# PREDICTIONS FOR SECONDARY STRUCTURES OF SIX PROTEINS FROM THE 50 S SUBUNIT OF THE *ESCHERICHIA COLI* RIBOSOME

B. WITTMANN-LIEBOLD, S. M. L. ROBINSON and M. DZIONARA
*Max-Planck-Institut für Molekulare Genetik, Abt. Wittmann, 1 Berlin 33 (Dahlem), Germany*

Received 27 June 1977

## 1. Introduction

In a previous publication, we presented predictions for the secondary structure of eleven proteins from the small subunit of the *Escherichia coli* ribosome [1]. The present work is an extension of these studies to the proteins deriving from the 50 S subunit. Predictive methods offer a means of assigning regions of highly probable conformation to proteins of known primary structure. The results can be related to other findings on the topography of the ribosomal subunits. Such an approach is valuable at a time when X-ray analyses of these proteins are not available. The predictions provide guidelines to the secondary structures of the proteins for directed studies of protein–protein neighbourhoods and the interactions of the proteins with the strands of ribosomal RNA.

For this purpose we applied four different predictive methods and, using the known amino acid sequences of these proteins, the calculated secondary structures were presented as diagrams showing the expected conformational states of each residue, i.e. for helix, extended structure, turn and random coil. Regions of highly probable secondary structure were derived from these histograms when at least three predictive methods agree about the conformational state of a residue [2].

## 2. Methods

The secondary structures of the ribosomal proteins were predicted according to the methods of Burgess et al. [3], of Chou and Fasman [4,5] and Chou et al. [6], of Nagano [7] and of Robson and Suzuki [8]. Fortran programmes 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 programmes. To these residues the state random coil was assigned.

## 3. Results

The secondary structures predicted by the four methods for the ribosomal proteins L5, L10, L12, L18, L25 and L29 are shown in figs 1–6. The predicted conformational states of the residues (helix, turn or bend, and extended structure of $\beta$-sheet) are represented symbolically. The definitions of the conformational states have been interpreted in their most general sense; i.e. the state helix encompasses all different types of helical state, and so on. This is necessary in order to obviate the differences in definitions given by the authors of the programmes.

In this report we investigated those L-proteins predicted to have a high helical content. Among these are the proteins L7/L12, L10 and L29 and, of particular interest, the 5 S RNA-binding proteins L5, L18 and L25. The secondary structure predictions of the other L-proteins of known primary structure are presented elsewhere [9].

### 3.1. *Protein L5*

The predictions for this 5 S RNA-binding protein are based on the sequence determined by Chen and Ehrke [10] and are shown in fig.1. The N-terminal



