The EMBO Journal vol.7 no.5 pp.1509–1513, 1988

The *Escherichia coli* 30S ribosomal subunit; an optimized three-dimensional fit between the ribosomal proteins and the 16S RNA

Dierk Schüler and Richard Brimacombe

Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, D-1000 Berlin-Dahlem, FRG

Communicated by H.G.Wittmann

We have generated a computerized fit between the 3-dimensional map of the *E.coli* 30S ribosomal proteins, as determined by neutron scattering, and the recently published 3-dimensional model for the 16S RNA. To achieve this, the framework of coordinates for RNA–protein cross-link sites on the phosphate backbone in the RNA model was related to the corresponding framework of coordinates for the mass centres of the proteins by a least squares fitting procedure. The resulting structure, displayed on a computer graphics system, gives the first complete picture of the *E.coli* 30S ribosomal subunit showing both the proteins and the double-helical regions of the RNA. The root mean square distance between cross-link sites and protein centres is 32 Å. The position of the mass centre of the combined double-helical regions was calculated from the model and compared with the position of the mass centre of the complete set of proteins. The two centres are displaced relative to one another by 20 Å in the model structure, in good agreement with the experimental value of 25 Å found by neutron scattering.

Key words: ribosome structure/computer graphics/RNA models/RNA–protein cross-linking/neutron scattering

Introduction

The ribosome is a complex array of protein and RNA molecules, and its structure has been the object of intensive study for many years, with particular attention being focused on the 30S subunit of the ribosome from *Escherichia coli*. A comprehensive description of the arrangement of the proteins from this subunit has come from the neutron scattering data of Moore *et al.* (1986), in the form of a 3-dimensional matrix of the protein mass centres, which has recently been extended to include all 21 of the 30S ribosomal proteins (Capel *et al.*, 1987). The corresponding arrangement of the 16S RNA is based on the phylogenetically established secondary structure of the molecule (Maly and Brimacombe, 1983; Noller, 1984), which can be constrained into three dimensions using intra-RNA cross-linking information (Atmadja *et al.*, 1986; Stiege *et al.*, 1986) as well as other types of data, such as the electron microscopic localization of specific nucleotides on the subunit surface (see for example Gornicki *et al.*, 1984).

A link between the respective RNA and protein structures has recently been provided by the determination of a network of RNA–protein cross-link sites within the 16S RNA, involving 13 of the ribosomal proteins (Osswald *et al.*, 1987; Greuer *et al.*, 1987). This latter set of data has played an important role in the construction of a detailed 'wire-and-tube' model for the arrangement of the 16S RNA *in situ* in the 30S subunit (Brimacombe *et al.*, 1988), the RNA–protein cross-linking sites being used to orient the various parts of the 16S RNA structure in relation to the neutron scattering map of the proteins (Moore *et al.*, 1986).

Fitting the ribosomal proteins to the RNA model via the RNA–protein cross-linking data is not a trivial matter, for three reasons. First, the RNA is itself highly constrained by virtue of its secondary structure as well as by the intra-RNA cross-links already mentioned (Atmadja *et al.*, 1986; Stiege *et al.*, 1986), and in particular several of the double-helical regions of the RNA contain cross-links to more than one protein. There are therefore, quite apart from considerations of the overall size and shape of the RNA model, severe limits to the extent to which the RNA can be manipulated to fit the protein arrangement. Secondly, whereas the neutron data give the relative positions of the protein centres of mass, the RNA–protein cross-links can be to any point on the surface of the protein concerned. Thirdly, since the detailed shapes of the proteins are not known, there is no simple or obvious way in which to build the proteins as such into the wire model of the RNA; it is clear that any attempt to fit the RNA model around a set of spherical objects representing the proteins would be a meaningless exercise. In consequence, in our description of the RNA wire model (Brimacombe *et al.*, 1988), we were not able to do more than measure the distances between the RNA–protein cross-link sites in the model and demonstrate that these distances were compatible with the inter-protein distances obtained by neutron scattering (Moore *et al.*, 1986).

