Journal of Supramolecular Structure 7:191–204 (1977)

Prediction for Secondary Structures of Ten Proteins From the 50S Subunit of the Escherichia coli Ribosome

Michael Dzionara, Stephen Michael Lewis Robinson, and Brigitte Wittmann-Liebold

Max-Planck-Institute für Molekulare Genetik, Abteilung Wittmann, Berlin-Dahlem, Federal Republic of Germany

Predictions of the secondary structures of the following 10 proteins from the large subunit of the E. coli ribosome were made using their known amino acid sequences: L6, L16, L19, L27, L28, L30, L31, L32, L33, and L34. The predictions were made according to 4 different methods and the results for each protein are presented as diagrams indicating the conformational states, helix, extended structure, turn, and random coil, of each residue. From these diagrams, regions of highly probable secondary structure for the proteins are calculated. Estimates are made of the maximum possible lengths of the proteins in order to correlate these with the results obtained from antibody binding sites in the 50S subunit as determined by electron microscopy.

Key words: ribosomes, proteins, predictions, secondary structures, topography

In previous publications, we presented predictions for the secondary structures of 11 proteins from the small subunit (1, 2) and 6 proteins from the large subunit (3) of the E. coli ribosome. In this paper, we present predictions for the secondary structures of 10 further proteins from the large ribosomal subunit whose sequences are known. In addition, we summarize the results of this series of studies on the ribosomal protein secondary structures.

In the absence of detailed x-ray analyses of these proteins, predictive methods were used in order to provide structural information which can be related to other findings on the topography of the ribosomal subunits, especially those obtained by immunoelectron microscopy (see 4). In this context, predictions offer a means of assigning highly probable conformations to proteins of known primary structure. These may be used as a starting point for energy minimization calculations and as guidelines to the complex interactions of proteins with each other and with the ribosomal RNA.

A more reliable picture of predicted protein conformations should be obtainable when the results of several predictive methods are taken together. In this study, we have used 4 different predictive methods to calculate the 4 conformational states, helix, turn or bend, extended structure, and random coil of the individual amino acids in the ribosomal

Received August 25, 1977; accepted August 25, 1977

proteins of known sequence. Histograms were constructed for each of these states and regions of highly probable structure were derived from these histograms when 3 predictive models agreed about the conformational state of a residue (2).

METHODS

The secondary structures of the ribosomal proteins were predicted according to the methods of Burgess et al. (5), of Chou and Fasman (6, 7) and Chou et al. (8), of Nagano (9), and of Robson and Suzuki (10). FORTRAN programs for the calculation of the predictive algorithms were supplied by Drs. H.A. Scheraga, K. Nagano, and B. Robson. The results of the calculations were treated as described in Refs. 1 and 2, where more detailed information is given. However, in this study helices predicted for less than 4 adjacent residues were neglected in the results of Robson's and Nagano's programs. To these residues the state "random coil" was assigned.

Where modified amino acids were encountered in the sequences (e.g., in L12, L16, and L33), they were treated as their nearest common amino acid, as indicated in the figures.

RESULTS

The secondary structures predicted by 4 methods for the ribosomal proteins L6, L16, L19, L27, L28, L30, L31, L32, L33, and L34 are shown in Fig. 1–10. In these figures, the predicted conformational states of the residues are represented symbolically, as indicated. The definitions of helix, turn or bend, extended structure, or β sheet have been interpreted in their most general sense, in order to overcome differences in definitions given by the authors of the programs.

With the exception of protein L30, the predictions of the different models generally corresponded well. The most probable secondary structure, obtained when 3 predictions are in agreement, is given in the figures in the line "PRE." The information so derived serves as a guideline to an expected structural feature and its position. It may involve further residues in either direction in the native folded-protein chain.

Protein L6

The predictions are based on the sequence determined by Chen et al. (11) and are illustrated in Fig. 1. The secondary structure of protein L6 is quite distinctive in appearance. Most outstanding are the following structural features: 1) The sequence of almost exclusively extended residues in the N-terminal end of the protein chain (positions, 7–10, 14–18/20, and 21–25). 2) The 2 elongated regions of turns in positions 51–60 and in the C-terminus (positions 152–160), each located between 2 helices. 3) The frequency of numerous other short turns along the protein chain. 4) The extensive region between the residues 88 and 141 for which a variety of secondary structures is predicted.

