Vol. 172, No. 1, 1990 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
October 15, 1990 Pages 35-41

THE PRIMARY STRUCTURE OF RAT RIBOSOMAL PROTEIN L12

Katsuyuki Suzuki, Joe Olvera, and Ira G. Wool*

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

Received August 24, 1990

SUMMARY: The covalent structure of the rat 60S subunit protein L12 which is a component of the ribosomal elongation factor binding domain was deduced from the sequence of nucleotides in a recombinant cDNA and confirmed from the NH_2-terminal amino acid sequence of the protein. L12 has 165 amino acids and a molecular weight of 17,834. Hybridization of the cDNA to digests of nuclear DNA suggests that there are 11-13 copies of the L12 gene. The mRNA for the protein is about 800 nucleotides in length. Rat L12 is homologous to *Saccharomyces cerevisiae* L15. The cDNA contains the highly repetitive DNA sequence, *R.dre.1*, in the 3' noncoding region.

Ribosomes are complex ribonucleoprotein particles that catalyze peptide bond formation and protein synthesis in all organisms in the biosphere; those in eukaryotic cells have 70 to 80 proteins and 4 molecules of RNA (1). An effort is being made to determine the primary structure of all of these molecules for a single mammalian species, the rat. The purpose is to establish a data base that will assist in solving the structure of the organelle. Knowledge of the structure of ribosomes is presumed to be essential (albeit perhaps not in itself sufficient) for a rational, molecular account of the function of the organelle. As a part of this endeavor we report here the covalent structure of rat ribosomal protein L12 which we have inferred from the sequence of nucleotides in a recombinant cDNA and which we confirmed by sequencing a portion of the protein.

MATERIALS AND METHODS

The recombinant DNA procedures and the methods used to determine the sequence of nucleotides in the nucleic acids were either described or cited before (2, 3). The probe for the cDNA encoding rat ribosomal protein L12 was based on an NH_2-terminal sequence PNEIKVVY derived directly from the protein by Edman degradation using an automated gas phase sequencer. The probe, which contained a mixture of 96 different oligodeoxynucleotides, 26 nucleotides in

*To whom correspondence should be addressed.

0006-291X/90 $1.50

length, was synthesized on a solid support by the methoxyphosphoramidite method using an Applied Biosystems, Model 380B, DNA synthesizer (4). The NH_2-terminal methionine was not found in the amino acid sequence derived from the protein, however, its codon was presumed to be in the cDNA sequence and its anticodon was therefore included in the probe. We discoverd later (see below) that protein L12 actually lacks the NH_2-terminal 6 residues encoded in the cDNA so the probe is for residues 7-14.

RESULTS AND DISCUSSION

The Sequence of Nucleotides in Recombinant cDNAs Encoding Rat Ribosomal Protein L12

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 (3) 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 rat ribosomal protein L12. One colony gave a positive hybridization signal with the probe. The DNA from the plasmid of this transformant was isolated and digested with restriction endonucleases. The insert was approximately 680 nucleotides in length and Southern blot hybridization with the probe indicated that the insert might encode L12. The sequence of nucleotides in both strands of the cDNA insert in the plasmid designated pL12-1 was determined using primers based either on nucleotide sequences in the vector or, as they were obtained, on sequences in the insert.

The cDNA insert in pL12-1 contains 664 nucleotides; it includes a 5' noncoding sequence of 81 bases, a single open reading frame of 498, and a 3' noncoding sequence of 85 (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 (TAA) at position 496; it encodes 165 amino acids (Fig. 1). The initiation codon occurs in the context ACC*AUG*C which is close to the optimum ACC*AUG*G (5). It is exceptional that the 5' and 3' noncoding sequence are more than 80 nucleotides long; the range for 35 other mammalian ribosomal protein mRNAs is 25-80 (1). Despite the length of the 3' noncoding sequence it lacks a polyadenylation signal (see later). The 5' noncoding sequence of pL12-1 starts with 5 consecutive pyrimidines CTTTC. Sequences of pyrimidines are found at the 5' end of many eukaryotic mRNAs and may play a role in the regulation of their translation (cf. 1 for references and discussion).

