SPATIAL CONFIGURATION OF RIBOSOMAL PROTEINS: A COMPUTER-GENERATED MODEL OF THE 30S SUBUNIT

Alex Bollen*, R.J. Cedergren, David Sankoff and Guy Lapalme

Département de biochimie et Centre de recherches mathématiques
Université de Montréal, Case Postale 6128, Montréal H3C 3J7

Received June 26, 1974

Summary: Much data are currently being produced pertinent to spatial relationships among ribosomal proteins. We propose the use of multidimensional scaling for the automated construction and updating of three-dimensional models of the ribosome, based on as much diverse data as becomes available. We examine an initial such model of the 30S subunit.

Ribosomes are now known to play a more active role in protein synthesis than previously thought, interacting with mRNA, tRNA and initiation factors (1,2), and knowledge of the three-dimensional arrangement of ribosomal components seems increasingly a prerequisite for the determination of the translation mechanism of the genetic message. Numerous studies relevant to the topography of the 30S subunit of the bacterial ribosome have appeared since the publication of the assembly map of Muzushima and Nomura (3). We gather here fourteen types of evidence for proximity relationships between the twenty-one different proteins comprising this subunit and from this construct estimates of most of the distances between protein pairs. These distances then serve as input to the multidimensional scaling program MDSCAL (4,5), which outputs the three-dimensional structure most consistent with the input distances.

We have collected those publications and personal communications most suggestive of spatial relationships between the proteins, but we have excluded the assembly map itself so as to reserve an independent source of validation of the final model. In assembling the data, we considered neither the influence of the other ribosomal subunit (50S), the position of the 16S RNA component of the 30S particle, nor questions of structural heterogeneity (6).

Table 1 lists each type of data, as well as the formula converting this

* Laboratoire de génétique, Université libre de Bruxelles, 1640 Rhode St. Genèse, Belgium.

into information about inter-protein distance. For example, experiments demonstrating the crosslinking of two proteins by bifunctional reagents provided a distance estimate between protein centers equal to the sum of their radii plus the average length (10 Å) of the crosslinking agent. To convert fragmentation data, the distance between any two proteins in a fragment is estimated as the sum of their radii plus 0, 1 or 2 average protein diameters (36 Å) depending on the size of the fragment. Each formula yields a conservative estimate of proximity, i.e. overestimates distances. These estimates are motivated by the necessity of combining several types of non-independent data on inter-protein distances. For each pair of proteins for which data is provided by two or more experiments, we choose the minimum distance estimate as the initial input to the scaling program. Thus, among all the protein pairs in a large fragment, those for which there is also crosslinking evidence are assigned the relatively small crosslink distance, while the other pairs are assigned a distance approaching the maximum diameter possible for such a fragment. This combination of small and large values yields a reasonable average over all pairs in the fragment.

The combined estimates resulting from these calculations serve as input for a preliminary analysis by MDSCAL. This program, which has proved extremely useful in many diverse applications in such fields as psychology, archaeology and geography, constructs an n-dimensional configuration of points (protein locations) which is as faithful as possible, according to a certain normalized least-squares criterion, to the relative values, or rank order, of all the distances in the input matrix. The criterion of fit is called stress (formula 2), and this was measured at 0.265 for the three-dimensional output of our preliminary analysis, a value which is considered "fair". We use the preliminary output as a heuristic device to better estimate fragment dimensions, and hence to provide more consistent estimates as input for a final analysis by MDSCAL. This is justified as simply a way to put more weight, although in an indirect way, on the stronger and more precise crosslinkage