*North-Holland Publishing Company – Amsterdam*

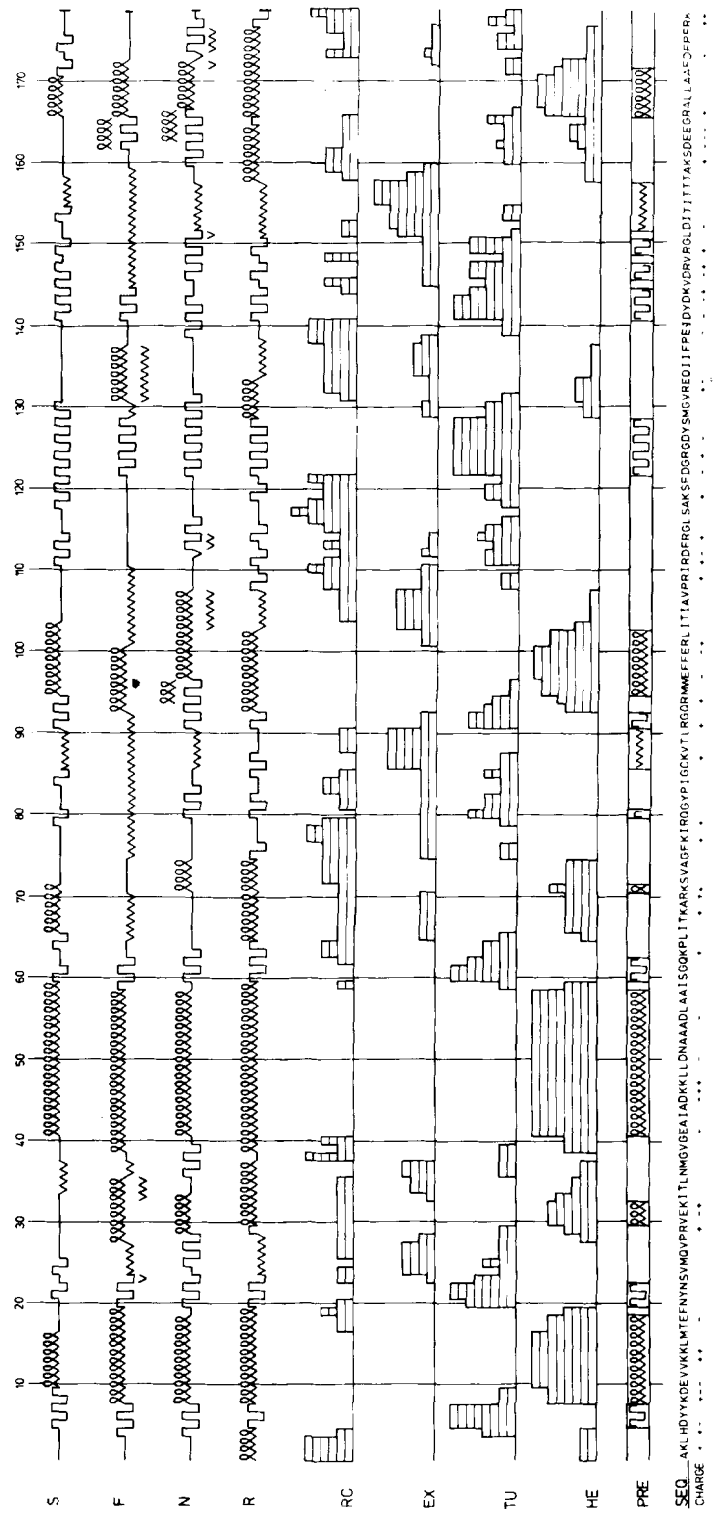


Fig. 1. Predicted secondary structures of protein L5 according to four different methods: (S), Burgess et al. [3]; (F), Chou and Fasman [4,5] and Chou et al. [6]; (N) Nagano [7]; (R) Robson and Suzuki [8]. The symbols used are: Helix (α), extended structure or β-sheet (v), turn or loop (τ) and random coil (——). In cases of ambiguity in the predictions both conformational states are indicated. The histograms represent random coil (RC), extended structure or β-sheet (EX), turn or loop (TU) and helix structure (HE). Half of the weight is given to the ambiguously predicted states. A most probable secondary structure is indicated for each residue in the line denoted by 'Pre', on the bases of three agreeing predictions (for more details see ref. [1]). SEQ means amino acid sequence (in the one letter code).

Table 1
Average values of secondary structure content and maximum lengths of ribosomal L-proteins based on at least three corresponding predictions

| Protein | Res[a] | Shape[b] | Binding sites[c] | Helix | | | Turn or bend | | | Extended | | | Random coil and not predictable | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Res | (%)[d] | Length[e] (Å) | Res | (%) | Length[f] (Å) | Res | (%) | Length[g] (Å) | Res | (%) | Length[h] (Å) | Maximum length[i] (Å) |
| L5 | 178 | | S | 47 | 26 | 71 | 27 | 15 | 47 | 11 | 6 | 38 | 93 | 52 | 335 | 491 |
| L10 | 165 | | | 78 | 47 | 117 | 14 | 9 | 25 | 5 | 3 | 17 | 68 | 41 | 245 | 404 |
| L12 | 120 | E | P | 77 | 64 | 116 | 4 | 3 | 7 | NC[j] | – | – | 39 | 33 | 140 | 263 |
| L18 | 117 | E | S,A | 51 | 44 | 77 | 17 | 15 | 30 | 11 | 9 | 38 | 38 | 32 | 137 | 282 |
| L25 | 94 | E | A | 39 | 42 | 59 | 9 | 10 | 16 | 9 | 10 | 31 | 37 | 39 | 133 | 239 |
| L29 | 63 | | | 42 | 67 | 63 | 3 | 5 | 5 | NC[j] | – | – | 18 | 29 | 65 | 133 |

[a] Res: number of amino acid residues
[b] Shape: (E) elongated
[c] Locations and number of antibody binding sites in the 50 S subunnit are indicated with: (S) seat region; (P) multiple binding sites central to the exposed protuberances; (A) anterior surface
[d] Percentage of residues in a given conformational state
[e] 1.5 Å/residue in helix conformation (assumed to be α-helix)
[f] 1.75 Å/residue in turn
[g] 3.47 Å/residue in extended structure
[h] 3.6 Å/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

region up to residue 60 is relatively rich in helix; two helices, positions 8–18 and 41–58, followed by turns are strongly predicted. The secondary structures of the central part of the protein chain, 59–140, are variously predicted. Extended structure is given for the region 86–90; a turn for 91–92/94 and a moderately long helix for 94/95–102. A strongly predicted turn region, given by all methods, is to be found between 122 and 128. A series of turns, 141–150, extended structure in positions 151/152–157, and a short helical region, 166–171, mark the C-terminal portion of the L5 protein chain.

