BNL--34681

Brookhaven Symposium Biology 32

DE84 012174

Brookhaven National Laboratory, June 1-4, 1982
NEUTRONS IN BIOLOGY, B. Schoenborn, Plenum Press NY

Conf Art
Bk Chpt

NEUTRON SCATTERING AND THE 30 S RIBOSOMAL SUBUNIT OF E. COLI

P.B. Moore,[a] D.M. Engelman,[b] J.A. Langer,[b]
V.R. Ramakrishnan,[a] ..G. Schindler,[a] B.P. Schoenborn,[c]
I-Y. Sillers,[a] and S. Yabuki[a]

[a]Dept. of Chemistry and
[b]Molecular Biophysics and Biochemistry
Yale University, New Haven, CT 06511

[c]Biology Dept., Brookhaven National Lab., Upton, NY 11973

INTRODUCTION

Ribosomes are nucleoprotein enzymes which catalyze the formation of polypeptide chains under mRNA control, using aminoacyl tRNAs as substrates--for reviews see Nomura et al. (22) and Chambliss et al. (2). While our knowledge of what these particles do in protein synthesis is satisfactory, our understanding of how they do it is minimal. We still have no idea, for example, what there is about the mechanism of protein synthesis that requires all ribosomes, whatever their source, to be two-subunit enzymes. It is most unlikely that mechanistic questions of even this simple kind will be answered until much more is known about the three-dimensional structure of these particles than is known today.

The barrier posed by our ignorance of ribosome structure to further understanding of protein synthesis has been recognized for a long time. Ten years ago it was pointed out that neutron scattering could make a useful contribution in this area (4), and the first results of the application of these ideas were reported at the 1975 Brookhaven Symposium (5,11).

The purpose of this paper is twofold: (i) It reviews the progress made in the study of the internal organization of the 30 S ribosomal subunit of E. coli by neutron scattering since 1975. A map of that particle showing the position of 14 of the subunit's 21 proteins will be presented, and the methods currently used for collecting and analyzing such data will be discussed. (ii) It also explores the possibility of extending the interpretation of neutron mapping data beyond the limits practical today.

THE EXPERIMENT

One of the most powerful ways to use neutron scattering in biological systems is to combine scattering measurements with specific deuterium labeling. Large numbers of non-exchangeable deuteriums can be incorporated into biological macromolecules. Consequently, because of the large difference in scattering length between ^{1}H and ^{2}D, labeling of quite small regions of a larger biological structure can result in a measurable alteration in its scattering. These changes can reveal the positions of the isotopically labeled sites, leading to a far better understanding of an object's internal organization than would be possible otherwise in the context of a scattering experiment.

The bacterial ribosome lends itself to site-specific deuterium labeling at the level of its constituent proteins. The small subunit, for example, is a complex of 21 unique protein molecules and a single RNA. Bacteria grow in perdeuterated media, and their ribosomes can be reassembled from their separated components in vitro. Thus, particles can be prepared with one or more proteins deuterium substituted at will, as described by Moore (19).

The measurements made on ribosomes containing deuterated proteins follow a scheme first proposed for measuring distances between pairs of specific sites on single molecules by x-ray scattering (9,13,29). X-ray scattering measurements can be done on solutions of particles in which the two sites are marked with heavy atoms. The heavy atom contribution to the overall scatter can be identified by comparison with that given by the unlabeled molecule. Because of interference between the scatter of the two atoms, the heavy atom contribution includes a conspicuous ripple which can be isolated experimentally from the rest of the scatter of the sample. The periodicity of this ripple reveals the distance between the heavy atoms and hence between the two labeled sites.

The neutron experiment done on ribosomes uses as its "heavy atoms" entire protein molecules labeled with deuterium. Otherwise the experiment is completely analagous. The techniques used for differencing scattering profiles to isolate the interference contribution from all other scattering contributions have been described at length elsewhere (6,23). Suffice it to say that the interference signal, $I_x(s)$, can be obtained from real data in an unambiguous fashion, and from it one would hope to deduce an interprotein distance.

In a multisubunit structure like the ribosome where many interference fringes can be measured, a set of pairwise distances could reveal the positions of its components in three dimen-

sions. Under almost no other circumstance can one imagine concatenating the results of a long series of solution scattering measurements to produce as much information about a structure as this. The possibility of obtaining such a three-dimensional structure is what brought us to apply this technique to the ribosome, where so little information was (is) available from other sources.

That data of this kind could be obtained from ribosomes by neutron scattering has been clear for a long time. That it could provide information about the separations between components has also been certain. Less evident has been the proper method for extracting information about the structure of the ribosome from these interference curves.

DATA ANALYSIS

a) The Interference Fringe

There is one experimental situation in which the interpretation of pairwise, interference fringes is straightforward. If the labeled entities have spherical symmetry, like atoms, $I_x(s)$ has the form

$$I_x(s) = 2f_1(s)f_2(s)\frac{\sin(2\pi ds)}{2\pi ds} \qquad (1)$$

where $f_1(s)$ and $f_2(s)$ are the form factors for the labeled regions, d is the distance between their centers, and s is the Bragg spacing at which the scattering is observed; $s = (2\sin\theta)/\lambda$.

