

## **Positions of Proteins S14, S18 and S20 in the 30 S Ribosomal Subunit of *Escherichia coli***

V. RAMAKRISHNAN, M. CAPEL, M. KJELDGAARD  
D. M. ENGELMAN AND P. B. MOORE

*Departments of Chemistry and  
Molecular Biophysics and Biochemistry  
Yale University*

*New Haven, CT 06511, U.S.A.*

*and*

*Biology Department  
Brookhaven National Laboratory  
Upton, L.I., U.S.A.*

*(Received 13 July 1983, and in revised form 31 October 1983)*

A map of the 30 S ribosomal subunit is presented giving the positions of 15 of its 21 proteins. The components located in the map are S1, S3, S4, S5, S6, S7, S8, S9, S10, S11, S12, S14, S15, S18 and S20.

### **1. Introduction**

In 1975 Engelman *et al.* demonstrated that intercomponent distances within macromolecular aggregates can be measured by techniques that rely on selective deuterium labeling and neutron scattering. Since then, a large number of pairwise distances have been estimated by the neutron method in the 30 S ribosomal subunit of *Escherichia coli* leading recently to a map of that structure showing the positions of 12 of its 21 proteins (Ramakrishnan *et al.*, 1981). This map was based on 45 data sets and yielded preliminary estimates for the radii of gyration of the 12 proteins mapped in addition to their positions.

A significant number of new data sets has been measured since 1981. Combined with existing information, the new data determine the positions of three additional proteins, raising the number mapped from 12 to 15. The “new” proteins, S14, S18 and S20, are joined in the map with proteins S1, S3, S4, S5, S6, S7, S8, S9, S10, S11, S12 and S15. These data and the map they engender are presented and discussed below.

### **2. Materials and Methods**

#### *(a) Distance finding, experimental outline*

The techniques used for measuring interprotein distances by neutron scattering in the ribosome have been described at both the theoretical and practical levels (Moore, 1979;

Engelman, 1979; Ramakrishnan & Moore, 1981). A general description of how the experiment is done will be given to orient the reader, but details will be provided only where the methods used here deviated from what has been reported before.

Neutron distance finding experiments depend on the fact that the substitution of  $^2\text{H}$  for  $^1\text{H}$  in a biopolymer substantially alters its neutron scattering properties. *E. coli* grows in predeuterated medium permitting one to obtain its ribosomal components in  $^2\text{H}$  form. In addition, the 30 S ribosomal subunit from *E. coli* can be reconstituted from purified components (Traub & Nomura, 1968). Thus it is conceptually a straightforward task to construct 30 S subunits with specific macromolecular components deuterium labeled in whatever pattern required, and this labeling is detectable by neutron scattering.

In these experiments, subunits are prepared with pairs of proteins  $^2\text{H}$ -labeled and the rest of the particle in  $^1\text{H}$  form. Each distance measurement requires a set of 4 samples: (1) one with both proteins of interest  $^2\text{H}$ -labeled, (2) and (3) the two possible samples with only one member of the pair labeled, and (4) a sample with no protein  $^2\text{H}$ -labeled. Scattering profiles are measured on solutions of an equimolar mixture of samples (1) and (4) and a similar mixture of samples (2) and (3) (Hoppe, 1972, 1973). After suitable normalization for differences in ribosome amount, data acquisition time, etc., subtraction of one profile from the other gives a scattering difference profile,  $I_x(s)$ .

$I_x(s)$  can be Fourier transformed to yield  $p_x(r)$ , the distribution of lengths of all vectors joining  $^2\text{H}$  atoms in the first labeled protein with  $^2\text{H}$  atoms in the second (Moore *et al.*, 1978). The second moment of  $p_x(r)$ ,  $M_{ij}$ , is related to the radii of gyration of the 2 proteins,  $R_i$  and  $R_j$ , and the distance between them,  $d_{ij}$  as follows:

$$M_{ij} = d_{ij}^2 + R_i^2 + R_j^2 \quad (1)$$

(Moore *et al.*, 1978; May, 1978; Stoekel *et al.*, 1979). Thus measurement of a large number of such second moments within a single object permits the construction of a model of its structure giving component positions and radii of gyration (Moore & Weinstein, 1979). In general, a minimum of 4 measurements is required to position a protein in the map.

#### (b) *Experimental details*

The techniques used for sample preparation and analysis were those described previously (Moore, 1979). All neutron data were collected at Brookhaven National Laboratory using the High Flux Beam Reactor (Engelman, 1979). About half the data were collected at the H4 satellite station, whose characteristics have been described (Schoenborn *et al.*, 1978). The neutron flux was  $2.5 \times 10^7$  neutrons  $\text{cm}^{-2} \text{min}^{-1}$ , and the wavelength was 2.37 Å with a full width at half maximum of 2%. Recently we have made use of the High Flux Beam Reactor cold moderator small-angle instrument (Schneider & Schoenborn, 1984). This spectrometer provides neutrons with wavelengths adjustable from 4 Å to 10 Å by virtue of a multilayer monochromator (Saxena & Schoenborn, 1977). We have used it at wavelengths around 5 Å, where it delivers a flux of  $4 \times 10^7$  to  $5 \times 10^7$  neutrons  $\text{cm}^{-2} \text{min}^{-1}$  with  $\Delta\lambda/\lambda = 0.08$ . The methods used for reducing the data have been described (Ramakrishnan & Moore, 1981).

### 3. Results

#### (a) *New data*

Figures 1, 2, 3 and 4 present the newly acquired data. Each Figure shows three difference scattering curves,  $I_x(s)$ , identified in the upper right-hand corner by the names of the two proteins whose distance was sought, and below them the three corresponding length distributions. Many of these data sets are "normal", which is to say similar in character to the better-determined data sets published in the past.