Antibodies specific to L5 have been shown by immune electron microscopy to bind in the so-called 'seat' region of the 50 S subunit in proximity to the antibody binding sites of the other 5 S RNA-binding proteins, L18 and L25 [11]. The maximum length of protein L5 calculated for a stretched conformation amounts to 491 Å (table 1).

### 3.2. *Protein L18*

The secondary structures predicted for the 5S RNA-binding protein L18 (fig.2) are based on the sequence determined by Brosius et al. [12]. This protein is predicted to contain five helices (14–21, 46–51/52, 53–62, 69–84/85 and 103–113). In contrast to protein L5, protein L18 contains 44% helix (see table 1). Also in contrast to protein L5, where most of the helix is predicted for the N-terminal region, the largest and most strongly predicted helices in L18 are located in the central part of the protein chain (46–84). In common with protein L5, there is a rather long region in protein L18 (87–102), with a high probability for turns, which is situated between two large helices. A stretch of extended structure, broken by a turn at about 32–34, is expected for the N-terminal region.

An elongated shape for this protein has been found by small-angle X-ray scattering studies [13].

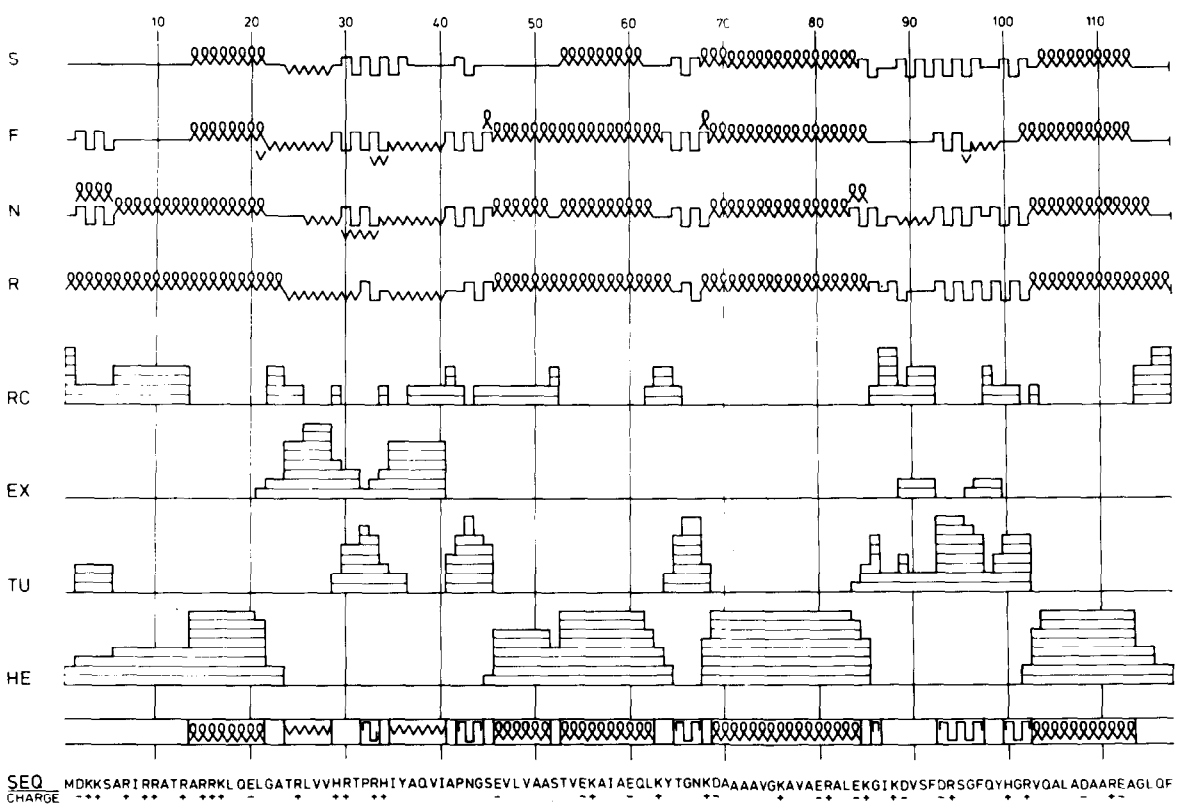


Fig.2. Predicted secondary structure of protein L18. For abbreviations see fig.1.

In accordance with these findings is the location of two antibody binding sites, in the lower protuberance (L18A) and in the seat region (L18B), in the 50 S subunit [11]. The L18 protein chain, of maximum calculated length of 282 Å, should be easily accomodated within the approximately 100 Å distance separating the two antibody binding sites.

