REVIEW ARTICLE

Structural studies of proteins by high-flux X-ray and neutron solution scattering

Stephen J. PERKINS

Department of Biochemistry and Chemistry, Royal Free Hospital School of Medicine, Rowland Hill Street, London NW3 2PF, U.K.

INTRODUCTION

Physical studies on biomolecules are of importance in that quantitative measurements of macromolecular structure and dynamics will provide a more rigorous understanding of the functional roles of these macromolecules. Protein crystallography (which uses diffraction phenomena) is the most powerful technique since three-dimensional models to atomic resolution are obtained (Blundell & Johnson, 1976). In a number of situations, it is essential to complement the crystallographic investigation by the use of low resolution structural methods, such as solution scattering. Solution scattering experiments are also based on diffraction properties, but the macromolecules are now viewed in random orientations.

In solution scattering, a solution of the sample is irradiated by a highly collimated beam of X-rays or neutrons (see Fig. 7 below). The intensity of the scattering curve $I(Q)$ is measured as a function of the scattering angle 2θ or Q . At small Q , such curves lead typically to the determination of the R_G , the radius of gyration, and the $I(0)$, the forward scattering at zero angle. The R_G is

a measure of macromolecular elongation. The $I(0)$ contains compositional information. At larger scattering angles, further macromolecular parameters can be determined, eventually to a nominal structural resolution of 2–4 nm (Fig. 1).

Solution scattering was initially developed using fixed-anode and then rotating-anode X-ray sources in the laboratory in the 1950's and 1960's. Neutron scattering experiments were started in the mid-1960's using so-called low-flux or medium-flux reactors. The five-fold advent of (a) synchrotrons to generate X-ray beams of very high intensities; (b) high-flux neutron reactors; (c) the increased capabilities of computers; (d) improved methods for the detection of X-ray and neutron scattering patterns and (e) new theoretical advances have together permitted the extension of the scattering technique to cover a wider range of more powerful applications since the late 1970's. This review will discuss the use of synchrotron X-ray and high-flux neutron beams in biology. Particular highlights of recent work include (a) the ability to survey usefully and comparatively the solution properties of several of the complement components of immune defence; (b) the monitoring of the allosteric properties of aspartate transcarbamoylase in solution; (c) the kinetic study of the assembly and disassembly of tubulin protofilaments and virus capsids under a range of conditions; (d) the dissection of glycoproteins, membrane protein-lipid complexes, chromatin, ribosomes and viruses in terms of distinct internal structural regions of differing chemical compositions; (e) the location of the 21 proteins of the 30 S ribosomal subunit by deuteration and label triangulation.

THEORETICAL ASPECTS

This section briefly outlines essential background; the reader is referred to Perkins (1988) and Glatter & Kratky (1982) for more details.

Bragg's Law, the Debye Equation, and scattering densities

Diffraction is described by Bragg's Law:

$$\lambda = 2d \sin \theta$$

where λ is the wavelength, d is the diffraction spacing, and 2θ is the angle of diffraction. In solution scattering, d corresponds to dimensions of 1–100 nm. Scattering angles are given in terms of the scattering vector Q , defined as $4\pi \sin \theta / \lambda$, where 2θ is the scattering angle from the direction of the main beam. This definition is the most widely adopted. X-ray crystallographers and

Fig. 1. General features of a solution scattering curve

This is based on the neutron solution scattering curve of C1q of complement measured in 100% $^2\text{H}_2\text{O}$ buffers (Perkins *et al.*, 1984) to show the Guinier region at low Q (out to Q of 0.16 nm^{-1}) and its continuation to a Q of 0.7 nm^{-1} . The experimental data are indicated as points, while the continuous line shows the simulated curve for a Debye sphere model of C1q (Perkins, 1985).

some solution scatterers use s in place of Q , where s is the reciprocal Bragg dimension and is defined by $2 \sin \theta / \lambda$ (e.g. see Fig. 8) (Blundell & Johnson 1976). The resolution of a scattering experiment is usually defined by $2\pi/Q_{\max}$, where $Q_{\max}$ is the largest scattering vector measured in the experiment (i.e. the shortest particle dimension).

The two chief elements of solution scattering theory are (a) the scattering properties of biological macromolecules, and (b) the geometrical relationship between individual scatterers within the macromolecule. Both are embodied in the Debye Equation of Scattering. For any pair of scattering elements (electrons in the case of X-rays; nuclei for that of neutrons), when 2θ is zero, the scattered waves are in phase and the intensity of scattering is at a maximum. At zero scattering angle, the intensity of scattering $I(0)$ is therefore a direct measure of the molecular mass. When Q and 2θ are non-zero, the intensity of scattering is given by the Debye equation (Debye, 1915; Guinier & Fournet, 1955):

$$I(Q) = \sum_p \sum_q f_p f_q \frac{\sin(rQ)}{rQ}$$

The terms f_p and f_q (or b_p and b_q for neutrons) are the scattering lengths of the electrons (or nuclei) at points P and Q in the molecule, and r corresponds to the distance between P and Q. Table 1 shows that f for an atom is proportional to the atomic number, while b is roughly similar for different nuclei with the important exception of ^1H (Bacon, 1975). The latter is the most important neutron scattering length for biological work, and plays the central role in all contrast variation work and in the use of deuteration.

Since solution scattering operates at structural resolutions of 1–4 nm at best, scattering densities $\rho(r)$ can be used to replace atomic scattering lengths. The contrast $\Delta\rho$ is defined as the difference between the mean scattering density of the macromolecule ρ_v and that of its buffer ρ_s (Fig. 2):

$$\Delta\rho = \rho_v - \rho_s$$

The terms in ρ_v and ρ_s are readily calculated from the total of scattering lengths $\sum f_p$ or $\sum b_p$ divided by the volume V (e.g. Perkins, 1986). In biology, distinct variations of scattering densities can occur within the macromolecule. Variation of the contrast $\Delta\rho$ permits their structural analysis, as illustrated for a simple protein–RNA virus in Fig. 2. Conceptually, it is useful to split $\rho(r)$ into two parts (Fig. 2) to correspond to the mean scattering density ρ_v multiplied by a step function $\rho_v(r)$ and the fluctuation of the scattering density $\rho_F(r)$ (Ibel & Stuhmann, 1975; Jacrot, 1976):

$$\rho(r) = \rho_v \cdot \rho_v(r) - \rho_F(r)$$

A compilation of the scattering densities of the main solvents and macromolecular classes is given in Table 2. Table 2 shows that the primary difference between X-rays and neutrons lies in the wider range of solvent contrasts in the latter case. Neutron contrast variation easily encompasses the scattering densities of all four macromolecular classes, lipids, proteins, carbohydrates and nucleic acids. Each of these has its own characteristic neutron scattering density, which reflects the different density of non-exchangeable protons in each one, so neutron scattering can discriminate between them (Fig. 2). X-ray contrast variation using sucrose is however limited to the density range between lipids and proteins.

Guinier analyses

At sufficiently low Q values, the Debye Equation becomes the Guinier approximation (Guinier & Fournet, 1955). The R_G and $I(0)$ parameters are obtained from a plot of $\ln I(Q)$ against Q^2 (Fig. 3):

$$\ln I(Q) = \ln I(0) - R_G^2 Q^2 / 3$$

The radius of gyration R_G characterizes the degree of elongation of the structure. It is analogous to the sedimentation or diffusion coefficients of hydrodynamic experiments. $I(0)$ is the intensity at zero Q and is proportional to M_r^2 . Guinier plots are valid only at low Q , but the lowest Q values cannot be measured for reason of overlap with the main beam and the beamstop (Fig. 1). The smallest Q required to measure the R_G is given by $Q \cdot D_{\max} \leq \pi$, where $D_{\max}$ is the maximum particle dimension (Glatter & Kratky, 1982). The largest Q permitted in Guinier plots depends on the particle shape (e.g. for spheres, the product of $Q \cdot R_G$ should be less than 1.3).

Other Guinier-type analyses can be performed for rod-like and platelet-like shapes. For example, curve analyses based on cross-sectional Guinier plots for rod-like macromolecules:

$$\ln[I(Q) \cdot Q] = \ln[I(Q) \cdot Q]_{Q \rightarrow 0} - R_{xs}^2 Q^2 / 2$$

will yield the cross-sectional radius of gyration R_{xs} . Combination of this analysis with that for the R_G will give the particle length L . If the shape is sufficiently simple, knowledge of the macromolecular volume V permits calculation of the triaxial dimensions of the particle.

Molecular mass M_r and matchpoint analyses are based on the $I(0)$ parameter of Guinier plots. Thus after normalization by the sample concentration c , the $I(0)/c$ values obtained in a given measurement session will yield relative M_r data. Absolute M_r values can be calculated following Kratky (1963), which is based on taking the ratio of the $I(0)$ value to the intensity of the primary beam. Since the latter is very high, standard X-ray scatterers such as calibrated Lupolen samples are required to reduce the primary beam intensity. In neutron scattering, the strong incoherent scattering of water I_{incoh} can be used instead as an absolute intensity standard (Jacrot & Zaccari, 1981). If neutrons are used in H_2O buffers, this minimizes the effect of errors in $\bar{v}$, so is therefore more advantageous than X-ray scattering for M_r determinations, despite the disadvantage of lengthy counting times caused by the high incoherent background in H_2O buffers. Accurate sample concentrations c are required also. For proteins, these can be determined from the 280 nm absorbance. Note that the absorption coefficient can be calculated from the sequence as a check (Perkins, 1986).