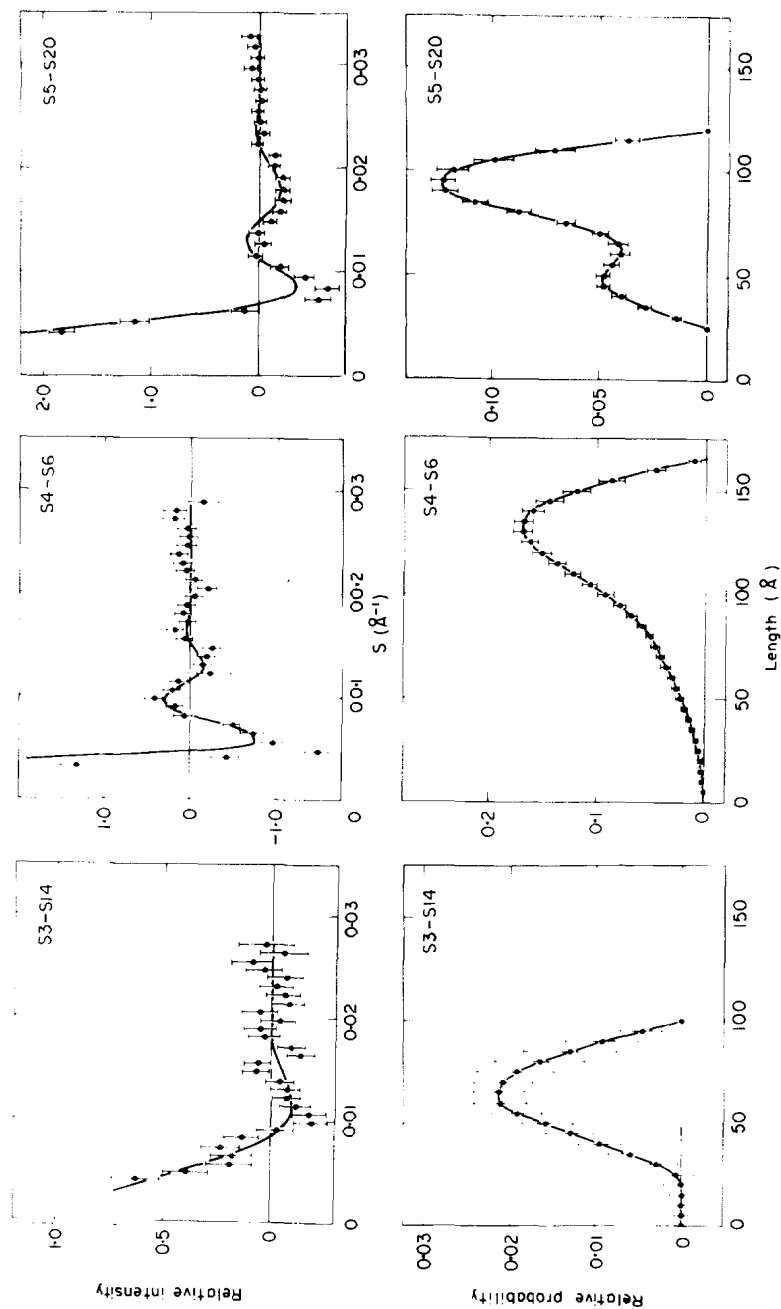


FIG. 1. Data for S3-S14, S4-S6 and S5-S20. Scattering intensity data and length distributions are shown for 3 experiments. Each panel is identified in its upper right-hand corner by the names of the pair of proteins whose relationship was sought in the experiment in question. The top row of panels show the scattering difference curves and the bottom row show the corresponding length distributions. In both cases, the error bars correspond to  $\pm 1\sigma$ . The smooth curves drawn through the scattering data are the scattering curves implied by the length distributions shown.

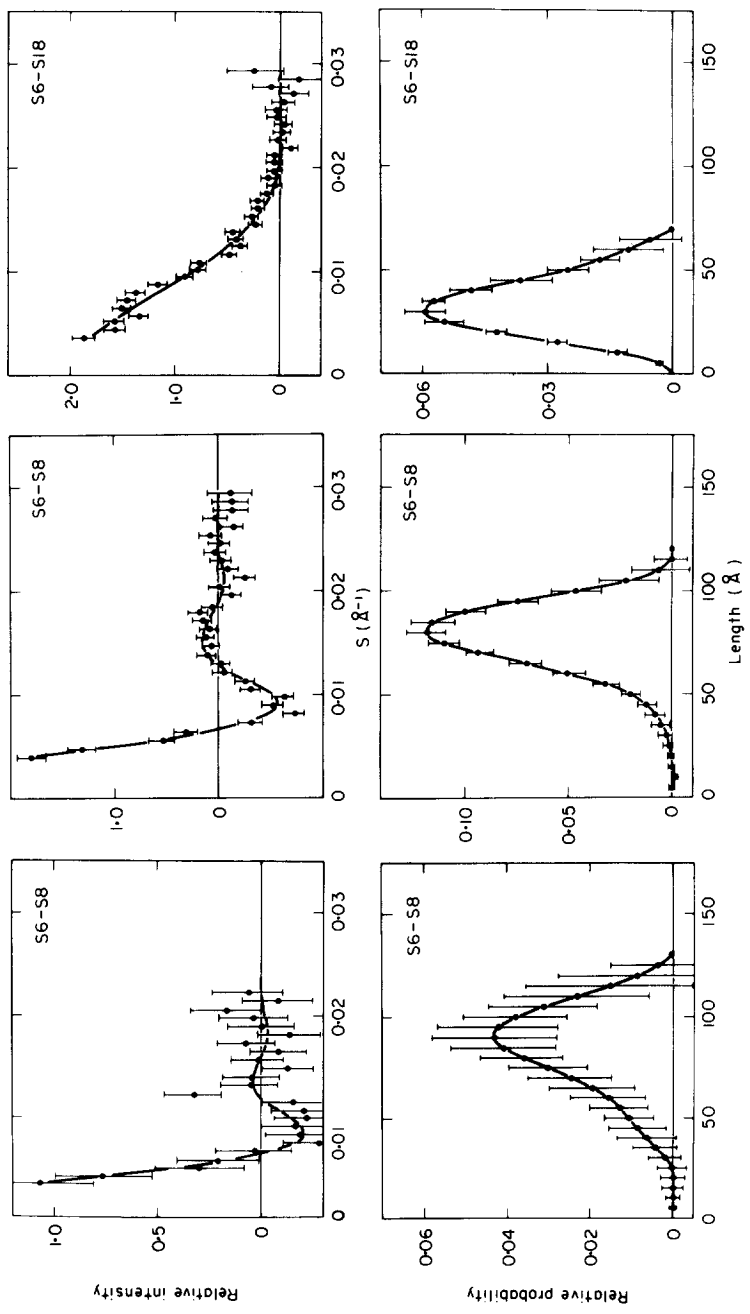


FIG. 2. Data for S6-S8 and S6-S18. See the legend to Fig. 1.

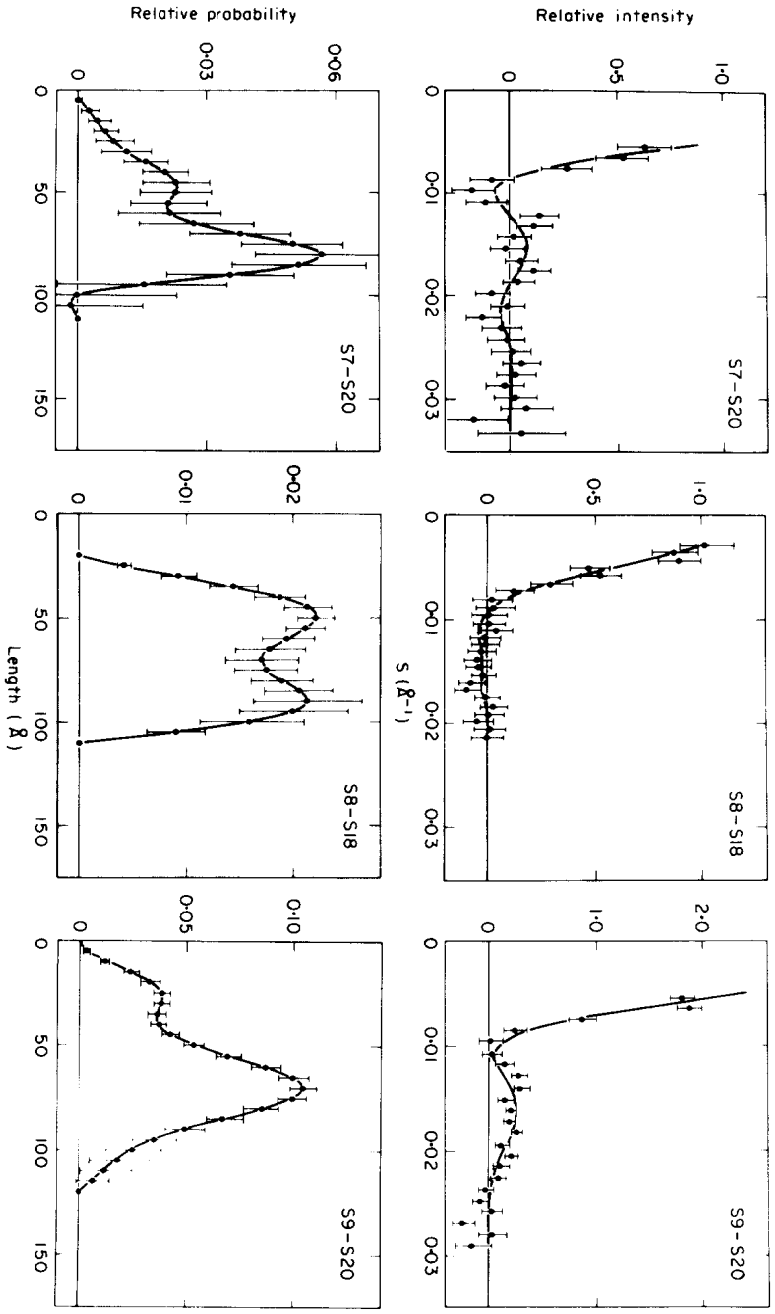


FIG. 3. Data for S7-S20, S8-S18 and S9-S20. See the legend to Fig. 1.