In this paper, we use the computer graphics version of the RNA model (Brimacombe *et al.*, 1988), in order to fit the proteins into the structure. For this purpose the coordinates of each RNA helix as well as those of the RNA–protein cross-link sites were measured from the wire model and fed into the computer (cf. Brimacombe *et al.*, 1988). The coordinates of the protein centres of mass were then transformed (as a rigid data set) by means of a least squares fitting procedure, so as to bring them into an optimal orientation with respect to the coordinates of the RNA–protein cross-link sites. The protein arrangement (represented as a set of spheres) is displayed in this optimal orientation simply by superimposing it onto the RNA structure, in which only the double-helical elements are visualized on the computer screen (as a set of cylinders). The plausibility of the resulting model is tested by calculating the positions of the overall centres of mass of the protein and RNA moieties, which have been found by neutron scattering studies to be displaced relative to one another by 25 Å (Ramakrishnan, 1986).

©IRL Press Limited, Oxford, England

Table I. Coordinates of mass centres of the ribosomal proteins, fitted to those of the RNA–protein cross-link sites[a]

Protein centres of mass				RNA–protein cross-link sites						Separation distance[d] (Å)
Protein	Coordinates (Å)[b]			Protein	Position in RNA[c]		Coordinates (Å)			
	x	y	z		Nucleotide	Helix	x	y	z	
S1	10	−1	29	–						
S2	14	−23	18	–						
S3	−38	2	1	S3	1155–1158	39–40	−42	34	−18	37
S4	−35	−55	−10	S4	413	16	−47	−72	−12	22
S5	−4	−53	16	S5	559–561	3–19	−5	−62	−5	23
S6	74	−44	−41	–						
S7	25	40	−28	S7-A	1238–1240	30–41	22	39	−21	8
				S7-B	1377–1378	28–43	39	16	−4	37
S8	19	−67	4	S8-A	629–633	21	18	−105	−16	43
				S8-B	651–654	21–22	58	−73	−2	39
S9	6	29	−9	S9-A	954	30	−2	25	16	27
				S9-B	1130–1131	39	−10	56	−6	32
S10	−31	33	−13	S10	1139–1144	39	−35	63	−10	30
S11	69	−8	4	S11-A	693–697	23	76	21	−29	45
				S11-B	702–705	23	81	10	−7	25
S12	−12	−65	−41	–						
S13	25	76	13	S13	1337–1338	29–42	29	41	18	36
S14	−56	64	−8	–						
S15	47	−65	19	–						
S16	−3	−33	−31	–						
S17	28	−89	−26	S17-A	278–280	11	19	−91	7	35
				S17-B	629–633	21	19	−107	−21	21
S18	42	−37	−45	S18	845–851	26	39	−25	−10	37
S19	3	40	41	S19	1223–1231	30	−1	46	12	30
S20	−33	3	−39	–						
S21	31	−18	−20	S21-A	693–697	23	71	21	−35	(58)
				S21-B	723–724	22–23	80	20	−9	(63)

[a]Both sets of coordinates are given to the nearest 1 Å in the Cartesian system of Brimacombe *et al.* (1988).
[b]Coordinates of the 30S proteins according to Capel *et al.* (1987), transformed according to the least squares fit with the RNA–protein cross-link sites.
[c]Taken from Brimacombe *et al.* (1988). Helix numbers give the 16S RNA helices within which (single numbers) or between which (pairs of numbers) the cross-links are located.
[d]Distance between RNA–protein cross-link site and corresponding protein centre of mass.