In this case, $I_x(s)$ is a damped sinusoidal ripple just as it would be if the labeled regions were point scatterers. The nodes of the ripple occur at intervals of $(2d)^{-1}$, and inspection reveals the distance between the centers of the scatterers.

There is a difficulty, however. Most proteins, ribosomal proteins included, are not spheres. Equation (1), therefore, is not an appropriate basis for interpreting the interference data they give. For non-spherical scatterers, a more general expression must be used:

$$I_x(s) = \sum_i \sum_j (2\pi s r_{ij})^{-1}\sin(2\pi s r_{ij}) \; . \qquad (2)$$

In Eq. (2), r_{ij} is the distance between the ith atom in the first labeled subunit and the jth atom in the second. The i summation runs over all atoms in the first subunits, and the j summation over all atoms in the second. (Constants such as scattering lengths are omitted for clarity.)

b) Length Distributions

Like any other scattering curve, $I_x(s)$ can be subjected to Fourier transformation. The transform of a solution scattering curve is (always) a length distribution--see Guinier and Fournet (7)--and in this case the distribution obtained, $p_x(r)$, is the distribution of lengths of all possible vectors joining isotopically substituted positions in the two labeled regions (17,18):

$$p_x(r) = r\int_0^\infty sI_x(s)\sin(2\pi sr)ds \ . \qquad (3)$$

The length distribution in this case reflects the distance between the centers of mass of the two labeled subunits, to be sure. But it is also influenced by their shapes and relative orientation. Thus length distributions should, and do, vary considerably in shape as well as in average distance from one pairwise experiment to the next. The interference profiles measured vary correspondingly, and often deviate significantly from the sinusoidal regularity of the spherical case.

It follows that, in general, inspection of such a ripple will not reveal the center-of-mass separation of a pair of proteins. All one can suggest is that the center-to-center distance is likely to be within the observed length distribution, probably somewhere around its average value, but even this need not always be true.

c) Second Moments

About five years ago it was recognized that a simple relationship exists between the second moment of a length distribution, M, and the center-to-center distance between subunits, d, which offers a way around the impasse described above:

$$M = d^2 + R_1^2 + R_2^2 \qquad (4)$$

where R_1 and R_2 are the radii of gyration of the labeled entities (16,18,27). (The second moment of a length distribution is twice the square of the radius of gyration one would derive by analysis of the low-angle region of the corresponding scattering profile.) Equation (4) is valid independent of the shapes and orientation of the two labeled regions.

A priori one does not know R_1 or R_2; Eq. (4) does not permit one to interpret a single, isolated data set. For structures like the ribosome with more than 8 subunits, however, the number of pairwise distance measurements possible within the object is greater than the number of positional coordinates and radii of gyration needed to specify it in the framework of Eq. (4). Therefore it is possible to derive a model for subunit positions

and radii of gyration of such an object by an analysis of a set of interference data which is independent of ad hoc assumptions, or any additional data.

The problem posed by a set of data of this kind can be solved satisfactorily by finding a model for the structure in which center-of-mass positions and radii of gyration are specified so as to minimize an objective function χ^2, which can be defined conveniently as follows:

$$\chi^2 = \Sigma \frac{1}{\sigma_{ij}^2} \left[M_{ij} - (x_i - x_j)^2 - (y_i - y_j)^2 - (z_i - z_j)^2 - R_i^2 - R_j^2 \right]^2 \qquad (5)$$

where x_i, y_i, z_i are the coordinates of the ith component, R_i its radius of gyration, M_{ij} the second moment of the length distribution found for the pair i and j, and σ_{ij}^2 the variance of M_{ij}. The sum is over all data sets.

The theory and the computational techniques required to find the optimal solution to Eq. (5) have been described in detail (20,23). It is these techniques that have led to the map of the 30 S subunit to which we now turn.

EXPERIMENTAL RESULTS

a) Samples

We started making ribosome interference measurements in the summer of 1974 using samples which were protonated in all but the protein subunits whose separation was sought. These were deuterium labeled to the extent of 80 to 85% in non-exchangeable positions, and the particles were suspended in 57% D_2O buffer to minimize ribosomal contributions to scatter by contrast matching ribosomal protonated components (19).

This pattern of labeling was chosen because it was the most economical one we could find which was adequate for the job. While it is a relatively cheap scheme to carry out, it has some experimental overhead. The buffer scatter of 57% D_2O is substantial because of the incoherent scattering from the ^1H atoms it contains; this is the primary source of background in the experiment. From the viewpoint of signal-to-noise ratio, it would have been better to use partially deuterated particles, matched to 100% D_2O, labeled with protonated proteins, like those being used at I.L.L. today (May, personal communication). The tripling in D_2O consumption such a strategy entails appeared more than our resources could bear, however, in 1974.

