

APPENDIX

Model Calculations of Protein Pair Interference Functions

P. B. MOORE AND D. M. ENGELMAN

A1. Introduction

We present below the results of some calculations done to illustrate the dependence of the interference cross term, derived from neutron scattering studies of structures containing pairs of deuterated proteins, on a variety of experimental factors (Engelman & Moore, 1972; Engelman *et al.*, 1975; main text). The factors considered are: (a) variation in the separation of the pair of substituted proteins, (b) fractional occupancy of the protein sites in the complex, (c) exchange of substituted proteins within samples, (d) multiple locations of a given protein in the structure, and (e) protein shape and orientation. These results provide guidance for the evaluation of the accuracy of pair-separation measurements obtained by the neutron scattering method, and should help clarify the interpretation of such data.

A2. Method

Our approach has been to proceed in 2 steps to obtain radial distribution functions and interference cross terms for models consisting of pairs of ellipsoids of revolution. The first step is to generate a pair-distribution function, that is, a radial distribution function describing the distribution of lengths of all vectors relating the 2 proteins of the pair. This is accomplished using Monte Carlo methods taking advantage of the fact that a point can be characterized as being inside an ellipsoid of revolution by comparing its distance to the 2 foci with the constant sum of such distances characteristic of the ellipsoidal surface. Thus one can locate points in each of 2 ellipsoids at random and determine the distance separating them. This process is repeated until a sufficient number of lengths have been determined to describe the radial distribution function with reasonable precision. In general, 10,000 points were computed for each ellipsoid pair. Once a distribution is obtained, radial Fourier inversion is done to obtain the pair interference function.

In the case where 2 spherical proteins are examined, the interference cross term is computed directly, taking advantage of the fact that it can be written in closed form. Thus:

$$I_{\text{cross}}(s) \propto \rho_1 \rho_2 (1/36\pi^4 s^6) (\sin 2\pi R_1 s - 2\pi R_1 s \cos 2\pi R_1 s) \cdot (\sin 2\pi R_2 s - 2\pi R_2 s \cos 2\pi R_2 s) \cdot \frac{\sin 2\pi s x}{2\pi s x} = \rho_1 \rho_2 f_1 f_2 \frac{\sin 2\pi x}{2\pi x}, \quad (\text{A1})$$

where R_1 and R_2 are the radii of the spheres in question, ρ_1 and ρ_2 are their respective scattering densities, s is the reciprocal space co-ordinate and x the distance separating the spheres. The quantities f_1 and f_2 are the transforms of the individual spheres.

A3. Results

(a) Separation

Figure A1 shows the effect on the interference function of varying the separation of two 34 Å spheres. The separations shown here are 35, 65 and 95 Å and a progressive

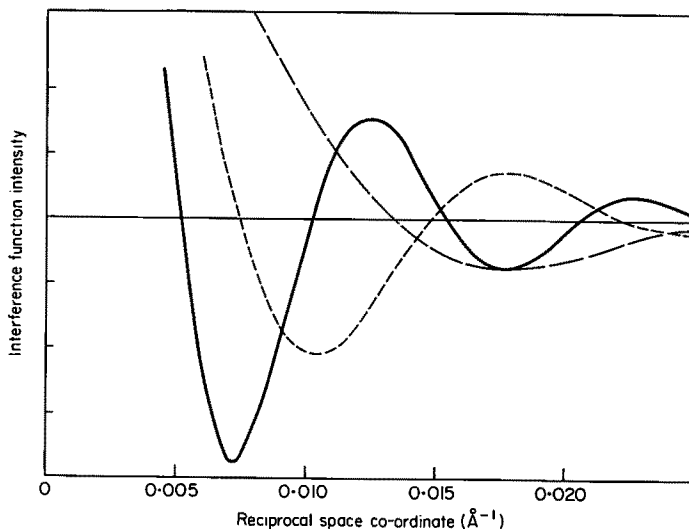


FIG. A1. Interference functions for spheres at various separations. Interference cross terms for the scattering for a pair of spheres of 34 Å diameter are shown for center-to-center separations of 35 Å (—), 65 Å (---) and 95 Å (-·-·-). The damping of the interference function by the transforms of the spheres is evident with increasing reciprocal space co-ordinate ($s = 2 \sin \theta/\lambda$); however, such damping has no effect on the point(s) at which the functions reach zero. Thus, the measurements of the separations based on the first zero are not affected by the magnitude of the separations.

damping of the amplitude of the interference ripple is apparent as the separation decreases. It is clear from equation (A1) that the closer the two spheres are, the more strongly damped the interference ripple will be, due to the envelope of the sphere transforms, f_1 and f_2 , which decrease with increasing s over the region considered. The locations of the nodes of the interference curves are not influenced by this damping, in the spherical case. They occur at intervals of $s = n/2x$, where n is an integer. An example of a strongly damped curve resulting from close proximity is found in the data for proteins S5 and S8 (Engelman *et al.*, 1975).