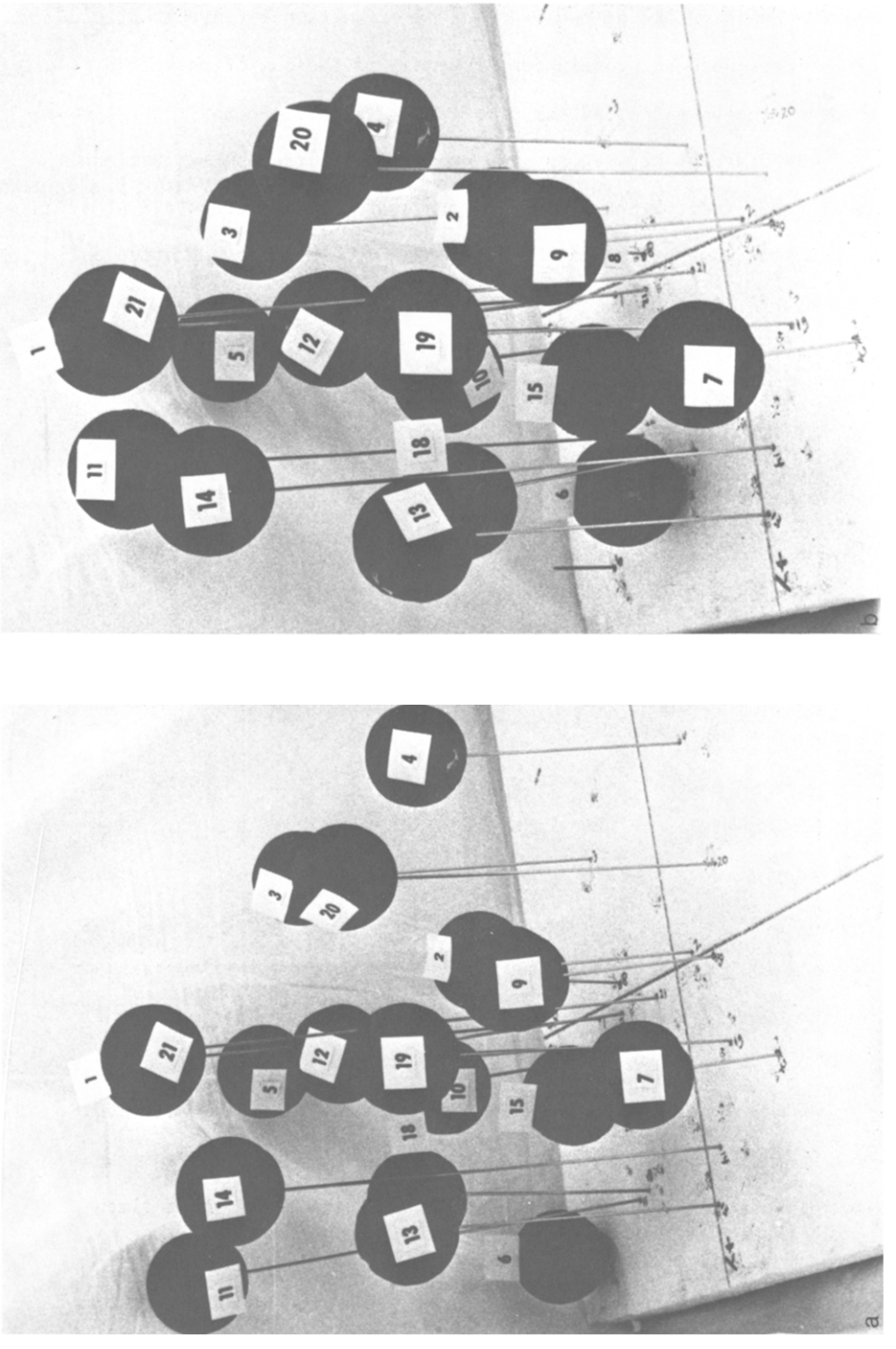

Fig. 1a. Model constructed with coordinates given in Table 2. Proteins S16 and S17 which are found anomalously far from the complex of the other proteins are not shown in the photograph. Protein S8 is blocked from view by S9 and S2. Fig. 1b Squeezed model constructed from model in Fig. 1a by approaching S3, S4, S11, S6, S20 towards the plane defined by S21, S19 and S7.

Table 1. Experiments used for proximity estimates

	Experiment (ref.)	Estimation Formula
A	Protein crosslinking (19-23)	Sum of radii + 10 Å
B	Crosslinking of initiation factors to 30S (24,25) Affinity labelling (26) Nuclease attack on reconstituted-intermediate (27) Protease attack on antibiotic-protected 30S (28-30) Binding of streptomycin to core particles (31)	Sum of radii + 0, 1 or 2 diameters *
C	Genetic studies (32-34)	Sum of radii + 1/2 maximum distances+
D	Nuclease attack on 30S ribosome (17,35)	1/2 maximal distance
E	Partial reconstitution of 30S ribosomes (3,36) Iodination of reconstituted intermediates (37)	
F	Antibody blocking of functions (38) Antibody blocking of antibiotic binding (38,39) Inhibition of 16S RNA methylation by proteins (40) Stimulation of ribosome function by proteins (41)	Sum of radii + 0 or 1 diameter

* 36 Å was used as an average protein diameter. + Maximum distance was calculated by assuming that proteins co-occuring in a fragment were arranged linearly.