In order to reduce the amount of buffer in the samples and hence the background penalty, the samples are concentrated by centrifugation to give a gel-like suspension. Highly concentrated samples are seldom suitable for solution scattering work because of interparticle interference effects. The reason concentrated samples should be tolerable in these experiments has been pointed out by Hoppe (9,10), and we tested his theory on two occasions with satisfactory results (24,25).

b) Data

So far 62 different protein pairs have been examined within the 30 S subunit, 8 of them more than once. The positions and radii of gyration for 14 of the 21 proteins in that structure can be estimated by using 54 of these data sets.

An exhaustive analysis of these results will not be presented here. All but 5 of the data sets required for the 14 protein maps have been discussed elsewhere (24,26) in a series of papers which presented a 12-protein map. Furthermore, the current 14-protein model contains some unresolved ambiguities and is therefore more than usually tentative. It should be replaced by something better quite soon. Instead of a full discussion, some pictures of this working model will be supplied, with comments on its strengths and weaknesses, and some remarks on comparisons between it and other data on the organization of the ribosome.

c) A 14-Protein Model of the 30 S Subunit

Figure 1 presents "front" and "back" stereo views of our current best estimate of protein locations. The array of proteins is fairly flat and is viewed roughly parallel to its thinnest direction. The maximum linear dimension of the array is about 170 Å, which should be compared with 220 to 250 Å, the maximum chord of the entire structure (Kearney and Moore, unpublished data). For convenience proteins are represented as spheres whose volumes are to scale.

Alternative sources of information about this protein arrangement have come from reassembly experiments, protein cross-linking experiments, fluorescence energy transfer studies, and finally electron microscopy. On the whole, the agreement between these different sources of information and the neutron model is quite good, as discussed in detail elsewhere (24). Perhaps the most striking comparison that can be made is between the neutron map and the data from electron microscopy. Maps of protein positions have been produced by examination of electron microscopic images of 30 S subunits stained with protein-specific antibody molecules (12,28). Staining with antibodies permits localization of the antigenic determinants of proteins within the larger structure.

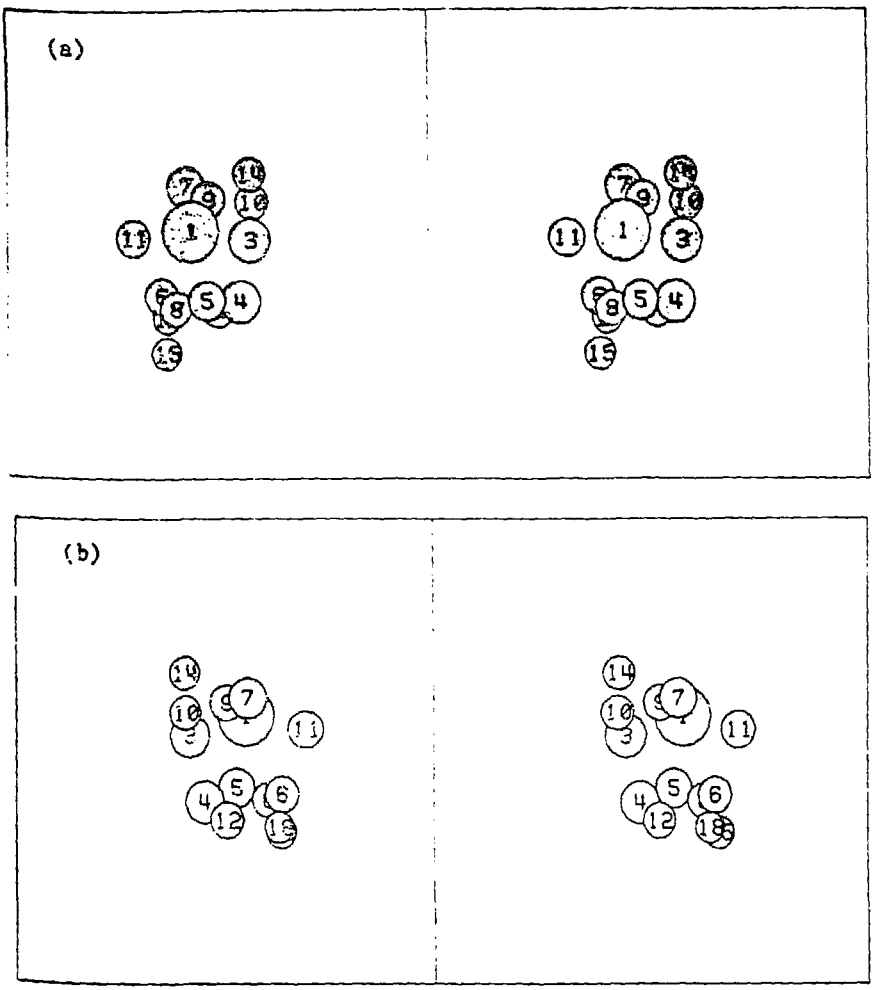

Fig. 1. Stereo views of the 14-protein model of the 30 S subunit of E. coli. Proteins S1, S3 to S12, S14, S15, and S18 have been located in the 30 S subunit by neutron interference techniques. Proteins are represented as spheres whose volumes are to scale and are those of the anhydrous proteins. The numbers in the spheres correspond to the standard protein identification numbers. (a) "Front" view of the subunit. (b) "Back" view, seen from the opposite side (180°).

