Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 7251–7255, October 1986
Biochemistry

Crystal structure of a prokaryotic ribosomal protein

(x-ray crystallography/protein–protein interaction/RNA binding site/common structural motif)

KEITH S. WILSON*†, KRZYSZTOF APPELT*‡, JOHN BADGER*§, ISAO TANAKA*¶, AND STEPHEN W. WHITE*||**

*Max Planck Institut fuer Molekulare Genetik, Abteilung Wittmann, Ihnestrasse 63-73, D-1000 Berlin 33, and †European Molecular Biology Laboratory Outstation, Deutsches Elektronen Synchrotron, Notkestrasse 85, D-2000 Hamburg 52, Federal Republic of Germany; ‡Department of Biology, University of California at San Diego, La Jolla, CA 92093; §Department of Physics, University of York, York, England; ¶Department of Applied Chemistry, Faculty of Engineering, Nagoya University, Chikusa, Nagoya 464, Japan; and ||Brookhaven National Laboratory, Department of Biology, 463, Upton, NY 11973

Communicated by R. B. Setlow, July 1, 1986

ABSTRACT **The structure of ribosomal protein L30 from *Bacillus stearothermophilus* has been solved to a resolution of 2.5 Å. The molecule is somewhat elongated and contains two helices and a three-stranded, anti-parallel β-pleated sheet. The protein fold, in which helices pack on the same side of the sheet, generates a simple helix-sheet, two-layered motif. It is possible to distinguish three hydrophobic patches on the molecular surface, and one end has six isolated arginine and lysine residues. It is proposed that these reflect sites of protein–protein and protein–RNA interaction, respectively. The protein fold is very similar to that of the only other known ribosomal protein structure, L7/L12 from *Escherichia coli*, and, based on this similarity, an attempt is made to align the amino acid sequences of the two proteins.**

The ribosome is the cellular organelle, coordinating protein synthesis, that consists of a large and a small subunit. In prokaryotes, the intact ribosome is a 70S particle with a large 50S and a small 30S subunit. The 50S subunit has a 23S and a 5S RNA molecule and about 33 ("L") proteins; the 30S subunit has a 16S RNA molecule and about 20 ("S") proteins. We describe here the three-dimensional structure of protein L30 from *Bacillus stearothermophilus*.

The ribosomal proteins from *Escherichia coli* have been the most extensively studied and are numbered according to their positions in two-dimensional gel electrophoresis (1). However, with the exception of L7/L12 (2), these proteins have failed to produce crystals suitable for x-ray diffraction studies, and we have concentrated on the proteins from the thermophile *B. stearothermophilus*, which are considerably more resistant to thermal denaturation (3). To avoid nomenclature confusion, homologies based on amino acid sequence data have been established between the available *B. stearothermophilus* proteins and those from *E. coli*, and the former are labeled accordingly. The L30 proteins have been fully sequenced (4, 5) and are identical at 53% of the positions (Fig. 1). This degree of homology is typical of the ribosomal proteins from the two organisms and clearly shows that the considerable biochemical information that has been accumulated on the *E. coli* proteins can be directly applied to their counterparts from *B. stearothermophilus*.

L30 is one of the smallest of the ribosomal proteins with a molecular weight of 7000. It is difficult to assign a particular biochemical function to L30 or, indeed, to any isolated ribosomal protein. It is not an essential protein since mutants that lack it have similar growth rates to the wild type (6). However, single protein omission experiments on reconstituted *E. coli* (7) and *B. stearothermophilus* (8, 9) ribosomes indicate a positive effect on peptidyltransferase activity. It appears, from reconstitution experiments (10), to be located near the surface of the large subunit and, from cross-linking experiments (11), to be close to the binding site of elongation factor Tu.

Structure Determination. L30 was purified using a mild, nondenaturing technique (12) that has successfully produced preparations of ribosomal proteins suitable for crystallization (13, 14). Well-ordered crystals with space group $P4_32_12$ (a = b = 46.3 Å, c = 61.4 Å) are grown in 3.85 M $(NH_4)_2SO_4$/0.1 M Tris·HCl, pH 8.4 at 18°C. There is one molecule in the asymmetric unit (13, 15).