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20
S2	-																			
S3	50	50																		
S4	66	-	40																	
S5	48	74	38	126																
S6	117	71	106	126	70															
S7	84	-	74	-	108	-														
S8	84	-	-	35	50	30	-													
S9	80	-	35	35	69	-	33	30												
S10	60	-	45	-	33	-	69	-	30											
S11	45	-	-	107	78	90	105	70	66	-										
S12	45	35	35	35	33	66	-	-	66	-	30									
S13	45	71	-	107	70	66	30	80	30	30	30	50								
S14	66	51	71	107	69	102	69	102	66	54	39	66	45							
S15	-	-	-	32	80	27	30	27	45	-	90	-	27	-						
S16	-	-	-	104	126	126	-	126	-	-	126	-	-	-	126					
S17	-	-	-	104	126	126	-	126	-	-	126	-	-	-	126	45				
S18	78	-	68	104	66	27	-	27	63	-	40	45	40	63	25	126	126			
S19	42	-	68	32	66	100	27	80	27	63	27	45	27	35	-	-	-	60		
S20	78	-	68	32	66	126	-	27	-	35	63	51	-	63	90	97	97	60	60	
S21	42	-	32	64	66	100	102	100	80	50	40	54	54	33	-	-	-	35	60	60

	X	Y	Z
S1	-53	53	3
S2	12	-34	-3
S3	-41	-1	-47
S4	27	-15	-63
S5	-97	-5	4
S6	-74	-83	75
S7	48	-55	38
S8	-36	-65	-4
S9	24	-25	-2

	X	Y	Z
S10	-39	-9	17
S11	-46	44	63
S12	-38	-3	5
S13	1	5	72
S14	9	54	57
S15	-18	-64	31
S16	177	44	-152
S17	136	94	-166
S18	-38	-27	61
S19	25	12	29
S20	35	17	-23
S21	-15	64	5

Table 2. Inter-protein distance estimates used for final analysis by MDSCAL (left), in Angstroms. Three-dimensional coordinates of model as output by program (upper right), in relative units.

data than on the large fragment and other less precise data. The stress of the final output is 0.223, a considerably improved value. The final input distances and the output coordinates of the proteins appear in Table 2, and Fig. 1a depicts a model built to these specifications.

Due to the paucity of consistent data involving S16 or S17, these appear anomalously far from the other proteins in the model and we must consider that their positions remain undetermined. We may validate the remainder of the model in two ways, first, by seeing how well it conserves the proximity relationships implicit in the experimental results on which we based our input, and second, by comparison with the assembly map and other independent results.

The model conserves the crosslinking data perfectly in that no third protein intervenes in direct line between any of S1-S12, S5-S8, S2-S3, S7-S9, S13-S19, S11-S18, S18-S21, and S11-S21. On the other hand other crosslinkage data (6), namely S11-S19 and S5-S9, which were not included in our input, are inconsistent with the model, unless markedly non-globular conformations are postulated for some of the proteins involved.

The nuclease fragmentation results also agree well with the model. The proteins S7, S9, S10, S13, S14, and S19 are neighbors as is expected from their occurence on a single RNA fragment. Similarly S8 and S15 which co-occur on a small RNA fragment are neighbors in the model.

The protein S12, gene product of the Str A locus, is surrounded in our model by S1, S3, S5, S10, S11, S18, S19, S20 and S21. It has thus a rather central position which could explain why the modification or the blocking of any of these surrounding proteins (and mainly these) distorts the drug receptor site and affects the binding of streptomycin.

Proteins S2, S5, S6, S12 and S13 are thought to compose the P (peptidyl-tRNA) site, and these can be seen to constitute a loose sequence of neighbors in the model. The A (aminoacyl-tRNA) site is thought to be made up of S8, S9, S11 and S18, which are all to be found on one side of the model. The relative

proximity of the two groups suggests a partial overlapping of the two sites.

