Vol. 201, No. 1, 1994 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
May 30, 1994 Pages 108-114

# THE PRIMARY STRUCTURE OF RAT RIBOSOMAL PROTEIN L15

Yuen-Ling Chan, Joe Olvera, and Ira G. Wool*

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

Received April 1, 1994

SUMMARY: The amino acid sequence of the rat 60S ribosomal subunit protein L15 was deduced from the sequence of nucleotides in two recombinant cDNAs. Ribosomal protein L15 has 203 amino acids, the $NH_2$-terminal methionine is removed after translation of the mRNA, and has a molecular weight of 24,000. Hybridization of the cDNA to digests of nuclear DNA suggests that there are 13 to 15 copies of the L15 gene. The mRNA for the protein is about 850 nucleotides in length. Rat L15 is related to ribosomal proteins from other eukaryotes. Rat L15 has the hexapeptide, TYKFFE, that also occurs in the amyloidogenic glycoprotein A4 which is associated with Alzheimer's disease and Down's Syndrome. 

An attempt is being made to accumulate a set of data which it is hoped will eventually encompass the sequence of amino acids in all the proteins in the ribosomes of a mammalian species, the rat (1). The motivation for this compilation is the value it is perceived too have in arriving at the solution of the structure of the organelle and, perhaps, in enabling one to provide a coherent account of the biochemistry underlying its function in protein synthesis. As a part of this undertaking, we report here the amino acid sequence of rat ribosomal protein L15

## *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 (2-4). Rat ribosomal protein L15 was isolated from a fraction of all the 60S subunit proteins (5) by high performance liquid chromatography in a size exclusion column (Bio-Rad, Bio-Sil TSK-125, 300 x 7.5 mm) with 0.1 M ammonium acetate (pH 4.1) as the mobile phase. The sequence of the $NH_2$-terminal 25 amino acids of L15, and the sequence of 37 amino acids in a peptide prepared from L15 by acid hydrolysis of an aspartylprolyl bond, were determined by Edman degradation in an Applied Biosystems, Model 470A, automated gas phase sequencer. The $NH_2$-terminal amino acid

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

0006-291X/94 $5.00

sequence was used to design an oligodeoxynucleotide probe - a mixture of 128 oligodeoxynucleotides, each 29 bases in length, that encoded residues 1-10 in L15. The oligodeoxynucleotides were synthesized on a solid support by the methoxyphosphoramidite method using an Applied Biosystems, Model 380B, DNA synthesizer.

## *RESULTS AND DISCUSSION*

### *The Sequence of Nucleotides in a Recombinant cDNA Encoding Rat Ribosomal Protein L15*

A random selection of 60,000 colonies from two cDNA libraries of 20,000 and 30,000 independent transformants that had been constructed from regenerating rat liver poly(A)$^+$mRNA (2, 3) was screened for clones that hybridized to an oligodeoxynucleotide probe that was related to the $NH_2$-terminal amino acid sequence in rat ribosomal protein L15. Five clones gave a positive hybridization signal with the probe. The DNA from the plasmids of two of the transformants was isolated and digested with restriction endonucleases. The sequence of nucleotides in both strands of the cDNA inserts in these two clones, pL15-1 and pL15-15, was determined. In pL15-15 the cDNA insert is 540 nucleotides long, has a 5' noncoding sequence of 16 bases, a single open reading frame of 524 nucleotides; but no termination codon and no 3' noncoding sequence (Fig. 1). The cDNA insert in pL15-1 is 197 nucleotides long, with a