Data were collected photographically to a resolution of 2.5 Å using an Enraf–Nonius Arndt–Wonacott oscillation camera. CuK_α x-radiation was provided by a Seifert stationary anode operating with a fine focus tube at 40 kV and 30 mA. The majority of the data were collected with the c axis of the crystals aligned with the rotation (ϕ) axis, and the blind region data were obtained by rotating the crystal about the a axis. The diffraction spot intensities were evaluated from the film using the MOSCO package of programs. The phases were determined by the method of multiple isomorphous replacement using the heavy atom derivatives reported (15) together with a new $K_2Pt(CN)_6$ derivative. The refined heavy atom parameters are listed in Table 1. The final mean figure of merit was 74%.

The protein chain was easily traced in the resulting electron density map and an unambiguous fit of the sequence (5) to the density was obtained on an Evans and Sutherland PS300 picture system using the FRODO package of programs (Fig. 2). The only residues for which there was little or no side chain electron density are all charged and exposed on the surface of the protein (lysine-2, arginine-24 and -29, and glutamic acid-60).

General Description of the Structure. The molecule is rather elongated with approximate dimensions of 40 Å × 30 Å × 25 Å. A stereoview of the backbone and a schematic representation of the molecule are shown in Fig. 3. L30 belongs to the α/β class of proteins being composed of two α-helices (α1, residues 17–26; α2, residues 41–51) and a three-stranded anti-parallel β-pleated sheet that displays the usual left-handed twist (strand 1, residues 3–7; strand 2, residues 33–37; strand 3, residues 54–59). α2 and the three strands of the sheet are approximately parallel to the long axis of the molecule, and α1 is oriented roughly 45° to this axis.

The distribution of the secondary structural elements within the sequence is shown in Fig. 1. The strands and helices are arranged in the order β-α-β-α-β, and both helices pack on the same side of the sheet to produce a helix-sheet, two-layered molecule. α1 connects to the sheet by two loops (residues 8–16 and residues 27–32), which together form one end of the molecule. For convenience, we shall refer to this as the "loop" end and the opposite end, which contains both the N and C termini, as the termini end.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