With respect to independent results, we may list the following:

a) S1, S14 and S21 may well be neighbors since these three proteins play an important role in the binding of aminoacyl-tRNA to the A site (7). The model confirms this proposal.

b) All antibodies for ribosomal proteins react with their cognate antigen on the 30S subunit (8), suggesting a flat or open structure for the subunit. The same conclusion was reached by Miller and Sypherd (9) in iodination studies of the ribosome. Although these results do not take into account the rate of reaction for each protein, our model shows all proteins to be rather accessible with the possible exceptions of S12 and S10 which are partially masked by S1 and S8-S6 respectively. Another study based on competitive labelling of ribosomal proteins is consistent with our model expecially with regards to the masking of S12 (10).

c) Cantor et al. (11), using fluorescent probes, reported recently that the distance between S20 and S4 is 40-60 Å, center to center. This estimate fits nicely with the value of 45 Å in the model (as measured in proportion to various crosslinkage distances).

d) Surface topography data obtained by Huang and Cantor (12) indicate that S1, S3, S4 and S21 are exposed in the 30S subunit and the 70S ribosome, while S5, S9, S12, S13, S14, and S19 are strongly shielded by the 50S particle. These two groups of proteins fall indeed into two layers in the model and their location indicates the top and bottom of the 30S particle with respect to the 50S subunit.

e) The assembly map gives rise to a number of proximity hypotheses. In many cases, we find agreement between the temporal assembly and our model, suggesting that the relative position of some proteins does not change drastically during assembly. For instance, the map suggests that S6, S11, S15 and S18 are topologically related and this fits perfectly with our model. Furthermore, the map indicates relationships between S7, S9 and S19 which agree well with

the model. The same holds true for S3, S5 and S21 as well as for S4, S8 and S20. There do appear to be some exceptions: S3 is not very close to S20, suggesting relocation of these proteins during assembly of the subunit.

f) The fact that S1 and S21 are close neighbors on our model correlates well with the fact that S1 can be crosslinked to the 3' end of the 16S RNA (13) and that S21 is missing in Colicin-treated 30S particles where 16S RNA lacks the 3' terminus (14). The 3' terminus of RNA is also known to be associated with several other proteins: S7, S9, S10, S13, S14 and S19 and the 5' end to another set: S4, S8, S15, and S20 (15). On the model the proteins in each group constitute a sequence of neighbors.

g) The overall dimensions suggested by low angle X-ray scattering of the 30S particles are 220 × 220 × 56 Å (16). The structure we propose (Fig. 1a) has no such thin dimension (56 Å). However, there is a grouping of most of the proteins in the vertical plane perpendicular to the plane of the photograph. By shifting S3, S6, S4, S11 and S20 towards this plane, we produce an alternate model (Fig. 1b) which satisfactorily fits the X-ray data, and conserves the proximity relations of the original model, although with greater overall stress. Assuming the average distance between crosslinked pairs to be about 45 Å (sum of the radius and length of the crosslinker) we calculate that the dimensions in our "squeezed" model (Fig. 1b) would be roughly 200 × 200 × 60 Å. Each ball is thus approximately one half the average protein diameter of 36 Å.

h) Morgan and Brimacombe (17) have proposed a model including S16 and S17 but not S1, S2 nor S12, based partly on the assembly map, but before most of the crosslinkage data became available. There is no major contradiction between their proposal and our model except for the pairs S5-S9, which they show to be rather close, and for the relative position of the group S8, S10, and S15. More detailed comparison is not possible since coordinates are not available for their model†.

† Another 30S ribosomal subunit model is being proposed by T.T. Sun, R. Heimark and R.R. Traut.

i) Crosslinking results (18) which appeared after our calculations were completed, namely the formation of a S6-S14-S18 trimer and a complex of S6-S13-S18-S19 all correspond very well with our model.

Several new pairs of proteins are close in our model; for example, S11-S14, S13-S14, S4-S20 and S1-S21. It remains for new proximity data to confirm or reject these predictions. It is a feature of our analysis by MDSCAL that without much effort new data can be routinely incorporated in the input to produce progressively better and testable three-dimensional structures of the 30S ribosomal subunit.

ACKNOWLEDGEMENT: - We thank Dr. J. Beaudoin for his photographs of the model. This work was carried out under contract, EURATOM-ULB 099-72-IBIAB with an agreement between the Belgium Government and U.L.B. and by a FCAC grant of the Ministère de l'Education of Quebec. A.B. is a chercheur qualifié of the Fonds National de la Recherche Scientifique of Belgium.