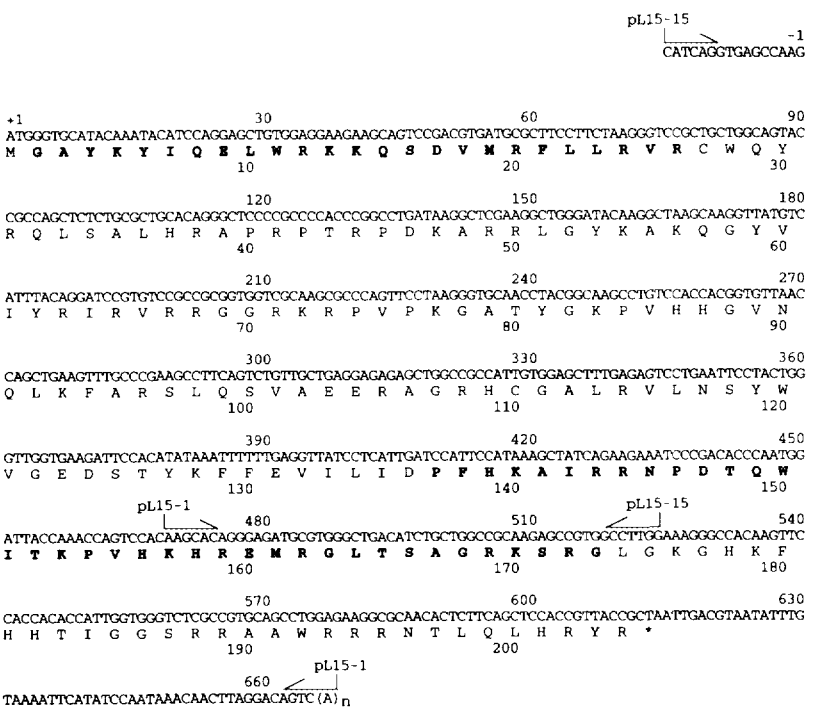


Fig. 1. *The sequence of nucleotides in the cDNA insert in pL15-1,15 and the amino acid sequence encoded in the open reading frame.* The positions of the nucleotides in the cDNA are given above the residue; the positions of the amino acids in protein L15 are designated below the residue. The initial and terminal nucleotides in pL15-15 and pL15-1 are designated by the vertical of the bent arrows. The amino acid residues in bold letters were determined also by Edman degradation of L15.

coding sequence of 147 nucleotide, and a 3' noncoding sequence of 50 bases and a long poly(A) stretch; it lacks a 5' noncoding sequence and an initiation codon (Fig. 1). The overlapping sequences in pL15-15 and pL15-1 are identical, hence, they are likely to be derived from the same gene. The two cDNAs together encode all of L15 and for convenience we shall refer to them as pL15-1,15.

The cDNA in pL15-1,15 is 681 nucleotides long, has 5' and 3' noncoding sequences of 16 and 50 bases, and an open reading frame of 615 nucleotides (Fig. 1). The open reading frame begins at an AUG codon at a position that we designate +1 and ends with a termination codon (TAA) at position 613; it encodes 204 amino acids (Fig. 1). The initiation codon occurs in the context AAG*ATG*G; this is similar to the consensus sequence, ACC*ATG*G (6). The hexamer AATAAA that directs posttranscriptional cleavage-polyadenylation of the 3' end of the precursor of the mRNA (7) is at position 645-650, 15 nucleotides upstream of the start of the poly(A) stretch.

*The Primary Structure of Rat Ribosomal Protein L15*

The rat ribosomal protein encoded in the open reading frame in pL15-1,15 was identified by the correspondence of the amino acid sequence encoded in the plasmid DNA with the $NH_2$-terminal 25 amino acids in L15, and with a sequence of 37 amino acids (positions 137-173) in a peptide prepared from L15, determined directly by Edman degradation (Fig. 1). In addition, the amino acid composition inferred from the sequence of nucleotides in pL15-1,15 and that determined from an hydrolysate of the protein (8) are in close agreement (Table I).

The molecular weight of rat ribosomal protein L15, calculated from the sequence of amino acids deduced from pL15-1,15, is 24,131. However, the $NH_2$-terminal methionine encoded in the L15 mRNA is removed after translation, i.e. it is not in the $NH_2$-terminal sequence determined directly from the protein. Thus, the mature processed L15 has 203 residues and the molecular weight is 24,000, close to the 24,500 estimated before (8) from SDS-PAGE of the purified protein.

Protein L15 has a large excess of basic residues (33 arginyl, 17 lysyl, and 11 histidyl) over acidic ones (5 aspartyl and 6 glutamyl) (Table I). There are several clusters of basic residues; for example, 7 of 11 amino acids at positions 63-73, and 5 of 7 residues at 156-162 (Fig. 1). There are two highly charged regions in L15; one is at positions 37-56, and the other is at 62-77. L15 does not have an unusually large number of hydrophobic amino acids for a ribosomal protein, 59 of 203 residues; there is, however, one notable hydrophobic district at positions 129-135. There are 11 histidines, i.e. 5.4 mole%, in L15; this is a relatively large amount for a ribosomal protein - the average is 2.4 mole%. There are 5 tryptophans (2.4 mole%) in L15; this too is an unusually large number - the average for a ribosomal protein is

