19.4.4. Distance measurements

19.4.4.1. *Theory and background*

By measuring distances between chemically specified points in a macromolecule or macromolecular complex, useful information may be obtained concerning its structure. In the case of a complex with many subunits, successive inter-subunit distance measurements can be combined by triangulation to generate a three-dimensional map specifying the relative positions of subunits and, perhaps, their individual radii of gyration. The only uncertainty is the handedness if more than three distances are combined. A number of approaches exist for making distance measurements, including NMR, fluorescence and electron-microscope techniques in which labelling is employed. Neutron scattering provides a particularly simple approach to such measurements in a conceptual sense. If we consider two labelled centres in a large macromolecule or complex, the interference of the scattering from the two labelled positions, $I_x(Q)$, can be separated from the scattering of the rest of the complex by measuring the scattering of the unlabelled complex, $I(Q)$, the two singly labelled complexes, $I_1(Q)$ and $I_2(Q)$, and the doubly labelled complex, $I_{12}(Q)$, as they are related by

$$I_x(Q) = [I_{12}(Q) + I(Q)] - [I_1(Q) + I_2(Q)]. \qquad (19.4.4.1)$$

Using the Debye relationship, it can be shown that the cross term will have the approximate form

$$I_x(Q) = 2f_1f_2 \sin(d_{12}Q)/d_{12}Q, \qquad (19.4.4.2)$$

where the interference intensity results from the correlation between radiation scattered from sites 1 and 2, with scattering strengths f_1 and f_2. It is a damped sinusoidal fringe with a periodicity reciprocally related to the separation between the scattering regions. If the scattering regions are very small compared with the object, there will be nodes at equal intervals in Q; if they are not, $I_x(Q)$ will be a sum of all cross correlations of label positions in the two regions.

19.4.4.2. *Neutron distance measurements*

In the neutron case, the usual application involves deuterium labelling of positions in the two regions, and the cross term becomes

$$I_x(Q) = 2(b_D - b_H)^2 \sum_i \sum_j (\sin d_{ij}Q)/d_{ij}Q, \qquad (19.4.4.3)$$

where the first sum runs over all labelled positions in region 1 and the second over all labelled positions in region 2.

The interference experiment as a technique for studying macromolecules in solution was first proposed in 1947 as an X-ray scattering measurement in which the labels are heavy atoms (Kratky & Worthmann, 1947). It has been discussed several times since and tested (Hoppe, 1972; Vainshtein *et al.*, 1970); however, no biological application has ever been published, since the signals are small (Hoppe, 1972). In 1972, the interference idea was discovered again, but in the context of neutron scattering (Engelman & Moore, 1972). Initially, it was thought that the distance measurement would follow from inspection of the damped sinusoidal term, but it soon became evident that the size and shape of the labelled regions would have an important influence. Using a power series expansion like that originally employed by Guinier, the previous expression becomes (for small Q)

$$I_x(Q) \propto [1 - (1/6)(R_i^2 + R_j^2 + d_{ij}^2)Q^2],$$

where R_i and R_j are the radii of gyration of the subunits *in situ*, and d_{ij} is the separation of their centres of mass. It follows that (Moore *et al.*, 1978; Stöckel *et al.*, 1979)

$$R_g^2 = (1/2)(R_i^2 + R_j^2 + d_{ij}^2). \qquad (19.4.4.4)$$

The second moment of a length distribution, M_{ij}, can be related to a radius of gyration. If $p_{ij}(r)$ is the length distribution of the cross term $I_x(Q)$, then

$$M_{ij} = \int r^2 p_{ij}(r)\, \mathrm{d}r = 2R_g^2.$$

Hence (Moore *et al.*, 1978)

$$M_{ij} = R_i^2 + R_j^2 + d_{ij}^2. \qquad (19.4.4.5)$$

Thus, M_{ij} is the parameter that contains the information in the difference experiment. If the radii of gyration of the labelled regions are small, then the distance is well measured in a single experiment. This would be the case, for example, if the labels were single heavy atoms in an X-ray experiment. However, in most complexes of macromolecules there will be many pairwise protein–protein relationships where the radii and the separation of centres have comparable magnitudes. One approach to extracting distance information is to know the radii of the subunits *in situ*, by estimation from their molecular weights, by measurement of the isolated subunits or from triple isotopic substitution (see above). The first approach is the least desirable but the easiest, and the last approach is correct but laborious.

If a complex has eight or more distinct subunits, the number of possible measurements of M_{ij}, $n(n-1)/2$, is sufficient to solve for the distances and radii, enriching the information obtained from the experiment. The minimum number of measurements required is $4n - 6$, and for large complexes, such as the ribosome, a great excess of possible measurements exists, permitting refinement of the information through redundancy. Moore & Weinstein (1979) have described analytical methods for solving the problem with correct error propagation.

19.4.4.3. *The statistical labelling method*

In many biological assemblies, multiple copies of a given subunit are found. In such cases, statistical labelling methods provide useful information. The basic concept is to produce two samples in which labelled subunits are randomly mixed with unlabelled subunits within the particles, and in which the ratios of labelled to unlabelled subunits is different. The difference in the scattering from the two samples will then contain shape and orientation information for the subunits (Fujiwara & Mendelson, 1996). A simple case is that of a dimer in a larger complex. Random introduction of equimolar labelled (D) and unlabelled (H) subunits will result in a mixture of complexes, where HH:HD:DH:DD = 1:1:1:1. A sample can be prepared with an equimolar mixture of all H and all D so that HH:DD = 1:1. The difference in scattering between these samples will be simply the interference cross term for the dimer, but at half the weight it would have in the case where the subunits were distinctly labelled.