### 3.3. *Protein L25*

The predictions for the secondary structure of this 5 S RNA-binding protein are based on the sequence determined by Dovgas et al [14] and Bitar and Wittmann-Liebold [15] and are illustrated in fig.3. As was given for protein L18, a relatively high proportion of helix, amounting to 42% (table 1), is predicted for protein L25. As in L18, the predicted regions of helix are clustered mostly in the central part of the protein chain (37–45 and 47–59). Other helical regions are predicted for positions 6–10, 69–77 and at the C-terminus (see fig.3). Adjacent to the central helical region at either end are two short stretches of extended structure (27–30 and 60–65). Three turns are given for protein L25 in positions 13/14–15/17, 31–33/35 and 80/81–84/85. The last, and most pronounced, of these is located in the C-terminal part of the protein chain, as was found for protein L18.

Only one site of antibody attachment in the seat area of the 50 S subunit has been demonstrated so far for protein L25 [11]. According to small-angle X-ray scattering studies the protein is elongated [13]. This result is consistent with the calculated maximal

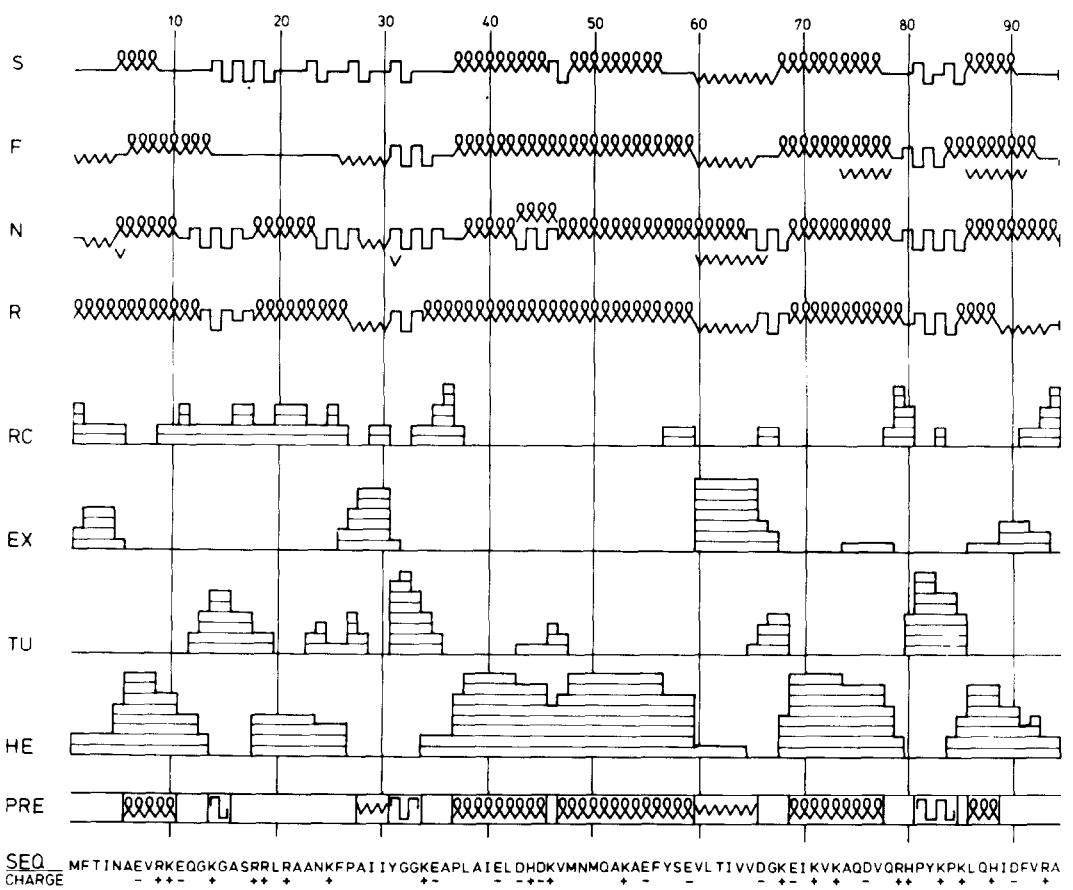


Fig.3. Predicted secondary structure of protein L25. For abbreviations see fig.1.

length of 239 Å for L25 (table 1). An attempt has been made [16] to predict the structure of L25 from the amino acid sequence with the assumption that L25 has a globular shape.