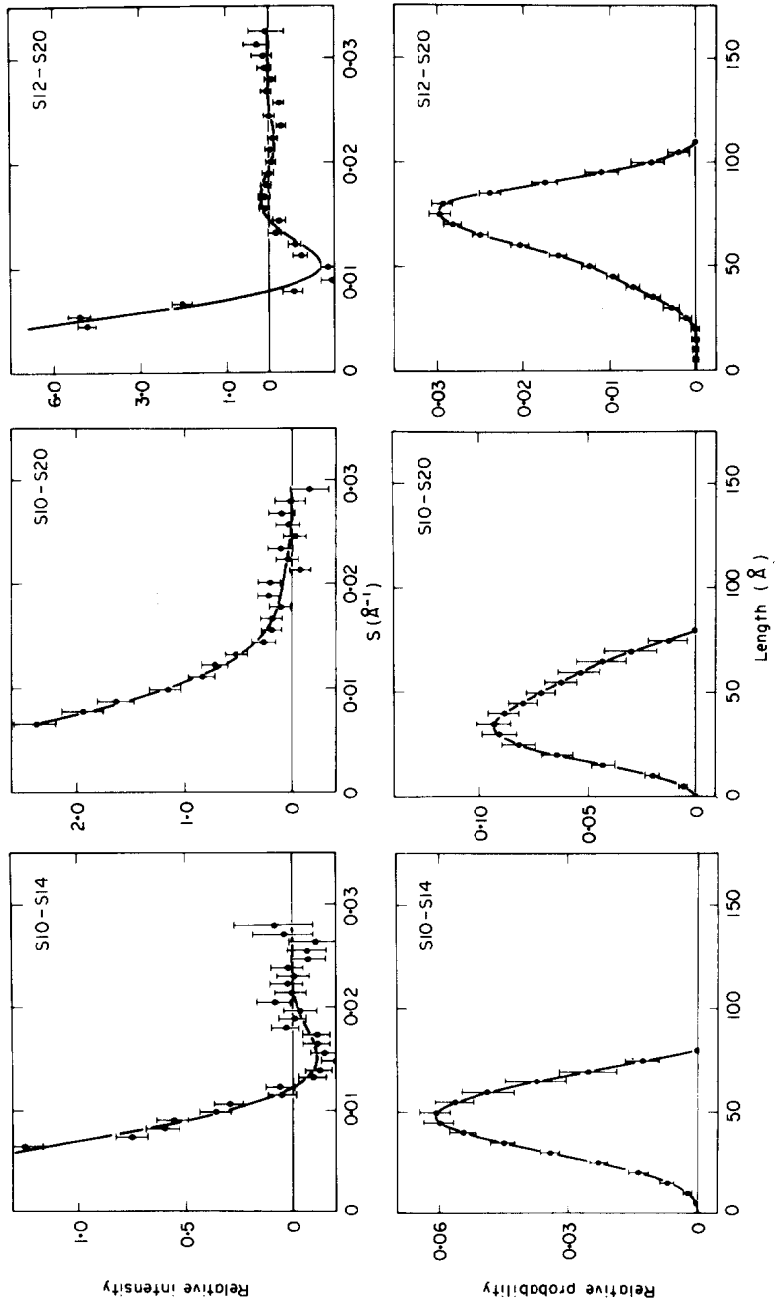


FIG. 4. Data for S10-S14, S10-S20 and S12-S20. See the legend to Fig. 1.

In Figure 1, the S3–S14 and S5–S20 data sets are “normal”; the “difficult” data set is that given by the pair S4–S6. An S4–S6 data set has been obtained before (Ramakrishnan *et al.*, 1981), but that set was poorly determined in the sense that a variety of length distributions could be found that corresponded well to the data statistically, but which had widely different second moments. The data presented here are far better determined, but not easily fit with an all positive length distribution. (Note that the length distributions in an experiment of this design are always positive.) The smooth curve superimposed on the measured data is the best fit to the data that could be found consistent with this constraint. The second moment for the S4–S6 pair derived from these data is larger than previously reported for the pair.

Two data sets for S6–S8 are presented in Figure 2. These represent data obtained from independently prepared samples. The first data set was collected using the H4 satellite small-angle apparatus. The sample size was small and the errors accordingly large. The second data set was measured using the cold moderator instrument and a larger sample. The conclusions from both experiments were essentially the same. The S6–S18 data set shown is also the second trial of the same pair. The first trial again involved a small sample but, as in the S6–S8 case, the conclusion drawn from the data was unaltered from the first to the second experiment (S. Yabuki, unpublished data).

The data sets shown in Figure 3 all lead to length distributions with irregular features. However, many of these features are small compared to error (e.g. the shoulder on the S7–S20 length distribution) and should be interpreted with caution. It is important to note that the algorithm used for calculating length distributions uses a sine function synthesis. “Bumpy” looking length distributions are likely to result. The data in Figure 4 are all “normal”.

From each data set, four parameters are estimated:  $M_{ij}$ , the second moment of the length distribution,  $\sigma_{ij}$ , the statistical error connected with that second moment,  $X^2$ , a parameter that describes the statistical quality of the fit of the proposed length distribution to the data, and  $EBW$ .  $X^2$  is given by:

$$X^2 = \frac{1}{N} \sum_{i=1}^N \frac{1}{\sigma_i^2} (d_i - f_i)^2, \quad (2)$$

where  $N$  is the number of data points,  $\sigma_i^2$  is the variance of the  $i$ th data point whose value is  $d_i$ , and  $f_i$  is the value of the fit at that position. The equivalent band width ( $EBW$ ) represents the range of the values of  $M_{ij}$  that could be accommodated by the data with equivalent goodness of fit, i.e. the same value of  $X^2$  (Ramakrishnan & Moore, 1981). The values obtained for these parameters are listed in Table 1.

(b) *A 15-protein model of the 30 S subunit*

Table 2 summarizes all the previously published data on distances within the 30 S subunit in the same format as Table 1. The data in Tables 1 and 2 were merged and used to derive a model for the 30 S subunit by means of the second moment method (Moore & Weinstein, 1979; Ramakrishnan & Moore, 1981).

TABLE I  
*Numerical summary of new data*

Pair	$X^2$	$M/\text{\AA}^2$	$\sigma/\text{\AA}^2$	$EBW/\text{\AA}^2$
S3-S14	1.14	4221	648	143
S4-S6 (2)	10.25	14,255	239	577
S5-S20	2.15	7144	226	230
S6-S8 <sup>1</sup>	0.64	7662	1098	548
S6-S8	1.21	6281	295	231
S6-S18 (2)	1.83	1280	106	36
S7-S20	0.86	4688	769	128
S8-S18	0.60	4968	422	435
S9-S20	4.93	4507	357	274
S10-S14	1.09	2429	105	100
S10-S20	0.62	1834	123	115
S12-S20	14.60	5200	151	144

$X^2$  is the value for the variance weighted residual of the fit of the best, all positive, length distribution to the measured data (see eqn (2)). Ideally,  $X^2$  should be of the order of 1.0.  $M$  is the second moment of the length distribution (see eqn (1)).  $\sigma$  is the error in  $M$  contributed by counting statistics.  $EBW$  is the effective contribution to the total standard error of  $M$  due to uncertainty in the maximum chord in the length distribution of a given protein pair (Ramakrishnan & Moore, 1981). A number in parenthesis following a pair name indicates the number of times a given distance measurement has been made. If no number is supplied, the number is 1. The S6-S8 value used for model building is the variance weighted mean of the 2 experiments listed.