The accumulation of extended structure predicted in the N-terminal region of protein L6 is unique among the ribosomal proteins studied to date. This may indicate a quite specific function for this part of the molecule. There is evidence for only one further extended region in this protein, i.e., in positions 75/77–79. In total, L6 is predicted to contain approximately 14% extended structure. This value is significantly above the average of calculated extended structure in the ribosomal proteins. Other proteins with a high content of predicted extended structure are proteins L28 (see below and Table I) and S12

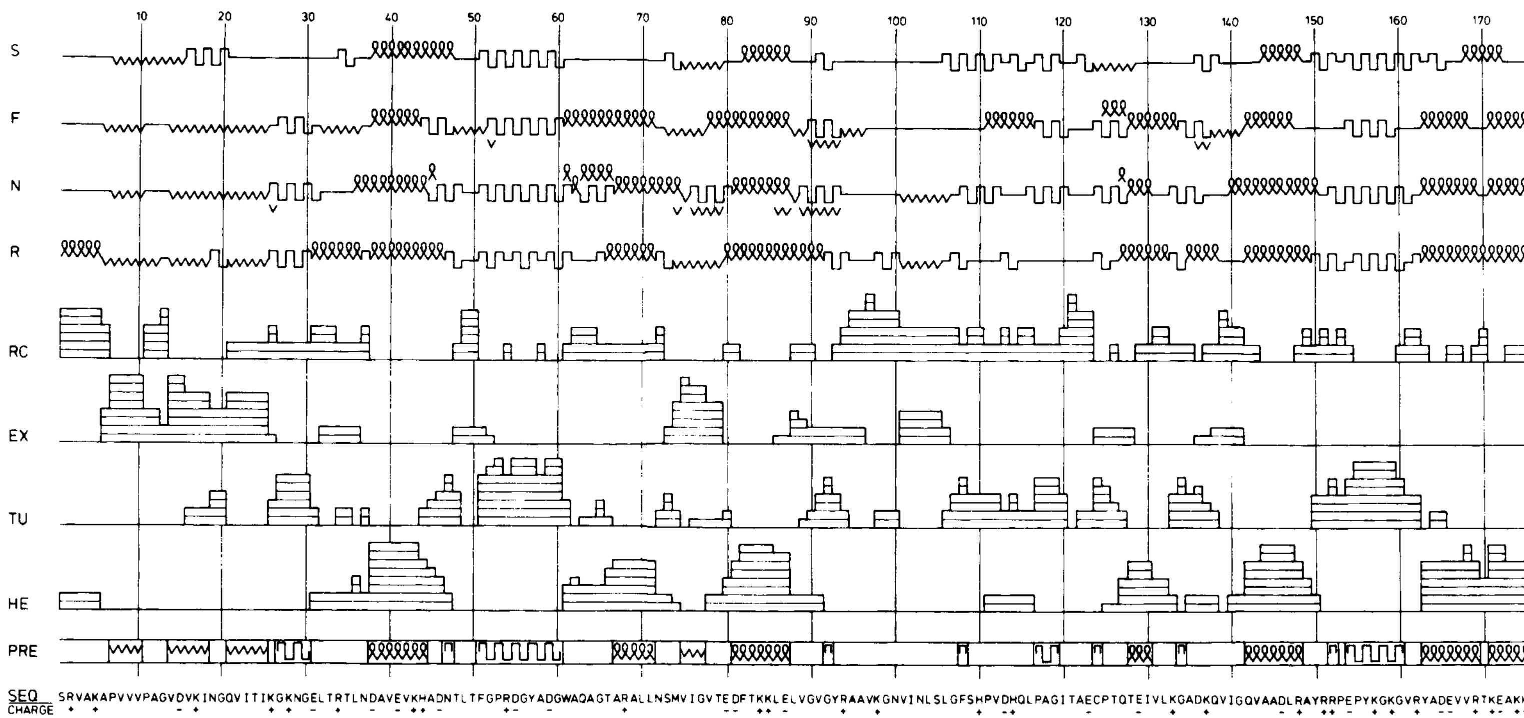

Fig. 1. Secondary structures predicted for protein L6 according to 4 different methods: S) Burgess et al. (5); F) Chou and Fasman (6, 7) and Chou et al. (8); N) Nagano (9); R) Robson and Suzuki (10). The symbols define residues in: Helix (ℓℓ), extended (⋀⋀⋀), turn or loop (⎴⎵) and random (——) conformations. In cases of ambiguity in the predictions, both conformation states are shown. The histograms represent the sum of residues in random coil (RC), extended structure (EX), turn or loop (TU) and helix (HE) conformations (see Methods). A most probable conformational state is assigned to each residue when 3 predictive methods agree; this is shown in the line denoted by "PRE." The amino acid sequence is given in the lowermost line, "SEQ."

TABLE I. Average Values of Secondary Structure Content and Maximum Lengths of Ribosomal L Proteins Based on at Least Three Corresponding Predictions

				Helix			Turn or bend			Extended			Random coil and not predictable				
Protein	Res[a]	Shape[b]	Binding sites[c]	Res	%[d]	Length[e] (Å)	Res	%[d]	Length[f] (Å)	Res	%[d]	Length[g] (Å)	Res	%[d]	Length[h] (Å)	Maximum length[i] (Å)
L6	176	E	S,S	42	24	63	30	17	53	17	10	59	87	49	313	488
L16	136	–	S	28	27	42	33	24	58	12	9	42	63	46	227	369
L19	114	–	–	29	25	44	13	11	23	10	9	35	62	54	223	325
L27	84	–	S	8	10	12	25	30	44	5	6	17	46	55	166	239
L28	77	–	–	17	22	26	12	16	21	11	14	38	37	48	133	218
L30	58	–	–	NC[j]	–	–	1	2	2	5	9	17	52	90	187	206
L31	62	–	–	NC[j]	–	–	15	24	26	3	5	10	44	71	158	194
L32	56	–	–	9	16	14	19	34	33	1	2	4	27	48	97	148
L33	54	–	–	18	33	27	8	15	14	3	6	10	25	46	90	141
L34	46	–	–	8	17	12	11	24	19	NC[j]	–	–	27	59	97	128