### 3.4. *Proteins L7/L12*

The predictions of the secondary structure of this protein pair are based on the sequence determinations of Terhorst et al. [17,18]. The two proteins differ only at their N-termini: protein L7 has an *N*-acetylated serine residue in the N-terminal position whereas, in protein L12, the acetyl group is absent. Both proteins also contain a partly *N*-ε-methylated lysine residue in position 81. The predictions given in fig.4 are for protein L12, and the *N*-ε-methylated derivative of lysine is treated as a lysine residue.

According to our predictions (table 1) proteins L7/L12 are characterised by a very high content of helix which amounts to approximately 64%. Helical conformations are predicted for residues 6–14/20, 21–28, 34–41, 45–57/60, 66–71, 82–95 and 102–118. The regions of helix appear to be distributed throughout the whole molecule and the intervening regions are either random coil or weakly predicted turns. One prominent turn area is predicted for the region 73/75–77/78. There is little suggestion of extended structure for proteins L7/L12 (see fig.4).

The high helix content of proteins L7/L12 as given by the predictions is in very good agreement with the values given for helix by earlier CD-measurements which range from 45–60% helical content [19–21]. More recent CD-measurements made with proteins isolated under denaturing [22] and non-denaturing [23] conditions are consistent with the above findings. According to theoretical evaluations a high degree of helicity for proteins L7/L12 was reported, namely

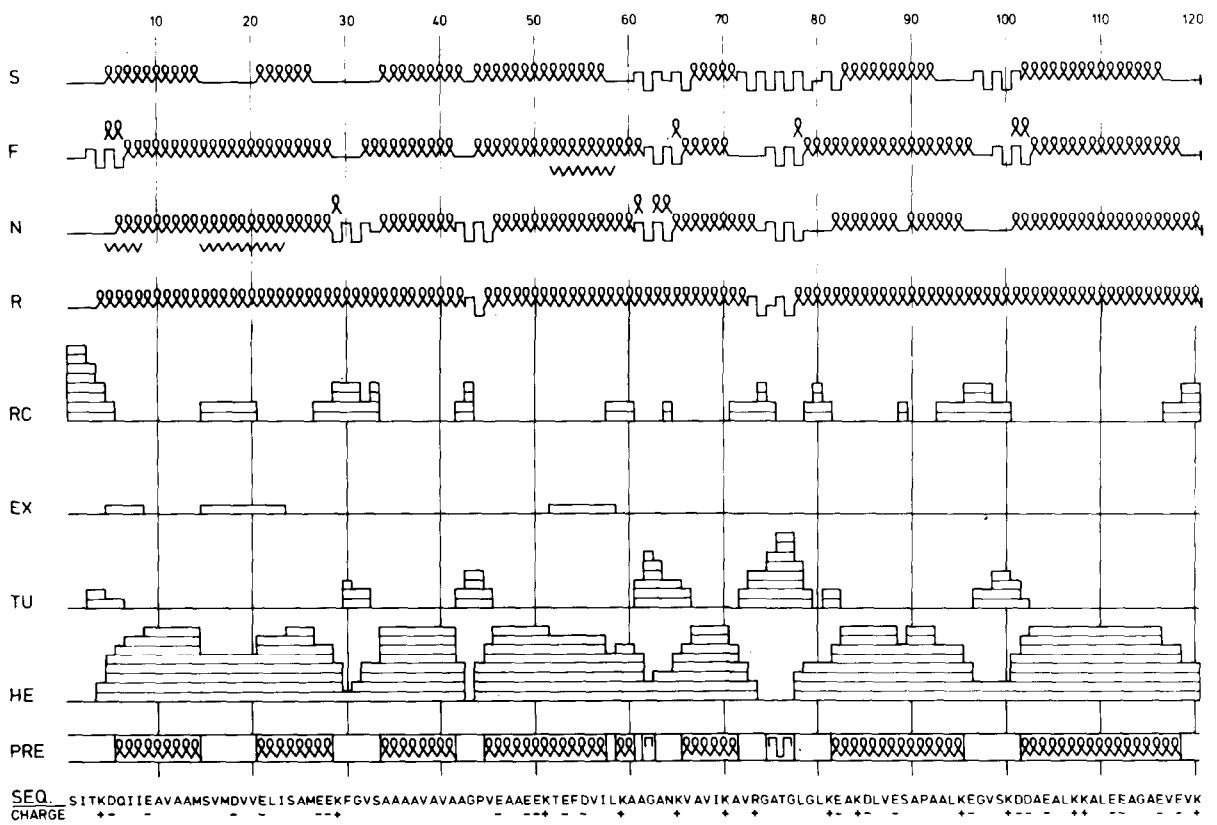


Fig.4. Predicted secondary structure of protein L12. For abbreviations see fig.1.

53% in a compact globular or 42% in an unfolded structure [24].