During the model building, the radii of gyration of the proteins were constrained to take values that were larger than the radii they would have as anhydrous spheres. These minimum values are readily calculated from published molecular weight and partial specific volume data (see Wittmann *et al.*, 1980).

The co-ordinates, radii of gyration and error estimates that emerged from the model building are given in Table 3. The model includes 61 data sets and specifies 54 independent parameters. The unreduced value for the sum of the squares of the variance weighted residuals for the model is 26.8. This value reflects, in part, the fact that eight of the radii of gyration are effectively determined by the minimum value constraint, not by the second moment data itself. (The radii determined by constraint are indicated by an asterisk in Table 3.) Thus the number of degrees of freedom is effectively 15, not 7, the value one would estimate comparing the amount of data with the number of parameters estimated. For the convenience of the reader, Table 4 lists all the interprotein distances in this map. Figure 5 is a stereo view of this model, in which proteins are represented as spheres whose volume is to scale.

### (c) *Error analysis*

To test the influence of errors on the map, Monte Carlo methods were used. Starting with the parameters of the map described in Table 3, a set of error-modified data sets was generated and used to produce a series of maps, the

TABLE 2  
Summary of earlier data

Pair	$X^2$	$M/\text{\AA}^2$	$\sigma/\text{\AA}^2$	$EBW/\text{\AA}^2$
S1-S3	1.13	6667	473	388
S1-S4	2.31	12,005	3821	577
S1-S5	1.56	6205	871	87
S1-S7	1.28	7447	862	866
S1-S9	0.88	7323	667	838
S1-S10	0.89	8182	1140	1073
S1-S14	2.43	6669	519	581
S3-S4	0.51	4973	205	68
S3-S5	0.62	4807	315	22
S3-S7 (3)	0.70	10,000	500	1000
S3-S8	1.90	8725	513	577
S3-S9	0.50	3528	242	39
S3-S10	1.02	2562	438	205
S3-S10	10.50	13,304	207	75
S3-S12	0.41	8370	856	661
S4-S5	0.70	2654	134	11
S4-S7	0.58	14,142	334	245
S4-S8	1.28	5770	350	349
S4-S9	1.60	10,532	367	200
S4-S10	3.46	8375	336	82
S4-S11	0.42	13,495	1460	401
S4-S12 (2)	2.49	2697	79	14
S4-S15	1.13	9150	636	71
S4-S18	0.37	7875	530	236
S5-S6	0.76	9593	1157	144
S5-S7	1.03	12,515	550	196
S5-S8 (2)	1.44	1531	57	0
S5-S9	0.88	8412	541	132
S5-S10	9.02	14,400	1231	58
S5-S11	1.13	8146	881	115
S5-S12	1.62	2897	167	126
S5-S15	0.84	3566	374	43
S6-S7	1.0	10,171	590	280
S6-S11	0.67	3302	312	79
S6-S15	1.35	4776	520	358
S7-S8 (2)	2.20	12,724	369	173
S7-S9	0.77	1445	115	26
S7-S10	1.15	4072	402	150
S7-S11 (2)	0.48	5325	385	234
S7-S12	1.08	11,655	778	288
S8-S11 (2)	0.70	5540	586	59
S8-S12	1.24	3053	358	35
S8-S15	0.89	2330	308	20
S9-S10	0.53	2018	187	40
S9-S14	1.30	3337	757	105
S11-S12	0.69	12,790	613	156
S11-S15 (2)	5.47	10,834	1556	10
S12-S15	0.51	7554	1274	447
S15-S18	0.84	6019	475	894

See Table 1 for an explanation of the symbols. The data shown are those given by Ramakrishnan *et al.* (1981) and Sillers & Moore (1981).

TABLE 3  
*Protein co-ordinates and radii of gyration*

Protein	$x$	$\sigma_x$	$y$	$\sigma_y$	$z$	$\sigma_z$	$R_g$	$\sigma_R$	$R_{min}$
S1	-3.6	15.6	52.1	22.9	23.2	11.1	53.5	19.2	21.02
S3	0	—	0	—	0	—	15.32	0.4	15.23
S4	62.7	3.4	0	—	0	—	28.2	7.3	14.67
S5	53.0	3.3	40.8	3.9	0	—	13.6	8.6*	13.39
S6	75.7	7.5	68.0	11.0	90.4	6.0	13.0	0.4*	12.89
S7	-9.4	4.7	19.1	8.1	87.0	3.5	13.5	0.4*	13.38
S8	65.6	5.9	54.5	6.4	19.0	4.7	25.6	4.9	12.41
S9	-15.2	4.2	17.5	6.4	53.7	3.4	12.7	0.3*	12.58
S10	-13.4	6.6	-20.5	6.0	38.6	6.0	12.1	0.8*	11.88
S11	22.0	8.4	82.5	6.7	72.1	7.7	12.4	0.4*	12.33
S12	86.1	6.9	5.5	11.0	32.6	6.6	15.0	30.7	12.3
S14	-53.3	21.9	-4.8	12.2	23.0	10.5	14.6	62.0	11.5
S15	81.8	30.8	84.8	11.3	23.4	12.8	17.3	52.8	11.09
S18	78.2	13.5	38.0	10.3	76.8	5.0	10.8	0.5*	10.66
S20	25.1	8.2	-22.9	7.2	45.1	8.0	12.9	20.8	10.92

Relative co-ordinates of protein centers of gravity ( $x, y, z$ ) are given for each protein along with the standard errors of these co-ordinates ( $\sigma_x, \sigma_y, \sigma_z$ ).  $R_g$  is the estimate for the protein's radius of gyration,  $\sigma_R$  the standard error of that estimate and  $R_{min}$  the radius of gyration expected for each protein if it were an anhydrous sphere. All numerical values are in Å. Error estimates are obtained from the covariance matrix generated during refinement of the map.  $\sigma_R$  values designated with an asterisk are values dominated by the  $R_{min}$  constraint. During refinement, no  $R_g$  value was permitted to fall below  $R_{min}$ .

relative properties of which should reveal the influences of experimental errors. To obtain a modified data set, the second moments were each assigned a random value within a Gaussian probability distribution specified by its experimentally observed variance. Each modified data set was then used to produce a map. The average values for protein positions and variances were computed from the set of maps, and the residual for each map was noted: 50 such trials were made.

Several conclusions can be drawn from this exercise. First, about 20% of the synthetic data sets examined had residuals larger than 26.8. Thus it is plausible, but of course not proven, that the observed data reflect an underlying distribution of protein within the ribosome like that shown, and that the errors in the data are reasonably well estimated. Second, the variances computed for the map's parameters from the Monte Carlo simulations are similar to those computed from the covariance matrix derived from the refinement of the model from the observed data. Parameters identified as poorly determined by the covariance matrix are poorly determined in the Monte Carlo series. The covariance matrix in most cases suggests somewhat larger (by ~30%) standard errors for co-ordinates than the Monte Carlo series. To be conservative, the co-ordinate errors given in Table 3 are those from the covariance matrix.

