Acta Cryst. (1994). D**50**, 790–792

Crystallization and preliminary X-ray diffraction studies of bacterial ribosomal protein L14. By

CHRISTOPHER DAVIES, *Department of Microbiology, Duke University Medical Centre, Durham, North Carolina* 27710, *USA*, SUE ELLEN GERCHMAN and J. H. KYCIA, *Biology Department, Brookhaven National Laboratory, Upton, New York* 11973, *USA*, KEBBLER MCGEE, *Department of Microbiology, Duke University Medical Centre, Durham, North Carolina* 27710, *USA*, VENKI RAMAKRISHNAN, *Biology Department, Brookhaven National Laboratory, Upton, New York* 11973, *USA*, and STEPHEN W. WHITE, *Department of Microbiology, Duke University Medical Centre, Durham, North Carolina* 27710, *USA*

(*Received* 5 *April* 1994; *accepted* 18 *April* 1994)

Abstract

Based on amino-acid sequence homology, it is predicted that ribosomal protein L14 is a member of a recently identified family of structurally related RNA-binding proteins. To verify this, the gene for *Bacillus stearothermophilus* L14 has been cloned, and the protein has been purified and crystallized. The crystals are in space group *C*2 with cell dimensions a = 67.0, b = 32.7, c = 49.4 Å, and β = 101.8°, and there is one molecule in the asymmetric unit (V_m = 2.0 Å^3 Da^{-1}). They are of high quality, and a native data set has been collected to a resolution of 1.6 Å with an R_{merge} of 5.3%.

Introduction

The problem of detecting structurally related proteins based on amino-acid sequence information has yet to be satisfactorily solved. The fact that even proteins which are known to be related can have extensive differences in primary structure (Creighton, 1983) clearly shows that a particular fold can be realized by many sequences. However, considering the relative ease with which sequence information can be obtained as compared to structural information, it is clearly a goal worth pursuing.

Several years ago, it was shown that two prokaryotic ribosomal proteins, the C-terminal half of L7/L12 (Leijonmarck, Eriksson & Liljas, 1980) (L12CTF) and L30 (Wilson, Appelt, Badger, Tanaka & White, 1986), have remarkably similar structures in which two α-helices pack on the same side of an anti-parallel three-stranded β-pleated sheet (Leijonmarck *et al.*, 1988). L12CTF contains an extra α-helix, but this is somewhat distant from the body of the molecule and can be regarded as an extension of a loop region of L30. It was proposed that this similarity between the first two ribosomal protein structures might indicate that a larger subset shares the same fold (Leijonmarck *et al.*, 1988). Recently, the structures of several additional ribosomal proteins have been solved by NMR and protein crystallography. These are S5 (Ramakrishnan & White, 1992), S17 (Golden, Hoffman, Ramakrishnan & White, 1993), L6 (Golden, Ramakrishnan & White, 1993), L9 (Hoffman *et al.*, 1994) and S6 (Liljas, 1993). The structural motif first observed in the L30 and L7/L12 structures has also been identified in L6, L9 and S6, and this supports the idea of a sub-family of related ribosomal proteins. The same motif has also been found in the U1 RNA-binding domain of the U1snRNP-A protein (Nagai, Oubridge, Jessen, Li & Evans, 1990; Hoffman, Query, Golden, White & Keene, 1991), which raises the interesting possibility that these RNA-associated proteins are derived from an ancient RNA-binding domain (Golden, Ramakrishnan & White, 1993; Hoffman *et al.*, 1994).

There have been several attempts to analyze the primary structures of ribosomal proteins for evidence of evolutionary relatedness (see, for example, Wittmann-Liebold, Ashman & Dzionara, 1984; Jue, Woodbury & Doolittle, 1980) but none predicted the homologies described above. However, now that several ribosomal proteins have been solved (see above), the new structural information provides a firm basis for repeating the analysis with the specific aim of identifying any other ribosomal proteins that might be in the same structural sub-family. Such information would not only help in our understanding of protein structure but would also simplify the task of determining the ribosome's structure at the molecular level.