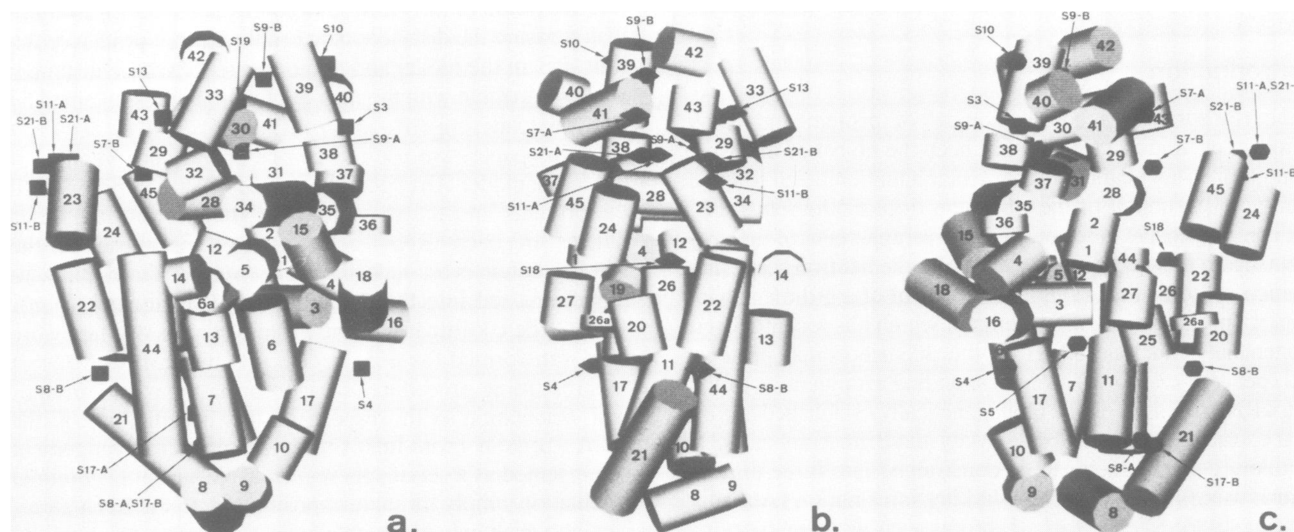

Fig. 1. Double-helical regions (cylinders) of the 16S RNA, together with positions of RNA–protein cross-link sites (small black polygons) on the RNA model. Nomenclature of both helices and cross-link sites is that of Brimacombe *et al.* (1988). Orientation of the model in the three views is (**a**) 190°, (**b**) 70° and (**c**) 315°.

Results

The positions of the various RNA–protein cross-link sites in the 16S RNA primary and secondary structure are listed in Table I, together with their coordinates measured from the 'phosphate backbone' of the RNA wire model. The nomenclature used for the cross-links and for the double-

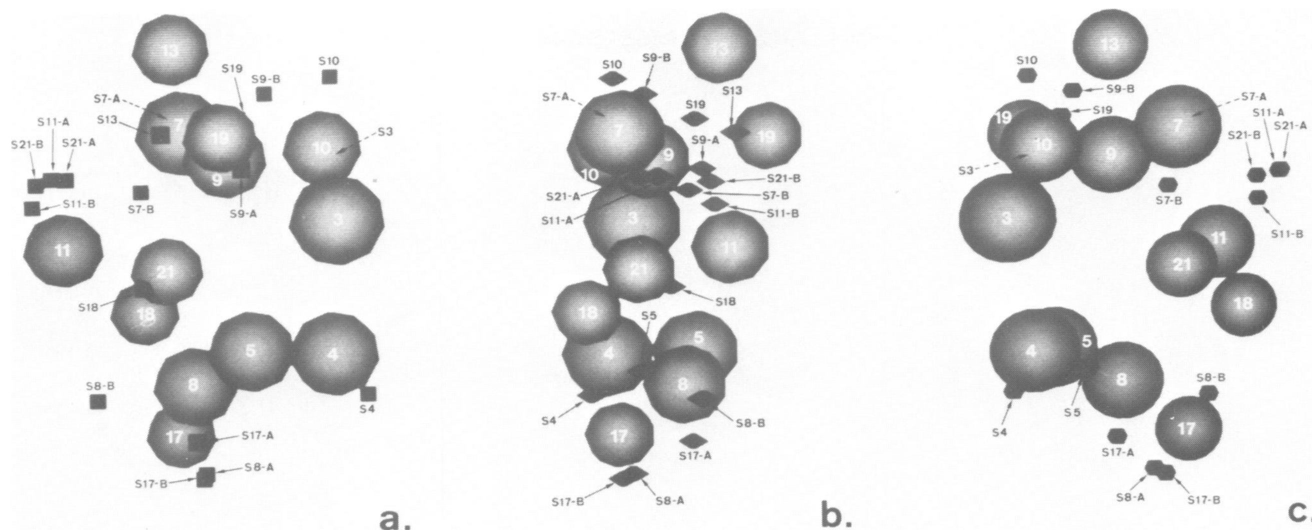

Fig. 2. Least squares fit of proteins (spheres) and RNA–protein cross-link sites (small polygons, cf. Figure 1). The 'S' prefixes for the names of the proteins are omitted from the set of white numbers on the protein spheres. Dotted lines signify that the cross-link site concerned lies 'within' the protein sphere indicated. Orientations of the cross-link sites in (**a**), (**b**) and (**c**) are the same as those of Figure 1.

helical regions of the RNA is the same as that of Brimacombe *et al.* (1988), and the coordinates are given in the same Cartesian system as that used to describe the RNA helices in the latter publication. Three views of the computer graphics model of the RNA, in which the RNA helices are represented as cylinders and the RNA–protein cross-link sites as small black polygons, are shown in Figure 1.