(b) Occupancy

A difference in the occupancy of the substituted protein sites among reconstituted particles used in the experiment can lead to errors in the estimation of the separation. If samples differ in occupancy, subtraction of their scattering profiles will not give the correct interference cross term. An additional contribution will be included in the difference, whose form is the transform of the shape of the low-occupancy protein, giving rise to a false background in the interference function. Concurrently, the cross term signal will be reduced in intensity. The consequences of a 10% difference in occupancy of deuterated proteins in the doubly substituted *versus* the two singly substituted reconstitution products used in the experiments are shown in Figure A2. The lower curve, that calculated for a low-occupancy experiment, lies below the ideal curve for all s values by an amount which decreases as s increases. The result is a small error in the estimate in the separation of proteins based on the first intercept with a zero level estimated from the high-angle difference value. This error, for a 10% difference in substitution, is approximately 3.5%. Differences between samples in

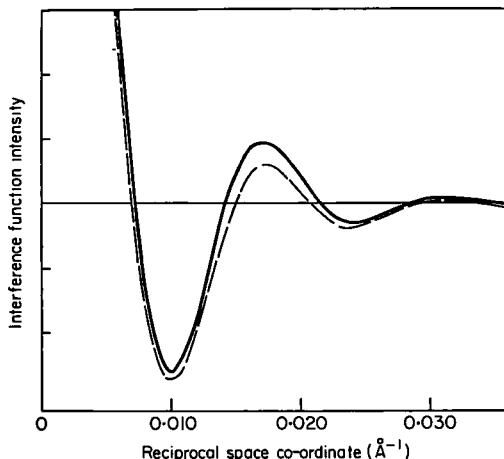


FIG. A2. Effect of occupancy on the interference function. A deficiency in the amount of the deuterated proteins bound to one of the reconstituted subunits adds a contribution to the interference signal having the form of the transform of a single protein. The Figure shows the effect of a 10% deficiency of one of the proteins in the ribosomal particle with 2 proteins deuterated. The transform of that protein is oversubtracted in obtaining the interference function, shifting the curve to a lower baseline and increasing the observed separation from 70 to 72.5 Å based on the first zero intercept. The calculation was done using the sphere transform method for 30 Å diameter spheres at 70 Å separation and scattering densities appropriate for our standard experimental conditions (Moore *et al.*, 1975). If the deficiency in occupancy were present in one of the reconstituted ribosomes containing a single deuterated protein, the result would be similar but in the opposite direction, that is, the single protein transform would add to the interference curve, shifting the curve up and slightly decreasing the value obtained for the separation of centers.

the occupancy of protonated components would be undetectable due to the small signal generated by hydrogenated proteins under conditions of contrast matching. Thus, small differences in occupancy among reconstituted particles used in the experiment can lead to errors in the estimate of the separation of centers; however, these errors are small. The shape features of the interference function are only modestly affected by such difficulties.

(c) *Exchange*

A different kind of occupancy problem is posed by proteins which may exchange between reconstituted particles. In such a case, the mixtures which were used to derive the pair interference function would have a different distribution of component from that expected based on the experimental design. For example, if we consider the all-hydrogen reconstituted particle which is mixed in equal proportions with the particle containing two deuterated proteins, exchange will produce the two reconstituted species containing single deuterated proteins paired with hydrogenated proteins. The converse will occur in the other sample. This will have the effect of simply reducing the contribution of the pair interference to the scattering interference curve and will be seen as a reduction in amplitude of J_{cross} (to zero if exchange is complete). If both mixtures exchange equally, then the effect will be a simple loss of signal strength, as is the case for any deviation which occurs equally in both samples. If the mixtures exchange to different extents, the effects will be similar to those seen as a consequence of occupancy. These difficulties, if encountered, can be overcome through the use of the method as we originally proposed it (Engelman & Moore,

1972), since this approach employs four separate reconstituted particles which are not mixed and therefore not subject to redistribution of deuterated substituents by exchange.