**To whom reprint requests should be addressed.

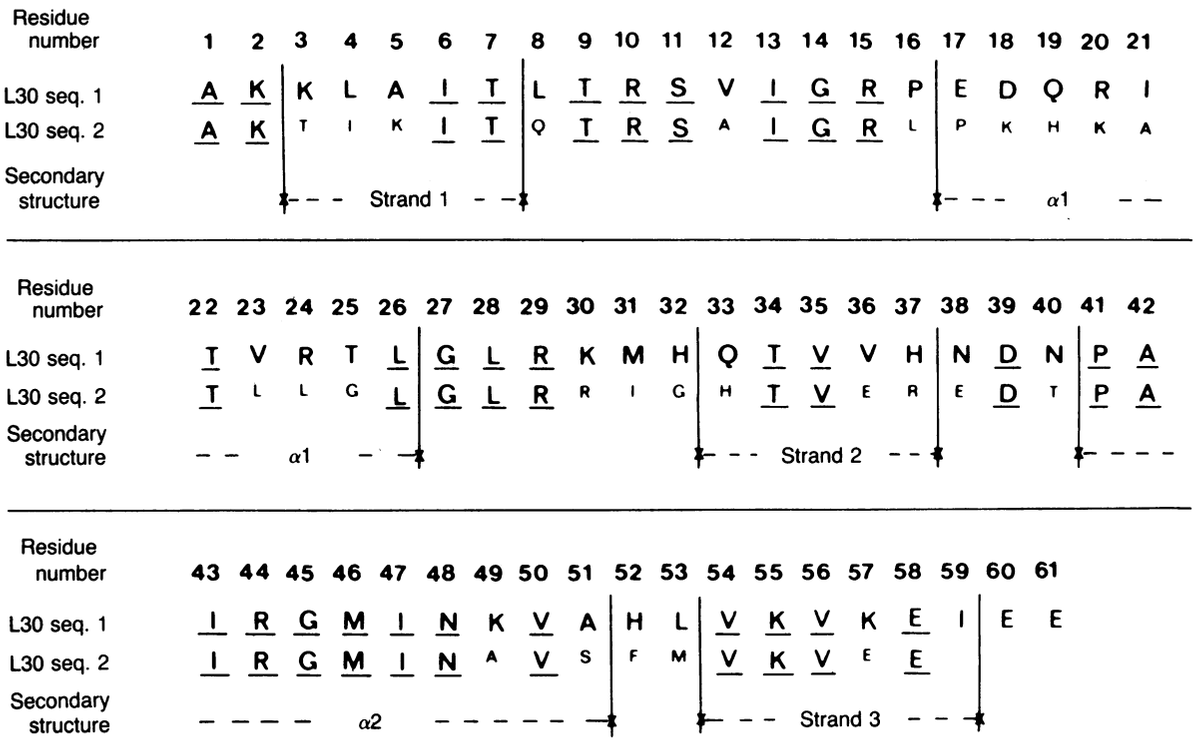

FIG. 1. The sequences of L30 from *B. stearothermophilus* (L30 seq. 1) and *E. coli* (L30 seq. 2) using the one-letter code. Identical amino acids are underlined and in heavy type. The distribution of secondary structure has been deduced from our electron density map.

The protein core is very hydrophobic and contains five leucines (residues 4, 8, 26, 28, and 53), three isoleucines (residues 6, 43, and 47), and four valines (residues 12, 23, 50, and 54). Isoleucine-43 and -47 are partially exposed and participate in the formation of hydrophobic patches on the surface (see below). All the polar and charged residues are exposed to the solvent. Thus, in a general sense, L30 has a typical globular protein structure. A somewhat curious feature is that both helices have proline residues at the N termini (Fig. 1) and not at their C termini where they are usually located and thought to act as helix breakers.

There appear to be salt bridges between lysine-49 and aspartic acid-18, arginine-20 and glutamic acid-17, and arginine-44 and aspartic acid-39; and there may be an additional one between lysine-2 and glutamic acid-60, but the side chains of these residues are not visible in our map. The salt bridge between residues 18 and 49 links the two helices together, and it may have an important stabilizing role at this, potentially, rather flexible region of the molecule (most

Table 1. The parameters, from the least-squares refinement of the heavy atom derivatives, used in the computation of the phases for L30

Derivative	Resolution	Occupancy	Axes x	Axes y	Axes z	R_c
$Pt(NH_3)_2Cl_2$	2.5 Å	1.795	0.144	0.476	0.426	40.1
$PtCl_4$	3.0 Å	1.545	0.141	0.475	0.423	40.0
$KAu(CN)_2$	3.0 Å	1.387	0.145	0.167	0.017	56.1
		0.749	0.082	0.031	0.297	
$K_2Pt(CN)_6$	2.5 Å	1.323	0.018	0.040	0.327	60.7
		1.001	0.283	0.134	0.314	

The occupancies are on an arbitrary scale, and the temperature values B are 15 and have not been refined. R_c is the R factor for the centric terms and is defined as

$$R_c = \sum_h \left| |F_{PH} - F_P| - f_H \right| / \sum_h |F_{PH} - F_P|,$$

where F_{PH} is the structure factor amplitude of the isomorphous derivative, F_P that of the native protein, and f_H is the calculated heavy atom amplitude. Note that the $Pt(NH_3)_2Cl_2$ and $PtCl_4$ derivatives are essentially identical.

