Vol. 187, No. 1, 1992 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
August 31, 1992 Pages 58-64

THE PRIMARY STRUCTURE OF RAT RIBOSOMAL PROTEIN L3

Yuh Kuwano and Ira G. Wool*

Department of Biochemistry and Molecular Biology,
The University of Chicago, Chicago, Illinois 60637

Received July 9, 1992

SUMMARY: The amino acid sequence of the rat 60S ribosomal subunit protein L3 was deduced from the sequence of nucleotides in a recombinant cDNA. Ribosomal protein L3 has 403 amino acids and has a molecular weight of 46,106. Hybridization of the cDNA to digests of nuclear DNA suggests that there are 7 to 9 copies of the L3 gene. The mRNA for the protein is about 1,400 nucleotides in length. Rat L3 is homologous to ribosomal proteins from other eukaryotes and to proteins from eubacterial, archaebacterial, and chloroplast ribosomes.

Solution of the structure, and understanding of the function, of eukaryotic ribosomes requires knowledge of the chemistry of the molecular components. For this reason we have undertaken to determine the sequence of nucleotides and of amino acids in the constituent nucleic acids and proteins in mammalian (rat) ribosomes (1). As part of this endeavor we have determined the primary structure of rat ribosomal protein L3 from the sequence of nucleotides in a recombinant cDNA.

Mutations in yeast ribosomal protein L3, which is homologous to rat L3, confers on ribosomes resistance to the antibiotic trichodermin (2). Trichodermin inhibits protein synthesis by interfering with the peptidyl transferase reaction (3); hence L3 is likely to be at the site of peptide bond formation in ribosomes. In addition, the ability of yeast to secrete the toxin K1 and immunity to the killer trait of the protein are determined by a double-stranded RNA called M; the replication and maintenance of M depends on ribosomal protein L3 (4). Finally, *Escherichia coli* ribosomal protein L3, which is also related to rat L3, participates in the initiation of the assembly of 50S ribosomal subunits (5).

*To whom correspondence should be addressed at Department of Biochemistry and Molecular Biology, The Universtiy of Chicago, 920 East 58th Street, Chicago, Illinois 60637. Telefax: 312-702-0439.

0006-291X/92 $4.00

MATERIAL AND METHODS

The recombinant DNA procedures and the methods used to determine the sequence of nucleotides in the nucleic acids have been described or cited (6, 7). An oligodeoxynucleotide probe for the cDNA encoding rat ribosomal protein L3 was synthesized predicated on a sequence of amino acids, RIAKEEGA, present at the carboxyl terminus of a mouse protein related to yeast ribosomal protein L3 (8); the termination codon ATT was added to this oligodeoxynucleotide. This probe was a mixture of 384 different oligodeoxynucleotides, each 27 bases long. The oligodeoxynucleotides were synthesized on a solid support by the methoxyphosphoramidite method using an Applied Biosystems, 380B, DNA synthesizer (9). The probe was made radioactive by labeling the 5' end with [^{32}P]ATP and T4 polynucleotide kinase (6, 7).

The *in vitro* transcription and translation procedures used to identify the protein encoded in the open reading frame in pcD-L3 also have been described or cited (10).

RESULTS AND DISCUSSION

The Sequence of Nucleotides in a Recombinant cDNA Encoding Rat Ribosomal Protein L3

A random selection of 20,000 cells from two cDNA libraries of 20,000 and 30,000 independent transformants that had been constructed from regenerating rat liver poly(A)$^+$mRNA was screened for clones that hybridized to an oligodeoxynucleotide probe that was synthesized to be complementary to the sequence of nucleotides predicted to be present in the mRNA for the rat ribosomal protein related to a mouse protein that is in turn related to yeast ribosomal protein L3 (8). Three clones gave a positive hybridization signal with the probe. The DNA from the plasmids of the 3 transformants was isolated and digested with restriction endonucleases. These clones had inserts that were about 150 nucleotides. Preliminary analysis of the sequences of nucleotides in these clones indicated that they encoded a rat protein related to mouse L3. The cDNA insert in one of these clones was made radioactive and was used as a probe to screen a second rat cDNA library (kindly provided by M. Brownstein, NIH). One clone hybridized to this probe. The cDNA inserts in this clone was about 1.4 Kb in length; the sequence of nucleotides in both strands of the cDNA was obtained.