(d) *Degeneracy*

A somewhat different situation is that of distinguishing between the effects of motion or degenerate location of a protein and the effect of its physical extent. An extreme situation of this kind would be the case of a protein which may occupy either one site or another in the structure *versus* a protein which has two lobes at equivalent locations. Such an example is shown in Figure A3. It is clear that the signals produced in the two cases differ slightly in amplitude over the range of interest, and virtually not at all in shape. Such an absolute difference would be very hard to detect. Thus, one cannot expect to distinguish between shape effects and degenerate locations, using interference data alone.

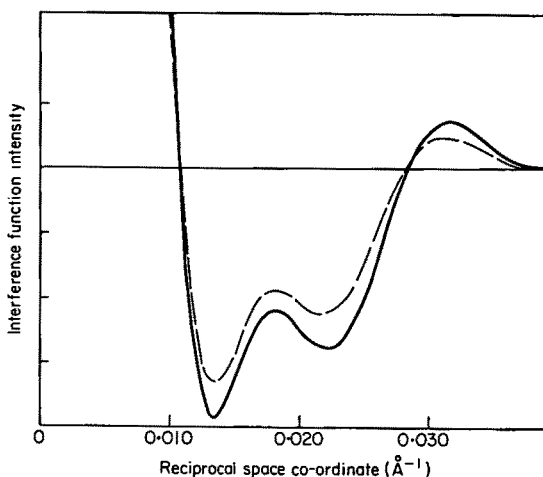


FIG. A3. Two locations for a protein, having equal probability of occupancy, compared to a protein of equal molecular weight with 2 lobes in the same positions. The sphere transform formula was used to calculate the interference functions for a 30 Å sphere, 35 Å from a reference sphere, and the same sphere, 90 Å from the reference. The 2 curves were divided by 2 and summed to give a curve for the 2 possible locations model (-----). Similar computations were done for two 27.2 Å spheres at the same positions relative to the reference. These 2 profiles were summed to represent the two-lobed model (—————).

It may be possible to resolve such problems by measuring the radius of gyration of the deuterated protein in singly substituted particles. Although such measurements would be hard to make due to contributions to scattering due to interference between the deuterated protein and the particle as a whole, radius of gyration measurements might be of sufficient accuracy to distinguish between a two-lobed structure and a spherical structure present in either of two locations. Examples of cases in which this problem of interpretation arises are all cases in which the radial distribution function is broad, e.g. S4–S7. To summarize, the pair interference measurement does not distinguish between physical extent and motion or degeneracy of one or both of the members of the protein pair being studied. The length-distribution function is averaged over the whole specimen population and averaged over time.

(e) *Shape*

We have refined and extended calculations concerning shape effects which we began late in 1971 and which were referred to in our earlier publication (Engelman & Moore, 1972). Figure A4 shows the influence of shape and orientation on the interference function and radial distribution function for each of several orientations of a pair of extended ellipsoids and for a sphere paired with an extended ellipsoid. In general, shape influences the damping of the interference curve and the breadth of the radial distribution function, but only weakly the position of the first node of the interference signal. More extended shapes in the direction of the line separating the centres of the

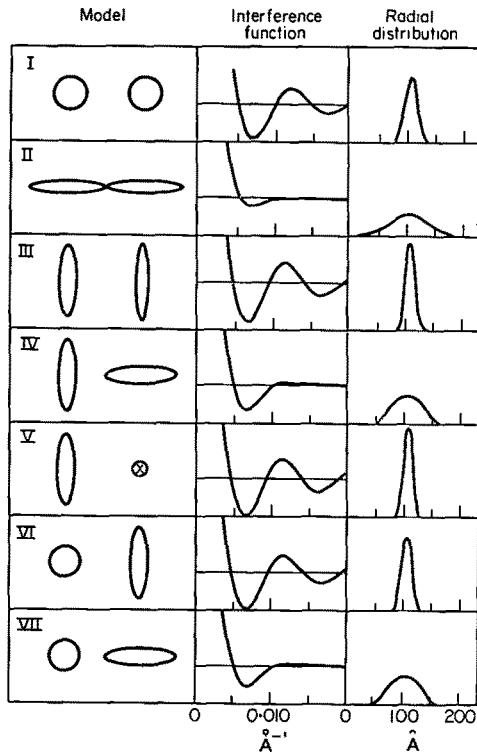


FIG. A4. Shape and orientation effects. Interference curves and pair distribution profiles were computed for 7 spherical or ellipsoidal models as described in Method. Ellipsoids considered were all 5:1 prolate ellipsoids of revolution. For each case, the models correspond to pairs of 25,000 dalton proteins separated by 100 Å. The shapes and orientations of the proteins are represented schematically and then the interference curve and length-distribution profile expected are shown.