The use of the covariance matrix leads to the assignment of errors to radii of gyration that are much larger than the Monte Carlo simulation suggests are proper in cases where the radii are data determined, but much smaller errors where the radii are determined by constraint. For consistency, the covariance

TABLE 4  
Protein-protein distances

	S1	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S14	S15	S18	S20
S1	0														
S3	<u>57</u>	0													
S4	<u>88</u>	<u>63</u>	0												
S5	<u>62</u>	<u>67</u>	<u>42</u>	0											
S6	<u>105</u>	<u>136</u>	<u>114</u>	<u>97</u>	0										
S7	<u>72</u>	<u>90</u>	<u>115</u>	<u>109</u>	<u>98</u>	0									
S8	<u>70</u>	<u>87</u>	<u>58</u>	<u>27</u>	<u>73</u>	<u>107</u>	0								
S9	<u>48</u>	<u>58</u>	<u>96</u>	<u>90</u>	<u>110</u>	<u>34</u>	<u>95</u>	0							
S10	<u>75</u>	<u>46</u>	<u>88</u>	<u>98</u>	<u>136</u>	<u>63</u>	<u>111</u>	<u>41</u>	0						
S11	<u>63</u>	<u>112</u>	<u>117</u>	<u>89</u>	<u>59</u>	<u>72</u>	<u>74</u>	<u>77</u>	<u>114</u>	0					
S12	<u>102</u>	<u>92</u>	<u>41</u>	<u>58</u>	<u>86</u>	<u>111</u>	<u>55</u>	<u>104</u>	<u>103</u>	<u>108</u>	0				
S14	<u>75</u>	<u>58</u>	<u>118</u>	<u>118</u>	<u>163</u>	<u>81</u>	<u>133</u>	<u>54</u>	<u>46</u>	<u>125</u>	<u>140</u>	0			
S15	<u>92</u>	<u>120</u>	<u>90</u>	<u>57</u>	<u>69</u>	<u>129</u>	<u>35</u>	<u>122</u>	<u>143</u>	<u>77</u>	<u>80</u>	<u>162</u>	0		
S18	<u>99</u>	<u>116</u>	<u>87</u>	<u>81</u>	<u>33</u>	<u>90</u>	<u>61</u>	<u>98</u>	<u>115</u>	<u>72</u>	<u>55</u>	<u>148</u>	<u>71</u>	0	
S20	<u>83</u>	<u>56</u>	<u>63</u>	<u>83</u>	<u>113</u>	<u>69</u>	<u>91</u>	<u>56</u>	<u>39</u>	<u>109</u>	<u>68</u>	<u>83</u>	<u>123</u>	<u>87</u>	0

The distances between the centers of gravity of all proteins in the map have been calculated (in Å). Underlined values are those for which a pairwise neutron measurement exists.

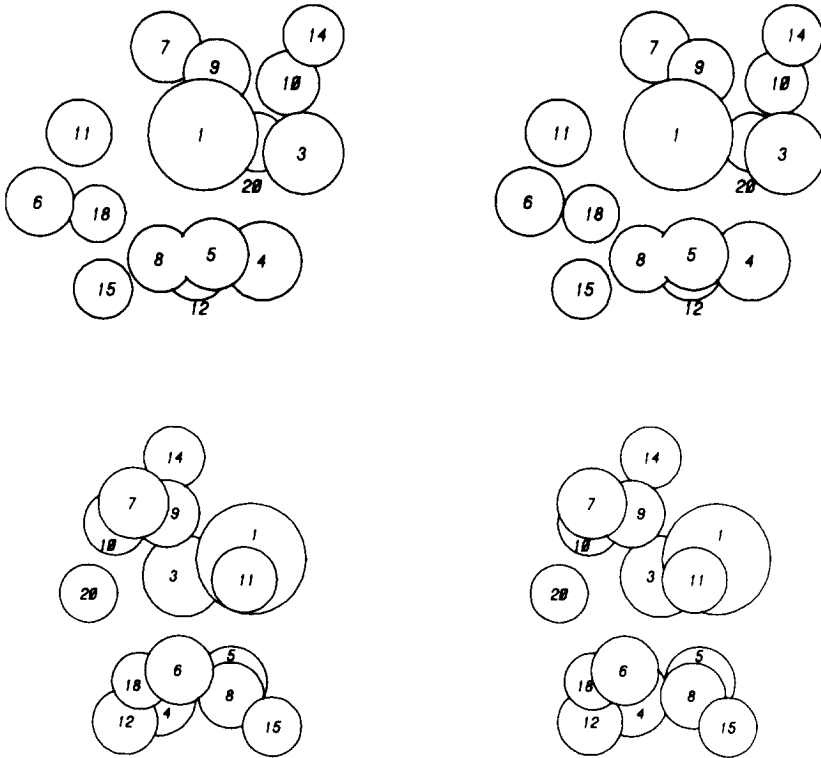


FIG. 5. A neutron map of the positions of proteins in the 30 S subunit of *E. coli*. The map implied by Table 3 has been drawn representing the proteins as spheres whose volumes represent the anhydrous volume of each species. Two stereo views are shown. The second is rotated by  $90^\circ$  counterclockwise around a vertical axis in the plane of the page.

values have again been used in Table 3. Whatever way errors are estimated, it is clear that only S1, S4 and S8 have radii of gyration that are determined well enough to be useful. All radii marked with an asterisk in Table 3 are determined primarily by the lower bound constraint. Realistically, their standard errors are at least  $\pm 5 \text{ \AA}$ . The other data-determined standard errors are extremely large. Many of the proteins appear to have modest radii of gyration, but the uncertainties for most of the estimates are too big to permit their values to be specified with any certainty.

The covariance matrix for the map is a 54 by 54 matrix whose elements are almost all non-zero, and significant in magnitude compared to the diagonal elements. This being the case, the standard errors given in Table 3 by no means constitute a full statement of the error properties of the map. Each protein position has an error structure that should be represented as a 54-dimensional hyperellipsoid. Such constructs are difficult to visualize. However, in the interest of showing the errors more fully, the 3 by 3 covariance matrices expressing the interrelations between errors in  $x$ ,  $y$  and  $z$  for each protein have been examined.

Error ellipsoids have been computed from each such matrix and these are shown in stereo in Figure 6. These ellipsoids represent the region of space in which there is a two-thirds probability of finding the center of gravity of the protein in question.