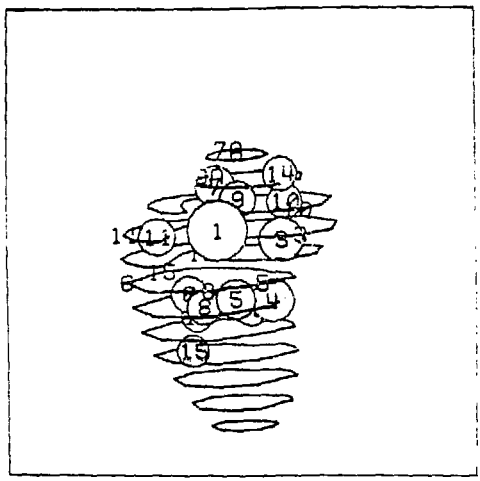

Fig. 2. Superposition of the neutron map on the electron microscopic image of the 30 S subunit (12). The contours in this drawing represent the outline of the 30 S subunit as visualized in negative stain in the electron microscope. The neutron map is superimposed on the EM image so as to place neutron-located proteins (circled numbers) as close as possible to the positions of their corresponding antigenic determinants as discovered by antibody staining (uncircled numbers).

Figure 2 is a superposition of the neutron and electron microscopic maps (12) done (by eye) so as to minimize the distance between the positions of proteins in the two maps. It is clear that a consistent superposition is possible; there is little doubt both techniques are describing the same particle.

The neutron map has some "weak" regions. Because of inconsistent data, it is not clear precisely how S6 should be placed relative to S8 and S4. Since S6 and S18 are close neighbors (by direct measurement), this difficulty with S6 strongly influences the position assigned to S18. The S3-S7 distance, one of the first we attempted to determine, is still unsettled. We have yet to obtain a fully satisfactory data set for that pair, and the data we do have violate the triangle inequality with respect to other distances in that part of the map. Besides these problems, we are not aware of any other major difficulties. All that is needed, as ever, is more (and better) data.

d) Radii of Gyration

The data also permit us to estimate radii of gyration for the individual proteins. As has been pointed out on theoretical

grounds (20), it is the nature of this kind of mapping that subunits with large radii of gyration will have their radii determined with useful precision, given data of the quality we can obtain, but, unfortunately, the radii for subunits with small radii will be determined very poorly; and so it is turning out in the event. The best determined radius is that for protein S1, which is about 55±10 Å, by far the largest radius encountered so far. The next best appears to be that of S4, 30±8 Å. All the rest have substantially smaller values with very large associated errors.

Because of the way error propogates in this system, it may not be possible for us ever to assign values having small errors for the majority of these proteins. However, it does seem reasonable to conclude, even in face of the errors, that most ribosomal proteins have shapes that are not radically extended, with the exceptions already noted. This finding is a useful one since there have been many claims in the past--e.g., Wittmann et al. (30)--that ribosomal proteins as a group have unusually extended configurations both as isolated molecules and in the ribosome.

It is also the nature of error propogation in this kind of mapping that coordinate errors are relatively small. The average error in this model is about ±10 Å in all three coordinates, x, y, and z. These errors make the neutron technique the most precise, by a substantial margin, of the ways currently known for determining the positions of proteins in this structure.

ANALYSIS OF DATA: FUTURE PROSPECTS

An impressive aspect of the work described above is the small number of parameters that have been specified about the 30 S structure (36 coordinates well determined, and 14 radii of gyration poorly determined) in return for the large amount of data collected. Each experiment involves the measurement of an entire scattering profile, about 30 values of intensity for each. Information content considerations suggest that these profiles should be able to specify four or five independent parameters apiece (14,21). For the purposes of model building, however, only a single number is used per data set, the second moment.

As pointed out in the section on data analysis, an interference profile has more in it than a single distance and some radii of gyration. It also reflects the shapes and relative orientation of the subunits. The problem that has confounded us for a long time is how to recover some of this additional information from the data. In the past few months an approach to this question has been explored which has illuminated the problem.

