escherichia coli* aspartate transcarbamylase. *Journal of Molecular Biology* **155**, 155–168.

Tauc, P., Vachette, P., Middleton, S. A. & Kantrowitz, E. R. (1990). Structural consequences of the replacement of Glu239 by Gln in the catalytic chain of

Escherichia coli aspartate transcarbamylase. *Journal of Molecular Biology* **214**, 327–335.

Taupin, D. & Luzzati, V. (1982). Informational content and retrieval in solution scattering studies I. Degrees of freedom and data reduction. *Journal of Applied Crystallography* **15**, 289–300.

Teixeira, J. (1994). The dynamics of water in small volumes probed by incoherent quasi-elastic neutron scattering. *Nuovo Cimento della Societa Italiana di Fisica D* **16**, 1433–1437.

Thomson, J. A., Schurtenberger, P., Thurston, G. M. & Benedek, G. B. (1987). Binary liquid phase separation and critical phenomena in a protein/water solution. *Proceedings of the National Academy of Sciences USA* **84**, 7079–7083.

Thuman-Commike, P. A., Tsuruta, H., Greene, B., Prevelige, P. E. Jr., King, J. & Chiu, W. (1999). Solution X-ray scattering-based estimation of electron cryomicroscopy imaging parameters for reconstruction of virus particles. *Biophysical Journal* **76**, 2249–2261.

Treiber, D. K. & Williamson, J. R. (2001). Beyond kinetic traps in RNA folding. *Current Opinion in Structural Biology* **11**, 309–314.

Tsuruta, H., Brennan, S., Rek, Z. U., Irving, T. C., Tompkins, W. H. & Hodgson, K. O. (1998a). A wide-bandpass multilayer monochromator for biological small-angle scattering and fiber diffraction studies. *Journal of Applied Crystallography* **31**, 672–682.

Tsuruta, H., Nagamura, T., Kimura, K., Igarashi, Y., Kajita, A., Wang, Z. X., Wakabayashi, K., Amemiya, Y. & Kihara, H. (1989). Stopped-flow apparatus for X-ray scattering at subzero temperature. *Review of Scientific Instruments* **60**, 2356–2358.

Tsuruta, H., Reddy, V. S., Wikoff, W. R. & Johnson, J. E. (1998b). Imaging RNA and dynamic protein segments with low-resolution virus crystallography: experimental design, data processing and implications of electron density maps. *Journal of Molecular Biology* **284**, 1439–1452.

Tsuruta, H., Sano, T., Vachette, P., Tauc, P., Moody, M. F., Wakabayashi, K., Amemiya, Y., Kimura, K. & Kihara, H. (1990). Structural kinetics of the allosteric transition of aspartate transcarbamylase produced by physiological substrates. *FEBS Letters* **263**, 66–68.

Tsuruta, H., Vachette, P. & Kantrowitz, E. R. (1998c). Direct observation of an altered quaternary-structure transition in a mutant aspartate transcarbamoylase. *Proteins* **31**, 383–390.

Tung, C. S., Wall, M. E., Gallagher, S. C. & Trewhella, J. (2000). A model of troponin-I in complex with troponin-C using hybrid experimental data: the inhibitory region is a beta-hairpin. *Protein Science* **9**, 1312–1326.

Tung, C. S., Walsh, D. A. & Trewhella, J. (2002). A structural model of the catalytic subunit-regulatory subunit dimeric complex of the cAMP-dependent protein kinase. *Journal of Biological Chemistry* **277**, 12423–12431.

Vanhoudt, J., Abgar, S., Aerts, T. & Clauwaert, J. (2000). A small-angle X-ray solution scattering study of bovine alpha-crystallin. *European Journal of Biochemistry* **267**, 3848–3858.

Velev, O. D., Kaler, E. W. & Lenhoff, A. M. (1998). Protein interactions in solution chracterized by light and neutron scattering: comparison of lysozyme and chymotrypsinogen. *Biophysical Journal* **75**, 2682–2697.

Vérétout, F., Delaye, M. & Tardieu, A. (1989). Molecular basis of eye lens transparency. Osmotic pressure and X-ray analysis of α-crystallin solutions. *Journal of Molecular Biology* **205**, 713–728.

Verwey, E. J. W. & Overbeek, J. T. G. (1948). *Theory of the Stability of Lyophobic Colloids*. Amsterdam: Elsevier.

Vigil, D., Gallagher, S. C., Trewhella, J. & Garcia, A. E. (2001). Functional dynamics of the hydrophobic cleft in the N-domain of calmodulin. *Biophysical Journal* **80**, 2082–2092.

Vivarès, D., Belloni, L., Tardieu, A. & Bonneté, F. (2002). Catching the PEG-induced attractive interaction between proteins. *European Physics Journal E, Soft Matter* **9**, 15–25.

Wakabayashi, K. & Amemiya, Y. (1991). Progress in X-ray synchrotron diffractrion studies of muscle contraction. In *Handbook on Synchrotron Radiation*, vol. 4 (eds. S. Ebashi, M. Koch & E. Rubenstein), pp. 597–678. North-Holland, Amsterdam: Elsevier Science Publishers.

Wall, M. E., Gallagher, S. C. & Trewhella, J. (2000). Large-scale shape changes in proteins and macromolecular complexes. *Annual Review of Physical Chemistry* **51**, 355–380.

Wang, J., Smerdon, S. J., Jager, J., Kohlstaedt, L. A., Rice, P. A., Friedman, J. M. & Steitz, T. A. (1994). Structural basis of asymmetry in the human immunodeficiency virus type 1 reverse transcriptase heterodimer. *Proceedings of the National Academy of Sciences USA* **91**, 7242–7246.

Weyerich, B., Daguanno, B., Canessa, E. & Klein, R. (1990). Structure and dynamics of suspensions of charged rod-like particles. *Faraday Discussions* **90**, 245–259.

Weyerich, B., Brunner-Popela, J. & Glatter, O. (1999). Small-angle scattering of interacting particles. II. Generalized indirect Fourier transformation under consideration of the effective structure factor for polydisperse systems. *Journal of Applied Crystallography* **32**, 197–209.

Wignall, G. D. & Bates, F. S. (1987). Absolute calibration of small-angle neutron scattering data. *Journal of Applied Crystallography* **20**, 28–40.

Wittmann, H. G. (1982). Components of bacterial ribosomes. *Annual Review of Biochemistry* **51**, 155–183.

Woenckhaus, J., Kohling, R., Thiyagarajan, P., Littrell, K. C., Seifert, S., Royer, C. A. & Winter, R. (2001). Pressure-jump small-angle X-ray scattering detected kinetics of staphylococcal nuclease folding. *Biophysical Journal* **80**, 1518–1523.

Yuzawa, S., Yokochi, M., Hatanaka, H., Ogura, K., Kataoka, M., Miura, K., Mandiyan, V., Schlessinger, J. & Inagaki, F. (2001). Solution structure of Grb2 reveals extensive flexibility necessary for target recognition. *Journal of Molecular Biology* **306**, 527–537.

Zhao, J., Hoye, E., Boylan, S., Walsh, D. A. & Trewhella, J. (1998). Quaternary structures of a catalytic subunit–regulatory subunit dimeric complex and the holoenzyme of the cAMP-dependent protein kinase by neutron contrast variation. *Journal of Biological Chemistry* **273**, 30448–30459.

Zheng, W. & Doniach, S. (2002). Protein structure prediction constrained by solution X-ray scattering data and structural homology identification. *Journal of Molecular Biology* **316**, 173–187.