The Primary Structure of Rat Ribosomal Protein L12

The rat ribosomal protein encoded in the open reading frame in pL12-1 was identified as L12 in the following manner: In the first instance, the recombinant cDNA clone pL12-1 was selected using oligodeoxynucleotide probes that were complementary to the codons for a sequence near

```
                                -60                                -30
          CTTTCGGTTCGGAGGAGGCAACGGTGCAACTTTCTTCGGTCGTCCCGAATCCGGGTTCATCCGACACCAGCCACCTCCACC


+1                               30                                 60                                 90
ATGCCGCCCAAGTTCGACCCCAACGAGATCAAAGTCGTGTACTTGAGGTGCACCGGAGGCGAGGTCGGCGCCACATCCGCCTTGGCCCCT
M  P  P  K  F  D  P  N  E  I  K  V  V  Y  L  R  C  T  G  G  E  V  G  A  T  S  A  L  A  P
1                              10                                 20                                 30

                               120                                150                                180
AAGATCGGTCCTCTGGGTCTGTCTCCCAAAAAAGTTGGTGATGACATCGCCAAGGCTACCGGTGACTGGAAAGGCCTCAGGATTACAGTG
K  I  G  P  L  G  L  S  P  K  K  V  G  D  D  I  A  K  A  T  G  D  W  K  G  L  R  I  T  V
                              40                                 50                                 60

                               210                                240                                270
AAACTGACCATCCAGAACAGACAGGCCCAGATTGAGGTGGTGCCCCTCTGCCTCTGCCCTGATCATCAAAGCCCTCAAGGAGCCACCAAGA
K  L  T  I  Q  N  R  Q  A  Q  I  E  V  V  P  S  A  S  A  L  I  I  K  A  L  K  E  P  P  R
                              70                                 80                                 90

                               300                                330                                360
GACAGGAAGAAGCAGAAAAACATTAAACACAATGGAAACATCACTTTTGATGAGATTGTCAACATTGCCCGGCAGATGAGACACCGGTCT
D  R  K  K  Q  K  N  I  K  H  N  G  N  I  T  F  D  E  I  V  N  I  A  R  Q  M  R  H  R  S
                             100                                110                                120

                               390                                420                                450
TTGGCCAGAGAACTTTCTGGAACTATCAAGGAGATCCTGGGTACTGCACAGTCTGTGGGCTGCAATGTGGACGGCCGCCACCCTCATGAC
L  A  R  E  L  S  G  T  I  K  E  I  L  G  T  A  Q  S  V  G  C  N  V  D  G  R  H  P  H  D
                             130                                140                                150

                               480                                510                                540
ATCATAGATGACATCAACAGTGGTGCGGTGGAGTGCCCAGCTAGTTAAGAAGCAACGAGAAGGGGTTGGGAATTTAGCTCAGTGGTAGAG
I  I  D  D  I  N  S  G  A  V  E  C  P  A  S  *
                             160

                               570
CGCTTGCCAAGCCCAAGGCCCTGGGTTCAGTCCCCAGCTCCGG
```

Fig. 1. The sequence of nucleotides in the cDNA insert in plasmid pL12-1 and the amino acid sequence encoded in the open reading frame. The positions of the nucleotides in the cDNA insert are given above the residue; the positions of amino acids in protein L12 are designated below the residue. The nucleotide sequence related to the rat *R.dre.1* family (10) is underlined.

the NH_2 terminus of L12. More positive identification was based on the congruence of the sequence of amino acids derived from the sequence of nucleotides in pL12-1 and of the NH_2-terminal 23 residues determined directly from protein L12 by Edman degradation using an automated gas phase sequencer (data not shown). The amino acid composition inferred from the cDNA, however, is different from that obtained before from an hydrolysate of purified L12 (6). We presume that the original determination was in error.

The molecular weight of L12, calculated from the sequence of amino acids deduced from pL12-1, is 17,834. However, the NH_2-terminal 6 residues, MPPKFD, encoded in the L12 mRNA are not found in the amino acid sequence derived from the protein. It is possible that L12 is either processed after it is synthesized or that it was degraded during isolation and purification. The latter is perhaps more likely since the peptide bond linking aspartyl to prolyl residues (as occurs in L12 between positions 6 and 7) can be cleaved in mildly acidic conditions (7). Thus, the number of residues in the processed or degraded protein is 159 and the molecular weight is 17,119 close to that of 18,700 estimated from the migration of the purified protein in sodium dodecyl sulfate gels (6).

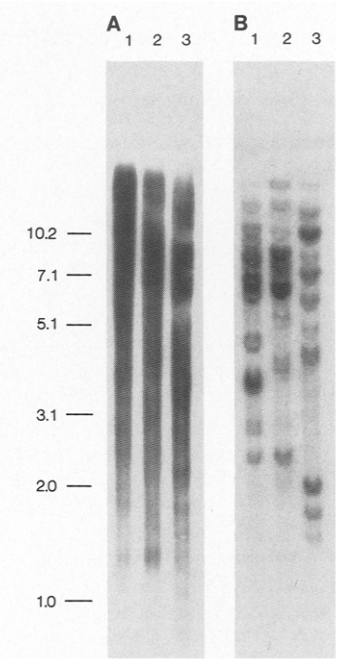

Fig. 2. Hybridization of ribosomal protein L12 cDNA to rat genomic DNA. Rat nuclear DNA (10 μg) was digested with restriction enzymes: *Bam*HI (lane 1); *Eco*RI (lane 2); or *Hind*III (lane 3). The digests were resolved by electrophoresis in 0.7% agarose gels and transferred to GeneScreen Plus nylon filters. In A, the uniformly labeled radioactive L12 cDNA insert from pL12-1 was hybridized to the immobilized restriction enzyme digest of genomic DNA; in B, the uniformly labeled radioactive *Pst*I - *Hinc*II fragment from pL12-1 was used for the hybridization. The position to which DNA standards of the size designated (in kilobase pairs) migrate is shown at the left.