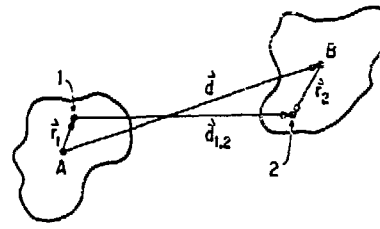

Fig. 3. Distance construction. Two arbitrary shapes are shown whose centers of mass are A and B; 1 and 2 are arbitrary points within these shapes. The vectors mentioned in the text are identified in the drawing.

a) Distance Geometry

It is useful to begin by considering the relationship between the vector that joins two (arbitrary) points in two labeled subunits, $\bar{d}_{12}$, the vector joining the centers of mass of the two subunits, $\bar{d}$, and the vectors between the two points and the centers of their respective subunits, $\bar{r}_1$ and $\bar{r}_2$.

$$\bar{d}_{12} = \bar{d} - \bar{r}_1 + \bar{r}_2 \qquad \text{(see Figure 3).}$$

$$d_{12} = [d^2 + r_1^2 + r_2^2 + 2dr_2\cos\theta_2 - 2dr_1\cos\theta_1 - 2\bar{r}_1\cdot\bar{r}_2]^{1/2} .$$

In these expressions, θ_1 and θ_2 are the angles between $\bar{d}$ and $\bar{r}_1$ and $\bar{r}_2$ respectively. (Vector amplitudes are given as the vector symbol without the superscript bar.)

For d large compared with r_1 and r_2,

$$d_{12} \simeq d + r_2\cos\theta_2 - r_1\cos\theta_1 .$$

Now r cosθ is the projection of $\bar{r}$ onto d. Furthermore all points within a subunit which lie on a plane perpendicular to d will have the same value of r cosθ. Thus in the limit of large d, the length distribution for a pair of subunits, $p_x(r)$, should be well approximated as follows:

$$p_x(r) = \int_{-\infty}^{+\infty} g_1(\delta)g_2(r - d + \delta)d\delta \qquad (6)$$

where g_1 and g_2 are one-dimensional density distributions formed by projecting the three-dimensional density distributions of the subunits onto the line whose direction is given by $\bar{d}$ (see Figure 4). The origin in both cases, $\delta = 0$, is the intersection of d with the subunit center of mass. The terms g_1 and g_2 can be called "line projections," and the length distribution is given as a cross correlation of line projections or "LPCC," Eq. (6).

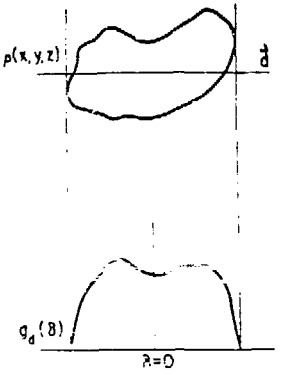

Fig. 4. [illegible] line projection. d is a vector of arbitrary direction passing through [illegible] center of mass of the shape described by $\rho(x_1,y_1,z)$. All densities within the shape lying on a plane perpendicular to d positioned at δ along d, relative to the object's center of gravity, are summed to give $g_d(\delta)$.

b) LPCCs Are Good Approximations to Length Distributions

An LPCC, of course, ignores the contribution to point-to-point distances of components perpendicular to d. It is reasonable to ask whether LPCCs usefully approximate the length distributions expected in biological assemblies in which subunit dimensions may not be much smaller than intersubunit separations. This issue has been explored computationally.

Figure 5A compares the LPCC and the true length distribution computed for a sphere of 15-Å radius 60 Å from a 3:1 prolate ellipsoid of revolution of the same volume. The ellipsoid axis is tipped 45° relative to the line joining their centers. The shapes of the two distributions are similar but not identical. The LPCC, not unexpectedly, predicts a length distribution displaced to shorter distances than the true distribution. From the standpoint of analysis of the shapes of profiles, however, that displacement is of no consequence. Figure 5B shows the two curves superimposed so that their centers of gravity coincide; the match is better. The result is typical. In most cases, the difference between the LPCC and the true $p_x(r)$ is within the error with which $p_x(r)$ can be determined experimentally in measurements of the kind under discussion here--see, e.g., Ramakrishnan et al. (24).

c) The Properties of Line Projections

What makes the line projection attractive is its simple relationship to the structure from which it is derived. The infor-

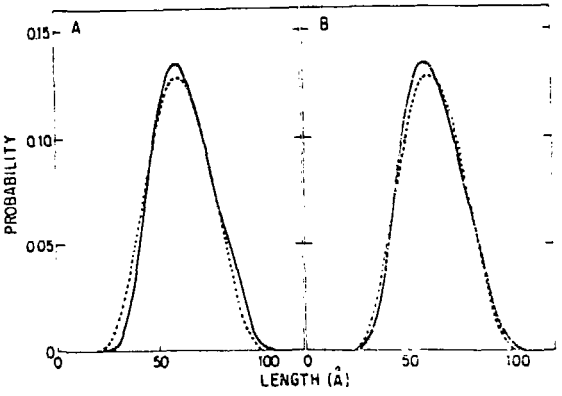

Fig. 5. A comparison of pairwise length distributions and length projection cross correlations. The length distribution was computed corresponding to the pair interference which would arise in a structure containing two differentially labeled subunits. The subunits are a sphere 15 Å in radius 60 Å from the center of a 3:1 prolate ellipsoid of the same volume. The ellipsoid's major axis makes a 45° angle with the line joining the centers of the two shapes. A: The length distribution (solid line) is compared with the corresponding line projection cross correlation (dashed line). B: The cross correlation function is shifted so that its centroid coincides with that of the length distribution.