In order to generate the least squares fit between the protein centres of mass and the RNA–protein cross-link sites, the protein coordinates of Capel *et al.* (1987) were used, with the sign of the 'z' coordinate being reversed so as to bring the coordinate system into the same handedness as that of Table I. The least squares programme (see Materials and methods) was then applied to the mass centres of those proteins for which there are also RNA–protein cross-link data (Table I). All the RNA–protein cross-links in Table I with the exception of those to protein S21 (see Discussion) were used in the fitting procedure, and otherwise no 'weighting' of individual cross-links was applied. It was furthermore not necessary to make a preliminary 'manual' fit of the data. Three views of the optimal fit obtained with the least squares programme, showing the RNA-protein cross-link sites and the corresponding proteins, are illustrated in Figure 2. In this figure the proteins are represented as spheres of a size which is in approximate relation to their molecular weights (cf. Capel *et al.*, 1987), and the three orientations correspond precisely to those of Figure 1.

The computer fit defines a rotation matrix and a linear translation vector, which must be applied (in that order) to the coordinates of Capel *et al.* (1987) so as to transform them into our coordinate system. The rotation matrix, written according to standard convention, is:

$$\begin{array}{ccc} 0.0546 & 0.7779 & -0.6260 \\ -0.9828 & -0.0688 & -0.1712 \\ -0.1762 & 0.6247 & 0.7608 \end{array}$$

and the translation vector is -37.8 Å (x), 2.3 Å (y), 0.6 Å (z). The transformed values for the coordinates of all the protein mass centres are given in Table I, which also shows the distance between each RNA–protein cross-link site and the corresponding protein centre. Calculation of the root mean square distance between cross-link sites and protein centres (excluding as above the data for protein S21) gives a value of 32.3 Å.

In Figure 3 all 21 of the ribosomal proteins are displayed, oriented in the transformed coordinates given by the least squares fit (Table I, Figure 2), and superimposed on the RNA structure. The RNA–protein cross-link sites are also included for easy reference to Figures 1 and 2, and the model is again shown in the same three orientations as those of Figures 1 and 2. Figure 3 thus gives a first complete picture of the *E.coli* 30S ribosomal subunit, showing both the protein and RNA moieties simultaneously.

Ramakrishnan (1986) has demonstrated by neutron scattering that the spatial distribution of protein and RNA in the 30S subunit is asymmetrical, with a separation between the respective centres of mass of 25 Å. The centre of mass of the proteins can be readily calculated from the neutron scattering data of Capel *et al.* (1987), and has coordinates of 7.8 Å, -12.0 Å, 0.0 Å (including protein S1) or 7.3 Å, -14.3 Å, -6.2 Å (excluding protein S1) in the transformed corodinate system of Table I. For the RNA moiety the calculation of the position of the centre of mass is not simple, but, since the single- and double-stranded regions of the 16S RNA are distributed very evenly throughout the secondary structure of the molecule (cf. Brimacombe *et al.*, 1988), a good approximation can be made by calculating the position of the centre of mass of the double-helical regions (i.e. the cylinders) in the RNA model (Figures 1 and 3). This mass centre has coordinates of 7.2 Å, -31.8 Å, -0.6 Å. With these values the separation of the respective RNA and protein centres of mass in the model (Figure 3) is 19.8 Å in the presence of protein S1 and 18.4 Å in the absence of the latter.

Discussion

As noted in the Introduction, the RNA–protein cross-link sites played an important role in the construction of the RNA