Structural analyses of R_G values

The R_G corresponds to the mean square distance of scattering elements from their centre of gravity. The degree of particle elongation can be quantified using the ratio of R_G/R_0 , where R_0 is the R_G of a sphere, the most compact shape of all (Kratky, 1963). The Parallel Axes Theorem will give the separation Δ between the centres of two subunits within a complex formed from their association. The experimental R_G can be compared with

Table 1. Scattering lengths of biologically important nuclei

	Atomic number	$f(2\theta = 0^\circ)$ (fm)	b (fm)
Hydrogen	^1H	1	2.81
	^2H	1	2.81
Carbon	^{12}C	6	16.9
Nitrogen	^{14}N	7	19.7
Oxygen	^{16}O	8	22.5
Phosphorus	^{31}P	15	42.3

Table 2. Scattering densities of solvents and macromolecules

	X-rays ($\text{e} \cdot \text{nm}^{-3}$)	Neutrons (% $^2\text{H}_2\text{O}$)
H_2O	334	0
$^2\text{H}_2\text{O}$	334	100
50% (w/w) sucrose in H_2O	402	13
Lipids	310–340	10–14
Detergents	300–430	6–23
Proteins	410–450	40–45
Carbohydrates	490	47
DNA	590	65
RNA	600	72

that calculated from the atomic co-ordinates of a crystallographic structure.

Since biological macromolecules generally do not have uniform scattering densities, the R_G depends on the contrast $\Delta\rho$. The Stuhmann equation (Ibel & Stuhmann, 1975) is a graph of R_G^2 versus $\Delta\rho^{-1}$ obtained from R_G measurements in several contrasts $\Delta\rho$, whereupon (Figs. 4 and 5):

$$R_G^2 = R_C^2 + \alpha/\Delta\rho - \beta/\Delta\rho^2$$

R_C is the radius of gyration at infinite contrast. This provides the absolute measure of the R_G of the external shape of the macromolecule free from density fluctuations. In principle, in neutron scattering, ^1H – ^2H exchange processes will mean that there will be internal regions where the macromolecule appears to be hollow. For proteins and glycoproteins, the effects of these can be shown to be negligible. It is however important in the

case of viruses, where the heavily hydrated DNA or RNA cores will cause the R_C to be larger than otherwise (Devaux *et al.*, 1983; Cusack *et al.*, 1985). The term in α reflects the radial distribution of scattering density fluctuations. It will be positive if the outermost scattering densities in a particle are higher than those nearer the centre (Fig. 4), and negative if they are lower. The term in β represents the curvature of the Stuhmann parabola (Fig. 5), and corresponds to the distance between the centres of two distinct components of very different scattering densities. In practice, β is not measurable with accuracy unless one of the two components has been heavily deuterated (Koch *et al.*, 1978; Perkins & Weiss, 1983; Ramakrishnan, 1986). Typical data for R_C , α and β are summarized in Table 3.

Fig. 2. Schematic variation of neutron scattering density for an object composed of a central sphere of RNA and a concentric outer shell of protein, i.e. a simple virus

The contrast difference $\Delta\rho$ is the difference between the scattering density of the solvent ρ_s and the solute ρ_v . High positive and negative $\Delta\rho$ are seen in 0% and 100% $^2\text{H}_2\text{O}$. Solution scattering is observed when the solute and solvent scattering densities are different. The protein shell is matched out in 43% $^2\text{H}_2\text{O}$ and the RNA core is matched out in 72% $^2\text{H}_2\text{O}$. Note that for reason of solvent ^1H – ^2H exchange, the average protein and RNA densities increase slightly on going from 0% to 100% $^2\text{H}_2\text{O}$. Note that the fluctuations in scattering densities $\rho_F(r)$ within each of the protein and RNA components do not disappear at their respective matchpoints. From Perkins (1988), with permission.

Scattering curve simulations and distance distribution functions

X-ray and neutron scattering curves $I(Q)$ measured beyond the R_G and R_{xs} ranges lead to further shape information. This is most readily extracted by com-

Fig. 3. Guinier plots of C3 of human complement and several derivatives of C3 measured by synchrotron X-ray scattering at Daresbury

The displacements on the vertical scale are arbitrary. In Perkins & Sim (1986), the R_G values are based on the Q^2 range out to 0.09 nm^{-2} . The $Q \cdot R_G$ values are given on the basis of $R_G = 5.2 \text{ nm}$, except for C3c where $R_G = 4.7 \text{ nm}$.

parisons with simulated scattering curves from models, such as those for spheres or ellipsoids. More sophisticated modelling strategies take advantage of powerful computers to create a molecular shape from an assembly of n small Debye spheres to follow the Debye equation (Rolbin *et al.*, 1973), and these can be readily visualized by computer graphics (e.g. Fig. 6). The calculated scattering curves may require corrections for instrumental effects due to wavelength polychromaticity and/or beam divergence.

Fourier transformation of the full curve $I(Q)$ yields the distance distribution function $D(r)$, which represents the particle shape in real space. This can be conceptually easier to visualize than $I(Q)$, and its main advantage is that D_{max} is readily obtained from the point where $D(r)$ drops to zero, since here $r = D_{\text{max}}$. The chief difficulty is that its computation requires the Q range between $0 \leq Q \leq \infty$. Termination errors are readily introduced in the calculation; however, the Glatter or the Moore Indirect Transform procedures are able to minimize these (Glatter & Kratky, 1982; Moore, 1980).

INSTRUMENTATION AND PRACTICAL ASPECTS

Several high-flux facilities for solution scattering are described (Lander & Price, 1985; Pynn & Fender, 1985; Winnick, 1987; Aigrain, 1987). These multiuser facilities are cost-effective in that ample scattering data for research can be obtained from one or two visits a year.

Daresbury synchrotron X-ray source

Synchrotron radiation sources are between 10^7 times more intense than the conventional laboratory anode source of X-rays. Their power is such that high-flux dedicated sources, starting with the SRS at the SERC Daresbury laboratory, are being constructed.

Fig. 4. Stuhrmann plots of R_G^2 versus $\Delta\rho^{-1}$ for α_1 -acid glycoprotein in buffers containing 1 M-NaCl (●) and 0.2 M-NaCl (○)

The native preparation is a mixture of the biantennary and tri/tetra-antennary forms. The unfolded conformation of carbohydrate chains occurs in 1 M salt for the biantennary form. The compact shape is found in 0.2 M-NaCl for both forms and in 1 M salt for the tri/tetra-antennary form. Since the Stuhrmann plots for the two forms intersect at negative $\Delta\rho^{-1}$ values, it is concluded that the hydrophilic part of α_1 -acid glycoprotein (i.e. the carbohydrate) with the higher scattering density is implicated in the conformational change. The hydrophobic protein core with the lower scattering density is dominant in the Stuhrmann plot at negative $\Delta\rho^{-1}$ (70 – 100% $^2\text{H}_2\text{O}$) and the R_G^2 values are similar in both 0.2 M and 1.0 M salt buffers. Computer models for α_1 -acid glycoprotein using spheres of 0.608 nm diameter are shown for the biantennary form at high salt (upper) and for the tri/tetra-antennary form (lower) (Perkins *et al.*, 1985).

Fig. 5. Stuhrmann plots of deuterated 70 S ribosomes ($\blacktriangle$) and half-deuterated 70 S ribosomes ($\bullet$) where the 30 S subunits are deuterated and the 50 S subunits are protonated (Koch *et al.*, 1978)

The positive slope α seen for the deuterated 70 S ribosomes reflects the higher matchpoint of the deuterated protein over that for the deuterated RNA, where the protein is located on average further from the centre of the particle. In protonated 70 S ribosomes, negative α values are observed (Table 3). The strong curvature β of $420 \times 10^{-14} \text{ nm}^{-2}$ seen for the half-deuterated 70 S ribosomes (Table 3) is a measure of the separation between the 30 S and 50 S subunits and their relative difference in scattering densities.

These are characteristically comprised of a synchrotron to accelerate an electron beam to high energies and a storage ring to maintain this beam for several hours. Since the electrons in the storage ring travel at very high speeds, and yet are constrained in a circular ring trajectory by electromagnets, this causes them to radiate energy, including X-rays. The lost energy is replenished by an oscillatory radiofrequency electric field. The use of magnetic devices such as undulators or wigglers causes higher beam brightnesses to be obtained, as exemplified by the installation of the High Brightness Lattice at the SRS Daresbury. A synchrotron is nominally characterized by its electron energy and current, and SRS Daresbury operates at 2 GeV with currents at 100–300 mA. The brilliance of a synchrotron source is its most important parameter for solution scattering. This is the flux divided by the angle through which the radiation is emitted and by the cross-sectional area of the source. It is usually expressed in units of number of photons $\cdot \text{s}^{-1} \cdot \text{mrad}^{-2} \cdot \text{mm}^{-2}$ per 0.1% relative bandwidth, and is often plotted as a function of photon energy or wavelength. Apart from the flux, other important properties of synchrotron radiation for solution scattering are the very good intrinsic collimation of the beam, and the range of wavelengths available.