TABLE I. Amino acid composition of rat ribosomal protein L15

| Amino Acid | A | B |
|---|---|---|
| Alanine | 15 | 14 |
| Arginine | 34 | 33 |
| Aspartic acid and asparagine | 10 | 5 + 4 |
| Cysteine | n.d. | 2 |
| Glutamic acid and glutamine | 15 | 6 + 9 |
| Glycine | 21 | 18 |
| Histidine | 14 | 11 |
| Isoleucine | 7 | 8 |
| Leucine | 16 | 15 |
| Lysine | 18 | 17 |
| Methionine | 1 | 3* |
| Phenylalanine | 6 | 6 |
| Proline | 9 | 9 |
| Serine | 9 | 9 |
| Threonine | 8 | 8 |
| Tryptophan | n.d. | 5 |
| Tyrosine | 9 | 10 |
| Valine | 11 | 12 |
| Residues | | 204* |

The amino acid composition (in numbers of residues) determined either (A) from a hydrolysate of purified L15 (8) or inferred (B) from the sequence of nucleotides in pL15-1,15.

*The $NH_2$-terminal methionine is removed after translation of the mRNA, thus the mature protein has 203 residues.

n.d., not determined.

0.7 mole%. There are several dipeptide repeats in the sequence at positions 181-194, HHTIGGSRRAAWRR (Fig. 1).

*The Number of Copies of the L15 Gene*

The cDNA insert in pL15-15 was made radioactive and used to probe separate digests of rat liver DNA made with the restriction endonucleases *Bam*HI, or *Eco*RI, or *Hin*dIII (3). The number of hybridization bands suggests that there are 13 to 15 copies of the L15 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 L15*

To determine the size of the mRNA coding for L15, poly(A)$^+$mRNA from rat liver was separated by electrophoresis and screened for hybridization bands using radioactive pL15-15 cDNA. One distinct band of about 850 nucleotides was detected (data not shown).

*Comparison of the Sequence of Amino Acids in Rat L15 with Ribosomal Proteins from Other Species*

The sequence of amino acids in rat L15 was compared, using the computer programs RELATE and ALIGN (9), to those in more than 1,400 other ribosomal proteins contained in a library that we have compiled; in addition, the program TFASTA (10) was used to search the GenBank DNA data base.

Rat L15 is related to *Saccharomyces cerevisiae* YL10 (11) - the RELATE score is 37 S.D. units and in an alignment of the amino acid sequences there are 146 identities in 203 possible matches (72%); the ALIGN score is 68 S.D. A *Chironomus tentans* protein (12), identified only as related to yeast YL10, is also related to rat L15 - the RELATE score is 48 S.D. and in an alignment of the amino acid sequences there are 151 amino acid identities in 204 possible matches (74%); the ALIGN score is 83 S.D. A human expressed sequence tag (13) identified only as related to yeast YL10 is the homolog of rat L15 - the RELATE score is 63 S.D. and in an alignment of the amino acid sequences there are 197 identities out of 201 possible matches (98%); the ALIGN score is 96 S.D. Inspection of the two amino acid sequences reveals there are 8 nonidentities, all at the carboxyl terminus; this is an unusually large number of differences for related mammalian ribosomal proteins and suggests a sequencing error. Indeed, examination of the nucleotide sequence for human L15 (13) reveals that 8 carboxyl-terminal amino acids identical to rat L15 are encoded out of frame in the DNA. Since the carboxyl-terminal amino acid sequence is conserved in the other homologs of rat L15, it is likely that there was an error in the determination of the nucleotide sequence for human L15 (13) that caused a frame shift. If one assumes the error, then there are 202 identities in 204 possible matches (99%). We have not found any prokaryotic or archaebacterial ribosomal proteins that are related to rat L15.

*C. tentans* L15 (12) has the amino acid sequence **GLTSAGKS** (positions 163-170) which approximates the consensus sequence, **GXXXXGKS/T**, for a phosphate binding (P-loop) motif (14); this motif is found in many proteins that bind ATP and GTP (14). The same motif occurs in rat L15 in a slightly altered form, **GLTSAGRKS** (Fig. 1; positions 163-171); a number of variants of the consensus sequence are known to be tolerated (14). The sequence in yeast YL10 and in human L15 is like that in rat L15. We have searched our library of ribosomal protein amino acid sequences and found the P-loop motif in the prokaryotic family of L2 ribosomal proteins; for example in *Escherichia coli* L2 the sequence is **GVQTKGKKT** (positions 248-256). In addition, the P-loop motif occurs in initiation and elongation factors. We do not know the functional significance, if any, of the P-loop motif in the L15 proteins.