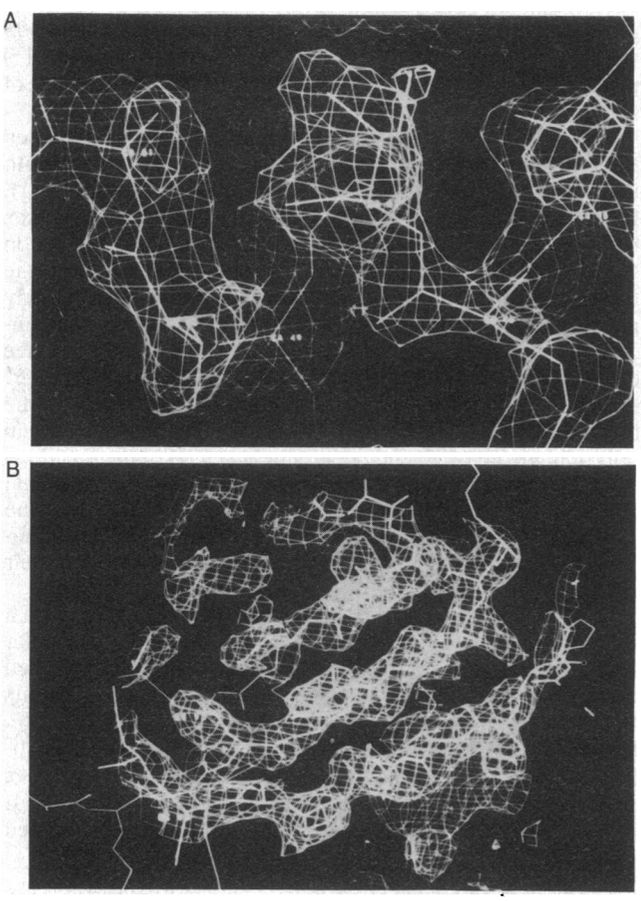

FIG. 2. Photographs of the L30 electron density from the Evans and Sutherland PS300. (*A*) Helix α2. (*B*) The β-pleated sheet.

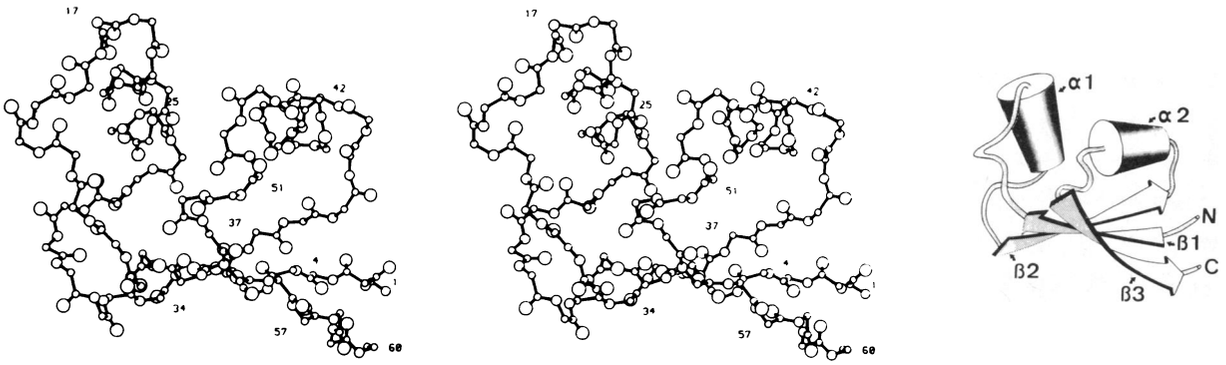

FIG. 3. A stereoview (*Left*) and a schematic representation (*Right*) of the backbone of protein L30.

clearly visible in the schematic representation in Fig. 3). We have built the *E. coli* sequence into our structure, and there appears to be no possibility of an equivalent interaction in the resulting molecule. Therefore, the greater thermal stability of the *B. stearothermophilus* protein may derive, in part, from this salt bridge.

L30 from both *E. coli* and *B. stearothermophilus* have been extensively studied in solution (unpublished results), and the results are consistent with our three-dimensional structure. Both proteins are resistant to proteases in the cold but are preferentially digested by trypsin at elevated temperatures at two positions in the chain, arginine-10 and arginine-29. These arginines are conserved and lie in the two exposed loops linking α1 to the sheet. The circular dichroic spectra are nearly identical for the two proteins, and the content of α-helix and β-sheet was estimated to be 30%, which agrees well with the crystal structure. The NMR spectra are also very similar and indicate the presence of a large number of weakly shifted ring current methyl resonances. These probably arise from the many hydrophobic amino acids in the proximity of histidine-52 (replaced by a phenylalanine in the *E. coli* protein).

In a completely independent two-dimensional Fourier transform NMR study of the *E. coli* protein (ref. 16, and F. J. M. van de Ven, S. H. de Bruin, and C. W. Hilbers, personal communication) the secondary structural elements and their relative topology were deduced and are in excellent agreement with our findings.