The design of cameras for synchrotrons is based on point collimation. The beam is brought to the sample after passing through a crystal monochromator and a curved mirror used for focussing. Pairs of vertical and horizontal slits are placed before and after the monochromator and the mirror to define the beam size and to reduce parasitic scatter. High fluxes also mean that the camera can be lengthened without significant intensity losses, thereby enabling smaller Q ranges to be accessed. The fluxes however mean that radiation levels are too high to permit work near the camera in the presence of beam, so remote control is required for camera alignments and operation. Samples can be as small as 0.05 ml

Fig. 6. Molecular modelling of the C4BP of human complement

The scattering curve was measured to $Q = 1 \text{ nm}^{-1}$ using synchrotron radiation at the SRS Daresbury at a protein concentration of 1.45 mg/ml. The Debye model was constructed using spheres contained within cubes of sides 1 nm, with seven arms of length 33 nm, joined together by a base (Perkins *et al.*, 1986).

Fig. 7. The layout of the neutron camera D11 at the Institut Laue-Langevin in Grenoble

A helical velocity selector is used for monochromatization. Movable neutron guides give collimation distances of 2 m, 5 m, 10 m, 20 m and 40 m. The beam size at the sample is defined by an adjustable diaphragm. The 64×64 element BF_3 multidetector is contained within a 40 m evacuated tube, and can be moved by remote control between 2 and 38 m sample-detector distances. The detector is interfaced with a local minicomputer system for data collection and storage. From Perkins (1988) with permission.

and are contained in 1 mm diameter quartz capillaries or in 1 mm thick Teflon cells with $10\text{--}15\ \mu\text{m}$ thick mica windows. Linear position-sensitive or area multiwire detectors with several hundred resolution elements are employed to record scattering curves. These are stored on a powerful minicomputer system for further treatment. A usable beam is generally available for 6–12 h, after which it is dumped and refilled.

I.L.L. high-flux neutron reactor

At the Institut Laue-Langevin in Grenoble, the fission of ^{235}U in a specially-designed reactor core produces fast high-energy neutrons. For neutron scattering, these are moderated using $^2\text{H}_2\text{O}$ and liquid $^2\text{H}_2$ to enhance the population of longer wavelength neutrons with λ of $0.5\text{--}1.2\text{ nm}$. Neutron guides from the reactor deliver the neutrons to the cameras D11 and D17 (Fig. 7). The beam is collimated using movable segments of beam guides. A velocity selector based on a rotating drum with a helical slot permits the monochromatization of the Maxwellian distribution of neutron wavelengths in the beam by allowing only those neutrons in a specific velocity range to be transmitted. The selector however leads to wavelength polychromaticity as the price for high flux, and these are 8–10 % as defined at full-width-half-maximum of the beam wavelength profile.

Samples of volumes 0.2 or 0.4 ml are contained in 1 mm or 2 mm thick Hellma quartz cells of the type used in spectrophotometers, and the advantage of these is that often optical densities can be measured in the same cells. Larger round 22 mm diameter quartz cells of the same thicknesses (0.3 or 0.6 ml) can be used to improve signal-noise ratios. Beam dimensions are typically $7\text{ mm} \times 10\text{ mm}$ for the rectangular cells and 16 mm diameter for the round cells. The scattering pattern is recorded on a 64×64 element area detector containing 4096 elements of area 1 cm^2 . Neutrons are detected using BF_3 in an ionisation chamber. To obtain the desired Q range, the detector position can be moved between 2–40 m sample-detector distances (camera D11) or rotated in an angular range between $0\text{--}30^\circ$ (camera D17). The cameras are interfaced to minicomputer systems with large data storage capacities to control data accumulations, storage, reduction and transfers to a mainframe computer for detailed analyses.

ISIS spallation neutron source

Spallation neutron sources use proton accelerators to generate neutrons. At the spallation neutron source ISIS at the Rutherford-Appleton Laboratory, a synchrotron is used to accelerate a proton beam through 800 MeV at currents of up to $200\ \mu\text{A}$, operating at 50 Hz, to strike a target made of depleted uranium. Each proton dislodges about 10–50 neutrons from a nucleus it hits, hence the term 'spallation'. Since the heat generated in the target can be dissipated between the pulses, pulsed sources have a potential for significant increases in neutron flux over reactor sources, where the removal of the intense heat in the reactor core is a limiting design factor. Nominally, neutrons are produced at very high intensities at the rate of $3 \times 10^{16}\text{ s}^{-1}$ from pulses containing 2.5×10^{13} protons each. Monochromatization of the neutron beam is achieved here by time-of-flight techniques. These are based on the time required for a neutron to reach the detector after leaving the source. Detector electronics assign a wavelength to each time slice of neutrons, thus permitting the effective usage of all the neutrons in each 'white' pulse. Cold sources are used once again to increase the proportion of the longer wavelength neutrons. By the nature of the spallation process, high fluxes of short wavelength neutrons are dominant in the beam. These contribute usefully to the scattering curve at large Q where the signal-noise ratios are at their weakest.

Samples are contained in quartz cells (as at the I.L.L.). The LOQ camera started operation in 1987 and currently operates at a sample-detector distance of 4.3 m with neutron wavelengths between $0.25\text{--}1.00\text{ nm}$. It has to run at 25 Hz, using choppers to remove every other neutron pulse in order to avoid frame overlapping between consecutive pulses. Very fast neutrons of wavelengths below 0.15 nm are removed by the use of a supermirror bender in the incident beam, thus removing background noise. Heavy shielding around the camera is required to reduce the effect of background neutron levels in the experimental hall. A multiwire BF_3 detector is employed, which is interfaced with a minicomputer, and therefrom to the ISIS mainframe computer for data handling.

Sample preparations

A single-contrast measurement by X-ray or neutron scattering requires about 0.5 ml of solution at 10 mg/ml

concentrations for the requisite dilution series to be carried out. A four-contrast neutron experiment will require 1.5–2.0 ml of solution. Sample concentrations must be accurately known prior to beam-time. Besides the obvious criterion of purity, which must be convincingly demonstrated, X-ray scattering work involves the risk of damage to the sample, unlike neutron scattering. X-ray samples should therefore be monitored for damage. In addition, the samples have to be monodisperse. This means that they are free from small contaminations of aggregates of comparable sizes to that of the system under study, and this is sample dependent. If present, aggregates can be removed by gel filtration. Since the buffer run has to be subtracted from the sample run in data reduction, samples have to be well-dialysed into their buffers prior to measurements. Note that X-ray sample transmissions diminish sharply with increased salt concentrations, unlike the corresponding neutron transmissions. Neutron transmissions are reduced in H_2O buffers due to incoherent scattering processes, and much improved counting statistics are obtained in $^2\text{H}_2\text{O}$ buffers. However the use of $^2\text{H}_2\text{O}$ solutions is on occasions a promoter of sample aggregation.

SYNCHROTRON X-RAY SCATTERING STUDIES

Four main applications of synchrotron X-ray scattering can be identified. (i) More rapid data collection strategies can be developed. (ii) The high quality of the scattering curves assists the structural interpretation of the data and their modelling. (iii) The high fluxes permit the development of kinetic experiments with time slices as short as 100 ms. (iv) The tunability of X-ray wavelengths permits anomalous solution scattering experiments to be carried out.

Data collection strategies

The utility of synchrotron radiation for detailed studies of proteins and glycoproteins is well illustrated by work carried out on several components of the complement cascade of immune defence. The C4b binding protein (C4BP) of M_r 550 000 was successfully studied in the very dilute range of concentrations of 0.32–1.45 mg/ml (Perkins *et al.*, 1986). Its M_r was determined by calibration using the C4u component of complement to show that the C4BP is constructed from 7.4 ± 1 subunits, in confirmation of biochemical work and electron microscopy studies that proposed seven subunits (Dahlbäck *et al.*, 1983). It was also shown that its A_{280} absorption coefficient is in fact 14.1 and not 9.3 as proposed earlier. This affects the presumed stoichiometry of the interaction with C4b, which becomes approx. 7:1 and not 5:1 as previously believed. Useful compositional data were thus obtained. The structural modelling of the C4BP was of interest in that two different but equivalent curve fits were obtained. The first corresponded to an artefactual oblate ellipsoid with axial proportions of 67:36:1 and a thickness of 0.66 nm. This was however physically implausible, and is in conflict with electron microscopy showing that there is a seven-armed substructure. The other fit corresponded to the electron microscopy model. However a good curve fit could only be obtained if the seven arms were positioned close to one another in the solution state (Fig. 6), and not splayed out in all directions as seen in the electron micrographs. This scattering model was fully consistent with hydrodynamic

data on C4BP, and is an illustration of how solution scattering can be used to improve the understanding of electron microscopy results. Each subunit of the C4BP is predominantly constructed from eight repeated domains, each of which has approximate dimensions of $2 \text{ nm} \times 2 \text{ nm} \times 4 \text{ nm}$ (Perkins *et al.*, 1988). Since detailed information on the structure of individual protein domains is available from n.m.r., protein crystallography or structure prediction methods, solution scattering can be of great value in offering an integrated view of the arrangement of such domains within the macromolecule in solution.