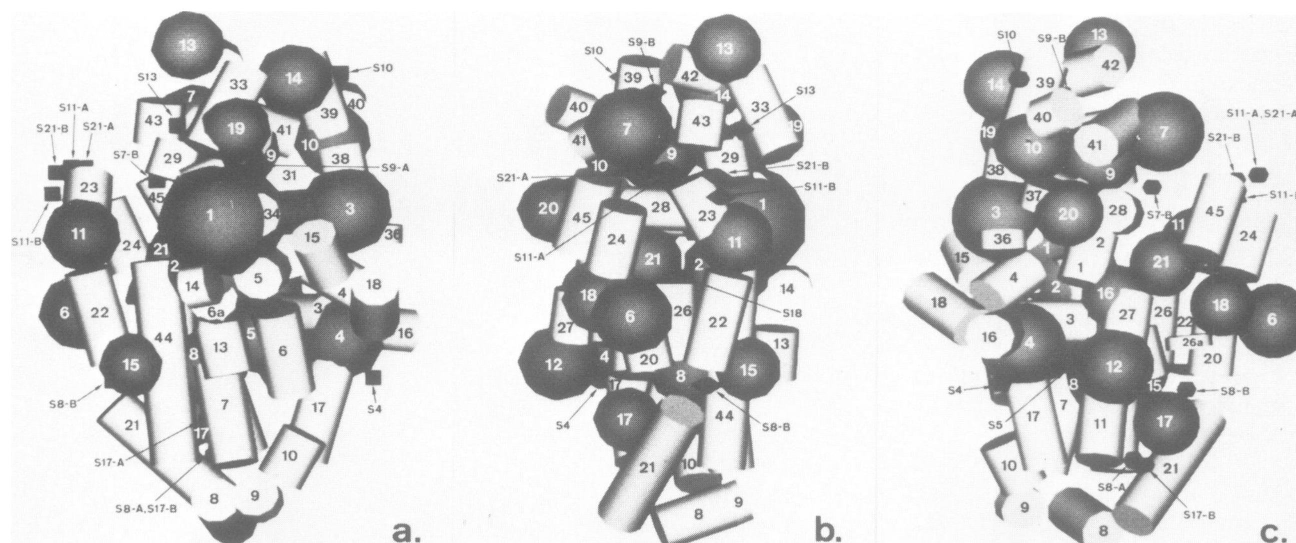

Fig. 3. Combination of Figures 1 and 2, showing all 21 ribosomal proteins, RNA helices, and RNA–protein cross-link sites. Orientations of the model in (**a**), (**b**) and (**c**) are the same as those of Figures 1 and 2.

model (Brimacombe *et al.*, 1988), in that these data form the link between the RNA structure and the neutron scattering map of the proteins (Moore *et al.*, 1986; Capel *et al.*, 1987). However, the 16S model is a wire structure representing the phosphate backbone of the RNA, whereas the neutron scattering data provide a set of coordinates for the mass centres of the proteins. An RNA–protein cross-link site can be to any point on the surface of the protein concerned (rather than to its mass centre), and the sites of cross-linking on the RNA side are most likely to be to the purine or pyrimidine bases of the respective nucleotides (rather than to points on the phosphate backbone). There is thus an inherent degree of uncertainty in the placement of the RNA–protein cross-link sites in the model relative to the corresponding protein mass centres.

A further (and more important) complication in fitting the RNA–protein cross-link sites to the neutron data is the fact that several helical elements in the RNA structure (helices 21, 23, 30 and 39) contain cross-link sites to two or even three different proteins. Since the helices are rigid elements, this means that the RNA–protein cross-link sites concerned cannot be manipulated independently of one another. Other interconnections and constraints in the RNA structure (such as the intra-RNA cross-link sites, or the secondary structure itself) enhance this effect to the point where alteration of the position of any one RNA–protein cross-link site in the model is almost certain to cause concomitant changes in the positions of several other sites. As a consequence of these factors, the positioning of the various RNA–protein cross-link sites with respect to the neutron data was achieved when building the model by a series of empirical and qualitative approximations. (At the same time it was by no means a foregone conclusion that it would *a priori* be feasible to fit all the RNA–protein cross-link data into a single coherent structure.) Only when the RNA model is considered as a complete entity is it possible to judge quantitatively how successful (or otherwise) the final correlation with the protein map (Capel *et al.*, 1987) has been, and this was the purpose of the least squares fitting procedure described in this paper.

Since the radius of an average ribosomal protein represented as a sphere would be of the order of 15 Å, then this is the 'ideal' distance that would be expected between a cross-link site on the surface of the protein and the corresponding protein centre of mass. The value of 32 Å found for the root mean square distance between the cross-link sites and protein mass centres (see above and Table I) is thus ~17 Å outside this ideal value. However, as already noted above (cf. Brimacombe *et al.*, 1988), the coordinates of the RNA–protein cross-link sites were measured to points on the phosphate backbone of the RNA and not to points on the purine or pyrimidine bases (a distance difference of up to ~8 Å), and also take no account of the length of the cross-linking reagents (~5 Å). Furthermore, we cannot assess the effects of deviations from sphericality in the shapes of the proteins, since the detailed shapes of the proteins (again as already noted) are still unknown. Thus a 17 Å deviation from the ideal value in the root mean square distance (32 Å) between cross-link sites and protein centres is very reasonable, for a first protein-fitting study at the current relatively low level of resolution.