The helicity of proteins L7/L12 is substantially higher than that observed in globular proteins, such as haemoglobins and sperm-whale myoglobin, and an elongated shape has been postulated from immune electron microscopy [25] and from physical studies [26,27] for proteins L7/L12.

From reconstitution experiments with fragments of proteins L7/L12 it has been demonstrated that the N-terminus up to position 26, including all methionine residues of the proteins is probably essential for their binding to the 50 S ribosomal subunit [28]. Thus, the N-terminal helix region (6–14/20 and 21–28) strongly predicted in this area should be involved in this function. In the 50 S model of Tischendorf et al. [25,29], proteins L7/L12 are found in the 50 S subunit folded around the exposed protuberance like a garland.

### 3.5. *Protein L29*

The predictions made for this protein are based on the sequence determination of Bitar [30] and are illustrated in fig.5. As can be seen from this figure, the protein is predicted to have a high degree of ordered secondary structure, leaving only about 29% for random coil (see table 1). With approximately 67% of predicted helix, L29 is comparable to proteins L7/L12. The distribution of the helices is restricted to three areas, one of which is quite lengthy (positions 2–31) whereas the others are located in positions 40–47 and 57–60 at the C-terminus of the chain.

The first two helix regions are interrupted by a turn 32/34–36. The area between the second and the third helix is predicted equivocally by the different methods.

Extended structure appears rather unlikely in this protein. Only the method of Chou et al. [4–6] gives some weak indication for this conformational state in the C-terminal region of the protein. The maximal length of 133 Å calculated from the predictions for this protein is shorter than that of some other proteins, e.g. S15, S18 and L28 [1,9]. Therefore, it can be suggested that L29 has a more globular structure than these three proteins.

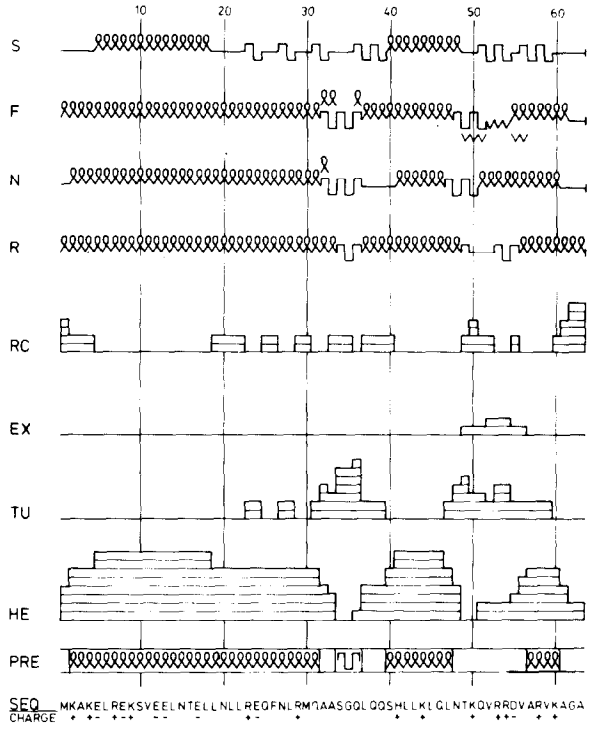


Fig.5. Predicted secondary structure of protein L29. For abbreviations see fig.1.

### 3.6. *Protein L10*

The predictions for this protein are based on the sequence determined by Heiland et al. [31] and are shown in fig.6. Predominant helices are calculated for the N- and C-terminal regions, the latter ones being more extensive than the N-terminal helix. The helical regions are located in positions 5–18/27 and in the C-terminal half from 92–114, 133–146 and 153–165. More weakly predicted, short helices are found in the region 36–72. The helix content of this protein amounts to 47% (table 1) and is comparable to proteins L18 and L25 which show a condensed helix region in the central part of the molecule. Several short areas of turn are indicated for protein L10 in positions 28–30, 44–45/48, around the proline residue in position 78, in positions 82/83–84/90, 119–121 and, less pronounced, in 146–150. A low content of extended structure is calculated in protein L10. This is found mainly in positions 50–54 and, less probably, for 73–81 and 122–132.