Work on C3 and C4 of complement indicates how extensive data collections in a few sessions can lead to useful results. C3 (M_r 187 000) is the most abundant complement component, and exists in several physiological forms with similar M_r values, namely C3u, iC3u, C3(a+b), C3b and iC3b, and also as fragments with reduced M_r , namely C3c, C3dg and C3a. Scattering curves for most of these forms could be compared with each other in the concentration range between 2 and 13 mg/ml (Perkins & Sim, 1986). Thus the R_g values of 5.2 nm for C3, C3u, C3(a+b) and C3b were found to be similar to one another (Fig. 3), while that for iC3u was larger at 5.9 nm. This result was verified by relative M_r calculations. Further comparisons based on R_{xs} data showed that these were similar at 2.4 nm for C3, C3u and C3(a+b), but is reduced for C3b at 2.1 nm. Together, the data show that gross conformational changes could be identified in C3b and iC3u relative to C3, C3u and C3(a+b). Calculation of the R_g/R_0 ratios shows that all these proteins have elongated shapes. Similar work has been carried out on C4 of complement and its fragments to show that there is a high degree of structural homology with C3 and its fragments (S. J. Perkins, A. S. Nealis & R. B. Sim, unpublished work). However C4c was found to be dimeric, unlike C3c. Debye scattering curve simulations were able to show that C3 (and C4 also) have a domain structure which can be subdivided into two distinct entities based on Debye models for the two chief fragments, C3c and C3dg (or C4c and C4d). In the case of C4, the simulations were improved such that these were able to offer an explanation of the conformational change between C4 and C4b (and that between C3 and C3b) in terms of the relative movement of the C4c and C4d domains within the molecule.

Resolution of scattering curves

X-ray solution scattering is superior in both accuracy and information content with improved resolution when synchrotron sources are employed. This is illustrated by the allosteric enzyme, aspartate transcarbamoylase (ATCase), which was studied by Moody and coworkers using both a rotating anode source and the synchrotron at LURE, Orsay, France (Fig. 5 of Moody *et al.*, 1979; Fig. 1 of Hervé *et al.*, 1985). Features in the wide-angle curve between $Q = 1.6$ and 3.1 nm^{-1} are observed with greater clarity in the synchrotron study. This was verified by controls in which these features were reproducibly observed in repeated measurements, and also in dilution series from 100 mg/ml to 10 mg/ml. R_g analyses show that these are 4.59 nm and 4.84 nm for the native and ligated forms respectively, and are correlated with distance distribution functions with maxima at 5.8 and 6.3 nm (Moody *et al.*, 1979). ATCase thus swells in the ligated form and has become more isometric in shape.

Fig. 8. Solution scattering studies of tubulin assembly

(a) Theoretical scattering curves of model structures for tubulin using the Debye equation. MT, microtubule with 13 protofilaments; DR, double concentric ring; SR, single concentric ring; OL, oligomer with four dimers; D, dimer. (b) Projection plot of the scattering traces obtained during a cycle of assembly and disassembly of microtubules in a temperature-jump experiment between 3 °C and 37 °C using synchrotron radiation and a position-sensitive linear detector. Note the changes of $I(s)$ at low s and also in the subsidiary maxima. At 3 °C, rings of tubulin predominate. At 37 °C, microtubules in the form of hollow cylinders predominate in the sample. From Spann *et al.* (1987) with permission. © 1987 American Chemical Society.

The maxima and minima in the wide-angle scattering curve can be used to monitor the proportions of the two forms of ATCase in solution as a function of the concentrations of ATP and CTP (Hervé *et al.*, 1985). It is of interest to note that in many scattering studies, the R_g of the protein is seen to decrease on substrate binding; however, conformational changes in allosteric multisubunit proteins on ligand binding do not always lead to decreased R_g values (Perkins 1988).

That similar marked improvements in curve resolution are obtained is also concluded by Miles *et al.* (1984, 1985) in studies of the 11S globulins, which are multisubunit storage proteins found in plant seeds with M_r values of $(3.0\text{--}3.6) \times 10^5$. Data were collected at the SRS at Daresbury at concentrations of 37–50 mg/ml out to a Q of 1.8 nm⁻¹, and more recently out to a Q of 6.29 nm⁻¹ (I'Anson *et al.*, 1987). The raw data, collected in 10 min, show clearly the subfeatures which are seen only after desmearing corrections have had to be applied in the conventional Kratky camera measurements (Badley *et al.*, 1975; Plietz *et al.*, 1983). These results should facilitate the testing of models for the quaternary structure of the seed globulins.

Kinetic experiments

Synchrotron X-ray radiation for reason of its high intensity has enabled detailed kinetic assembly and disassembly experiments on tubulin to be carried out (Mandelkow *et al.*, 1980, 1983; Bordas *et al.*, 1983;

Spann *et al.*, 1987). This work was performed using the DESY ring in Hamburg. Tubulin is a heterodimer of M_r 100 000 which can form linear arrangements of protofilaments. A ring of 13 parallel protofilament rods arranged as a hollow cylinder forms the microtubule wall. Firstly, solution scattering was used to identify the scattering curves and structures of the stable intermediates, i.e. the heterodimer, single rings with varying numbers of heterodimers, protofilament rods, and microtubules. This work showed that the intensities recorded at the lowest Q values reflect the weight-averaged degree of polymerization, while those at high Q are independent of the state of aggregation. The proportions of each form could be readily identified since the scattering curves of rings and hollow cylinders exhibit characteristic maxima and minima at high Q (Fig. 8). Next, kinetic experiments could thus be set up to record complete scattering curves in time slices as short as every 2–15 s in order to follow tubulin assembly and disassembly (Fig. 8). The actual intermediates in transition during a temperature jump could be identified using difference spectra, which were required since the samples were usually complex mixtures of subpopulations. The nucleation pathway could be elucidated to show that rings disappear prior to microtubule assembly, and do not act as nucleation centres as once believed. Instead, the rings break into small fragments of protofilaments which can elongate by the endwise addition of subunits. The protofilaments associate by lateral inter-

actions until 13 of them close into a cylinder to form the microtubule. Further experiments tested this scheme, where it was shown that chemical conditions which stabilize tubulin rings, such as excess Ca^{2+} or GDP, act as inhibitors of microtubule assembly. The most recent results have succeeded in distinguishing the prenucleation events from the subsequent microtubule assembly by the appropriate adjustments of pH, ionic strength, GTP concentration and temperature jumps to show that the transitions between prenucleation oligomers occur independently of the interactions between the nucleating oligomers (Spann *et al.*, 1987).

Other time-resolved experiments have shown that time resolutions of the order of 100 ms are possible in studies of ATCase dissociation by mercurials (Moody *et al.*, 1980) and the aggregation of bovine serum albumin after disulphide bond cleavage with dithiothreitol (Ueki *et al.*, 1985). When the reactions are too rapid to be followed, such as in the kinetics of the quaternary changes in ATCase, the use of acetyl phosphate in place of carbamoyl phosphate and the presence of 10% or 30% ethylene glycol were effective in slowing down the changes (Kihara *et al.*, 1987).

Anomalous contrast variation

X-ray contrast variation is usually achieved by additions of sucrose to the buffer in order to probe the internal structure, and X-ray synchrotron radiation is useful here in that it overcomes the poorer signal-noise ratios at high sucrose concentrations. Anomalous X-ray scattering however leads to a different approach to contrast variation (Stuhrmann, 1981). When the wavelength λ is close to an atomic absorption edge, the scattering length of the atom becomes strongly dependent on λ . In the case of ferritin, changes in the scattering density of the iron core are achieved by working near the K-edge of iron. Alternatively, ferritin can be studied in a 30% caesium chloride solution by working near the L_3 -edge of caesium. These experiments thus lead to an analysis of the internal structure of ferritin. Synchrotron radiation is required for this work since it offers a wide range of tuneability of λ . The R_G data can be analysed as a function of λ using expressions similar to those of the Stuhrmann plot above.

NEUTRON SCATTERING STUDIES

Three types of applications of high-flux neutron scattering can be described. (i) Rapid data collection strategies can be elaborated to take advantage of the very low backgrounds when working with $^2\text{H}_2\text{O}$ buffers. Sometimes the neutron sample environment permits more unusual scattering or diffraction experiments in $^2\text{H}_2\text{O}$ to be performed, e.g. those involving the use of high hydrostatic pressures, which require a titanium-zirconium alloy container of zero net scattering length to hold the specimens (Braganza & Worcester, 1986), or those involving high salt concentrations where X-ray sample transmissions are adversely affected (Perkins *et al.*, 1985). (ii) Contrast variation experiments using a range of $\text{H}_2\text{O}/^2\text{H}_2\text{O}$ mixtures enable absolute M_r calculations to be made, the shape at 'infinite contrast' to be determined, and the internal arrangement of chemical constituents to be probed. (iii) Label triangulation using specifically-deuterated subunits leads

to the determination of the relative location of these subunits within a multisubunit structure as well as their shapes.