Functional Sites of L30. In Fig. 1, the *E. coli* L30 sequence is aligned with that from *B. stearothermophilus* and, from the pattern of conservation, one can begin to predict which regions of the molecule are likely to be important to the structure and function of L30. Thus, for example, all of the hydrophobic core amino acids (see above) are conserved or display conservative changes.

Locating the functional sites on L30 is rather problematical since its function is unknown (see above) and, also, the protein may be but a single component of a larger functional region of the ribosome involving other proteins and/or RNA. However, it is almost certain to be subject to specific interactions with other proteins and RNA, and it is possible to search for evidence of these in the structure and sequence conservation. It should be noted that there are differences between all the components of the *B. stearothermophilus* and *E. coli* ribosomes and, therefore, protein–protein and protein–RNA binding sites need not be identical.

Protein–protein binding is often mediated through exposed hydrophobic patches, and the relatively small L30 structure displays no less than three such areas on its surface (Fig. 4). Patch 1 includes valine-12, isoleucine-13, and methionine-31 and is at the loop end of the molecule. Patch 2, on the helix side of the molecule, consists of isoleucine-21 and -43, alanine-42, and methionine-46. Finally, patch 3 lies on the sheet side of L30 and comprises alanine-5 and -51, valine-35, -36, and -56 and isoleucine-47 and -59. Patch 1 and patch 2 are conserved (Fig. 1) and are likely to be of greater significance than the rather more diffuse and variable patch 3.

The necessarily close association between ribosomal proteins and the highly negatively charged RNA is generally evident from their large number of positively charged residues (17). In this respect L30 is typical and contains 12 arginines plus lysines and only 6 aspartic and glutamic acids. Lysine-2 and -49 and arginine-20 and -44 appear to be involved in ion pair formation (see above) but the remaining eight are isolated and are ideal candidates for binding to the phosphate backbone of one or more RNA molecules. Lysine-3 and -57 are not conserved and are located at the termini end of the molecule. Arginine-10, -15, -24, and -29 and lysine-30 and -55 are grouped together at the loop end and are almost completely conserved. It would, therefore, seem reasonable

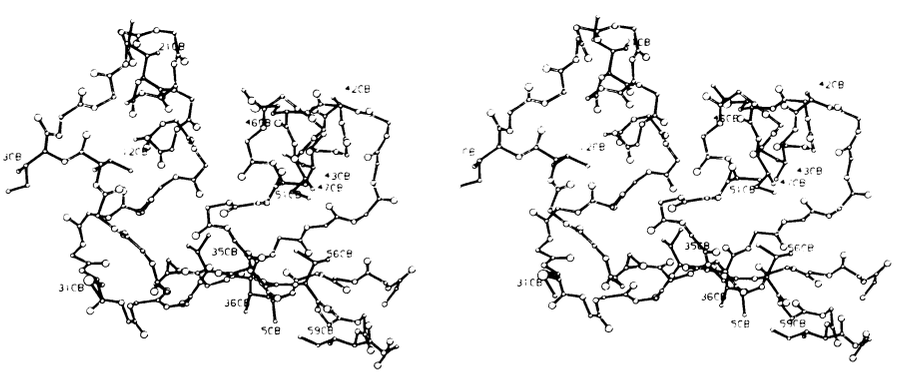

FIG. 4. A stereoview of the exposed hydrophobic amino acids (solid lines) on the α-carbon backbone. The residues can be grouped into three patches—patch 1, bottom left; patch 2, top; patch 3, bottom right. For clarity, only the CB atoms are labeled.