The cDNA insert in pcD-L3 is 1,316 nucleotides long and has a 5' noncoding sequence of 30 bases, a single open reading frame of 1,212, a 3' noncoding sequence of 55 and a long poly(A) stretch (Fig. 1). In the other two reading frames the sequence is interrupted by termination codons. The open reading frame begins at an ATG codon at a position that we designate +1 and ends with a termination codon (TGA) at position 1,210; it encodes 403 amino acids (Fig. 1). The initiation codon occurs in the context GAG*ATG*T which deviates from the consensus ACC*ATG*G (11). The hexamer AATAAA that directs post-transcriptional cleavage-polyadenylation of the 3' end of the precursor of the mRNA (12) is at position 1250-1255, 12 nucleotides upstream of the start of the poly(A) stretch.

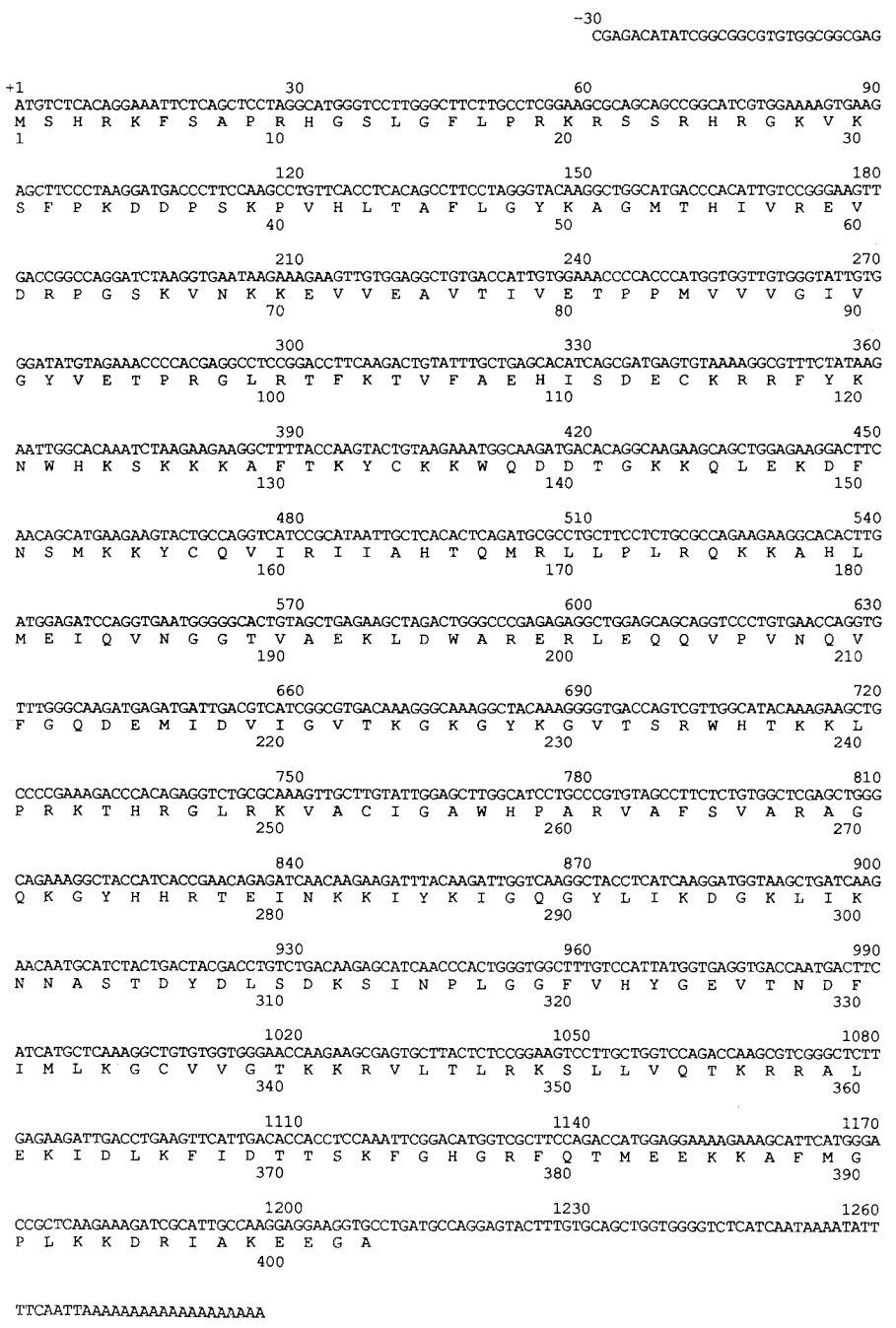

Fig. 1. *The sequence of nucleotides in the cDNA insert in plasmid pcD-L3 and the amino acid sequence encoded in the open reading frame.* The position of the nucleotide in the cDNA is given above the residue; the position of the amino acid in protein L3 is designated below the residue.