Data collection strategies

Since proteins match out between 40 and 45% $^2\text{H}_2\text{O}$, data collection can proceed rapidly in 100% $^2\text{H}_2\text{O}$ solutions. For example, ^1H - ^2H exchange in myoglobin can be followed by rapidly dissolving the protein in a $^2\text{H}_2\text{O}$ buffer and recording scattering curves in short time intervals (Haas, 1977). The kinetics of self-assembly of a virus protein capsid can be followed in seconds using pH jumps or changes in LiCl molarity with neutrons (Cuillel *et al.*, 1983, 1987). The studies on the complement components illustrate two general strategies of data collection and analyses.

In the first of these, when the glycoproteins are available only in dilute solutions, neutron data collection of the full scattering curve $I(Q)$ is feasible only in $^2\text{H}_2\text{O}$ solutions (Fig. 1). C1q of complement was successfully studied at concentrations between 0.7 and 1.8 mg/ml out to a Q of 0.8 nm^{-1} on the D11 camera at the I.L.L. in Grenoble (Gilmour *et al.*, 1980; Perkins *et al.*, 1984). Molecular models using Debye simulations showed that a slight increase in the length of the C1q molecule was required (as could be predicted from the primary structure of C1q) to bring its shape into complete agreement with electron microscopy data (Perkins, 1985). Of interest is that the simulated curves for C1q are sensitive to its molecular form. Curve fits to the solution data could not be obtained unless it was postulated that C1q has in fact a flexible structure in solution. This is not apparent at first from electron microscopy images (Schumaker *et al.* 1981).

A joint X-ray and neutron scattering investigation was applied to C3 and C4 of complement (see above), and also to several globular subunits found in proteoglycan aggregates of cartilage. In order to ensure that internal inhomogeneities of scattering densities do not cause artefacts in the scattering curve analyses, data in both high positive and high negative contrasts are required (cf: Fig. 2). Since neutron data collection times in H_2O buffers can be excessively long, synchrotron X-ray scattering curves can be used as an effective substitute. No significant differences were observed for C3 and C4 under these conditions (Perkins & Sim, 1986; S. J. Perkins, A. S. Nealis & R. B. Sim, unpublished work), thus justifying the use of uniform density Debye models to model their scattering curves. Distinct differences were however seen for the hyaluronate-binding region of proteoglycan aggregates. A full neutron contrast variation study thus showed that there was a carbohydrate-rich domain on the long axis of binding region which significantly affected the R_{xs} data, but not the R_G data as much (S. J. Perkins, D. G. Dunham, T. E. Hardingham & H. Muir, unpublished work). This could be incorporated into a two-density Debye model for curve simulations.

Contrast variation methods

Internal macromolecular structures are determined by neutron contrast variation (Table 3). The selective deuteration to enhance the scattering properties of one component inside the macromolecule is often of great value.

Table 3. Comparisons of R_G , α and β for macromolecules

	$10^{-3} \times M_r$	Non-protein	R_G (nm)	$10^5 \times \alpha$	$10^{14} \times \beta$ (nm $^{-2}$)
Myoglobin	18	0 %	1.48	3.5	—
α_1 -Acid glycoprotein	37	44 % CHO	2.24	7 ± 2	—
C3 of complement	187	4 % CHO	5.1	30 ± 10	—
C1q of complement	460	7 % CHO	12.8	-100 ± 100	—
Low density lipoprotein	1900	79 % lipid	8.0	280	—
Cytochrome reductase					
Triton X-100	670	10 % detergent	5.73	48 ± 6	$\approx 80-240$
Deuterated Cemulsol	660	10 % detergent	5.67	45 ± 7	280 ± 40
Chromatin nucleosomes	235	53 % DNA	4.04	45 ± 2	—
30 S ribosomal subunit	850	59 % RNA	6.9–7.4	-100	—
30 S ribosomal subunit with deuterated RNA	850	59 % RNA	6.9	0	290 ± 20
70 S ribosomes	2300	65 % RNA	8.1	-190	—
70 S ribosomes with deuterated 30 S	2300	65 % RNA	8.5	50	420
Southern bean mosaic virus	6700	21 % RNA	12.3	-280	—
Tomato bushy stunt virus	8700	17 % RNA	13.3	-240	—
Influenza B virus	176000	5 % RNA, 27 % lipid	44.8	500	—
Frog virus 3	520000	20 % DNA, 10 % lipid	72	—	—

Proteins and glycoproteins. Proteins are inhomogeneous in scattering densities. In Stuhmann plots, the R_G measured in H_2O is larger than that measured in 2H_2O . Hydrophilic amino acid residues on the surface of the protein have higher scattering densities than the hydrophobic ones located at the centre of the protein (Ibel & Stuhmann, 1975). The radial distribution of this inhomogeneity is described by positive Stuhmann α values which are typically of the order of 5×10^{-5} . Even though α is proportional to R_G^2 , the α values for very elongated proteins and glycoproteins with high R_G values are often close to zero. This corresponds to the more even radial distribution of hydrophilic and hydrophobic residues which is imposed by the elongated macromolecular shape (Perkins *et al.*, 1984).

Contrast variation studies on carbohydrate–protein moieties as in glycoproteins are characterized by neutron matchpoints at 49 % 2H_2O for the carbohydrate component, to be compared with an average matchpoint of around 30 % 2H_2O for most hydrophobic residues and 50 % 2H_2O for most hydrophilic residues (Perkins, 1986). Since carbohydrates are found on the surfaces of glycoproteins with the hydrophilic residues, α values are larger than those for proteins. Work on α_1 -acid glycoprotein (44 % carbohydrate) shows how shape information on the protein and carbohydrate components can be obtained. In H_2O buffers, its R_G increased from 2.2 nm to 2.7 nm on going from 0.2 M-NaCl buffers to 1.0 M-NaCl buffers (Li *et al.*, 1983). Stuhmann plots of the compact and expanded forms are linear with different slopes, and these intersect at negative $\Delta\rho^{-1}$ values (Fig. 4). It is concluded that the hydrophobic part has a constant shape in the two salt buffers, so that the conformational change has taken place in the hydrophilic part. Sphere models were developed to show that this could be satisfactorily explained in terms of the five carbohydrate chains of α_1 -acid glycoprotein (Perkins *et al.*, 1985). This study suggests as a general hypothesis that glycan chains are usually found in intimate contact

with protein surfaces under normal physiological conditions.

Protein–lipid and protein–detergent complexes. Neutron studies on protein–lipid complexes will yield data on the relative arrangement of the two components. In one approach, membrane proteins can be incorporated into lipid vesicles, and the Stuhmann R_G and α parameters of this complex and the protein-free vesicle can be determined. For cytochrome b_5 , comparisons with model calculations suggested that the hydrophobic domain of the protein penetrated both the inner and outer monolayers of the vesicle bilayer (Gogol & Engelman, 1984); see also Torbet & Freyssinet (1987). Another approach is based on Stuhmann analyses of the R_G of the membrane thickness. This has been employed with the reaction centre protein of photosynthetic membranes (Sadler & Worcester, 1982; Sadler *et al.*, 1984). The lipoproteins of plasma constitute an important class of protein–lipid complexes. Neutron studies on the human low density lipoproteins have confirmed their quasi-spherical shapes, with the protein located primarily at the surface (Stuhmann *et al.*, 1975). Experiments using deuterated cholesterol and with deuterated phospholipids have identified by R_G analyses and by difference curves the location of these deuterated labels within the lipoprotein core (Laggner *et al.*, 1981, 1984).

The use of protein–detergent complexes offers an alternative strategy for the study of membrane proteins. This has been applied to several systems, including rhodopsin, the acetylcholine receptor and cytochrome reductase. Contrast variation is best applied using both protonated and deuterated detergents which respectively have matchpoints below and above that of the protein (Osborne *et al.*, 1978; Perkins & Weiss, 1983). Determination of the two Stuhmann plots for the complexes will identify two R_G values at positive and negative $\Delta\rho^{-1}$ values for each component at which the other has been matched out. The four R_G values can then be interpolated

to obtain the R_G values for the two components. The importance of this procedure is to minimize the effect of internal density fluctuations within the protein on the observed R_G for the detergent at protein matchout conditions. Otherwise the R_G of protonated Triton X-100 in its complex with the acetylcholine receptor is too small to be physically plausible (Wise *et al.*, 1979). For cytochrome reductase, two isomorphous complexes were successfully used to verify the results of two-dimensional electron microscopy studies on the protein, to locate the position of the membrane in this structure (which was invisible in the electron microscopy work), and to identify the way in which this protein structure was cleaved into two major parts (Perkins & Weiss, 1983).

Protein-DNA and protein-RNA complexes. Many protein-nucleic acid complexes have been studied by neutron contrast variation. DNA forms extended structures, while RNA differs in being able to fold into globular structures. This difference is reflected in the applications of solution scattering to these systems.