mation contained in such a projection is best appreciated by considering what it corresponds to in the reciprocal space of the parent object. If $F(R,\theta,\Phi)$ is the Fourier transform of the object's density distribution, $p(r,\theta,\phi)$, it is easy to show that the transform of the line projection of $p(r,\theta,\phi)$ is simply $F(R,\theta,\Phi)$ evaluated along the line running through the point $F(0,0,0)$ in the same direction as the line projection vector. Clearly if one were to obtain enough line projections for an object, taken in many different directions, a sufficient number of coefficients for F would be available in three dimensions to permit recovery of $p(r,\theta,\phi)$ by Fourier inversion. Reconstruction from line projections is the one-dimensional counterpart of the two-dimensional reconstruction technique used by electron microscopists (3).

Line projections are not an efficient means for defining structures. Sampling considerations show that it would take about 100 line projections, evenly distributed in space, to define the structure of an arbitrarily shaped object with linear dimensions of 40 Å to a resolution of 10 Å. The 30 S ribosome, on the other

hand, contains only 21 subunits, so that no more information about a given subunit than that number of line projections can normally be obtained. Clearly only a limited representation of a real subunit in such a structure can be achieved under the best of circumstances, and even that, only if a way is found to recover line projections from pairwise length distributions.

Given the severe limitation in the power of the data, it is reasonable to confine efforts at subunit shape modeling to simple shapes like equivalent ellipsoids of revolution. By accepting this limitation, the number of shape parameters to be extracted from the data is reduced to two angles (to specify the orientation of the major axis) and an axial ratio. Volume can be estimated from subunit molecular weight.

d) An Algorithim for Shape and Orientation Modeling

The information in a length distribution is not a line projection, but the cross correlation of two line projections, something a good deal less informative. Cross correlations, however, are relatively easy to deal with in reciprocal space (1):

$$FT(LPCC) = FT[g_1(\delta)] \cdot FT[g_2(\delta)] \qquad (7)$$

where FT stands for Fourier transform. The transform of an ellipsoid of revolution is known in closed form as a function of axial ratio and orientation (2), so that the generation of these products for pairs of arbitrarily oriented ellipsoids is straightforward.

These facts suggest an analysis of the data as follows. The measured length distributions are Fourier transformed with their origins taken at the distributions' centroids. One then attempts to find orientations and axial ratios for a set of ellipsoids of appropriate volume, located at the positions revealed by second moment analysis, which will produce a set of LPCCs whose transforms match the observed transforms. (In this fitting process, only the real parts of the transforms of the measured data need be considered. The LPCCs of ellipsoids are all even functions with respect to their centroids.)

In a data set for a structure of 8 or more subunits, the burden put on the minimum data needed to locate the subunits by this further analysis amounts to about one parameter per data set. It is not much to ask.

A program has been written to implement these ideas. It is organized as a nonlinear least-squares minimization based on the Marquardt algorithm (15). The purpose is to find the model which minimizes the residual χ^2 where

$$\chi^2 = \sum_m \left[\sum_s \frac{1}{\sigma(s)^2} \{FT[p_x(r)] - FT(LPCC)\} \right] \quad (8)$$

and $\sigma(s)^2$ is the variance of $FT[p_x(r)]$ at a given point in its profile. The summation in s runs over all values for which $FT[p(r)]$ has been calculated in reciprocal space, and the summation in m runs over all data sets. The program starts with a user-specified model and refines it to produce a model in which χ^2 is minimized.

In order to explore the practicality of this approach, model data were computed by using techniques described previously (17). The object of the exercise was to recover from the model data the shape and orientations of the ellipsoids used to generate the data set.

e) Computational Results

A number of conclusions have emerged from the testing of this program.

First, the algorithm is convergent. If the starting model for the refinement is the model that generated the test data, the fit of the ellipsoid LPCCs to the "data" is very good to begin with, as expected, and upon refinement is improved still further at the expense of some small adjustments in angular orientations and axial ratios. These adjustments reflect the fact that the LPCCs only approximate true length distributions. Furthermore, the algorithm will always find a configuration substantially better than the starting model, whatever it may be.

Second, the residual space--i.e., χ^2 space, Eq. (8)--being explored here has many minima. Each distinct starting model tested so far has led to a different "best" model for the data. These "best" models differ substantially both in the final value of χ^2 obtained and in the subunit orientations and axial ratios they suggest. The "right" model, i.e., the model derived by refinement of the model that generated the data, has the lowest residual found so far. The presence of many minima appears to be related to the fact that inter-ellipsoid length distributions are not very distinctive. The strongest influence on the distribution is that of the angles made by the major axes of a pair of ellipsoids with the line joining their centers of mass. All ellipsoid orientations around this line which have the same "tip" angles give length distributions that are hard to distinguish. Their corresponding LPCCs are identical.