[a]Res: number of amino acid residues.
[b]Shape: E, elongated.
[c]Locations and number of antibody binding sites in the 50S subunit:S, seat region.
[d]Percentage of residues in a given conformational state.
[e]1.5 Å per residue in helix conformation (assumed to be α helix).
[f]1.75 Å per residue in turn conformation.
[g]3.47 Å per residue in extended conformation.
[h]3.6 Å per residue in random coil.
[i]Summation of the respective lengths in the different conformations (assumes a stretched form of the protein chain).
[j]NC, not calculated according to the above assumption.

(1). In the case of protein S12 the extended structures are more uniformly distributed throughout the protein chain. The occurrence of turns among these suggests that pleated-sheet formation in protein S12 is more likely than in protein L6.

Protein L6 is identified with the aminoacyl-tRNA binding site of the E. coli ribosome (reviewed in Ref. 12). Furthermore, it is located on the lip of the "seat" area of the 50S subunit and has been shown to have an elongated conformation in situ (reviewed in Ref. 4). Hydrodynamic studies, on the other hand, indicate that the isolated protein L6 is only slightly elongated (13). With a maximum length of 488 Å (Table I), protein L6 can be readily accommodated within the 200-Å distance separating the individual sites of antibody attachment in the 50S subunit. Further, it may be speculated that the 2 prominent regions of turns, 51–60 and 151/152–160, and their adjacent helices are situated at or near the observed antibody-binding sites (cf. Ref. 1). At present, investigations with antibodies raised to specific fragments of L6 are in progress to assist in its orientation in the large subunit and to test this hypothesis.

Protein L16

The predictions are based on the sequence determined by Brosius and Chen (14) and are presented in Fig. 2. In the N-terminal region of the protein chain, several turns are indicated (positions 5/6, 13/14–19, 22–24, 28–32, and 37–40). Besides these structures, 2 models predict short helices in this region (see Fig. 2). The central part of the molecule is characterized by 1 helix of some 13–18 residues in length (positions 45–57) and by an extended region (positions 63–67) which is followed by a strongly predicted turn centred on residue 70. The C-terminal part of the molecule is predicted to have a variety of structures. Of these, the turns in positions 83–88, 98–99, and 106–107 are predominant. At the C-terminus of protein L16, a long helix (positions 110–124) and extended structure (129–135/136) are given. These 2 regions may be separated by a turn at about position 126. This secondary structure at the C-terminus (110–136) seems to repeat that of the central part (45–67) of the molecule.

Protein L16 has been identified as a 23S RNA-binding protein (15) and is located at the peptidyltransferase center of the ribosome (reviewed in Ref. 12). Corresponding with this function, one antibody-binding site has been located for protein L16 in the upper region of the "seat" of the 50S subunit (4).

Protein L19

The predictions are based on the sequence determined by Brosius and Arfsten (16) and are shown in Fig. 3. In the N-terminal third of the protein chain a helix of some 10 residues and 2 turns at about positions 17 and 21 are given. The following stretch (24–36) is not reliably predicted and is designated as random coil. The models indicate that extended structure is possible for this region. The central part of the molecule is predicted to have a short helix (positions 37–43, extended structure (44–47) and a turn in position 50–53. Further extended structure is given for residues 57/59–62/64. It is possible that the 2 regions of extended structure are involved in pleated-sheet formation with each other. The protein chain passes through another turn in position 65–66. Various conformations are predicted for the region 67–88, among which is again extended structure. The C-terminal third of the molecule is indicated to contain 2 short helices (positions 90–95/96 and 106–111) connected by a turn in positions 103–104.