Amyloidogenic glycoprotein, A4 protein, is an integral, glycosylated brain membrane protein (15) that is associated with Alzheimer's disease (16); the latter association stems from

the finding that a peptide of 43 residues derived from A4, called the amyloid ß protein, is a major constituent of amyloid deposits in Alzheimer's disease and in Down's syndrome. The amyloid ß protein both precedes and forms a part of the unique transmembrane region of A4. The function of the A4 protein is not known but it has been suggested that it mediates cell-cell interactions (15). The amino acid sequence of A4 proteins from mammalian species and from Drosophila is conserved (17). A hexapeptide, TYKFFE, in rat L15 (positions 126-131) is also found in the cytoplasmic domain at the carboxyl terminus of all A4 proteins and in no other protein. The threonine in the hexapeptide is a potential protein kinase C phosphorylation site, i.e. S/TXK/R (18). There are two other regions where L15 and A4 may have significant similarities in their amino acid sequences (dots identify conservative changes):

| | |
|---|---|
| Rat L15 | KYIQELWRKKQSDVMRF |
| Rat A4 | KYVRAEQKDRQHTLKHF |
| Common | KY. . .Q . F |
| Rat L15 | LLRVRCWQY |
| Rat A4 | LVMLKKKQY |
| Common | L. .. QY |

We do not know if the similarities in the amino acid sequences in rat ribosomal protein L15 and in the A4 proteins imply a relationship in their evolutionary origins and/or in their function, or if the putative similarities are merely fortuitous. Intuitively, it seems unlikely, that the conservation of the amino acid sequences, especially of the TYKFFE peptide, is only fortuitous. Nonetheless, if the association has significance we do not know what it is.

*ACKNOWLEDGMENTS*

This work was supported by National Institutes of Health Grant GM 21769. We are grateful to Anton Glück for assistance with the computer analyses, to Veronica Paz for technical help, and to Arlene Timosciek for aid in the preparation of the manuscript. The sequence is in the EMBL/GenBank/DDBJ Nucleotide Sequence Database; the accession number is X78167.

*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.*, eds.), pp. 203-214, Amer. Soc. Microbiol., Washington, D.C.
2. Chan, Y.L., Lin, A., McNally, J., and Wool, I.G. (1987) J. Biol. Chem. 262, 12879-12886.
3. Chan, Y.L., and Wool, I.G. (1988) J. Biol. Chem. 263, 2891-2896.
4. Glück, A., Chan, Y.L., Lin, A.L., and Wool, I.G. (1989) Eur. J. Biochem. 182, 105-109.
5. Collatz, E., Lin, A., Stöffler, G., Tsurugi, K., and Wool, I.G. (1976) J. Biol. Chem. 251, 1808-1816.

6. Kozak, M. (1986) Cell 44, 283-292.
7. Proudfoot, N.J., and Brownlee, G.G. (1976) Nature 263, 211-214.
8. Tsurugi, K., Collatz, E., Todokoro, K., and Wool, I.G. (1977) J. Biol. Chem. 252, 3961-3969.
9. 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.
10. Pearson, W.R., and Lipman, D.J. (1988) Proc. Natl. Acad. Sci. USA 85, 2444-2448.
11. Suzuki, W., accession number D14675.
12. Galli, J., and Wieslander, L. (1992) Nucleic Acids Res. 20, 5473.
13. Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R.F., Kelley, J.M., Utterback, T.R., Nagle, J.W., Fields, C., and Venter, J. (1992) Nature 355, 632-634.
14. Saraste, M., Sibbald, P.R., and Wittinghofer, A. (1990) Trends Biochem. Sci. 15, 430-434.
15. Dyrks, T., Weidemann, A., Multhaup, G., Salbaum, J.M., Lemaire, H.G., Kand, J., Muller-Hill, B., Masters, C.L., Beyreuther, K. (1988) EMBO J. 7, 949-957.
16. Glenner, G.G. (1988) Cell 52, 307-308.
17. Rosen, R.R., Martin-Morris, L., Luo, L., and White, K. (1989) Proc. Natl. Acad. Sci. USA 86, 2478-2482.
18. Woodget, J.R., Gould, K.L., and Hunter, T. (1986) Eur. J. Biochem. 161, 177-184.