The Primary Structure of Rat Ribosomal Protein L3

The rat ribosomal protein encoded in the open reading frame in pcD-L3 was identified as L3 by transcription of the L3 cDNA, which was subcloned in pGEM-2-L3, and translation of

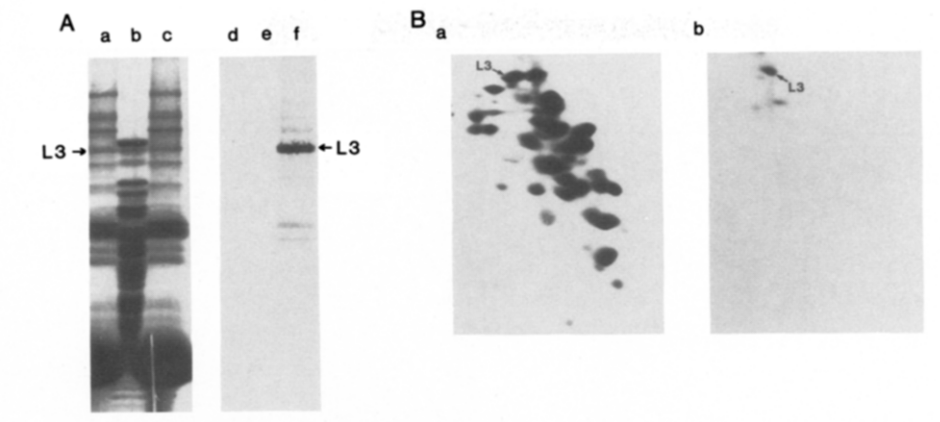

Fig. 2. *Analysis of the product of the translation of the pGEM-2-L3 cDNA transcript.* A reticulocyte lysate (50 µl) was incubated with the RNA transcript of the pGEM-2-L3 cDNA (1 µg). A sample (15 µl) of the lysate containing the product of the translation, which was labeled with [^{35}S]methionine, was extracted with 67% acetic acid and the protein was precipitated with 90% acetone (10). In *A*, electrophoresis was in one-dimension in SDS-polyacrylamide gels: *a* and *d*, the protein from 15 µl of reticulocyte to which no transcript was added; *b* and *e*, 20 µg of the proteins from rat 60S ribosomal subunits; *c* and *f*, the protein from 15 µl of reticulocyte lysate which had been programmed with transcript from pGEM-2-L3. In *a-c* the gels were stained with Coomassie brilliant blue; in *d-f* radioautographs were made and visualized by fluorography. In *B*, electrophoresis in polyacrylamide gels containing urea was from left to right in the first dimension and from top to bottom in the second: *a* and *b*, the protein from 15 µl of reticulocyte lysate which had been programmed with transcript from pGEM-2-L3 and supplemented before analysis with 20 µg of the proteins from rat 60S ribosomal subunits. In *a* the gel was stained with Coomassie brilliant blue and in *b* visualization of the radioautograph was by fluorography. The location of rat ribosomal protein L3 (designated by arrows) was determined with purified L3 (9).

the RNA transcript in a nuclease-treated reticulocyte lysate (10); identification of the radioactive product of translation was from its migration in one-dimensional SDS and in two-dimensional polyacrylamide-urea gels (Fig. 2).

The molecular weight of rat ribosomal protein L3, calculated from the sequence of amino acids deduced from pcD-L3 is 46,106 somewhat greater than the 37,800 estimated before (13) from SDS-PAGE of the purified protein. We do not know whether the NH_2-terminal methionine encoded in the L3 mRNA is removed after translation. However, the residue next to the initial methionyl in L3 is seryl which has been reported (14) to favor NH_2-terminal processing.