Using the *Sequence Analysis Software Package* of the Genetics Computer Group (GCG) (Devereux, Haeberli, & Smithies, 1984), a search was made for ribosomal proteins that are similar to L30, the smallest intact member of the family (White, unpublished results). Two proteins had significant homology, the C-terminal half of L7/L12 and residues 36–91 of L14. When analyzed in more detail, the L30 structure could be built into L14 without disturbing the L30 motif (see Fig. 1). The correct prediction of the L30–L7/L12 homology supports the L14 result. To test the result directly, however, the gene for L14 was cloned from *Bacillus stearothermophilus* and a crystallographic analysis of the protein was undertaken.

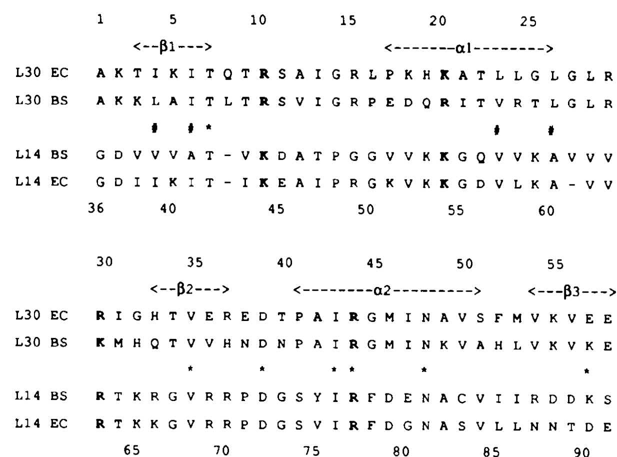

Fig. 1. An alignment of L14 and L30 sequences from *B. stearothermophilus* (BS) and *E. coli* (EC) using one-letter amino-acid codes. The four most crucial hydrophobic core residues in the L30 structure are labelled with a #. Conserved residues between the two *B. stearothermophilus* sequences are labelled with a * (residue 57 of L14 belongs to both categories). The bold residues are conserved lysine and arginine residues that may be involved in RNA binding. The secondary structure shown is that of L30.

© 1994 International Union of Crystallography
Printed in Great Britain – all rights reserved

ISSN 0907-4449 ©1994

Fortunately, L14 is also one of the more important ribosomal proteins, and its high-resolution structure will be of great value in understanding the translation process. It is a 13.3 kDa protein from the large subunit of the ribosome and consists of 122 residues. It is one of the most conserved ribosomal proteins with 69% identity between the *Escherichia coli* and *B. stearothermophilus* sequences (Kimura, Kimura & Ashman, 1985). In common with two other highly conserved ribosomal proteins (L2 and L5), L14 has been shown to be important for the poly(U) and peptidyl transferase activities of the ribosome (Auron & Fahnestock, 1981). Immunoelectron microscopy has localized L14 on the surface of the 50S ribosomal subunit at the 30S–50S subunit interface (Hackl, Stöffler-Meilicke & Stöffler, 1988), and it can be crosslinked to L19 and L32 (Walleczek, Schüler, Stöffler-Meilicke, Brimacombe & Stöffler, 1988). L14 binds directly to 23S ribosomal (r) RNA (Herold & Nierhaus, 1987), and can be crosslinked to a site between bases 1990 and 2000 in helix 61 (Brimacombe *et al.*, 1990). This region of 23S rRNA forms part of the proposed tRNA-binding site (Mitchell, Osswald, Schüler & Brimacombe, 1990).

Experimental

Gene cloning and sequencing

The procedure for cloning and sequencing the L14 gene from *B. stearothermophilus* was the same as that described previously (Gerchman, Graziano & Ramakrishnan, 1994; Ramakrishnan & Gerchman, 1991) with the following modifications.