two proteins produce stronger damping and broader distribution functions. It is to be noted that even extended shapes can give interference functions and radial distribution functions which are difficult to distinguish from those of spheres, if the axes of the extended proteins are oriented perpendicular to the line separating the centers of the proteins. Table A1 shows the values obtained for the separations based on the location of the first zero of the interference cross term and those based on the location of the peak in the radial distribution function. In all cases, the errors are relatively modest, being less than 10% of the distance separating the centers. The use of the first zero of the cross term to establish the separation of the proteins is to be preferred

TABLE A1

Parameters of model functions shown in Figure A4

Model (see Fig. A1)	Distance obtained from 1st zero pt. (Å)	Peak of radial distribution (Å)	Radial distribution half-width (Å)
I	102	104	27
II	90	102	83
III	108	107	20
IV	101	103	70
V	107	108	20
VI	104	106	24
VII	95	105	65

since it is a more direct property of the data than is the radial distribution function which is obtained only by smoothing and manipulating the data. The parameter which is most sensitive to shape and orientation is the width of the distribution. The width of such a distribution is conveniently characterized using a parameter such as the full-width of the distribution at half-maximum height (FWHH). The relationship between FWHH and axial ratio and orientation has been explored for prolate ellipsoids, by model computations. The relationship we find can be represented as follows:

$$\text{FWHH} \sim (R_1 + R_2) \cdot 0.8 + [(|\vec{A}_1 \cdot \vec{D}_{1,2}| - R_1) + (|\vec{A}_2 \cdot \vec{D}_{1,2}| - R_2)] \cdot 1.2, \quad (\text{A2})$$

where R_1 and R_2 are the radii the two proteins would have if they were spheres, \vec{A}_1 and \vec{A}_2 are vectors having the length and orientation of the major axes of the two ellipsoids and $\vec{D}_{1,2}$ is a unit vector in the direction connecting the centers of mass of the two ellipsoids. If either of the two expressions of the form: $|\vec{A} \cdot \vec{D}| - \bar{R}$ is less than zero, it is replaced by zero.

What equation (A2) means is that it is very difficult to generate a geometry having a narrower distribution than the distribution found for two spheres of the appropriate size, and that the full-width at half-maximum reflects primarily the projected length of the major axes of the two ellipsoids along the line joining centers of mass. This relationship was derived from a series of computations in which one of the two objects was modeled as a sphere and the second object was an ellipsoid whose major axis was pointed towards the center of the reference sphere and whose axial ratio was varied. What is surprising is that this expression also predicts the full-width at half-maximum observed for situations in which both subunits are ellipsoids, regardless of their axial ratios or relative orientation. For these more complex situations, the expression will give an estimate of the full-width at half-maximum accurate to about $\pm 15\%$.

A4. Conclusions

On the basis of the model calculations conducted using ellipsoids of revolution and spheres to test various influences on the radial distribution functions and pair interference functions obtained from neutron scattering experiments, several conclusions concerning the interpretation of such data may be reached. The first is that the

influences of protein shape and occupancy on the measurement of the distance separating the centers of mass of a pair of proteins are modest. The use of the first zero of the interference cross term to obtain the distance separating the pair of proteins is a direct and reasonably accurate method for extracting this information from the data. As a guide, one would be unlikely to be misled in assuming an error of $\pm 10\%$ as the accuracy for distance measurements obtained in this way.

“Shape” information, while contained in the interference cross term and radial distribution functions, can equally represent motion, degeneracy, shape, or a combination of all these factors. Thus a model produced by interpreting pair distribution data in terms of equivalent ellipsoids would represent the average of all possible positions of every protein in the population. It might not be a good representation of the structure of any of the conformations being averaged. In summary, the neutron scattering method is a relatively accurate and sound method for obtaining values for the average distances separating the centers of a pair of proteins, but that the extraction of further information concerning the shape of the proteins must be done with full appreciation of the implications of the averaging involved.

REFERENCES

- Engelman, D. M. & Moore, P. B. (1972). *Proc. Nat. Acad. Sci., U.S.A.* **69**, 1997–1999.
Engelman, D. M., Moore, P. B. & Schoenborn, B. P. (1975). *Proc. Nat. Acad. Sci., U.S.A.* **72**, 3888–3892.
Moore, P. B., Engelman, D. M. & Schoenborn, B. P. (1975). *J. Mol. Biol.* **91**, 101–120.