Protein L3 has a large excess of basic residues (30 arginyl, 56 lysyl, and 16 histidyl) over acidic ones (18 aspartyl and 20 glutamyl) (Table I). The basic residues, as is common for ribosomal proteins, tend to be clustered; for example, 8 of the 12 residues at positions 19-30; 12 of the 22 between positions 115 and 136; and 13 of the 27 at positions 224-250 (Fig. 1). L3 has a number of hydrophilic regions; for example, 17 of the 37 residues at positions 3-39 and 23 of the 42 residues at positions 108-149 are charged (Fig. 1).

TABLE I. Amino acid composition of rat ribosomal protein L3

Amino Acid	A	B
Alanine	23	21
Arginine	30	30
Aspartic acid and asparagine	29	18 + 10
Cysteine	n.d.	5
Glutamic acid and glutamine	39	20 + 14
Glycine	36	33
Histidine	13	16
Isoleucine	19	21
Leucine	27	26
Lysine	38	56
Methionine	9	10
Phenylalanine	18	17
Proline	19	15
Serine	21	18
Threonine	25	24
Tryptophan	n.d.	5
Tyrosine	10	11
Valine	35	33
Residues		403

The amino acid composition (in numbers of residues) was determined either (A) from an hydrolysate of purified L3 (9) or inferred (B) from the sequence of nucleotides in pcD-L3.

The Number of Copies of the L3 Gene

The cDNA insert in pcD-L3 was made radioactive and used to probe digests of rat liver DNA made with restriction endonucleases *Bam*HI, *Eco*RI, or *Hin*dIII (7). The number of hybridization bands suggest that there are 7 to 9 copies of the L3 gene (data not shown). Many other mammalian ribosomal protein genes have been found to be present in multiple copies (cf. (1) for references and discussion). However, in no instance has it been shown that more than one of the genes is functional; the presumption is that the other copies are retroposon pseudogenes.

The Size of the mRNA Encoding Rat Ribosomal Protein L3

To determine the size of the mRNA coding for L3, poly(A)$^+$mRNA from rat liver was separated by electrophoresis and screened for hybridization bands using radioactive pcD-L3 cDNA. One distinct band of about 1,400 nucleotides was detected (data not shown).

```
          1        10        20
RL3       MSHRKFSAPRHGSLGFLPRKR
ScL3      MSHRKYEAPRHGHLGFLPRKR
common    MSHRK  APRHG LGFLPRKR
```

Fig. 3. The NH_2-terminal amino acid sequences of yeast and rat ribosomal proteins L3 containing a putative nuclear localization signal.

Comparison of the Sequence of Amino Acids in Rat L3 with Ribosomal Proteins from Other Species

The sequence of amino acids in rat L3 was compared, using the computer programs RELATE and ALIGN (15), to those in more than 1,000 other ribosomal proteins contained in a library that we have compiled. The comparison that yielded the closest identity was with mouse L3 (8); the RELATE score is 102.3 S.D. units. In an alignment of the amino acid sequences there are 401 identities out of 403 possible matches (the ALIGN score is 170.6). Rat L3 is also related to the following proteins: to *Saccharomyces cerevisiae* L3 (16) - the RELATE score is 75.5 and in an alignment of the amino acid sequences there are 260 identitities in 387 possible matches (the ALIGN score is 109.1); to *Arabidopsis thaliana* ARP1 (17) - the RELATE score is 65.1 and in an alignment there are 253 identities in 388 possible matches (the ALIGN score is 111.0); *Halobacterium marismortui* HmaL3 (18) - the RELATE score is 20.0 and in an alignment there are 119 identities in 334 possible matches (the ALIGN score is 41.0); *Mycoplasma capricolum* L3 (19) - the RELATE score is 8.3 and in an alignment there are 45 identities in 185 possible matches (the ALIGN score is 9.4); *Yersinia pseudotuberculosis* L3 (20) - the RELATE score is 6.9 - and in an alignment there are 25 identities in 108 possible matches (the ALIGN score is 7.4); and to *Escherichia coli* L3 (21) - the RELATE score is 6.1 and in an alignment there are 42 identities in 156 possible matches (the ALIGN score is 10.4). Thus there is strong conservation of this ribosomal protein across the three kingdoms suggesting it was a component of an early ribonucleoprotein ribosome.