References

1. M. Nomura, Bact. Rev. 34, 228 (1970).
2. C.G. Kurland, Ann. Rev. Biochem. 41, 377 (1972).
3. S. Mizushima and M. Nomura, Nature (Lond.) 226, 1214 (1970).
4. J.B. Kruskal, Psychometrika 29, 1 (1964).
5. J.B. Kruskal, Psychometrika 29, 115 (1964).
6. P. Voynow and C.G. Kurland, Biochemistry 10, 517 (1971).
7. I. Ginzburg, R. Miskin, A. Zamir, J. Mol. Biol. 79, 481 (1973).
8. G. Stöffler, R. Hasenbank, M. Lütgehaus, R. Maschler, C.E. Morrison, H. Zeichardt and R.O. Jarret, Mol. Gen. Genet. 127, 89 (1973).
9. R.V. Miller and P.S. Sypherd, J. Mol. Biol. 78, 539 (1973).
10. L.P. Visentin, M. Yaguchi, H. Kaplan, Can. J. Biochem. 51, 1487 (1973).
11. C.R. Cantor, K.H. Huang, M.R. Fairclogh in The Ribosome, P. Lengyel, M. Nomura, A. Tissières, Ed. (Cold Spring Harbor Press, 1974). In preparation.
12. K. Huang and C.R. Cantor, J. Mol. Biol. 67, 265 (1972).
13. R.A. Kenner, Biochem. Biophys. Res. Comm. 51, 932 (1973).
14. C.M. Bowman, J.E. Dahlberg, T. Ikemura, J. Konisky, M. Nomura, Proc. Natl. Acad. Sci. 68, 964 (1971).
15. R.A. Zimmerman, A. Muto, P. Fellner, C. Ehresmann, C. Branlant, Proc. Natl. Acad. Sci. 69, 1282 (1972).
16. W.S. Smith, Ph.D. Thesis, University of Wisconsin (1971).
17. J. Morgan and R. Brimacombe, Eur. J. Biochem. 37, 472 (1973).
18. C. Clegg and D. Hayes, Eur. J. Biochem. 42, 21 (1974).
19. R. Heimarck and R.R. Traut, personal communication.
20. T.T. Sun, A. Bollen, L. Kahan, R.R. Traut, Biochemistry, in press.
21. T.A. Bickle, J.W.B. Hershey, R.R. Traut, Proc. Natl. Acad. Sci. 69, 1327 (1972).
22. L.C. Lutter, H. Zeichardt, C.G. Kurland, Mol. Gen. Genet. 119, 357 (1972).
23. C.T. Shih and C.H. Craven, J. Mol. Biol. 78, 651 (1973).
24. A. Bollen, L. Kahan, A. Cozzone, J.W.B. Hershey, R.R. Traut, personal communication.

25. R. Heimarck and R.R. Traut, personal communication.
26. O. Pongs and V.A. Erdmann, FEBS Letters 38, 47 (1973).
27. R. Zimmerman, personal communication.
28. F.N. Chang and J.G. Flaks, Proc. Natl. Acad. Sci. 67, 1321 (1970).
29. A. Bollen, R. Maschler, G. Stöffler, personal communication.
30. D.P. Rummel and H.F. Noller, Nature New Biology 245, 72 (1973).
31. G. Schreiner and K.H. Nierhaus, J. Mol. Biol. 81, 71 (1973).
32. D. Apirion, G. Stöffler, E. Deusser, H.G. Wittmann, Mol. Gen. Genet. 111, 334 (1971).
33. R.A. Zimmerman, R.T. Gauvin, L. Gorini, Proc. Natl. Acad. Sci. 68, 2263 (1971).
34. M. Kuwano, H. Endo, Y. Ohnishi, J. Bact. 97, 940 (1969).
35. H.E. Roth and K.H. Nierhaus, FEBS Letters 31, 35 (1973).
36. M. Nomura, S. Mizushima, M. Ozaki, P. Traub, C.V. Lowry, Cold Spring Harbor Symp. of Quant. Biol. 34, 49 (1969).
37. G. Craven, personal communication.
38. J.C. Lelong, D. Gros, F. Gros, A. Bollen, R. Maschler, G. Stöffler, Proc. Natl. Acad. Sci. 71 (1974).
39. A. Bollen, R. Maschler, G. Stöffler, personal communication.
40. P. Thammana, personal communication.
41. L.L. Randall-Hazelbauer and C.G. Kurland, Mol. Gen. Genet. 115, 234 (1972).