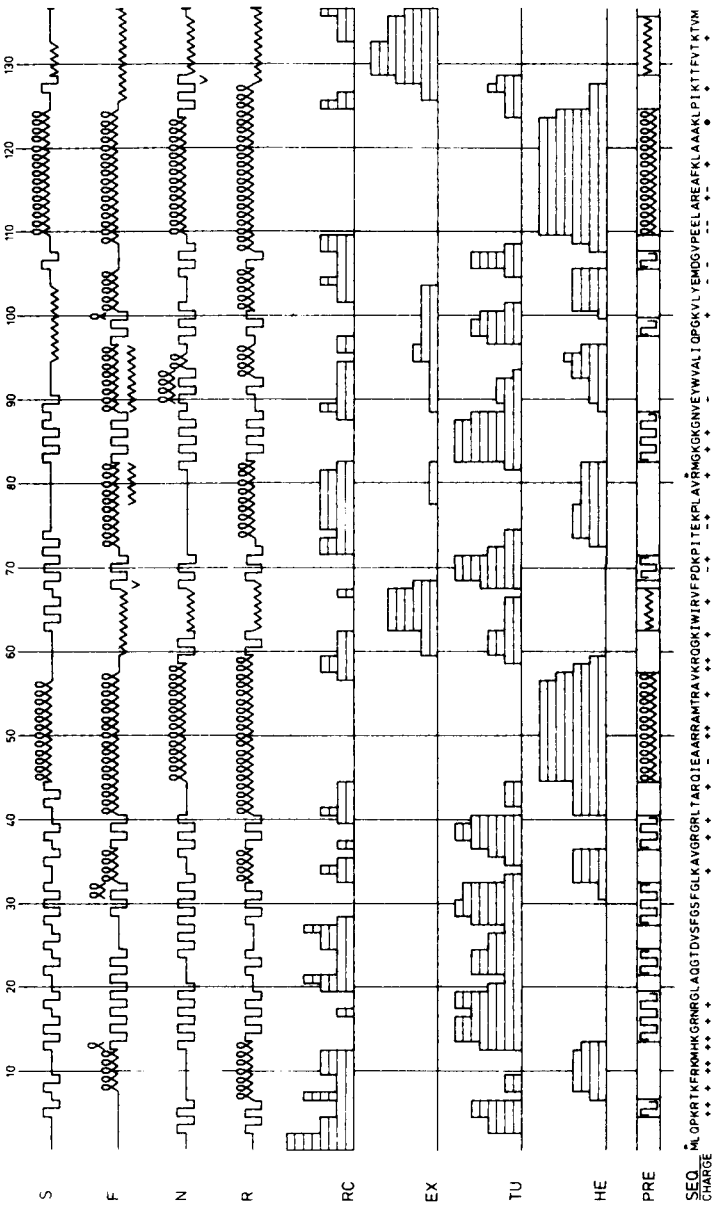

Fig. 2. Predicted secondary structure of protein L16. Abbreviations as for Fig. 1. M) N-monomethylmethionine treated as methionine; R) unknown basic amino acid derivative treated as arginine.

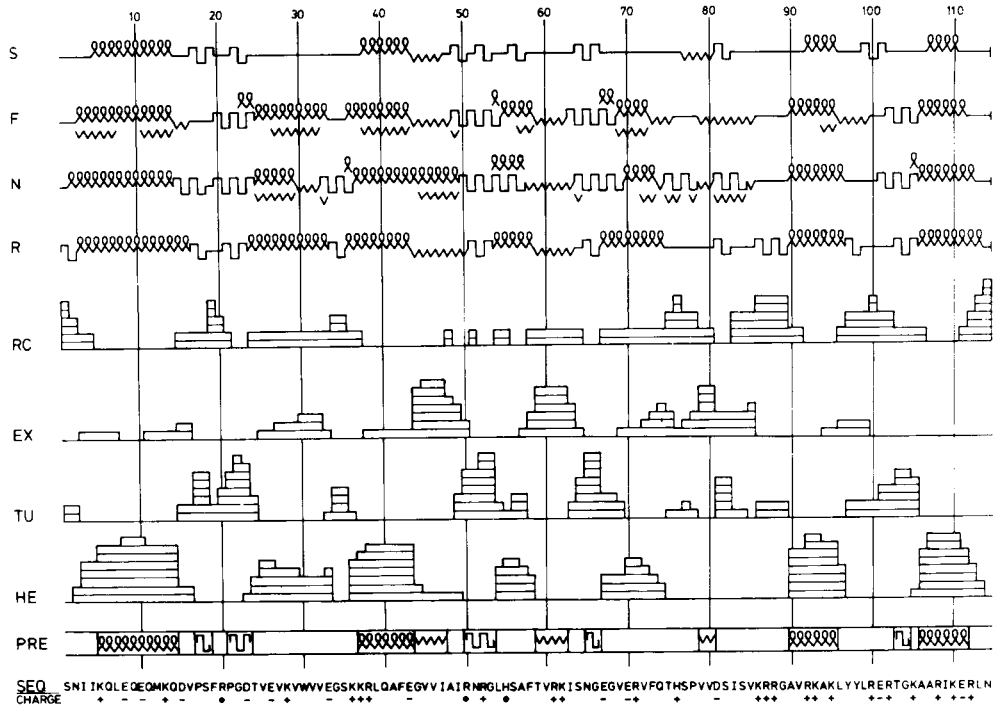

Fig. 3. Predicted secondary structure of protein L19. Abbreviations as for Fig. 1.

Protein L27

The predictions are based on the sequence determined by Chen et al. (17) and are shown in Fig. 4. The N-terminal 5 residues are equivocally predicted, both helix and random structure being equally possible. From positions 6–14 a region of turns is indicated. This is succeeded by a short helix of some 3–8 residues and a turn at about position 26. Extended structure is given for the sequence, Gly-Ser-Ile-Ile-Val-Arg in the middle of the L27 protein chain. The C-terminal half of the molecule contains 4 regions of turns (40–42, 51–54, 63–65, and 71–75). Another very short helix is indicated, centered at about position 60. The helix content of this L protein is predicted to be very low (Table I).