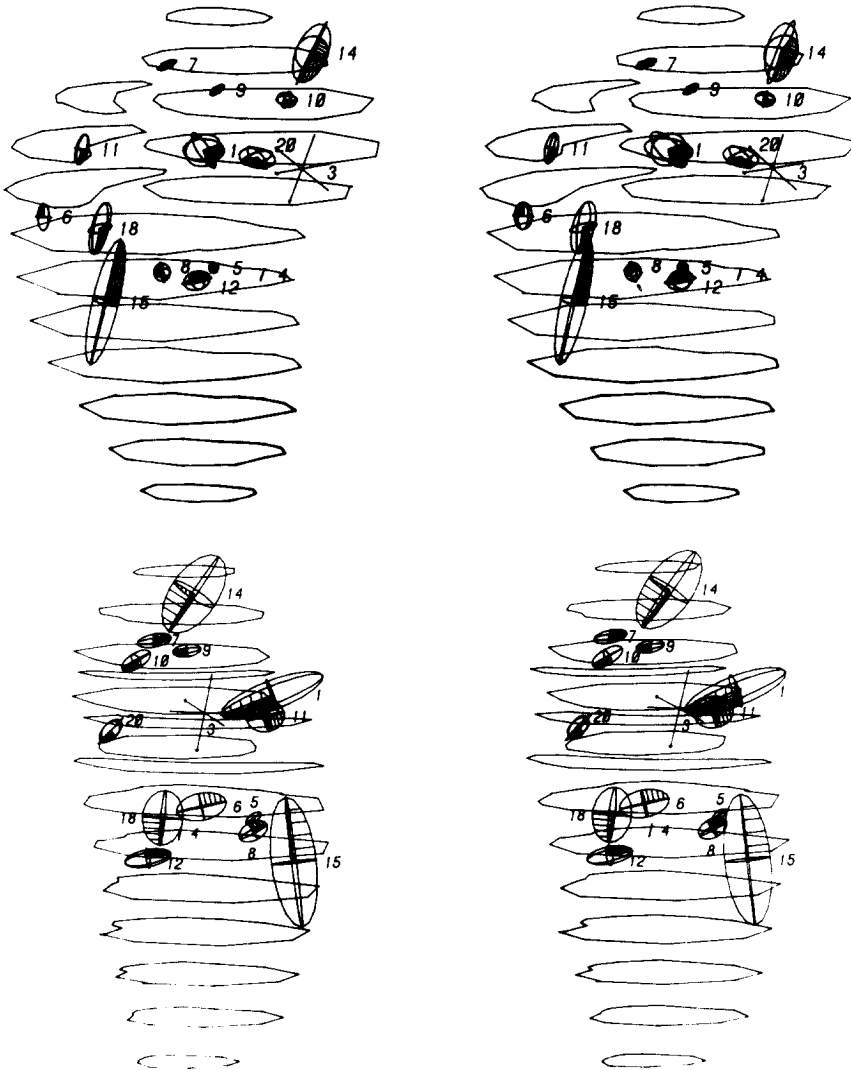


FIG. 6. Error ellipsoids for the positions of protein centers of mass. The 3 by 3 submatrix within the covariance matrix corresponding to the  $x$ ,  $y$  and  $z$  co-ordinates was identified for each protein. Each submatrix was subjected to a principal axis transformation to identify the directions of maximum variance. The variance ellipsoids so computed are shown in stereo, superimposed on the electron microscopic contours of the 30 S particle (Kahan *et al.*, 1981). Two views similar to those in Fig. 5 are shown. The superposition and the view are the same as in Fig. 7. Note that since the position of S3 is the origin of the co-ordinate system, its error ellipsoid is a point. Correspondingly, the ellipsoid of S4 is a line and that of S5 is a planar ellipse.

#### 4. Discussion

##### (a) *Comparisons with earlier neutron maps*

Of the 15 proteins mapped above, 12 were also in the map published in 1981 (Ramakrishnan *et al.*, 1981). Since the amount of data has increased by more than a third between that map and this, it is reasonable to ask whether the conclusions drawn then about this set of 12 have been altered materially.

Eight of the 12 (S3, S4, S5, S7, S9, S10, S11 and S12) have positions and radii of gyration in the new map within one standard error of those estimated for them previously. The  $x$  co-ordinate of S1 has been altered by somewhat more than  $1\sigma$ , but its radius of gyration and  $x$  and  $z$  co-ordinates have been maintained within  $1\sigma$ . (It should be noted, however, that the co-ordinates for S1 were poorly determined in the 1981 map and remain so.) S8, whose error ellipsoid is much smaller, moved by slightly more than  $1\sigma$  in  $x$ , but its radius of gyration is now estimated to be 25.6 Å.

The remaining two proteins, S6 and S15, both have positions, but not radii of gyration, significantly different from those assigned them in 1981. These changes reflect the new value used for the S4–S6 distance in this map. (It should be noted that reservations were expressed about the earlier S4–S6 data at the time, and the sensitivity of the position of S6 to the S4–S6 data noted (Ramakrishnan *et al.*, 1981).)

##### (b) *Critique*

The mathematics of the triangulation method show that while four distances measured from known locations to a new protein component are necessary to locate it, they also show that four may not be sufficient. It can happen that two positions will be possible for a given component, mapped by four distances, both accompanied by similar residuals, i.e. statistically indistinguishable. These situations are hard to recognize until the data have been examined extensively, and can evade notice even then. A problem of this kind was discovered previously in the case of S11 (Schindler *et al.*, 1979); once recognized, it was resolved by making more distance measurements. The problem has arisen again in this map for S14. S14 can be mapped as shown in Figure 5. It also can be fit in a position close to S20 with only minor rearrangements in the rest of the map, mainly involving S1, and a residual comparable to that of the map shown. At this point, the choice between the two positions is arbitrary. The ambiguity will be resolved, as always, by further measurements.

##### (c) *Prospects for further development of the map*

Six proteins remain to be located: S2, S13, S16, S17, S19 and S21. The procedures currently in use should suffice to locate all of these except S2. S2 is exceptional because it exchanges between ribosomes at an appreciable rate (Robertson *et al.*, 1977; Subramanian & van Duin, 1977). S2 could be mapped by

stabilizing its association with the ribosome by crosslinking, as was done with S1, which also exchanges (Sillers & Moore, 1981). It could be mapped by an experimental strategy relying on unmixed samples, such as that being used for the 50 S subunit (Nierhaus *et al.*, 1983).

Beyond placing the remaining proteins, completing the map at the present level of description, there has long existed the hope that it would be possible to interpret the length distribution data obtained in this work to arrive at a map that specifies the orientations of the proteins that are elongated. The possibility of developing an algorithm to accomplish this analysis of the data has been explored in detail recently (Moore *et al.*, 1984). The results are not encouraging. The configurational space that has to be explored has multiple minima. There are many configurations that will have the property of fitting the observed data better than any in their immediate neighborhood. Unfortunately, there are no economical algorithms for finding the global minimum in non-linear systems with multiple minima. The dimensionality of the problem is large enough to discourage direct searches for the minimum. Furthermore, the error inevitably present in the input data makes it possible that, even if the global minimum were found by a search, the configuration identified would differ significantly from the real one, i.e. the wrong local minimum would be identified as the global one. Thus, unless methods can be found for combining neutron data with other information in a systematic way, a 21 protein map like that shown, with radii of gyration included, will be the end product of this work.

#### (d) *Protein placement in the 30 S subunit*

The neutron protein map can be oriented within the electron microscopic image of the 30 S subunit by reference to protein antigenic sites located by immune electron microscopy (e.g. Fig. 7). A consistent superposition of the two models is readily obtained when the data of Kahan *et al.* (1981: on 8 proteins) are employed. It is interesting that their model does not include S4, a protein whose antigenic determinants they cannot detect in intact subunits. Recently (Winkelman *et al.*, 1982), they have obtained a position for S4 by immune electron microscopy done on particles reconstituted lacking S12 and S5, both known from neutron data to be close to S4. The position they find for S4 in these particles is consistent with that predicted by the superposition in Figure 7.

In the last year, the Berlin group has proposed an electron microscopic model of the 30 S subunit including all proteins except S1 and S8 (Stöffler-Meilicke & Stöffler, 1982). A consistent placement of the neutron model within the Berlin electron microscopic model is possible, which is similar to that shown in Figure 7. There are obvious conflicts for only two proteins. The Berlin model puts S20 near S4 in the middle of the particle; the neutron data place S20 higher in the head. The Berlin model locates S15 further down in the body than the neutron map suggests. The error ellipsoid for S15 in the neutron map, however, is elongated in a direction that would allow placement of S15 closer to the Berlin position than its current optimum position without much violence to the data.