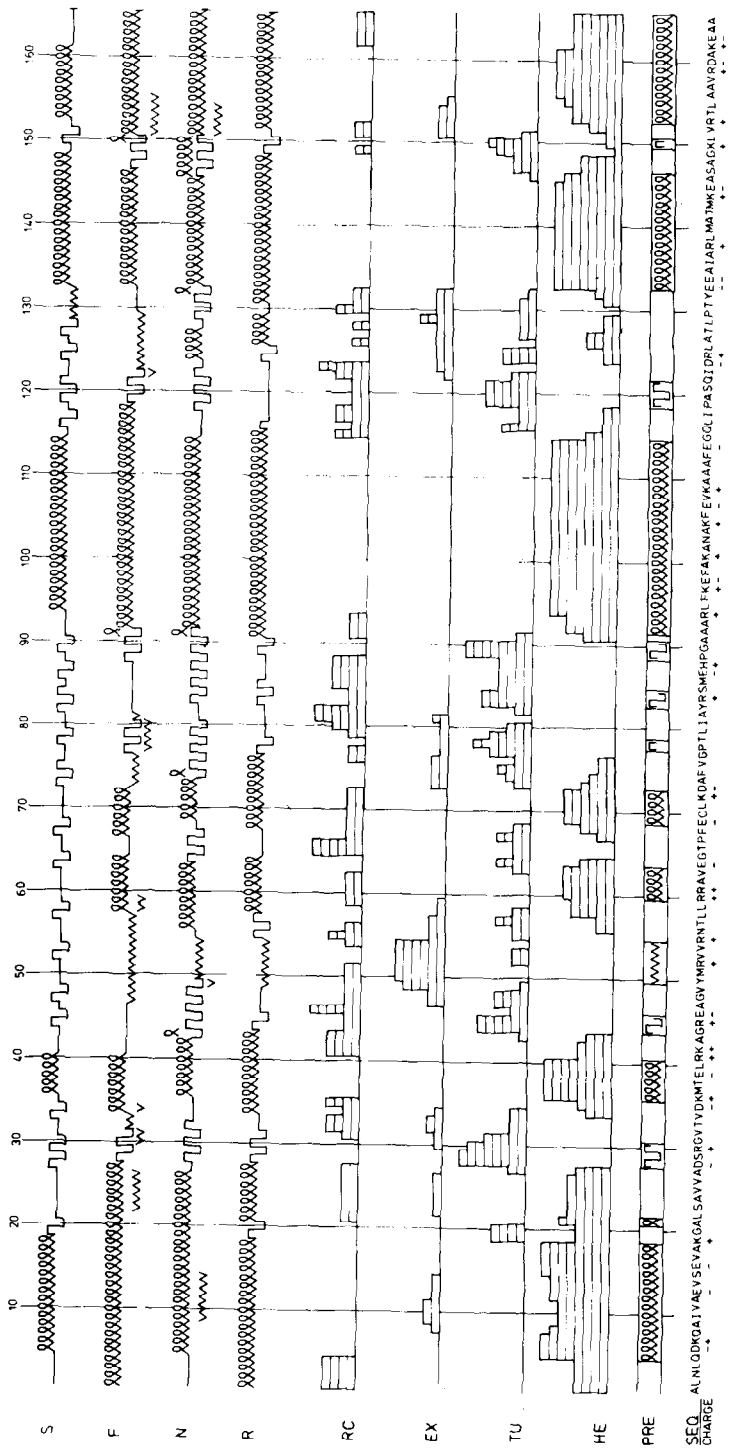


Fig.6. Predicted secondary structure of protein L10. For abbreviations see fig.1.

Protein L10 is located in the 50 S particle near L7/L12 as can be concluded from the following findings:

(a) L10 forms a complex with L7/L12 [32].
(b) L10, L11 and L7/L12 can be crosslinked to each other [33].
(c) L10 is necessary for the incorporation of L7/L12 into 50 S core particles [34,35].
(d) L11 and L7/L12 are located relatively close to each other in the 50 S model deduced from immune electron microscopic studies [25].

## 4. Conclusions

The predicted secondary structures of eleven 30 S proteins [1,2] and fourteen 50 S proteins (given in this report and elsewhere [9]) differ from each other considerably. Some of the ribosomal proteins are predicted to have a high degree in helix, which amounts to 50–67% for proteins S20, S21, L7/L12 and L29 and to 40–50% for proteins L10, L18 and L25. This tendency for helix formation is due to the high percentage of certain amino acids in these proteins: i.e. S20, L10 and L7/L12 contain 22%, 20% and 23% of alanine, respectively. Protein L29 has 19% of leucine and 11% of glutamine, whereas protein L25 consists of 13% glutamic acid and 11% of valine.

The predicted amount of extended structure, on the other hand, is relatively low. It does not exceed 10% for proteins L5, L10, L18 and L25. No significant extended structure was found for proteins L7/L12 and L29; this was also the case with proteins S13, S15, S20 and S21 from the small subunit [1]. A rather high degree of turn areas (10–20%) predicted for many S-proteins has been found in proteins L5, L18 and L25. These turn regions, in preference to random coil, very frequently interrupt helical stretches or follow them.