One site of antibody attachment has been found for protein L27 in the "seat" region of the 50S subunit (4). By affinity labeling experiments it was found that protein L27 belongs to the group of proteins which are very close to the peptidyl-tRNA-binding site (reviewed in Ref. 12).

Protein L28

The predictions are based on the sequence determined by Wittmann-Liebold and Marzinzig (18) and are presented in Fig 5. Three to 6 residues at the N-terminus are expected to be in extended structure. Between residues 7 and 17 is a region of 1 or 2 turns. Three short helices (positions 19–23, 39–44, and 67–72) are predicted to be separated by an approximately equal number of residues. A single and quite pronounced turn is found at position 30, where a proline residue occurs in the sequence. High probabilities

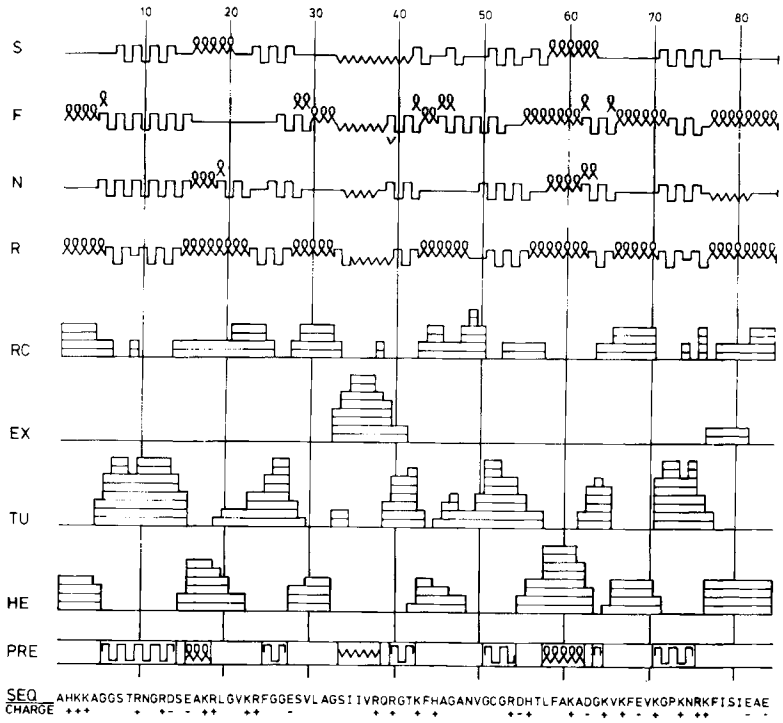

Fig. 4. Predicted secondary structure of protein L27. Abbreviations as for Fig. 1.

for extended structure are given for positions 46–50 and 56–58. These 2 regions are separated by a strongly predicted turn at about position 53 which could allow them to form pleated-sheet structure. Protein L28 has an above average content of extended structure for the ribosomal proteins so far investigated (Table I and see Refs., 1–3).

Protein L30

The predictions are based on the sequence determined by Ritter and Wittmann-Liebold (19) and are shown in Fig. 6. For protein L30, the different predictive methods give a variety of secondary structures which agree very poorly and in this case the derivation of a highly probable secondary structure may be misleading. Only 1 region of extended structure (positions 3–7) and a turn at about position 32 can be assigned on the basis of at least 3 models in agreement. Helix structure is weakly and ambiguously predicted for 3 regions (see Fig. 6) but the correspondence between the models is not sufficient for an assignment of any confidence. Hence, protein L30 seems to contain little or no helical structure (Table I). Similarly, other regions of structure are not well emphasized.

Protein L31

The predictions are based on the sequence determined by Brosius (20) and are shown in Fig. 7. This protein contains 4 cysteine residues at positions 16, 18, 37, and 40. Apparently, protein L31 contains no helical structures; short regions of helix are given by only 1 model (see Fig. 7). The N-terminal region up to residue 20 is given essentially as coil structure with turns predicted for positions 7–8 and about residue 17. The sequence Cys-Ser-Cys is found in the second of these turns. Similarly, the sequence of the other