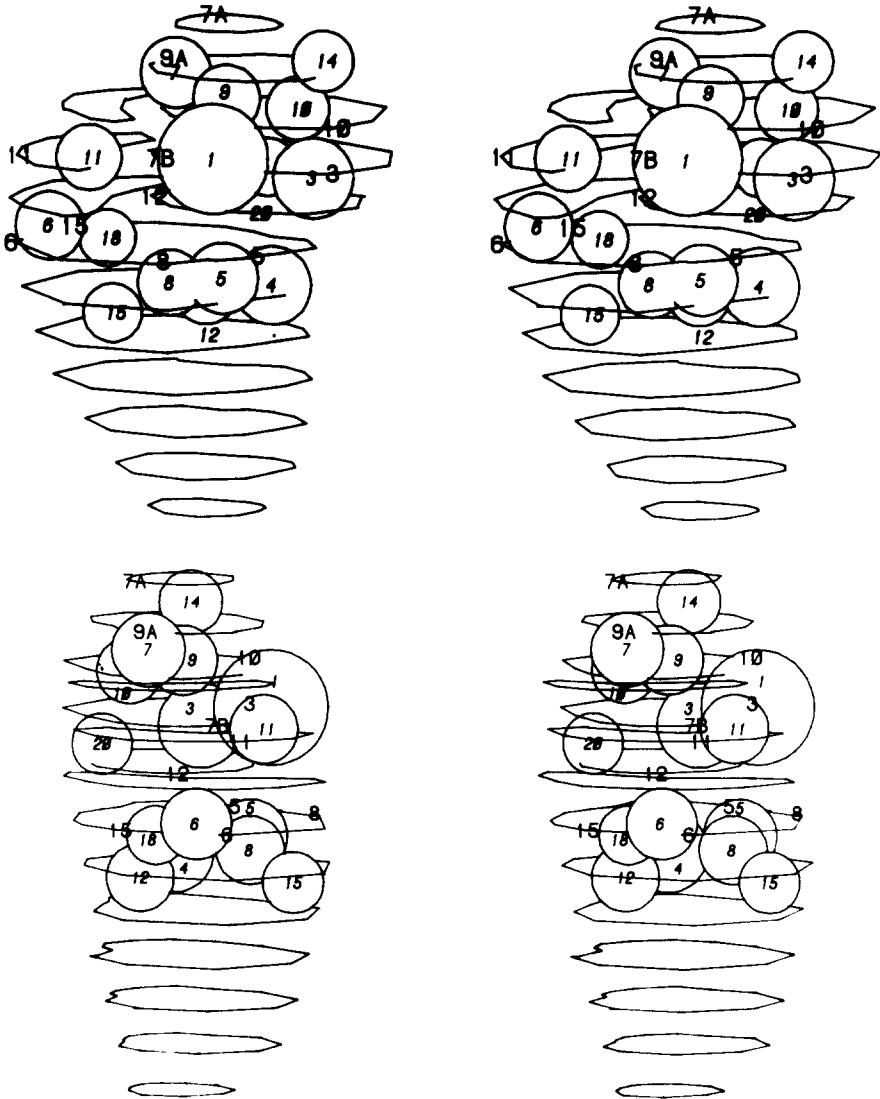


FIG. 7. Superposition of the neutron map of the 30 S subunit with the immune electron microscopic map produced by Kahan *et al.* (1981). The neutron map has been rotated and translated so as to bring the centers of gravity of the proteins as close as possible to the positions of specific antigenic determinants in the electron microscopic map. These have been marked with the larger numerals. Effort is also made to keep the proteins within the contours of the particle as far as possible. Again, the two views of Fig. 5 are presented.

(e) *Bulk distribution of protein in the 30 S subunit*

The superposition of the neutron map on the electron microscopic image of the subunit plainly implies that the top of the particle is protein-rich, while the bottom is protein-poor. The ribosome is 40% (v/v) protein. If the top two-thirds

contain all the protein, then that portion will be 60% (v/v) protein, while the bottom third is devoid of protein. Since 80% of the total 30 S protein has now been located, this conclusion will not materially alter whatever the locations found for the remaining six species, and there is every reason to believe they will be located within or adjacent to the region defined by the proteins whose portions are already known.

The radius of gyration of the mapped protein array is  $\sim 60$  Å and its center of gravity is at  $x = 32$  Å,  $y = 24$  Å,  $z = 36$  Å, which is in the middle of the array in the region devoid of protein. Thus, there is a substantial separation between the center of gravity of the protein distribution and that of the whole particle, let alone that of the RNA. The longest chord of the protein array, allowing for protein diameter, is about 180 Å (S14 to S15). For comparison, the longest chord in the whole particle is about 230 Å and its radius of gyration is around 70 Å (see Kearney & Moore, 1983, and references therein).

Contrast variation studies were done in the early 1970s to determine parameters for the bulk distribution of RNA and protein in these particles (for a review, see Koch & Stuhmann, 1979). The conclusion reached was that in the 30 S subunit, RNA and protein have concentric distributions, with the protein radius of gyration being greater than that of the whole particle. These results are simply incompatible with the findings here, as was pointed out at the 12-protein stage of this project (Ramakrishnan *et al.*, 1981). The reason for this conflict is still unclear.

#### (f) *Comparisons with other structural data*

As the data of Table 4 are examined, one becomes impressed with the degree to which they correlate with the fluorescent energy transfer data reported by Huang *et al.* (1975). Seven of their 18 pairwise distances can be matched with neutron-determined ones, six of them measured directly and one of them derived from other data. Five of the seven fluorescence distances are within a standard error of the corresponding neutron center-to-center distance. The sixth, S4-S15, was reported as 66 Å by fluorescence and appears to be 90 Å by neutrons. The neutron length distribution, however, includes 66 Å. The last, S4-S20, was reported as 46 Å by fluorescence and a derived value of 63 Å on the current map, again a plausible result.

The results of an extensive protein crosslinking study of the 30 S subunit have been reported recently (Lambert *et al.*, 1983), which extends and corrects earlier data. Three previously reported crosslinks have been withdrawn: S4-S6, S5-S9 and S7-S8. In addition, a fourth cross-link of interest here, S12-S20, reported earlier, has not been observed in the recent work. These four crosslinks were all extremely difficult to reconcile with neutron data. Their removal from the list of cross-links has resulted in a substantially improved, but still imperfect, correlation between the two kinds of data. Proteins that are far apart in the neutron map crosslink either not at all, or with low yield. Pairs that are close together often crosslink with high yield.

*(g) The newly placed proteins*

The placements given S14, S18 and S20 in this map are compatible with existing data. There is extensive literature on S18 suggesting its location. Its proximity to S6 is indicated by crosslinking results (Traut *et al.*, 1980), the fluorescence data already referred to, and the fact of the strong interdependence of the two molecules in assembly (Held *et al.*, 1974). (It has been observed consistently that proteins which interact strongly in assembly are neighbors in the mature ribosome.) The placement of S6 relative to other proteins can be confirmed from electron microscopy (Kahan *et al.*, 1981; Stöffler-Meilicke & Stöffler, 1982; Olson *et al.*, 1980). S14 is placed by electron microscopy in the head of the subunit. That it "belongs" there is suggested by the fact that it interacts with proteins S9, S10 and, by mediation of S19, with S7 as well during assembly (Held *et al.*, 1974). The neutron data place S14 in the neighborhood of these molecules.

Of the three proteins, there is least information about the position of S20. The recent Berlin data make it a neighbor of S4. In assembly, S20 interacts weakly with a large number of proteins, some of them in the "head" group (e.g. S7) and others from the middle of the particle (e.g. S4, S8; Held *et al.*, 1974); it is not clear where S20 "belongs".

*(h) Functional correlations*

Ribosomal components function at two levels; first, in the assembly process that leads to ribosome biogenesis, and second, in the mature particle to assist in the catalysis of protein synthesis. The protein role in assembly has been studied in detail only in the case of reconstitution *in vitro*, which is a rather artificial analogue of the process *in vivo*. The placements found for S14, S18 and S20 further verify the trend noted previously; namely, that proteins which interact strongly in assembly are close neighbors in the mature particle (see Ramakrishnan *et al.*, 1981, and references therein). It is interesting that the assembly map (Held *et al.*, 1974) has within it two sequences of interactions starting at S8 proceeding to S5 and then to S3, and another starting at S7 running from S9 to S10 and then to S3. These relationships correspond to crossing the belt from one side of the particle to the other and then going back across the top of the head. S14 fits in as a side branch related to S10 and through S19, whose position is unknown, to S7.

The assembly interaction of S18 with S6 has already been noted. S20 makes no strong interaction with any protein in assembly; it interacts weakly only with S7 among the proteins in this map.

The absence of strong interactions in assembly between proteins that are far apart in the structure suggests that the proteins act by stabilizing relatively local structures in the RNA and hints that whatever long-range interactions are needed for ribosome assembly may be RNA-mediated.