19.4.5. Practical considerations

19.4.5.1. *Feasibility*

For a biological scientist, the first question is whether knowledge of the parameters that can be obtained from neutron scattering is of interest for the biological problem under consideration. If the answer is affirmative, the best course of action is to make contact with individuals who have conducted experiments in the past, as well as with biologists at a neutron-scattering facility. As a preliminary guide, a few general points are pertinent.

For a solution-scattering experiment on an unlabelled sample, typical sample volumes are 0.1–0.2 ml at concentrations of around

Copyright © 2006 International Union of Crystallography

10 mg ml^{-1}. For contrast variation, samples dialysed to different D_2O levels will be required, so significant amounts of material need to be at hand. Making a set of measurements will require a few hours at a modern facility.

Longer collection times and significantly greater biochemical efforts will be involved for measurements with labelled material, especially if reconstitution is part of the strategy. Labelled biomolecules are most often produced by growth of organisms in D_2O (Moore & Engelman, 1976; Vanatalu *et al.*, 1993), but strategies using chemical synthesis or providing labelled precursors have also been employed. Preparation of samples with labelled ligands is usually more straightforward than the generation of reconstituted complexes, but still requires tests of homogeneity.

19.4.5.2. *Homogeneity and stability*

For measurements to be made, the samples must be transported to a neutron facility and measured for many hours. This requires stability in the usual biochemical sense, and proper experiments will include assays performed after the measurements. Damage from the neutrons themselves is minimal, but handling and transportation impose greater challenges.

Where mixed samples are employed, stability also involves resistance to exchange of subunits or ligands between complexes in the mixture. This is hard to assess in advance, the first sign of trouble often being an absence of an expected difference signal. Given shipment and measuring times, even a slow exchange is significant. With a consequent qualification of the results, cross-linking may be used to stabilize the complex.

In all scattering measurements, homogeneity of the sample is a great advantage in interpretation. However, some variation is tolerable for many purposes, and the level of purity typically sought for crystallization experiments is not usually required.

19.4.5.3. *Solvent conditions*

Background arises from the incoherent scattering of hydrogen. Consequently, the use of high levels of D_2O in the buffer improves the background, but may not be innocuous, since high levels of D_2O have been seen to induce aggregation. In cases where complexes with labelled subunits are measured, it is advantageous to suppress small-angle background scatter by contrast matching the unlabelled regions *via* solvent adjustment.

19.4.6. Examples

19.4.6.1. *Contrast variation*

An early result that changed ideas about nucleosome organization came from measurements relying on the difference in protein and DNA densities (Baldwin *et al.*, 1975; Bradbury *et al.*, 1976). Information on the organization of serum lipoproteins was also based on intrinsic scattering differences (Stuhrmann *et al.*, 1975; Atkinson & Shipley, 1984). The contrast between RNA and protein was used in early ribosome measurements (Crichton *et al.*, 1977; Moore *et al.*, 1974). Recent examples include detergent binding to membrane proteins (Timmins *et al.*, 1991) and the study of membrane protein–lipid complexes (Jeanteur *et al.*, 1994). Elegant use of contrast variation resulted in a structural explanation of the anti-cooperative binding of tRNA to synthetase, which had resisted study by other methods (Dessen *et al.*, 1978).

19.4.6.2. *Contrast matching*

In some applications, use has been made of contrast matching a large part of a complex to study a component with a contrasting scattering density. Examples include studies of ribosomal proteins *in situ* (Nierhaus *et al.*, 1983; Nowotny *et al.*, 1994), proteins of the DNA-dependant RNA polymerase (Stöckel *et al.*, 1980*a*), studies of muscle proteins (Stone *et al.*, 1998) and a view of a membrane protein in a lipid bilayer (Hunt *et al.*, 1997).

19.4.6.3. *Spin contrast variation*

Some applications of the spin contrast variation method have appeared (Junemann *et al.*, 1998; Nierhaus *et al.*, 1998).

19.4.6.4. *Specific deuteration, combination with X-ray measurements*

Many recent measurements use specific labelling and combine information with data from small-angle X-ray scattering. Studies of ligand binding (Bilgin *et al.*, 1998), hydration layers (Svergun *et al.*, 1998), troponin complexes (Olah *et al.*, 1994) and ribosomes (Svergun *et al.*, 1996) are examples that illustrate the approaches.

19.4.6.5. *Distance measurements and triangulation*

By measuring distances and radii of gyration, models of some large complexes have been created. These include a model of the relative positions of all 21 proteins in the small ribosomal subunit from *E. coli* (Capel *et al.*, 1987), a model of seven proteins from the large subunit (May *et al.*, 1992) and a model of the DNA-dependant RNA polymerase (Stöckel *et al.*, 1980*b*). Intramolecular conformational states of cholesterol esters have been observed using distance measurements on molecules chemically labelled in different regions (Burks & Engelman, 1981).

references