Polymerase chain reaction using *Taq* DNA polymerase (Promega) was carried out with genomic *B. stearothermophilus* DNA as the template in a 20 μl reaction. Amplification consisted of 40 cycles of melting at 367 K for 30 s, hybridization at 315 K for 1.5 min, and primer extension at 345 K for 1.5 min. The amplified DNA was cloned into the *Nde*I–*Bam*HI site of the plasmid vector pET-13a (Gerchman, Graziano & Ramakrishnan, 1994).

Protein purification

E. coli cells containing the expression vector were grown in a medium that contained 25 g l^{-1} of Luria Broth (Gibco) and 25 mg l^{-1} of kanamycin. Two 1 l flasks each containing 750 ml of medium were inoculated with 3 ml of an overnight culture. When the OD_{550} had reached 0.6, the cells were induced by adding isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.4 m*M*. After 3 h, the cells were harvested by centrifugation at 500*g*, and stored at 253 K in a 50 ml buffer containing 20 m*M* Tris–HCl, pH 7.0, 10 m*M* $MgCl_2$ and 2 mg each of lysozyme and DNase. To purify L14, the cell sample was defrosted and spun at 30 000*g* for 20 min, and the resulting supernatant was applied to an S-sepharose column (Pharmacia) equilibrated in 20 m*M* Tris–HCl, pH 7.0. L14 was obtained by eluting the column with a salt gradient of 0–1 *M* NaCl. The fractions containing L14 were determined by their absorbance at 276 nm and confirmed by SDS–PAGE.

Crystallization

The protein was dialysed against 20 m*M* Tris–HCl, pH 7.5 and concentrated to 30 mg ml^{-1} using Centricon-10 microconcentrators (Amicon). The crystallization method used was that of vapor diffusion using hanging drops (Davies & Segal, 1971). In these experiments, 5 μl of the protein solution was mixed with 5 μl of well solution containing precipitant and buffer. A variety of precipitants and buffers were tested, and all trials were performed at 295 K.

X-ray analysis

The crystals were initially characterized on an Enraf–Nonius precession camera mounted on a Rigaku RU300 rotating-anode generator with a copper target and a nickel filter, operating at 40 kV and 80 mA. Diffraction data were collected on a Rigaku RAXIS-II image-plate system (Molecular Structures Corporation) mounted on the same generator and at the same power. A total of 102° of data were collected in steps of 3° using standard oscillation geometry and processed using the RAXIS-II data-processing software.

Results

The sequence alignment between L30 and residues 36–91 of L14 from both *E. coli* and *B. stearothermophilus* is shown in Fig. 1. Over this stretch, 8 and 18 residues are identical between the *B. stearothermophilus* and *E. coli* sequences, respectively. Where changes do occur many are conservative. The alignment did not introduce any gaps into the secondary structural elements of L30, and modelling has shown that after mutating residues in the structure of L30 to those of L14, no hydrophilic residues point to the inside of the protein and prolines remain in turns. Furthermore, the four most important core residues in L30 and L7/L12 are conserved (Leijonmarck *et al.*, 1988). Finally, four of the conserved arginine and lysine residues in L30 that may mediate RNA binding (Wilson, Appelt, Badger, Tanaka & White, 1986) are also conserved in L14 thus indicating a similar RNA-binding mechanism.

Crystals of *B. stearothermophilus* L14 were obtained using a variety of conditions, namely 1.8–2.3 *M* ammonium sulfate, 100 m*M* Tris–HCl pH 8.2–8.8; 14–24% polyethylene glycol 4000, 100 m*M* Tris–HCl pH 8.2–8.8; 0.8–1.3 *M* phosphate,