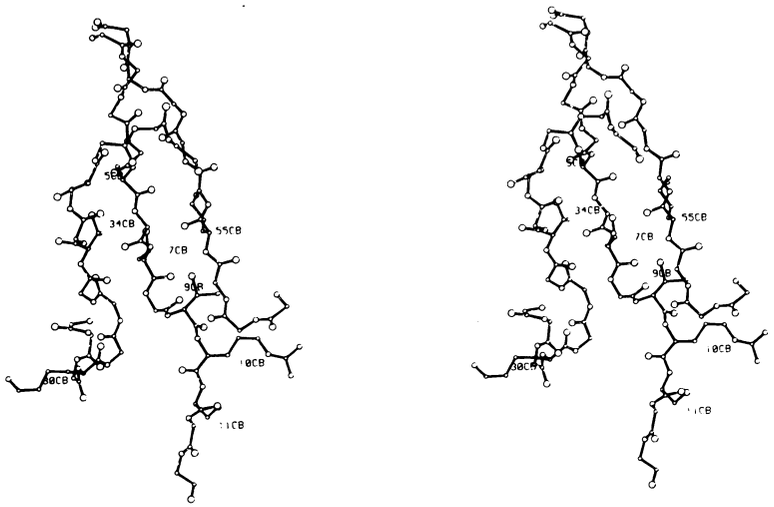

FIG. 5. A stereoview of a possible specific RNA binding site at the loop end of the β-pleated sheet. Serine-11 and threonine-7, -9, and -34 form a line of potential hydrogen bond acceptors and/or donors and are surrounded by the isolated positively charged residues arginine-10 and lysine-30 and -55. The CB atoms of these amino acids are labeled, and their bonds are drawn with solid lines.

that the loop end is the principal site of attachment to the ribosomal RNA.

The specificity of RNA binding probably derives from hydrogen bonds between the bases and protein side chains as has been suggested for protein–DNA interactions (18). Since there are no data concerning the L30 RNA binding site, we are unable to perform any meaningful model-fitting studies. However, the five threonines and the serine residue may be important since they are highly conserved and surrounded by the isolated arginines and lysines. In particular, threonine-7, -9, and -34 and serine-11 form a line of potential hydrogen bond donors and/or acceptors across the surface of the sheet (Fig. 5).

Relationship to Ribosomal Protein L7/L12. It is interesting that one of the few other known protein structures that displays the L30 helix-sheet, two-layered motif is the C-terminal fragment of the ribosomal protein L7/L12 from *E. coli* (2). Furthermore, apart from the addition of an extra helix between strands A and B in L7/L12, the topological relationship between the secondary structural elements is identical (Fig. 6). There is also the possibility that the proteins are functionally related since both have been associated with elongation factors G and Tu (11, 19). There have been several attempts to find evolutionary relationships between the ribosomal proteins based on sequence comparison techniques (20, 21), and L30 and L7/L12 do appear to be distantly related. In light of our structural information, we decided to repeat the comparison.

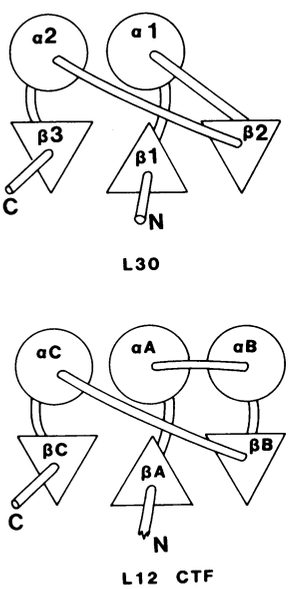

FIG. 6. Topological relationships between the secondary structural elements of proteins L30 and L7/L12. The circles and triangles represent helices and β-strands, respectively. If the point of the triangle is upward, the strand is oriented such that the N terminus is toward the reader.

The program RELATE†† was used to search for the optimal alignment of the complete sequences of the two proteins. We then used the program ALIGN†† to investigate this alignment more closely. ALIGN obtains the best match of two input sequences and introduces gaps wherever they appear to be appropriate.

Two comparisons were made, one between the sequences of the *E. coli* proteins (4, 22) and the other between the *B. stearothermophilus* proteins (ref. 5 and A. T. Matheson and A. Liljas, personal communication). In both cases, RELATE showed that L30 most closely resembles the C-terminal half of L7/L12. ALIGN, however, produced different optimal alignments of these regions, each with questionable significance. Clearly, if the proteins are related, the relationship can be detected at the level of tertiary structure but is barely discernible at the level of primary structure. This has also been found to be the situation with other potential families of proteins such as the dehydrogenases (23).

Assuming that L30 and L7/L12 are related, we suggest in Fig. 7 a possible alignment of the two sequences from *E. coli*. In this arrangement we have attempted to optimize the alignment at all levels of the structure. The gap in L30 between residues 29 and 30 is to accommodate helix αB, which appears from the structure to have no equivalent in L30.