There is, however, a problem with the data for protein S21, which was not included in the 19-protein neutron scattering map of Moore *et al.* (1986). This 19-protein map was the one used to help build the RNA model, and when the final 21-protein map (Capel *et al.*, 1987) appeared it was immediately obvious that the positions of the cross-link sites to S21 in the RNA model were not reconcilable with the position of this protein in the neutron map of the latter authors. For this reason, as already noted above, the cross-link data for S21 were not included in the least squares fitting procedure (cf. Figure 2), and the distances of the two cross-link sites from the centre of protein S21 are 58 and 63 Å respectively in the optimal fit (Table I). We had already expressed some reservations about the positioning of the cross-link sites for S21 (Brimacombe *et al.*, 1988). One of these sites (S21-B), at positions 723–724, is located on a rather long (and hence flexible) single-stranded region of the RNA between helices 22 and 23, and this single-stranded region could readily be moved so as to bring the cross-link

site close to the neutron position of protein S21 in the model (Figure 3b, centre). The other cross-link site (S21-A) is to positions 693–697 in the loop-end of helix 23, a location which in contrast is relatively rigidly fixed in the RNA model. Thus it is this latter cross-link to protein S21 which is incompatible with the neutron data, and its localization is currently being-reinvestigated.

In the final neutral map of Capel *et al.* (1987) the positions of proteins S14 and S19 are revised relative to the earlier data of Moore *et al.* (1986), and—in contrast to the situation with S21—this has led to a marked improvement in the fitting of the proteins to the RNA model. Not only is the distance between the cross-link site to S19 and the corresponding protein mass centre now considerably reduced (see Figure 2, Table I, and cf. Brimacombe *et al.*, 1988), but also protein S14 has become very close to helix 33 (Figure 3a, b, top). This latter neighbourhood is in precise agreement with recent studies made in our laboratory on the protection of regions of the 16S RNA by specific proteins (Wiener *et al.*, 1988). Another very satisfactory correlation is the position of protein S4, which lies at the junction of helices 3, 4, 16, 17 and 18 (Figure 3a, right), in precise agreement with the protein 'foot-printing' data of Stern *et al.* (1986).

The protein fit of Figure 3 allows for the first time a direct comparison to be made with the locations of the various proteins on the surface of the 30S subunit, as determined by immune electron microscopy (Stöffler and Stöffler-Meilicke, 1986; Oakes *et al.*, 1986). If Figure 3 is compared with Figure 2.5 of Stöffler and Stöffler-Meilicke (1986), then the locations of all the protein antigenic sites (with the exception of a minor discrepancy in the location of protein S19) show a remarkable correspondence with the positions of the proteins in the model (Figure 3). A similar comparison with Figure 3.4 of Oakes *et al.* (1986) shows again a very high level of agreement, although in this data set the antigenic sites for proteins S4, S5, S8 and S6 would appear to lie on the 'upper' sides of the proteins concerned, as compared with Figure 3. It is the location of these latter antigenic sites which has given rise to the prevalent notion among ribosomologists that the entire lower part of the 30S subunit is 'protein free'.

We have already disputed the existence of such an extensive protein-free zone (Brimacombe *et al.*, 1988), and the protein fit of Figure 3 supports the view that this zone is indeed not very large. The location of the antigenic site for protein S17 low down in the 30S subunit by Stöffler and Stöffler-Meilicke (1986) is in full agreement with the location of this protein in the model (Figure 3), as well as with the positions of cross-link sites to S17 almost at the bottom of the subunit. S8 is the only protein which needs to be 'stretched' in order to acommodate both our data and the data of Oakes *et al.* (1986), and this discrepancy could be significantly reduced by rotating helix 21 upwards slightly in the model (see for example Figure 3a, bottom left). Such a shift would have the effect of shortening the distances between the S8 mass centre and the two cross-link sites to this protein in helix 21 (Table I, Figure 2), without worsening the corresponding distances for S17. The lower part of the subunit is on the other hand clearly richer in RNA than the upper regions, and it is noteworthy that helices 6, 10 and 17 (which are deleted in the 16S RNA from chloroplasts (see Brimacombe *et al.*, 1988, for discussion) appear to form a large part of the most RNA-rich region (Figure 3a, lower right).