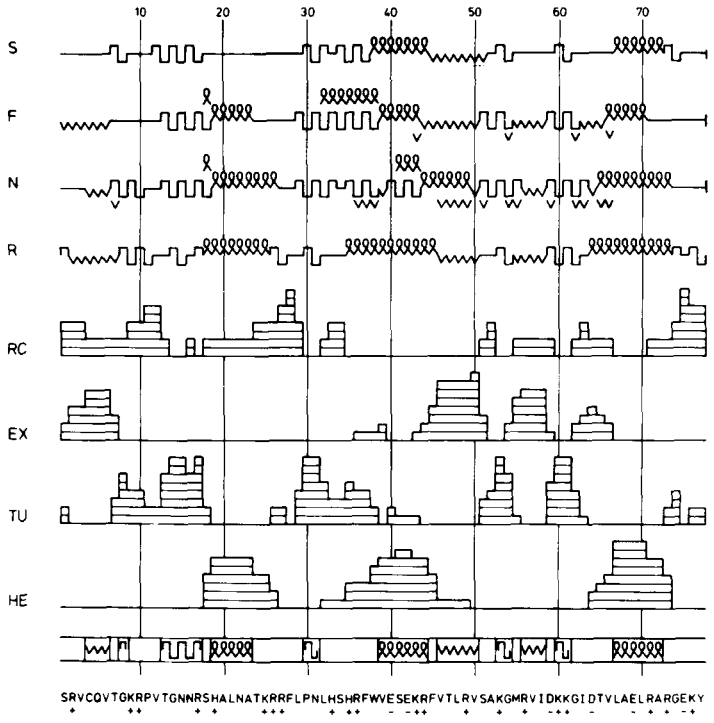

Fig. 5. Predicted secondary structure of protein L28. Abbreviations as for Fig. 1.

cysteine residues, Cys-Ser-Lys-Cys, is also predicted to occur in a turn. Further turns are given in positions 41–42 and in the C-terminus of the molecule (53–56 and 58–61).

Extended structure is given by the different models in the region 12–37 and is predicted for positions 21–24. Pleated-sheet formation may be possible in this protein.

Protein L32

The predictions for protein L32 are based on the sequence determined by Wittmann-Liebold et al. (21) and are shown in Fig. 8. The N-terminus is marked by a region of 2 turns, 6–7 and 9–13, followed by a short, poorly developed helix, 17/18–21/25. Extended structure is given by 2 models for the stretch of residues 23/24–29/30. A turn in positions 30–34, a helix of some 6–10 residues, and another series of turns (45–51) are obtained for the C-terminal region of protein L32. This protein has the highest content of turns (34%) calculated for all the ribosomal proteins so far studied. A similarly high turn content is shown only by protein L27 (30%, see Table I).

Protein L33

The predictions are based on the sequence determined by Wittmann-Liebold and Pannenbecker (22) and are presented in Fig. 9. Three helices, each comprising 6 residues are given for positions 5–10 in the N-terminal region, for 31–36 in the middle part, and/or in 49–54. Protein L33 is predicted to have a moderate content of helix (Table I). Turns in positions 13–15 and 24–27 appear to be separated by 3–6 residues in an extended

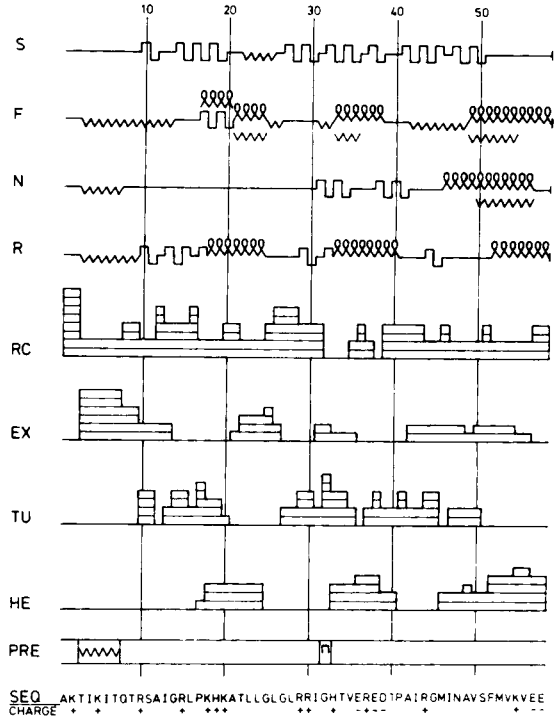

Fig. 6. Predicted secondary structure of protein L30. Abbreviations as for Fig. 1.

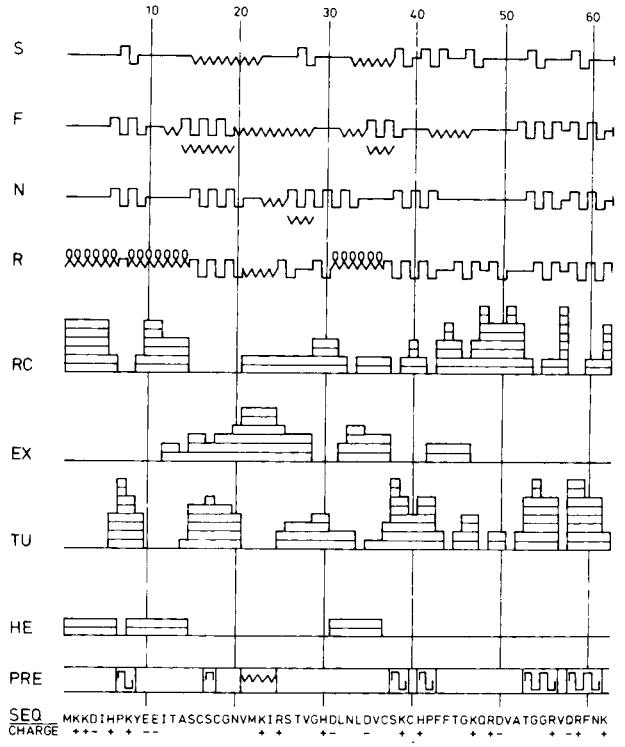

Fig. 7. Predicted secondary structure of protein L31. Abbreviations as for Fig. 1.