## References

[1] Dzionara, M., Robinson, S. M. L. and Wittmann-Liebold, B. (1977) Hoppe-Seyler's Physiol. Chem. 358, 1003–1019.
[2] Wittmann-Liebold, B., Robinson, S. M. L. and Dzionara, M. (1977) FEBS Lett. 77, 301–307.
[3] Burgess, A. W., Ponnuswamy, P. K. and Scheraga, H. A. (1974) Israel J. Chem. 12, 239–286.
[4] Chou, P. Y. and Fasman, G. D. (1974) Biochemistry 18, 211–222.
[5] Chou, P. Y. and Fasman, G. D. (1974) Biochemistry 18, 222–249.
[6] Chou, P. Y., Adler, A. J. and Fasman, G. D. (1975) J. Mol. Biol. 96, 29–45.
[7] Nagano, K. (1977) J. Mol. Biol. 109, 251–274.
[8] Robson, B. and Suzuki, E. (1976) J. Mol. Biol. 107, 327–356.
[9] Dzionara, M., Robinson, S. M. L. and Wittmann-Liebold, B. (1977) J. Supramol. Structure, to be submitted.
[10] Chen, R. and Ehrke, G. (1976) FEBS Lett. 69, 240–245.
[11] Stöffler, G. and Wittmann, H. G. (1977) in: Protein Synthesis, (Weissbach, H. and Pestka, S. eds) pp. 117–202, Academie Press, London and New York.
[12] Brosius, J., Schiltz, E. and Chen, R. (1975) FEBS Lett. 56, 359–361.
[13] Österberg, R., Sjöberg, B. and Garrett, R. A. (1976) FEBS Lett. 65, 73–76.
[14] Dovgas, N. V., Markova, L. F., Mednikova, T. A., Vinokurov, L. M., Alakhov, Y. B. and Ovchinnikov, Y. A. (1975) FEBS Lett. 53, 351–354.
[15] Bitar, K. G. and Wittmann-Liebold, B. (1975) Hoppe-Seyler's Z. Physiol. Chem. 356, 1343–1352.
[16] Finkelstein, A. V., Kozitsyn, S. A. and Ptitsyn, O. B. (1975) FEBS Lett. 60, 137–140.
[17] Terhorst, C., Möller, W., Laursen, R. and Wittmann-Liebold, B. (1972) FEBS Lett. 28, 325–328.
[18] Terhorst, C., Möller, W., Laursen, R. and Wittmann-Liebold, B. (1973) Eur. J. Biochem. 34, 138–152.
[19] Möller, W., Castelman, H., Terhorst, C. P. (1970) FEBS Lett. 8, 192–196.
[20] Dzionara, M. (1970) FEBS Lett. 8, 197–200.
[21] Boublik, M., Brot, N. and Weissbach, H. (1973) Biopolymers 12, 2083–2092.
[22] Heiland, I., Dzionara, M. and Snatzke, G., unpublished results.
[23] Morrison, C. A., Dijk, J. and Littlechild, J., in preparation.
[24] Ptitsyn, O. B., Denesyuk, A. I., Finkelstein, A. V. and Lim, V. I. (1973) FEBS Lett. 34, 55–57.
[25] Tischendorf, G. W., Zeichhardt, H. and Stöffler, G. (1974) Mol. Gen. Genet. 134, 187–208.
[26] Wong, K. P. and Paradies, H. H. (1974) Biochem. Biophys. Res. Commun. 61, 178–184.
[27] Österberg, R., Sjöberg, B., Liljas, A. and Pettersson, I. (1976) FEBS Lett. 66, 48–51.
[28] Kittler, R., Tischendorf, G. W. and Stöffler, G., unpublished results.
[29] Tischendorf, G. W., Zeichhardt, H. and Stöffler, G. (1975) Proc. Natl. Acad. Sci. USA 72, 4820–4824.
[30] Bitar, K. G. (1975) Biochim. Biophys. Acta 386, 99–106.

[31] Heiland, I., Brauer, D. and Wittmann-Liebold, B. (1976) Hoppe-Seyler's Z. Physiol. Chem. 357, 1751–1770.
[32] Pettersson, I., Hardy, S. J. S. and Liljas, A. (1976) FEBS Lett. 64, 135–138.
[33] Expert-Bezançon, A., Barritault, D., Milet, M. and Hayes, D. H. (1976) J. Mol. Biol. 108, 781–787.
[34] Schrier, P. I., Maassen, J. A. and Möller, W. (1973) Biochem. Biophys. Res. Commun. 53, 90–98.
[35] Stöffler, G., Hasenbank, R., Bodley, J. W. and Highland, J. H. (1974) J. Mol. Biol. 86, 171–174.