However, the most important factor in any discussion of the 'protein-free zone' is the experimental finding by Ramakrishnan (1986) that the protein and RNA centres of mass are displaced relative to one another by 25 Å. This parameter is entirely independent of any of the factors taken into account in the construction of the RNA model, and as already mentioned above, the protein fit of Figure 3 gives a value for this mass centre displacement of 19.8 Å (in the presence of protein S1) and 18.4 Å (in the absence of the latter protein). This is in very close agreement to Ramakrishnan's value. Thus, although our model will undoubtedly require a great deal of refinement in its detailed topography before it accurately reflects the actual structure of the 30S subunit (and some specific suggestions for minor improvements have already been made in the foregoing discussion), the overall distribution of RNA and protein in the final structure must be very similar to that shown in Figure 3.

Materials and methods

The RNA model and the data used in its construction have been described in detail by Brimacombe *et al.* (1988). Coordinates measured from the model (of RNA helices and RNA–protein cross-link sites) were fed into an Evans and Sutherland computer graphics system, again as described in the latter publication. The RNA–protein cross-link sites were fitted to the protein centres of mass (Capel *et al.*, 1987) using the least squares program 'Exifit' of A.D.Mclachlan. This program is part of the CCP4 package from the SERC Dalesbury (UK) laboratory (1985).

Acknowledgements

The authors are very grateful to Dr P.Moore for sending us his latest neutron data prior to publication. We also thank Dr H.Lemke of the Technical University, Berlin, for making the Evans & Sutherland PS300 system available to us, M.Engelhorn for his advice with the programming, and Drs W.D.Bennett and H.G.Wittmann for many valuable discussions.

References

Atamdja,J., Stiege,W., Zobawa,M., Greuer,B., Osswald,M. and Brimacombe,R. (1986) *Nucleic Acids Res.*, **14**, 659–673.

Brimacombe,R., Atmadja,J., Stiege,W. and Schüler,D. (1988) *J. Mol. Biol.*, **199**, 115–136.

Capel,M.S., Engelman,D.M., Freeborn,B.R., Kjeldgaard,M., Langer,J.A., Ramakrishnan,V., Schindler,D.G., Schneider,D.K., Schoenborn,B.P., Sillers,I.Y., Yabuki,S. and Moore,P.B. (1987) *Science*, **238**, 1403–1406.

Gornicki,P., Nurse,K., Hellmann,W., Boublik,M. and Ofengand,J. (1984) *J. Biol. Chem.*, **259**, 10493–10498.

Greuer,B., Osswald,M., Brimacombe,R. and Stöffler,G. (1987) *Nucleic Acids Res.*, **15**, 3241–3255.

Maly,P. and Brimacombe,R. (1983) *Nucleic Acids Res.*, **11**, 7263–7286.

Moore,P.B., Capel,M., Kjeldgaard,M. and Engelman,D.M. (1986) In Hardesty,B. and Kramer,G. (eds), *Structure, Function and Genetics of Ribosomes*. Springer-Verlag, New York, pp. 87–100.

Noller,H.F. (1984) *Annu. Rev. Biochem.*, **53**, 119–162.

Oakes,M., Henderson,E., Scheinman,A., Clark,M. and Lake,J.A. (1986) In Hardesty,B. and Kramer,G. (eds), *Structure, Function and Genetics of Ribosomes*. Springer-Verlag, New York, pp. 47–67.

Osswald,M, Greuer,B., Brimacombe,R., Stöffler,G., Bäumert,H. and Fasold,H. (1987) *Nucleic Acids Res.*, **15**, 3221–3240.

Ramakrishnan,V. (1986) *Science*, **231**, 1562–1564.

Stern,S., Wilson,R.C. and Noller,H.F. (1986) *J. Mol. Biol.*, **192**, 101–110.

Stiege,W., Atmadja,J., Zobawa,M. and Brimacombe,R. (1986) *J. Mol. Biol.*, **191**, 135–138.

Stöffler,G. and Stöffler-Mielicke,M. (1986) In Hardesty,B. and Kramer,G. (eds), *Structure, Function and Genetics of Ribosomes*. Springer-Verlag, New York, pp. 28–46.

Wiener,L., Schüler,D. and Brimacombe,R. (1988) *Nucleic Acids Res.*, **16**, 1233–1250.

Received on January 14, 1988; revised on February 15, 1988