The complex of DNA with the Gene 5 protein of the filamentous phage fd is rod-like, so the contrast variation studies are based on Stuhmann analyses of the R_{xs} parameter. These experiments were performed using hydrogenated DNA and 87% deuterated DNA (Torbet *et al.*, 1981; Gray *et al.*, 1982). Both gave negative α values of -5×10^{-5} and -42×10^{-5} respectively. This clearly shows that the DNA lies near the centre of the cross-section of the rod-like complex. This result usefully complemented crystallographic studies on the Gene 5 protein itself, since in the absence of such evidence it had been proposed that the DNA was located on the outer periphery of the complex.

Work with RNA-protein complexes usually involves the R_G and $I(0)$ parameters. For example, in the study of the synthetases in complexes with their cognate transfer RNA, the stoichiometry of association can be 2:1, or 1:1, or 1:2, depending on the system under investigation. Study of the $I(0)$ parameter in H_2O (where it is insensitive to $\bar{v}$) or in 70–77% 2H_2O (RNA matchout conditions) permits the identification of the stoichiometry in question. Structural studies are best carried out when the protein and RNA moieties are of similar sizes, such as in the case of the complex of the elongation factor EF-Tu·GTP with aminoacyl-tRNA. Measurements of R_G values of 2.0 nm for EF-Tu·GTP, 2.3 nm for aminoacyl-tRNA, and 2.6 nm for their complex in H_2O buffers have led to the proposition of a compact structure for the complex, as well as confirming the 1:1 stoichiometry of the interaction and identifying some dimerization of free EF-Tu·GTP (Antonsson *et al.*, 1986).

Chromatin is the DNA-protein complex of eukaryotic cell nuclei. The basic structural unit is the nucleosome (53% DNA) which gives an R_G of 4.0 nm and an α of 45×10^{-5} in Stuhmann plots (Pardon *et al.*, 1975; Hjelm *et al.*, 1977). The positive α value shows that DNA is external to a protein core in the nucleosome. A full contrast variation study based on $I(Q)$ measured in several contrasts has led to the three basic scattering functions $I_v(Q)$, $I_{vf}(Q)$ and $I_f(Q)$ (Hjelm *et al.*, 1977; Suau *et al.*, 1977). These summarize all the data on the external I_v and internal I_f structures, together with their cross-term I_{vf} . These permitted the development of Debye sphere models to define the basic structural model for the nucleosome in terms of a DNA superhelix wound

1.75 times around an octamer of eight core histone proteins (Braddock *et al.*, 1981). Further experiments have studied other solution properties of this protein-DNA complex. The experiments have been extended to the study of assemblies of nucleosomes to form rod-like higher order structures (Suau *et al.*, 1979). Analyses of the cross-sectional R_{xs} and $[I(Q) \cdot Q]_{Q \rightarrow 0}$ values show that for dilute chromatin fragments containing over 50 nucleosomes, a clear structural transition takes place between buffers in low ionic strengths with EDTA to those in higher ionic strengths or in the presence of divalent cations. The R_{xs} increased from about 3 nm to about 11 nm, while $[I(Q) \cdot Q]_{Q \rightarrow 0}$ rose from about 0.1 nucleosome/nm to 0.6 nucleosomes/nm. Contrast variation showed that the DNA is on the outside in these complexes. At the present time, these experiments are being reinterpreted to take account of the complexity of chromatin higher-order structures using both X-ray and neutron scattering (Bordas *et al.*, 1986; Greulich *et al.*, 1987; Gerchmann & Ramakrishnan, 1987; Koch *et al.*, 1987). Other experiments have studied intact chromosomes and cell nuclei, although molecular interpretations of the curves are difficult at the present time.

Contrast variation experiments to study the 30 S, 50 S and 70 S ribosomal subunits from *E. coli* have given R_G values of 6.9–7.4 nm, 6.9–7.4 nm and 8.1–8.5 nm in that order from Stuhmann plots (Koch & Stuhmann, 1979). The α are -90×10^{-5} , -260×10^{-5} and -190×10^{-5} . These values correspond to a compact core of ribosomal RNA which is stabilized by an outer surface containing the ribosomal proteins. This is the reverse of the situation found for chromatin. The β are approximately zero, thus initially suggesting that the distributions of proteins and RNA are concentric. The use of deuterated RNA in the 30 S subunit in fact indicated an asymmetric distribution of the two components (Ramakrishnan, 1986), which is in better agreement with the results from label triangulation of the protein subunits (below). Likewise a pronounced β could be measured (Fig. 5) by deuterating either the 30 S or 50 S subunits and reassembling the 70 S ribosome (Koch *et al.*, 1978). Other studies have examined the interactions of the ribosomal RNAs and their fragments with ribosomal proteins, where for example the role of the individual ribosomal proteins S4, S7, S8, S15 and S17 in assisting the folding of the 16 S RNA into its final conformation could be followed by contrast variation (Serdyuk *et al.*, 1983).

Multicomponent systems and viruses. Macromolecular complexes containing more than two components have only been studied in detail in the case of viruses. Many viruses are often highly isometric in structure. Neutron scattering is able to analyse such structures in terms of shells of various chemical compositions by contrast variation (Jacrot, 1981). Unlike X-ray scattering, the degree of hydration of the viral core does not affect the structural interpretation of the scattering curves. Guinier analyses lead to $I(0)$ and matchpoint data which provide information on the relative amounts and proportions of protein, lipid and nucleic acids within the virus. Stuhmann plots give negative α values in reflection of the lower scattering densities of the protein and lipid components close to the viral surface, compared with the nucleic acids and proteins present in the core. Beyond the Guinier region, the scattering curves of isometric viruses exhibit minima and maxima, and these can be analysed

Fig. 9. Spherical shell models fitted to the scattering curves of the influenza virus B

A four-shell model with an outer diameter of 117 nm and based on matchpoints of 41.0% (outermost), 7.6%, 42.8% and 45.0% (innermost) gave curve fits to data collected in five contrasts. These matchpoints are consistent with an outer glycoprotein surface separated by a lipid bilayer from a RNA/protein core. The deviations seen in the low percentage $^2\text{H}_2\text{O}$ data are attributed to the breakdown of the shell models caused by the existence of protein spikes on the viral surface. From Cusack *et al.* (1985) with permission.

using a shell model constructed from concentric hollow spheres (Fig. 9). A different scattering density is allocated to each shell, and the scattering curve can be computed for each contrast. The final analysis is based on one set of shells for all contrasts, and only the scattering densities are allowed to vary. As checks, the scattering density within each shell is required to be linear as a function of the contrast, and the matchpoint of each shell should conform to known compositional data. Modelling by this approach has been successfully applied to tobacco bushy stunt virus and southern bean mosaic virus (Krüse *et al.*, 1982*a,b*) as well as to lipid-containing viruses such as the frog virus 3 or influenza virus (Cuillel *et al.*, 1979; Cusack *et al.*, 1985). In cases where the virus crystal structure has been solved, and crystallographic disorder has caused part of this structure to become invisible, the neutron analysis has provided useful information on the location of the 'missing' component. Other benefits of the technique include the clarification of the true molecular weights of several viruses, and the characterization of the solution properties of the viral shape, including the study of conformational changes and the kinetics of assembly of the virus or its capsid (Cuillel *et al.*, 1987; Berthet-Colominas *et al.*, 1987).

Label triangulation

The logical extension of neutron contrast variation is the selective deuteration of two protein components in a multisubunit complex in order to identify their relative

Fig. 10. The 21-protein neutron map of the 30 S ribosomal subunit from *E. coli* in two views

Each protein is represented by a sphere whose volume is the same as that of the protein. The numbering of the spheres corresponds to the standard nomenclature for the ribosomal proteins. The maximum linear dimension of the array is about 19 nm. The 30 S subunit bears these proteins mostly in the upper half as viewed, while the RNA is predominant in the lower half (Moore *et al.*, 1986). From Moore & Capel (1988) with permission. © 1988 Annual Reviews Inc.

location within the complex. These proteins have to be incorporated into the complex by dissociation and reconstitution methods. Label triangulation employs interference analysis to determine the distance between the two subunits. The interference function I_{12} is calculated from the difference:

$$(x^2 - 2x)I_{12} = (I_A + I_D) - (I_B + I_C)$$

where I_A is the scattering curve of the unlabelled complex A, I_D is that for the bilabelled complex D, and I_B and I_C are those for the two monolabelled complexes B and C. The degree of deuteration is denoted by x . In practice, two samples only need to be measured, namely an equimolar mixture of A and D, and another one of B and C. The advantage is that high concentrations can be used since interparticle interference effects are eliminated (Hoppe, 1973). Signal-noise ratios are improved, which is important since otherwise the concentration of the label in the sample becomes extremely low. The function I_{12} exhibits zeroes at multiples of πQ , so the separation Δ between the two labels is given directly by $n/\pi Q$, where n is an integer (Moore *et al.*, 1978). By successive reconstitutions and I_{12} measurements, Δ can be determined for all pairs of subunits within the structure. If the Guinier region of I_{12} can be obtained, the $R_{G,12}$ can be measured to give Δ from:

$$2R_{G,12}^2 = R_{G,1}^2 + R_{G,2}^2 + \Delta^2$$

where $R_{G,1}$ and $R_{G,2}$ are the individual R_G values of the two labels (Moore *et al.*, 1978; Stöckel *et al.*, 1979). Given a full enough determination of I_{12} , $R_{G,12}$ can be derived from the interference curve by Patterson inversion (Moore & Weinstein, 1979):

$$2R_{G,12}^2 = \int_0^\infty r^2 \cdot \rho(r) dr$$

Since Δ is known from the nodes of I_{12} , shape information on $R_{G,1}$ and $R_{G,2}$ can be additionally obtained by label triangulation. The price of this is that four additional measurements of interference functions are now necessary beyond the minimum required for simple triangulation.