The NH_2-terminal 21 amino acids in yeast L3 are responsible for the transport of the protein from the cytoplasm where it is synthesized to the nucleus where it is assembled into ribosomes (22). Of the 21 amino acids, 18 are identical in rat L3 (Fig. 3), suggesting that the NH_2-terminal sequence in the rat protein is also responsible for nuclear localization.

The sequence of amino acids in L3 was searched for internal duplications. None were found.

The determination of the sequence of amino acids in rat L3 is a contribution to a data set which it is hoped will eventually include the structure of all the proteins in the ribosomes of this mammalian species. The primary purpose for the accumulation is its anticipated use in arriving

at a solution of the structure of the organelle. However, the information may also help in understanding the evolution of ribosomes, in unraveling the function of the proteins, in defining the rules that govern the interaction of the proteins and the rRNAs, and in uncovering the amino acid sequences that direct the proteins to the nucleolus for assembly on nascent rRNA.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health Grant GM 21769. We are grateful to Yuen-Ling Chan for advice and for technical assistance, to Anton Glück and Katsuyuki Suzuki for computer analyses, and to Arlene Timosciek for aid in the preparation of the manuscript. The sequences are in the EMBL/Gen Bank/DDBJ Nucleotide Sequence Databases under the accession number X62166.

REFERENCES

1. Wool, I.G., Endo, Y., Chan, Y.L., and Glück, A. (1990) in The Ribosome: Structure, Function, and Evolution (Hill, W.E., *et al.* ed.), pp. 203-214, Amer. Soc. Microbiol., Washington, D.C.
2. Fried, H.M., and Warner, J.R. (1981) Proc. Natl. Acad. Sci. USA 78, 238-242.
3. Carrasco, L., Barbacid, M., and Vazquez, D. (1973) Biochim. Biophys. Acta 312, 368-376.
4. Wickner, R.D., Ridley, S.P., Fried, H.M., and Ball, S.G. (1982) Proc. Natl. Acad. Sci. USA 79, 4706-4708.
5. Franceschi, F.J., and Nierhaus, K.H. (1988) Biochemistry 27, 7056-7059.
6. Chan, Y.L., Lin, A., McNally, J., and Wool, I.G. (1987) J. Biol. Chem. 262, 12879-12886.
7. Chan, Y.L., and Wool, I.G. (1988) J. Biol. Chem. 263, 2891-2896.
8. Peckham, I., Sobel, S., Comer, J., Jaenisch, R., and Barklis, E. (1989) Genes Dev. 3, 2067-2071.
9. Beaucage, S.L., and Caruthers, M.H. (1981) Tetrahedron Lett. 22, 1859-1862.
10. Chan, Y.L., Devi, K.R.G., Olvera, J., and Wool, I.G. (1990) Arch. Biochem. Biophys. 283, 546-550.
11. Kozak, M. (1986) Cell 44, 283-292.
12. Proudfoot, N.J., and Brownlee, G.G. (1976) Nature 263, 211-214.
13. Tsurugi, K., Collatz, E., Todokoro, K., and Wool, I.G. (1977) J. Biol. Chem. 252, 3961-3969.
14. Flinta, C., Persson, B., Jörnvall, H., and von Heijne, G. (1986) Eur. J. Biochem. 154, 193-196.
15. Dayhoff, M.O. (1978) in Atlas of Protein Sequence and Structure (Dayhoff, M.O., ed.) Vol. 5, Suppl. 3, pp. 1-8, National Biomedical Research Foundation, Washington, D.C.
16. Schultz, L.D., and Friesen, J.D. (1983) J. Bacteriol. 155, 8-14.
17. Kim, Y., Zhang, H., and Scholl, R.L (1990) Gene 93, 177-182.
18. Arndt, E., Kroemer, W., and Hatakeyama, T. (1990) J. Biol. Chem. 265, 3034-3039.
19. Ohkubo, S., Muto, A., Yamao, F., and Osawa, S. (1987) Mol. Gen. Genet. 210, 314-322.
20. Gross, U., Chen, J-H., Kono, D.H., Lobo, J.G., and Yu, D.T.Y. (1989) Nucleic Acids Res. 17, 3601-3602.
21. Wittmann-Liebold, B. (1984) Adv. Protein Chem. 36, 56-78.
22. Moreland, R.B., Nam, H.G., Heresford, L.M., and Fried, H.M. (1985) Proc. Natl. Acad. Sci. USA 82, 6561-6565.