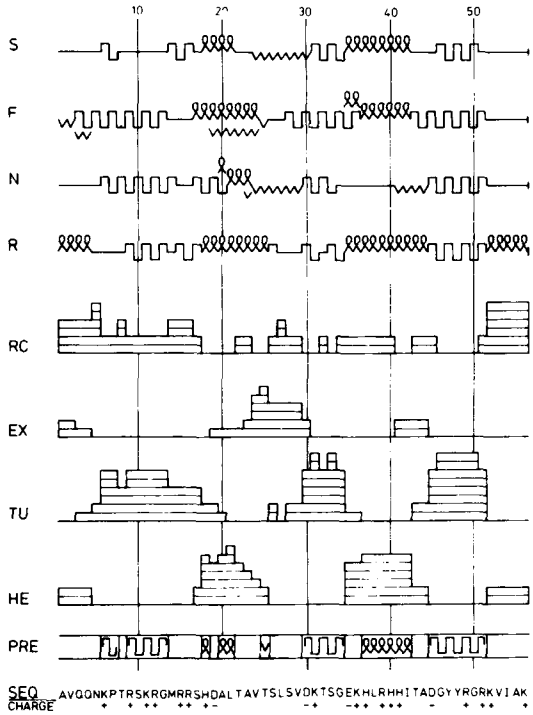

Fig. 8. Predicted secondary structure of protein L32. Abbreviations as for Fig. 1.

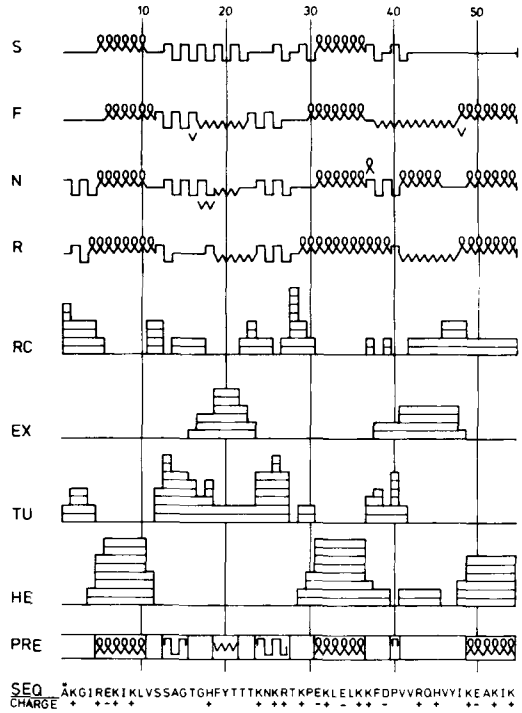

Fig. 9. Predicted secondary structure of protein L33. Abbreviations as for Fig. 1. *Å) N-monomethyl-alanine treated as alanine.

conformation. The helix in the middle part of the protein chain is terminated by a turn at position 40. Two models predict this to be followed by further extended structure (see Fig. 9).

Protein L34

Predictions for the secondary structures of protein L34 are based on the sequence determined by Chen and Ehrke (23) and Chen (24). They are shown in Fig. 10. Various structures are indicated for the first 12 residues. A region of turns is given for positions 13–17/19 followed by a helix of about 4 residues. In the center of the protein chain a strongly predicted turn is given at about position 26. The C-terminal region contains another short helix (32–36) and a pronounced turn in positions 37–38. As for the N-terminus, various structures are predicted for the last 8 residues. In both cases, at least 2 models suggest that these regions could be in an extended conformation. However, on the basis of 3 agreeing predictions protein L34 seems to lack any developed extended structure (Table 1).

CONCLUSIONS

In general, the secondary structure predicted by the 4 different methods agreed well. This result encouraged us to study all the ribosomal proteins of known primary structure (1–3). There is no evidence so far for homologous sequences among the E. coli ribosomal proteins contributing to homologous regions of secondary structure.

In this paper, we have focused our interest on the ribosomal L proteins which were predicted to have low to moderate helical content (up to 30%; see Table I). Seven of these proteins are comparatively short in sequence and, in these, structural features appear to be less well developed. Helices exceeding 6–8 residues are infrequently predicted. This may be compared with other small proteins whose crystal structures are known, e.g., insulin, bovine pancreatic trypsin inhibitor, and ribonuclease S, in which helices of more than 10 residues are seldom found. In some proteins, for example L27, L30, and L31, the helical content is predicted as extremely low. In particular, for protein L30 this may be unrepresentative because the predictions are in each case rather individual and themselves ambiguous.

The ribosomal proteins have, as a special class of proteins, been predicted to contain relatively little extended structure (see Refs. 1–3) and this may be attributed to underprediction. However, experimental observations of intact ribosomes with infrared spectroscopy indicate that in fact the proteins have little extended structure (25).