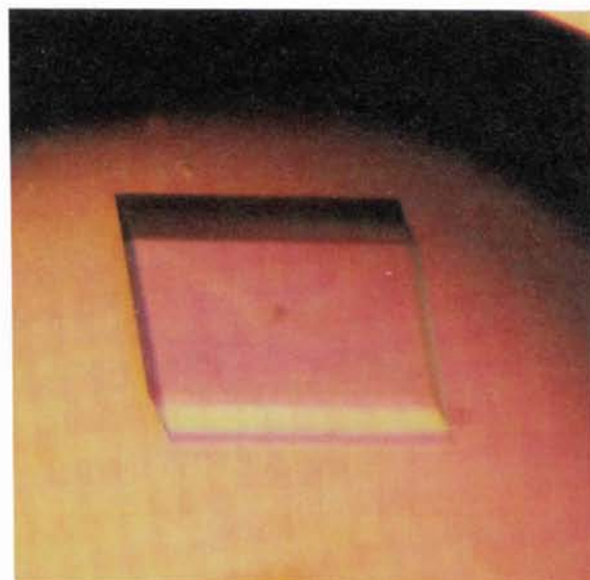

Fig. 2. A crystal of ribosomal protein L14 from *B. stearothermophilus* grown by the hanging-drop method using sodium citrate as a precipitant. The dimensions are 0.7 × 0.5 × 0.3 mm. The monoclinic angle of 101.8° is evident from the external morphology.

pH 5.2–5.8; and 0.9–1.3 *M* sodium citrate, 100 m*M* Tris–HCl pH 6.8–8.0. However, most grow as thin plates with a high degree of mosaicity and only the crystals grown in citrate are suitable for X-ray analysis. The latter initially appeared in 2–3 weeks from protein precipitate, but large single crystals are now grown reproducibly in 2–3 d by seeding from showers of microcrystals. The crystals are rhomboid shaped with typical dimensions 0.7 × 0.5 × 0.3 mm, and the monoclinic β angle of 101.8° (see below) is clearly visible (Fig. 2).

The crystals diffract to at least 1.6 Å and are stable in the X-ray beam for more than a week. Precession photographs showed the space group to be *C*2 with cell dimensions $a = 67.0$, $b = 32.7$ and $c = 49.4$ Å and $\beta = 101.8°$ (Fig. 3). Assuming that there is one molecule in the asymmetric unit, the crystal density V_m is 2.0 Å^3 Da^{-1}. This is low for typical protein crystals (Matthews, 1968) and corresponds to a solvent content of 38.2%. However, it is entirely consistent with their excellent quality of diffraction and resistance to radiation damage. A native data set to a resolution of 1.64 Å has been collected which is 92.4% complete. A total of 24 918 reflections were collected, and these reduced to 11 827 independent reflections with an R_{merge} of 5.3% ($R_{\text{merge}} = \sum_h \sum_i |I_{ih} - \langle I_h \rangle| / \sum_h \sum_i \langle I_h \rangle$ where $\langle I_h \rangle$ is the mean intensity of the i observations of reflection h). An attempt to solve the structure by molecular replacement using Patterson correlation refinement in *X-PLOR* (Brünger, 1990) with the structure of L30 as a search model was unsuccessful and a search for suitable heavy-atom derivatives is currently underway.

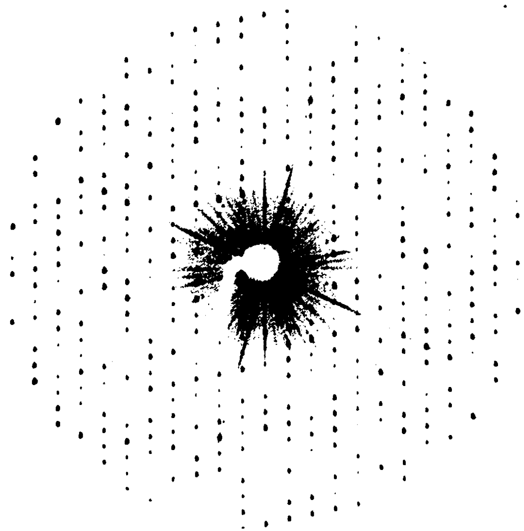

Fig. 3. A 15° screened precession photograph of the *h*0*l* zone from a crystal of *B. stearothermophilus* L14. The exposure time is 12 h. The edge of the photograph corresponds to a resolution of about 3 Å.