The most extensive application of label triangulation has been to the ribosomal proteins of the 30 S and 50 S subunits of the *E. coli* ribosome. In the 30 S subunit, all 21 proteins have now been mapped by Moore and Engelman (Fig. 10) using 105 measurements of Δ , mostly at the Brookhaven high-flux reactor (Moore *et al.*, 1986; Capel *et al.*, 1987). Fully-deuterated labels were re-incorporated into protonated 30 S subunits, which were then measured in 57% $^2\text{H}_2\text{O}$ buffers, where the protonated matrix is on average matched out. The R_G of many of the individual proteins were found to be low, and similar to those for the proteins obtained after purification, provided that this had been done using gentle extraction techniques. Others were more elongated. The hand of the map was determined by comparisons with data from immune electron microscopy methods. The value of this quaternary structure is that it can be correlated with the functioning of the ribosomal proteins, given that crystal structures of the individual subunits are beginning to be available. In the 50 S subunit, using a different deuteration strategy, 11 of the 32 proteins have been mapped to date (Nierhaus *et al.*, 1983; Nowotny *et al.*, 1987). Other applications have been described for DNA-dependent RNA polymerase

(Stöckel *et al.*, 1980a,b) and tryptophan synthase (Ibel *et al.*, 1985).

CONCLUSIONS

The advent of high-flux beam sources over the past decade has led to a considerable widening of the scope of solution scattering in biological studies. This has been highlighted in the preceding two sections with illustrations of the detailed investigations that are now becoming routine. A large number of good-quality spectra can be recorded in short times. These can be applied to an extensive set of macromolecular comparisons, or to time-resolved kinetic experiments. Contrast variation by X-ray scattering or neutron scattering can give valuable and unique information about the internal structure within the macromolecule, particularly by neutron scattering where many elegant applications have been demonstrated. These techniques are a powerful complement to other biophysical methods for structural or compositional analyses, particularly when utilized in a multidisciplinary strategy.

The benefits of the use of solution scattering can be enumerated as follows.

(a) It is a solution technique, and the overall structure is thus measured in conditions close to physiological. This is an advantage over electron microscopy methods, which while useful in being able to visualize images directly is prone to artefacts caused by the need to work *in vacuo*, beam damage, and the use of stains.

(b) It is a multiparameter method, which is an advantage over hydrodynamic methods which yield only a single measurement of macromolecular elongation through sedimentation or diffusion coefficients.

(c) Molecular masses are routinely determined. Such controls are important, since unexpected oligomerization or dissociation phenomena can be readily identified. The extent of macromolecular complexation, the integrity of the sample, or the validity of the protein 280 nm absorption coefficients can also be monitored, depending on circumstances.

(d) Conformational changes in the solution state can be detected in the appropriate circumstances, and molecular modelling calculations of the scattering data can be used to describe these changes. Kinetic experiments can be used to follow the time course of these changes in time slices as short as 100 ms.

(e) The structural results can be quantitatively compared with data obtained by other means, i.e. electron microscopy, hydrodynamic studies, protein crystallography, and molecular masses from sequences. This emphasizes the usefulness of the technique as a tool in multidisciplinary structural work. Even though the information content of a scattering curve is sharply reduced in comparison with that of a crystal structure, and unique structure determinations are not possible from a scattering curve, modelling of the scattering curve is much constrained by knowledge of these independent lines of evidence. In particular, solution scattering is effective at ruling out models that are incompatible with the scattering data, such as incorrect electron microscopy structures.

(f) The important advantage of neutron scattering over X-ray scattering is that internal structures can be comprehensively probed by contrast variation. The structures of distinct structural domains within the

macromolecule, such as lipids, proteins, carbohydrates or nucleic acids, can therefore be investigated. The use of selective deuteration techniques offers a powerful extension of this approach, since the disposition and shapes of individual proteins within a complex system can be determined.

I gratefully acknowledge beamtime support from the Institut Laue-Langevin in Grenoble, the SRS at Daresbury, DESY at the EMBL-Hamburg, the ISIS at the Rutherford-Appleton Laboratory, and the Argonne National Laboratory in Chicago during the past decade. The literature cutoff date is December 1987. The author's work has been supported by the European Molecular Biology Laboratory, the Medical Research Council, the Lister Institute of Preventive Medicine, the Science and Engineering Research Council, and the Wellcome Trust.