Maximum lengths, based on the most probable secondary structure, were calculated for all the ribosomal proteins studied. These are especially useful in testing an elongated conformation proposed when multiple antibody-binding sites have been found for a small protein in the ribosome. For these proteins, only a rather stretched conformation would most satisfactorily explain their accommodation, assuming they are present in unimolecular stoichiometry. Therefore, it may be expected that they have a high proportion of coil or elongated structure in nature. This is as predicted, e.g., for S15 and S18 (1). In connection with specific antibody-binding studies to protein fragments, such information may indicate not only the orientation of a protein but also its manner of spatial accommodation. Further, a high proportion of helical as well as turn structure was predicted for certain proteins whose globular shape has been determined experimentally, e.g., S8 (13, 26).

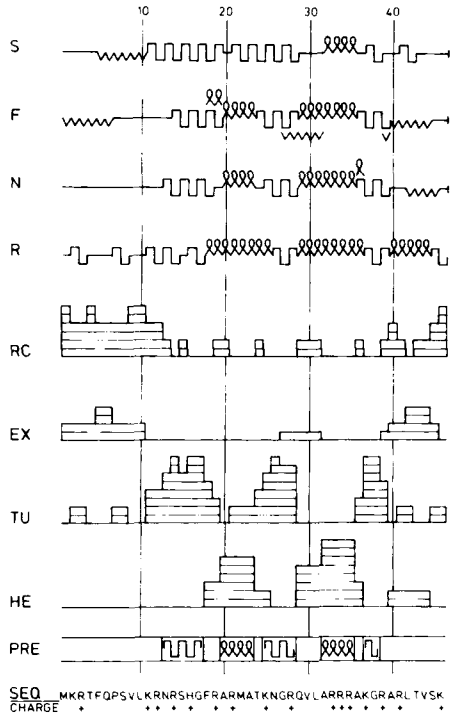

Fig. 10. Predicted secondary structure of protein L34. Abbreviations as for Fig. 1.

As has been discussed for mutant proteins of the 30S ribosomal particle (1), it will be worthwhile to examine the predicted structural changes induced in mutants of the L proteins as soon as the alterations in the mutant proteins become available.

ACKNOWLEDGMENTS

We are indebted to Drs. H.A. Scheraga, K. Nagano, and B. Robson for kindly providing us with their computer programs. We express our thanks to Mrs. Eva Philippi for her careful drawing of the diagrams.

REFERENCES

1. Dzionara M, Robinson SML, Wittmann-Liebold B: Hoppe-Seyler's Z Physiol Chem 358:1003–1019, 1977.
2. Wittmann-Liebold B, Robinson SML, Dzionara M: FEBS Lett 77:301–307, 1977.
3. Wittmann-Liebold B, Robinson SML, Dzionara M: FEBS Lett 81:204–213, 1977.
4. Stöffler G, Wittmann HG: In Weissbach H, Pestka S (eds): "Protein Synthesis," London and New York: Academic Press, 1977, pp 117–202.
5. Burgess AW, Ponnuswamy PK, Scheraga HA: Israel J Chem 12:239–286, 1974.
6. Chou PY, Fasman GD: Biochemistry 18:211–222, 1974.
7. Chou PY, Fasman GD: Biochemistry 18:222–249, 1974.
8. Chou PY, Adler AJ, Fasman GD: J Mol Biol 96:29–45, 1975.
9. Nagano K: J Mol Biol 109:251–274, 1977.
10. Robson B, Suzuki E; J Mol Biol 107:327–356, 1976.

11. Chen R, Arfsten U, Chen-Schmeisser U: Hoppe-Seyler's Z Physiol Chem 358:531–535, 1977.
12. Brimacombe R, Nierhaus KH, Garrett RA, Wittmann HG: Prog Nucleic Acid Res 18:1–324, 1977.
13. Giri L, Littlechild J, Dijk J: FEBS Lett 79:238–244, 1977.
14. Brosius J, Chen R: FEBS Lett 68:105–109, 1976.
15. Stöffler G, Daya L, Rak KH, Garrett RA: Mol Gen Genet 114:125–133, 1971.
16. Brosius J, Arfsten U: Manuscript submitted.
17. Chen R, Mende L, Arfsten U: FEBS Lett 59:96–99, 1975.
18. Wittmann-Liebold B, Marzinzig E: FEBS Lett 81:214–217, 1977.
19. Ritter E, Wittmann-Liebold B: FEBS Lett 60:153–155, 1975.
20. Brosius J: Manuscript submitted.
21. Wittmann-Liebold B, Greuer B, Pannenbecker R: Hoppe-Seyler's Z Physiol Chem 356:1977–1979, 1975.
22. Wittmann-Liebold B, Pannenbecker R: FEBS Lett 68:115–118, 1976.
23. Chen R, Ehrke G: FEBS Lett 63:215–217, 1976.
24. Chen R: Hoppe-Seyler's Z Physiol Chem 357:873–886, 1976.
25. Cotter RJ, Gratzer WB: Eur J Biochem 8:352–356, 1969.
26. Engelman DM, Moore PB, Schoenborn BP: Proc Natl Acad Sci USA 72:3888–3892, 1975.