If this relationship is real, the question of whether this motif is a common feature of ribosomal proteins immediately arises. A definitive answer must await the determination of the structures of other ribosomal proteins such as S5 and L6 (24, 25).

††National Biomedical Research Foundation (1985) *Protein Sequence Data Base of the Protein Identification Resource* (Washington, DC), Release No. 1.0.

We thank Professor H.-G. Wittmann for his support of this work; Dr. Jan Dijk for protein purification; and Volker Brachvogel,

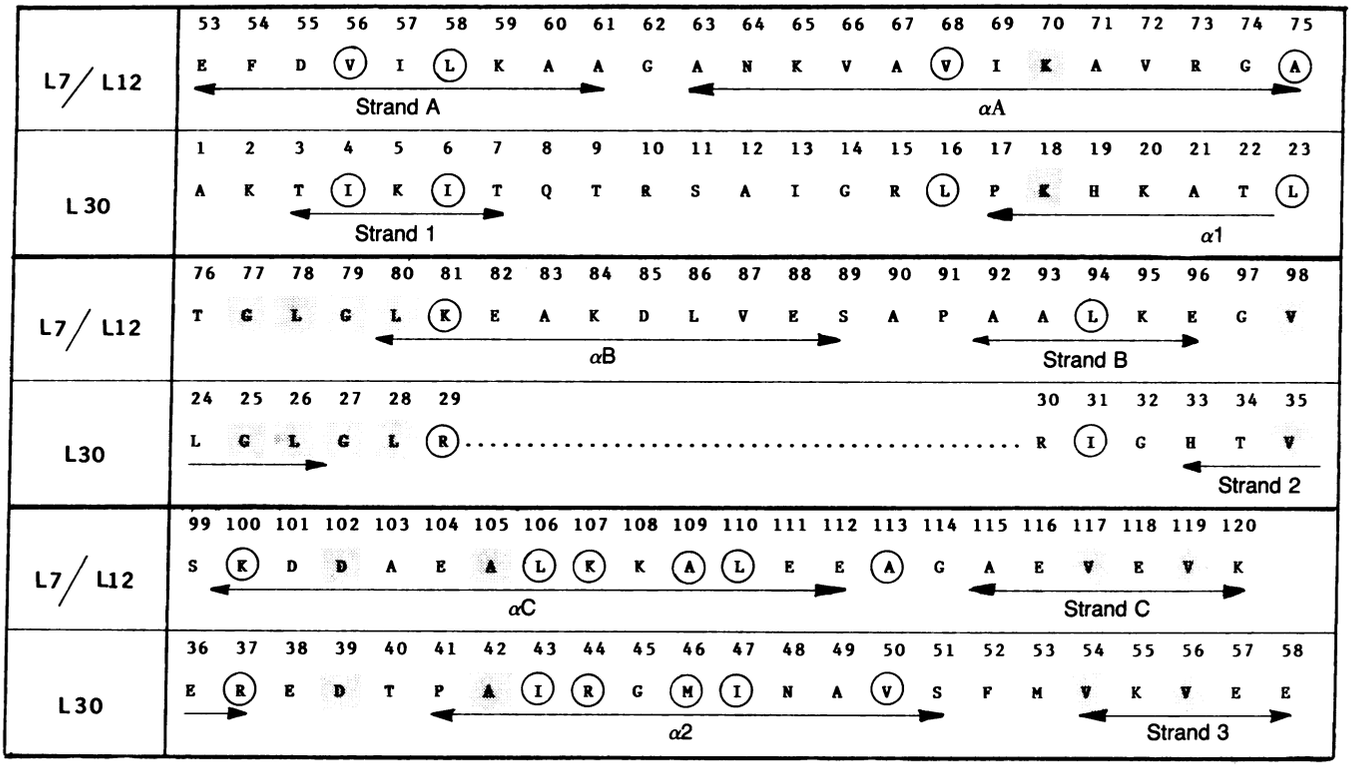

FIG. 7. A suggested alignment of the *E. coli* L30 and L7/L12 amino acid sequences using the one-letter code. Identical residues are boxed and shaded, and similar residues are circled. The alignment is purely hypothetical and represents a compromise between the optimal alignments of the primary, secondary, and tertiary structures. A gap has been introduced into the L30 sequence to accommodate helix αB of L7/L12, which appears to be an extra secondary structural element.