Third, the addition of external information is remarkably unhelpful. The program was modified to allow the specification of certain model parameters by the operator and limitation of refinement to the remainder, in the belief that the algorithm would do better if constrained. The fixing of various combinations of parameters was tested, including fixing of the entire set of axial ratios at their correct values, all to no effect. This case corresponds to the situation that would obtain if reliable radii of gyration were available for all subunits from second moment analysis or other sources, and the modeling made use of this information. The multi-minima property remained.

f) Outlook for Shape Modeling

This experience leaves us discouraged about prospects for solving the general shape problem. The line projection concept seems, to us, to correspond to a "best case." With large subunits, closely spaced, when LPCCs fail to give satisfactory results, the modeling of length distributions in terms of simple shapes will be harder--not easier--to carry out, and the refinement algorithms that emerge will be even more nonlinear and probably worse behaved. Thus, if the problem cannot be solved by using LPCCs, with data sets for which LPCCs approximate the data satisfactorily, we think it may be insoluble altogether.

Clearly, the reason the problem is so intractable is that the residual surface has many minima. This means that gradient optimization techniques, which are the most powerful methods available for solving nonlinear problems, cannot be used unless very good starting models can be suggested. The difficulties here are twofold. First, the same multi-minima character that makes convergence to the global minimum problematic must make the generation of a good starting model equally difficult. Second, no information is available on the range of convergence, hence no estimate of how "good" a starting model must be in order to ensure global convergence.

The only alternative to gradient algorithms for resolving this problem is "grid searching." A grid search is simply a systematic trial-and-error search through parameter space to find the model that gives the lowest minimum. Given that there are 3N shape and orientation parameters to specify and N is about 20, even a coarse search would be an extravagant exercise indeed.

Finally, it must not be forgotten that the data to be analyzed include error, both systematic and non-systematic. Given error in the data, and a multi-minima residual space, there can be no guarantee that the global minimum discovered in a real data set will be recognizably related to the true structure.

CONCLUSIONS

It is hard for non-mathematicians like ourselves to prove that a problem cannot be solved. Certainly, one can imagine large numbers of computational experiments that could be tried which might cast some additional light on the matter. Additionally, we have been changing our ways of testing ellipsoidal conformations against measured data which are much more efficient than the currently used way. However, in our estimation none of these expedients avoids the traps presented by the many minima in this problem, and therefore we have not tested any.

We would like nothing better than to have the challenge implicit in these remarks taken up and to have our gloomy assessment proven wrong. Until the problem is solved, however, we are left with the conclusion that neutron scattering techniques can provide us only with subunit positions and radii of gyration.

Within the limitations of the data analysis we are able to carry out today, however, the picture is bright. It is clear from the experimental work summarized above that the neutron model of the distribution of proteins in the 30 S ribosomal subunit of E. coli is nearing completion. Comparisons between the neutron model and the other data available (gratifyingly) support the view that the neutron model is reliable.

We think the 30 S subunit problem will be completely solved in the next few years, and that it will prove possible to include in the map tRNA, initiation factors, etc.--the auxiliary molecules and substrates in protein synthesis. These measurements should provide the biochemical community with some useful insights into the mechanism of protein synthesis.

ACKNOWLEDGMENTS

We have received able technical assistance over the years from Betty Rennie, Darcy Fazio, Theresa Dougan, Pamela Saha, Donna Tabayayon, and Kathy Cullen. The neutron data summarized here were collected primarily at the High Flux Beam Reactor at Brookhaven National Laboratory under the auspices of the U.S. Department of Energy. We are also indebted to the Institute Laue-Langevin at Grenoble, France, for access to D11 for the collection of some of these data. This work has been supported by grants from the NIH (AI-09167) to PBM and from the NSF (PCM78-10361) to DME and PBM. Computational facilities have been supported by NIH (GM-22778).