REFERENCES

- Aigrain, P. (1987) ESRF Foundation Phase Report, European Synchrotron Radiation Facility
- Antonsson, B., Leberman, R., Jacrot, B. & Zaccai, G. (1986) *Biochemistry* **25**, 3655–3659
- Bacon, G. E. (1975) *Neutron Diffraction*, 3rd edn., Oxford University Press, Oxford
- Badley, R. A., Atkinson, D., Hauser, H., Oldani, D., Green, J. P. & Stubbs, J. M. (1975) *Biochim. Biophys. Acta* **412**, 214–218
- Berthet-Colominas, C., Cuillel, M., Koch, M. H. J., Vachette, P. & Jacrot, B. (1987) *Eur. Biophys. J.* **15**, 159–168
- Bordas, J., Mandelkow, E. M. & Mandelkow, E. (1983) *J. Mol. Biol.* **164**, 89–135
- Bordas, J., Perez-Grau, L., Koch, M. H. J., Vega, M. C. & Nave, C. (1986) *Eur. Biophys. J.* **13**, 157–173
- Blundell, T. L. & Johnson, L. N. (1976) *Protein Crystallography*, Academic Press, London
- Braddock, G. W., Baldwin, J. P. & Bradbury, E. F. M. (1981) *Biopolymers* **20**, 327–343
- Braganza, L. F. & Worcester, D. L. (1986) *Biochemistry* **25**, 2591–2596
- Capel, M. S., Engelman, D. M., Freeborn, B. R., Kjeldgaard, M., Langer, J. A., Ramakrishnan, V., Schindler, D. G., Schneider, D. K., Schoenborn, B. P., Sillers, I. Y., Yabuki, S. & Moore, P. B. (1987) *Science* **238**, 1403–1405
- Cuillel, M., Tripier, F., Braunwald, J. & Jacrot, B. (1979) *Virology* **99**, 277–285
- Cuillel, M., Berthet-Colominas, C., Krop, B., Tardieu, A., Vachette, P. & Jacrot, B. (1983) *J. Mol. Biol.* **164**, 645–650
- Cuillel, M., Berthet-Colominas, C., Timmins, P. A. & Zulauf, M. (1987) *Eur. Biophys. J.* **15**, 169–176
- Cusack, S., Ruigrok, R. W. H., Krygsmann, P. C. J. & Mellema, J. E. (1985) *J. Mol. Biol.* **186**, 565–582
- Dahlbäck, B., Smith, C. A. & Müller-Eberhard, H. J. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3461–3465
- Debye, P. (1915) *Ann. Physik* **46**, 809–823 and 2927–2929
- Devaux, C., Timmins, P. A. & Berthet-Colominas, C. (1983) *J. Mol. Biol.* **167**, 119–132
- Gerchmann, S. E. & Ramakrishnan, V. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7802–7806
- Gilmour, S., Randall, J. T., Willan, K. J., Dwek, R. A. & Torbet, J. (1980) *Nature (London)* **285**, 512–514
- Glatter, O. & Kratky, O. (eds.) (1982) *Small Angle X-ray Scattering*, Academic Press, London
- Gogol, E. P. & Engelman, D. M. (1984) *Biophys. J.* **46**, 491–495
- Gray, D. M., Gray, C. W. & Carlson, R. D. (1982) *Biochemistry* **21**, 2702–2713
- Greulich, K. O., Wachtel, E., Ausio, J., Seger, D. & Eisenberg, H. (1987) *J. Mol. Biol.* **193**, 709–721
- Guinier, A. & Fournet, G. (1955) *Small-Angle Scattering of X-rays*, Wiley, New York
- Haas, J. A. (1977) Ph.D. Thesis, Institut Laue-Langevin, Grenoble
- Hervé, G., Moody, M. F., Tauc, P., Vachette, P. & Jones, P. T. (1985) *J. Mol. Biol.* **185**, 189–199
- Hjelm, R. P., Kneale, G. G., Suau, P., Baldwin, J. P. & Bradbury, E. M. (1977) *Cell* **10**, 139–151
- Hoppe, W. (1973) *J. Mol. Biol.* **78**, 581–585
- I'Anson, K. J., Bacon, J. R., Lambert, N., Miles, M. J., Morris, V. J., Wright, D. J. & Nave, C. (1987) *Int. J. Biol. Macromol.* **9**, 368–370
- Ibel, K. & Stuhmann, H. B. (1975) *J. Mol. Biol.* **93**, 255–266
- Ibel, K., May, R. P., Kirschner, K., Lane, A. N., Szadkowski, H., Dauvergne, M. T. & Zulauf, M. (1985) *Eur. J. Biochem.* **151**, 505–514
- Jacrot, B. (1976) *Rep. Progr. Phys.* **39**, 911–953
- Jacrot, B. (1981) in *Comprehensive Virology* (Fraenkel-Conrat, H. & Wagner, R. R., eds.), vol. 17, pp. 129–181, Plenum, New York
- Jacrot, B. & Zaccai, G. (1981) *Biopolymers* **20**, 2413–2426
- Kihara, H., Takahashi-Ushijima, E., Amemiya, Y., Honda, Y., Vachette, P., Tauc, P., Barman, T. E., Jones, P. T. & Moody, M. F. (1987) *J. Mol. Biol.* **198**, 745–748
- Koch, M. H. J. & Stuhmann, H. B. (1979) *Methods Enzymol.* **59**, 670–706
- Koch, M. H. J., Parfait, R., Haas, J., Crichton, R. R. & Stuhmann, H. B. (1978) *Biophys. Struct. Mech.* **4**, 251–262
- Koch, M. H. J., Sayers, Z., Vega, M. C. & Michon, A. M. (1987) *Eur. Biophys. J.* **15**, 133–140
- Kratky, O. (1963) *Progr. Biophys. Mol. Biol.* **13**, 105–173
- Krüse, J., Timmins, P. A. & Witz, J. (1982a) *Virology* **119**, 42–50
- Krüse, J., Krüse, K. M., Witz, J., Chauvin, C., Jacrot, B. & Tardieu, A. (1982b) *J. Mol. Biol.* **162**, 393–417
- Laggner, P., Kostner, G. M., Rakusch, U. & Worcester, D. L. (1981) *J. Biol. Chem.* **256**, 11832–11839
- Laggner, P., Kostner, G. M., Degovics, G. & Worcester, D. L. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 4389–4393
- Lander, G. H. & Price, D. L. (1985) *Physics Today*, 1–8
- Li, G. Z., Perkins, S. J. & Loucheux-Lefebvre, M. H. (1983) *Eur. J. Biochem.* **130**, 275–279
- Mandelkow, E. M., Harmsen, A., Mandelkow, E. & Bordas, J. (1980) *Nature (London)* **287**, 595–599
- Mandelkow, E., Mandelkow, E. M. & Bordas, J. (1983) *J. Mol. Biol.* **167**, 179–196
- Miles, M. J., Morris, V. J., Carroll, V., Wright, D. J., Bacon, J. R. & Nave, C. (1984) *Int. J. Biol. Macromol.* **6**, 291–292
- Miles, M. J., Morris, V. J., Wright, D. J. & Bacon, J. R. (1985) *Biochim. Biophys. Acta* **827**, 119–126
- Moody, M. F., Vachette, P. & Foote, A. M. (1979) *J. Mol. Biol.* **133**, 517–532
- Moody, M. F., Vachette, P., Foote, A. M., Tardieu, A., Koch, M. H. J. & Bordas, J. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 4040–4043
- Moore, P. B. (1980) *J. Appl. Crystallogr.* **13**, 168–175
- Moore, P. B. & Capel, M. S. (1988) *Annu. Rev. Biophys. Biophys. Chem.* **17**, 349–367
- Moore, P. B. & Weinstein, E. (1979) *J. Appl. Crystallogr.* **12**, 321–326
- Moore, P. B., Langer, J. A., & Engelman, D. M. (1978) *J. Appl. Crystallogr.* **11**, 479–482
- Moore, P. B., Capel, M., Kjeldgaard, M. & Engelman, D. M. (1986) *Biophys. J.* **49**, 13–15
- Nierhaus, K. H., Lietzke, R., May, P., Nowotny, V., Schulze, H., Simpson, K., Wurmbach, P. & Stuhmann, H. B. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2889–2893

- Nowotny, V., Nowotny, P., May, R. P. & Nierhaus, K. H. (1987) 10th Discussion Conference on Small-Angle Scattering and Related Methods, Prague, abstr. no. P79
- Osborne, H. B., Sardet, C., Michell-Villaz, M. & Chabre, M. (1978) *J. Mol. Biol.* **123**, 177–206
- Pardon, J. F., Worcester, D. L., Wooley, J. C., Tatchell, K., Van Holde, K. E. & Richards, B. F. (1975) *Nucleic Acids Res.* **2**, 2163–2176
- Perkins, S. J. (1985) *Biochem. J.* **228**, 13–26
- Perkins, S. J. (1986) *Eur. J. Biochem.* **157**, 169–180
- Perkins, S. J. (1988) in *New Comprehensive Biochemistry* (Neuberger, A. & Van Deenen, L. L. M., eds.), vol. 18B, part 2, pp. 143–264
- Perkins, S. J. & Sim, R. B. (1986) *Eur. J. Biochem.* **157**, 155–168
- Perkins, S. J. & Weiss, H. (1983) *J. Mol. Biol.* **168**, 847–866
- Perkins, S. J., Villiers, C. L., Arlaud, G. J., Boyd, J., Burton, D. R., Colomb, M. G. & Dwek, R. A. (1984) *J. Mol. Biol.* **179**, 547–557
- Perkins, S. J., Kerckaert, J. P. & Loucheux-Lefebvre, M. H. (1985) *Eur. J. Biochem.* **147**, 525–531
- Perkins, S. J., Chung, L. P. & Reid, K. B. M. (1986) *Biochem. J.* **233**, 799–807
- Perkins, S. J., Haris, P. I., Sim, R. B. & Chapman, D. (1988) *Biochemistry* **27**, 4004–4012
- Plietz, P., Damaschun, G., Müller, J. J. & Schwenke, K. D. (1983) *Eur. J. Biochem.* **130**, 315–320
- Pynn, R. & Fender, B. E. F. (1985) *Physics Today*, 47–53
- Ramakrishnan, V. (1986) *Science* **231**, 1562–1564
- Rolbin, Y. A., Kayushina, R. L., Feigin, L. A. & Schredin, B. M. (1973) *Kristallografiya* **18**, 701–705
- Sadler, D. M. & Worcester, D. L. (1982) *J. Mol. Biol.* **159**, 485–489
- Sadler, D. M., Rivas, E., Gulik-Krzywicki, T. & Reiss-Hiessan, F. (1984) *Biochemistry* **23**, 2704–2712
- Schumaker, V. N., Poon, P. H., Seegan, G. W. & Smith, C. A. (1981) *J. Mol. Biol.* **148**, 191–197
- Serdyuk, I. N., Agalarov, S. C., Sedelnikova, S. E. & Spirin, A. S. (1983) *J. Mol. Biol.* **169**, 409–425
- Spann, U., Renner, W., Mandelkow, E. M., Bordas, J. & Mandelkow, E. (1987) *Biochemistry* **26**, 1123–1132
- Stöckel, P., May, R. P., Strell, I., Cejka, Z., Hoppe, W., Heumann, H., Zillig, W., Crespi, H. L., Katz, J. J. & Ibel, K. (1979) *J. Appl. Crystallogr.* **12**, 176–185
- Stöckel, P., May, R. P., Strell, I., Cejka, Z., Hoppe, W., Heumann, H., Zillig, W. & Crespi, H. (1980a) *Eur. J. Biochem.* **112**, 411–417
- Stöckel, P., May, R. P., Strell, I., Cejka, Z., Hoppe, W., Heumann, H., Zillig, W. & Crespi, H. (1980b) *Eur. J. Biochem.* **112**, 419–423
- Stuhrmann, H. B. (1981) *Q. Rev. Biophys.* **14**, 433–460
- Stuhrmann, H. B., Tardieu, A., Mateu, L., Sardet, C., Luzzati, V., Aggerbeck, L. & Scanu, A. M. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 2270–2273
- Suau, P., Kneale, G. G., Braddock, G. W., Baldwin, J. P. & Bradbury, E. M. (1977) *Nucleic Acids Res.* **4**, 3769–3786
- Suau, P., Bradbury, E. M. & Baldwin, J. P. (1979) *Eur. J. Biochem.* **97**, 593–602
- Torbet, J., Gray, D. M., Gray, C. W., Marvin, D. A. & Siegrist, H. (1981) *J. Mol. Biol.* **146**, 305–320
- Torbet, J. & Freyssinet, J. M. (1987) *Biochemistry* **26**, 7791–7798
- Ueki, T., Hiragi, Y., Katoaka, M., Inoko, Y., Amemiya, Y., Izumi, Y., Tagawa, H. & Muroga, Y. (1985) *Biophys. Chem.* **23**, 115–124
- Winick, H. (1987) *Sci. Am.* **257**(5), 72–81
- Wise, D. S., Karlin, A. & Schoenborn, B. P. (1979) *Biophys. J.* **28**, 473–496