Howard Terry, and Pauline Wales for technical assistance. We also thank Drs. A. T. Matheson, A. Liljas, and C. W. Hilbers for their unpublished results. This work was in part funded by the Deutsche Forschungsgemeinschaft (SFB9), and I.T. was a Humboldt Fellow.

1. Kaltschmidt, E. & Wittmann, H. G. (1970) *Proc. Natl. Acad. Sci. USA* **67**, 1276–1282.
2. Leijonmark, M., Ericksson, S. & Liljas, A. (1980) *Nature (London)* **286**, 824–826.
3. Dijk, J., White, S. W., Wilson, K. S. & Appelt, K. (1983) *J. Biol. Chem.* **258**, 4003–4006.
4. Ritter, E. & Wittmann-Liebold, B. (1975) *FEBS Lett.* **60**, 153–155.
5. Kimura, M. (1984) *J. Biol. Chem.* **259**, 1051–1055.
6. Dabbs, E. R. (1979) *J. Bacteriol.* **140**, 734–737.
7. Hampl, H., Schulze, H. & Nierhaus, K. (1981) *J. Biol. Chem.* **256**, 2284–2288.
8. Cohlberg, J. A. & Nomura, M. (1976) *J. Biol. Chem.* **251**, 209–221.
9. Auron, P. E. & Fahnestock, S. R. (1981) *J. Biol. Chem.* **256**, 10105–10110.
10. Nierhaus, K. (1980) in *Ribosomes: Structure, Function and Genetics*, eds. Chambliss, G., Craven, G. R., Davies, J., Davis, K., Kahan, L. & Nomura, M. (University Park Press, Baltimore), pp. 267–294.
11. San Jose, C., Kurland, C. G. & Stoffler, G. (1976) *FEBS Lett.* **71**, 133–137.
12. Dijk, J. & Littlechild, J. (1979) *Methods Enzymol.* **59**, 481–502.
13. Appelt, K., White, S. W. & Wilson, K. S. (1983) *J. Biol. Chem.* **258**, 13328–13330.
14. Appelt, K., Dijk, J., White, S. W. & Wilson, K. S. (1983) *FEBS Lett.* **160**, 75–77.
15. Appelt, K., Dijk, J., White, S. W., Wilson, K. S. & Bartels, K. (1983) *FEBS Lett.* **160**, 72–74.
16. van de Ven, F. J. M., de Bruin, S. H. & Hilbers, C. W. (1984) *FEBS Lett.* **169**, 107–111.
17. Nierhaus, K. H. (1982) *Curr. Top. Microbiol. Immunol.* **97**, 81–155.
18. Ohlendorf, D. H., Anderson, W. F., Fisher, R. G. & Matthews, B. W. (1982) *Nature (London)* **298**, 718–723.
19. Koteliansky, V. E., Domogatsky, S. P., Gudkov, A. T. & Spirin, A. S. (1977) *FEBS Lett.* **73**, 6–11.
20. Wittmann-Liebold, B., Ashman, K. & Dzionara, M. (1984) *Mol. Gen. Genet.* **196**, 439–448.
21. Jue, R. A., Woodbury, N. W. & Doolittle, R. F. (1980) *J. Mol. Evol.* **15**, 129–148.
22. Terhorst, C., Moller, W., Laursen, R. & Wittmann-Liebold, B. (1973) *Eur. J. Biochem.* **34**, 138–152.
23. Rossman, M. G., Liljas, A., Bränden, C.-I. & Banaszak, L. J. (1975) *The Enzymes*, ed. Boyer, P. D. (Academic, New York), 3rd Ed., Vol. 11, pp. 61–102.
24. White, S. W., Appelt, K., Dijk, J. & Wilson, K. S. (1983) *FEBS Lett.* **163**, 73–75.
25. Appelt, K., White, S. W., Wilson, K. S. & Tanaka, I. (1984) *FEBS Lett.* **165**, 43–45.