The "division of labor" between RNA and protein in the ribosome remains remarkably obscure. Starting in the early 1970s, large numbers of experiments were done in an effort to assign functions to the proteins of the ribosome (for reviews, see Cooperman, 1980; Liljas, 1982). The evidence available so far falls

short of proving that any protein is an essential component of a binding or catalytic site on the 30 S subunit. Most of the data in the literature, at best, show which proteins are geographically close to whatever site is being probed. Provided one has an independent source of information about the positions of proteins within the structure, chemical proximity evidence of this kind localizes the regions where interactions take place. Several examples of such localizations are given in the review by Liljas (1982). (An alternative approach, of course, is to attempt to visualize the ligand bound to the ribosome by electron microscopy (e.g. Bernabeu & Lake, 1982).) In studies such as these, the protein map serves as a reference frame relative to which sites can be located.

It is clear that the protein reference frame will be of considerable interest in the next few years as efforts develop to work out the organization of 16 S RNA in the 30 S subunit. The determination of the nucleotide sequences of a number of 16 S RNAs has led to the proposal of a general secondary structure for all 16 S RNAs (Noller & Woese, 1981; for a review, see Wittmann, 1982). Information has been accumulating for a long time about where specific proteins bind to 16 S RNA (for a review, see Zimmerman, 1980). Obviously, the combination of this information with knowledge of the spatial distribution of proteins should lead to an understanding of how the RNA is arrayed in the particle (e.g. Noller & Woese, 1981).

We thank Drs D. Schneider and B. Schoenborn for their help and encouragement. Neutron data were collected at Brookhaven National Laboratory under the auspices of the Department of Energy. We acknowledge the expert technical assistance of Mrs Betty Freeborn. The work has been supported by grants from the National Institutes of Health (AI-09167 to P.B.M.), and the National Science Foundation (PCM78-10361 to D.M.E. and P.B.M.). The computational facilities used were supported by an N.I.H. grant (GM-22778). M.K. is the recipient of a grant from the Danish Science Research Council.

## REFERENCES

- Bernabeu, C. & Lake, J. A. (1982). *Proc. Nat. Acad. Sci., U.S.A.* **79**, 3111-3115.
- Cooperman, B. S. (1980). In *Ribosomes: Structure, Function and Genetics* (Chambliss, G., Craven, G. R., Davies, J., Davis, K., Kahan, L. & Nomura, M., eds), pp. 531-554. University Park Press, Baltimore.
- Engelman, D. M. (1979). *Methods Enzymol.* **59**, 656-669.
- Engelman, D. M., Moore, P. B. & Schoenborn, B. P. (1975). *Proc. Nat. Acad. Sci., U.S.A.* **72**, 3888-3892.
- Held, W. A., Ballou, B., Mizushima, S. & Nomura, M. (1974). *J. Biol. Chem.* **119**, 391-397.
- Hoppe, W. (1972). *Israel J. Chem.* **10**, 321-333.
- Hoppe, W. (1973). *J. Mol. Biol.* **78**, 581-585.
- Huang, K. H., Fairclough, R. H. & Cantor, C. R. (1975). *J. Mol. Biol.* **97**, 443-470.
- Kahan, L., Winkelman, D. A. & Lake, J. A. (1981). *J. Mol. Biol.* **145**, 193-214.
- Kearney, K. & Moore, P. B. (1983). *J. Mol. Biol.* **170**, 381-402.
- Koch, M. H. J. & Stuhmann, H. B. (1979). *Methods Enzymol.* **59**, 670-706.
- Lambert, J. M., Boileau, G., Cover, J. A. & Traut, R. R. (1983). *Biochemistry*, **22**, 3913-3920.
- Liljas, A. (1982). *Prog. Biophys. Mol. Biol.* **40**, 161-228.
- May, R. (1978). Ph.D. thesis. Tech. Univ., München.
- Moore, P. B. (1979). *Methods Enzymol.* **59**, 639-655.

- 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., Engelman, D. M., Langer, J. A., Ramakrishnan, V. R., Schindler, D. G., Schoenborn, B. P., Sillers, I. Y. & Yabuki, S. (1984). *Brookhaven Symp. Biol.* **32**, in the press.
- Nierhaus, K. H., Lietzke, R., May, R. R., Nowotny, V., Schulze, H., Simpson, K., Wurmbach, P. & Stuhmann, H. B. (1983). *Proc. Nat. Acad. Sci., U.S.A.* **80**, 2889–2893.
- Noller, H. F. & Woese, C. R. (1981). *Science*, **212**, 403–411.
- Olson, H. M., Grant, P. G., Glitz, D. G. & Cooperman, B. S. (1980). *Proc. Nat. Acad. Sci., U.S.A.* **77**, 890–894.
- Ramakrishnan, V. R. & Moore, P. B. (1981). *J. Mol. Biol.* **153**, 719–738.
- Ramakrishnan, V. R., Yabuki, S., Sillers, I.-Y., Schindler, D. G., Engelman, D. M. & Moore, P. B. (1981). *J. Mol. Biol.* **153**, 739–760.
- Robertson, W. R., Dowsett, S. J. & Hardy, S. J. S. (1977). *Mol. Gen. Genet.* **157**, 205–214.
- Saxena, A. N. & Schoenborn, B. P. (1977). *Acta Crystallogr. sect. A*, **33**, 805–813.
- Schindler, D. G., Langer, J. A., Engelman, D. M. & Moore, P. B. (1979). *J. Mol. Biol.* **134**, 595–620.
- Schneider, D. & Schoenborn, B. P. (1984). *Brookhaven Symp. Biol.* **32**, in the press.
- Schoenborn, B. P., Alberi, J., Saxena, A. M. & Fisher, J. (1978). *J. Appl. Crystallogr.* **11**, 455–459.
- Sillers, I.-Y. & Moore, P. B. (1981). *J. Mol. Biol.* **153**, 761–780.
- Stoeckel, P., May, R., Strell, I., Geha, Z., Hoppe, W., Henmann, H., Zillig, W. & Crespi, H. L. (1979). *J. Appl. Crystallogr.* **12**, 176–185.
- Stöffler-Meilicke, M. & Stöffler, G. (1982). In *Electron Microscopy 1982*, vol. 3, pp. 99–100, Deutsche Gesellschaft für Electronmikroskopie, Frankfurt.
- Subramanian, A.-R. & van Duin, J. (1977). *Mol. Gen. Genet.* **158**, 1–9.
- Traub, P. & Nomura, M. (1968). *Proc. Nat. Acad. Sci., U.S.A.* **59**, 777–784.
- Traub, R. R., Lambert, J. M., Boileau, G. & Kenny, J. W. (1980). In *Ribosomes: Structure, Function and Genetics* (Chambliss, G., Craven, G. R., Davies, J., Davis, K., Kahan, L. & Nomura, M., eds), pp. 51–58, University Park Press, Baltimore.
- Winkelman, D. A., Kahan, L. & Lake, J. A. (1982). *Proc. Nat. Acad. Sci., U.S.A.* **79**, 5184–5188.
- Wittmann, H. G. (1982). *Annu. Rev. Biochem.* **51**, 155–183.
- Wittmann, H. G., Littlechild, J. A. & Wittmann-Liebold, B. (1980). In *Ribosomes: Structure, Function and Genetics* (Chambliss, G., Craven, G. R., Davies, J., Davis, K., Kahan, L. & Nomura, M., eds), pp. 51–58, University Park Press, Baltimore.
- Zimmerman, R. A. (1980). In *Ribosomes: Structure, Function and Genetics* (Chambliss, G., Craven, G. R., Davies, J., Davis, K., Kahan, L. & Nomura, M., eds), pp. 531–554, University Park Press, Baltimore.

*Edited by C. R. Cantor*