This work was supported by NIH grant GM 44973 (to SWW and VR) and by a grant from the Office of Health and Environmental Research of the US Department of Energy (to VR).

References

AURON, P. E. & FAHNESTOCK, S. R. (1981). *J. Biol. Chem.* **256**, 10105–10110.

BRIMACOMBE, R., GORNICKI, P., GREUER, B., MITCHELL, P., OSSWALD, M., RINKE-APPEL, J., SCHÜLER, D. & STADE, K. (1990). *Biochim. Biophys. Acta*, **1050**, 8–13.

BRÜNGER, A. T. (1990). *Acta Cryst.* A**46**, 46–57.

CREIGHTON, T. E. (1983). *Proteins – Structure and Molecular Properties*. New York: W. H. Freeman.

DAVIES, D. R. & SEGAL, D. M. (1971). *Methods Enzymol.* **22**, 266–269.

DEVEREUX, J., HAEBERLI, P. & SMITHIES, O. (1984). *Nucleic Acids Res.* **12**, 387–395.

GERCHMAN, S. E., GRAZIANO, V. & RAMAKRISHNAN, V. (1994). *Protein Expr. Purif.* **5**, 242–251.

GOLDEN, B. L., HOFFMAN, D. W., RAMAKRISHNAN, V. & WHITE, S. W. (1993). *Biochemistry*, **32**, 12812–12820.

GOLDEN, B. L., RAMAKRISHNAN, V. & WHITE, S. W. (1993). *EMBO J.* **12**, 4901–4908.

HACKL, W., STÖFFLER-MEILICKE, M. & STÖFFLER, G. (1988). *FEBS Lett.* **233**, 119–123.

HEROLD, M. & NIERHAUS, K. N. (1987). *J. Biol. Chem.* **262**, 8826–8833.

HOFFMAN, D. W., DAVIES, C., GERCHMAN, S. E., KYCIA, J. H., PORTER, S. J., WHITE, S. W. & RAMAKRISHNAN, V. (1994). *EMBO J.* **13**, 205–211.

HOFFMAN, D. W., QUERY, C. C., GOLDEN, B. L., WHITE, S. W. & KEENE, J. K. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 2495–2499.

JUE, R. A., WOODBURY, N. W. & DOOLITTLE, R. F. (1980). *J. Mol. Evol.* **15**, 129–148.

KIMURA, M., KIMURA, J. & ASHMAN, K. (1985). *Eur. J. Biochem.* **153**, 491–497.

LEIJONMARCK, M., APPELT, K., BADGER, J., LILJAS, A., WILSON, K. & WHITE, S. W. (1988). *Proteins*, **3**, 243–251.

LEIJONMARCK, M., ERIKSSON, S. & LILJAS, A. (1980). *Nature (London)*, **286**, 824–826.

LILJAS, A. (1993). Personal communication.

MATTHEWS, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

MITCHELL, P., OSSWALD, M., SCHÜLER, D. & BRIMACOMBE, R. (1990). *Nucleic Acids Res.* **18**, 4325–4333.

NAGAI, K., OUBRIDGE, C., JESSEN, T. H., LI, J. & EVANS, P. R. (1990). *Nature (London)*, **348**, 515–520.

RAMAKRISHNAN, V. & GERCHMAN, S. E. (1991). *J. Biol. Chem.* **266**, 880–885.

RAMAKRISHNAN, V. & WHITE, S. W. (1992). *Nature (London)*, **358**, 768–771.

WALLECZEK, J., SCHÜLER, D., STÖFFLER-MEILICKE, M., BRIMACOMBE, R. & STÖFFLER, G. (1988). *EMBO J.* **7**, 3571–3576.

WILSON, K. S., APPELT, K., BADGER, J., TANAKA, I. & WHITE, S. W. (1986). *Proc. Natl Acad. Sci. USA*, **83**, 7251–7255.

WITTMANN-LIEBOLD, B., ASHMAN, K. & DZIONARA, M. (1984). *Mol. Gen. Genet.* **196**, 439–448.