REFERENCES

1. Bracewell, R.N., "The Fourier Transform and Its Application," 2nd ed., McGraw-Hill, New York (1978).
2. Chambliss, G., Craven, G.R., Davies, J., Davis, K., Kahan, L., and Nomura, M., eds., "Ribosomes--Structure, Function and Genetics," University Park Press, Baltimore, MD (1980).
3. DeRosier, D.J. and Klug, A., Nature 217:130-4 (1968).
4. Engelman, D.M. and Moore, P.B., Proc. Natl. Acad. Sci. USA 69:1997-9 (1972).
5. Engelman, D.M., Moore, P.B., and Schoenborn, B.P., in: "Neutron Scattering for the Analysis of Biological Structures," B.P. Schoenborn, ed., Brookhaven Symp. Biol. 27:IV 20-37 (1975).
6. Engelman, D.M., Methods Enzymol. 59:656-69 (1979).
7. Guinier, A. and Fournet, G., "Small Angle Scattering of X-Rays," Wiley, New York (1955).
8. Guinier, A., in: "International Tables for X-Ray Crystallography," 2nd ed., Vol. 3, pp. 324-9, K. Lonsdale, ed., Kynoch Press, Birmingham (1968).
9. Hoppe, W., Israel J. Chem. 10:321-33 (1972).
10. Hoppe, W., J. Mol. Biol. 79:581-5 (1973).
11. Hoppe, W., May, R., Stöckel, P., Lorenz, S., Erdmann, V.A., Wittmann, H.G., Crespi, H.L., Katz, J.J., and Ibel, K., Brookhaven Symp. Biol. 27:IV 38-48 (1975).
12. Kahan, L., Winkelman, D., and Lake, J.A., J. Mol. Biol. 145:193-214 (1981).
13. Kratky, O. and Worthmann, W., Monatsh. Chem. 76:263-81 (1947).
14. Luzzati, V., Tardieu, A., and Aggerbeck, L.P., J. Mol. Biol. 131:[illegible]73 (1979).
15. Marquardt, D.W., SIAM J. Appl. Math. 11:431-41 (1963).
16. May, R., Ph.D. Thesis, Technical University, München (1978).
17. Moore, P.B., Langer, J.A., Schoenborn, B.P., and Engelman, D.M., J. Mol. Biol. 112:199-234 (1977).
18. Moore, P.B., Langer, J.A., and Engelman, D.M., J. Appl. Crystallogr. 11:479-82 (1978).
19. Moore, P.B., Methods Enzymol. 59:639-55 (1979).
20. Moore, P.B. and Weinstein, E., J. Appl. Crystallogr. 12:321-6 (1979).
21. Moore, P.B., J. Appl. Crystallogr. 13:168-75 (1980).
22. Nomura, M., Tissieres, A., and Lengyel, P., "Ribosomes," Cold Spring Harbor Lab., Cold Spring Harbor, NY (1974).
23. Ramakrishnan, V.R. and Moore, P.B., J. Mol. Biol. 153:719-38 (1981).
24. Ramakrishnan, V.R., Yabuki, S., Sillers, I-Y., Schindler, D.G., Engelman, D.M., and Moore, P.B., J. Mol. Biol. 153:739-60 (1981).
25. Schindler, D.G., Langer, J.A., Engelman, D.M., and Moore, P.B., J. Mol. Biol. 134:595-620 (197[illegible]).

26. Sillers, I-Y. and Moore, P.B., J. Mol. Biol. 153:761-80 (1981).
27. Stöckel, P., May, R., Strell, ..., Galka, Z., Hoppe, W., Hermann, H., Zillig, W., Crespi, H., Katz, J.J., and Ibel, K., J. Appl. Crystallogr. 12:176-85 (1979).
28. Tischendorf, G.W., Zeichhardt, H., and Stöffler, G., Proc. Natl. Acad. Sci. USA 72:4820-4 (1975).
29. Vainstein, B.K., Sasfenov, N.I., and Feigin, L.A., Doklady Akad. Nauk SSSR 190:574-7 (1970).
30. Wittmann, H.G., Littlechild, J.A., and Wittmann-Liebold, B., in: "Ribosomes--Structure, Function and Genetics," pp. 51-88, G. Chambliss et al., eds., University Park Press, Baltimore (1980).

DISCUSSION

ZACCAI: On your map most of the proteins are grouped at one end of the particle. Could you comment with relation to the nearly R_E measurements?

MOORE: There is little doubt that our model does imply a concentration of protein at one end of the ribosome. Not very much protein, mass-wise remains to be found, and it seems hardly possible the verdict will change when it is found. The EM people also see the particle the same way. Our model shows a clear contradiction with the original bulk distribution data, or more correctly, the interpretation given that data. We do not know its cause.

SCHIFFER: Can you explain the source of the discrepancy in the position of protein 15 as observed in the electron micrograph and with neutron scattering?

MOORE: I think the problem is imperfect data on both sides.

MAY: Is there a global minimum in the "line projection program," or are there only many local minima?

MOORE: It seems there is. If you ask the program to "refine" the model used to generate the test data set, it finds a very low residual for it and adjusts some parameters by a small amount to "improve" it. This improvement reflects the lack of perfect correspondence between line projection cross correlations and the length distributions.

KOEPPE: Immuno-electron microscopy offers the possibility of yielding information related to protein shape, if the same protein can be labeled at more than one antigenic site. Would you comment on this possibility?

MOORE: In principle you are right. In the early days, the EM people thought they had identified multiple antigenic sites for many proteins. In the last two years, they have been able to show that their antibody preparations have sometimes been contaminated. Most of the multiple sites have been withdrawn, and no generally accepted multiple site protein is left. The EM people, therefore, agree with us at this time that ribosomal proteins are mostly compact.

MARTEL: You showed a superposition of EM and neutron data for the ribosome subunit which illustrated substantial agreement. The time taken in acquiring the neutron data is in excess of seven years. How long did it take to acquire the EM data?

MOORE: The neutron data collection has taken about six months (24 hours a day) since we started in 1974. When the EM work started, also in about 1974, we thought it would be finished in a year or so and were a bit discouraged by the prospect. In fact, it has proven much harder than first expected. They are still at it and know about as much as we do. We are surprised!