The *Saccharomyces cerevisiae* ribosomal proteins L15 (8), which is homologous to rat L12 (see later), and S31 (9) also lack NH_2-terminal amino acids. The NH_2-terminal 16 residues of yeast L15 and the NH_2-terminal 14 of yeast S31 are encoded in the genes but are not in the amino acid sequence of the isolated proteins. It is remarkable that these ribosomal proteins have the same NH_2-terminal sequences, MPPK, encoded in the genes (yeast L15 and S31) or in the mRNA (rat L12).

Protein L12 has an excess of basic residues (10 arginyl, 14 lysyl, and 4 histidyl) over acidic ones (9 aspartyl and 8 glutamyl). As has been noted before for ribosomal proteins, the basic residues tend to be clustered; for example, 7 of 11 residues at position 90-100. Indeed, there is a prominent hydrophilic region in L12; 11 of 18 residues at positions 83-100 are charged.

The Number of Copies of the L12 Gene

The cDNA insert in pL12-1 was made radioactive and used to probe restriction endonuclease (*Bam*HI, *Eco*RI, and *Hind*III) digests from rat liver DNA (3). Contrary to expectation the hybridization pattern indicated an enormous number of copies of the L12 gene. It seemed more likely that the cDNA insert in pL12-1 has a repetitive sequence (Fig. 2,A). A search of the

EMBL nucleotide sequence library revealed that the 3' noncoding region in pL12-1 has a nucleotide sequence similar to the *Alu*-like retroposon *R.dre.1* family (10); there are 64 identities in 71 possible matches in an alignment of the 3' noncoding region of pL12-1 (positions 512-582) and the *R.dre.1*. There are 1-1.5 x 10^5 copies of *R.dre.1* in the rat genome. The sequence has been found in introns and in flanking regions of other rat genes; however, this is the first time it has been identified in a mRNA for a ribosomal protein.

To determine the number of L12 genes it was necessary to delete the 3' untranslated region of pL12-1. A radioactive *Pst*I-*Hinc*II fragment of about 440 nucleotides was prepared from pL12-1. This restriction fragment includes the 5' noncoding region of the L12 cDNA and 330 nucleotides of the 5' part of the coding sequence. The number of hybridization bands obtained with this probe suggest that there are 11-13 copies of the L12 gene (Fig. 2,B). 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 pseudogenes.

The Size of the mRNA Encoding Rat Ribosomal Protein L12

To determine the size of the mRNA for L12, total poly(A)$^+$mRNA from rat liver was separated by electrophoresis and screened for hybridization bands using either radioactive pL12-1 or the truncated cDNA. With either one a single band of about 800 nucleotides was detected (data not shown).

Comparison of the Sequence of Amino Acids in Rat L12 with Ribosomal Proteins From Other Species

The sequence of amino acids in rat ribosomal protein L12 was compared, using the computer programs RELATE and ALIGN (11), to the sequences of amino acids in more than 500 other ribosomal proteins contained in a library that we have compiled. The comparison that yielded the highest RELATE score (43.1 S.D. units) was with *S. cerevisiae* L15 (8). In an alignment there are 115 amino acid identities in 165 possible matches (the ALIGN score is 78.8). The two proteins obviously are homologous, i.e. derived from a common ancestral gene. Rat L12 is also related to *Sulfolobus solfataricus* L11e (12) (the RELATE score is 9.3; the ALIGN score is 19.3 with 42 identities in 156 possible matches) and to *Halobacterium cutirubrum* L11e (13) (the RELATE score is 5.4; the ALIGN score is 10.9 with 25 identities in 157 possible matches). In addition, a group of eubacterial L11 ribosomal proteins may be related to rat L12. These include *Escherichia coli* L11 (14) (the ALIGN score is 3.5 with 28 identities in 140 possible matches); *Proteus vulgaris* L11 (15) (the ALIGN score is 5.4 with 28 identities in 141 possible matches); and *Serratia marcescens* L11 (15) (the ALIGN score is 4.2 with 28 identities in 141 possible matches). The RELATE scores, however, are not significant.

The sequence of amino acids in rat L12 was searched for internal repeats but none were found.

The Function of Rat Ribosomal Protein L12

Yeast ribosomal protein L15 binds to 26S rRNA at a conserved site that is related to the region of *E. coli* ribosomal 23S rRNA where L11 binds; moreover, yeast L15 and *E. coli* L11 bind to the same site in mouse 28S rRNA (16). The apparent functional relationship of yeast L15 and *E. coli* L11 is reflected in immunological cross-reactivity and in amino acid sequence similarities (8). Thus, it would not be surprising if it is found that rat L12 binds to this same conserved site in 28S rRNA; in addition, it is likely that rat L12 and *E. coli* L11 are related proteins (see above).

Rat L12 can be crosslinked to the elongation factors 1α (EF-1α) (17) and 2 (EF-2) (18). This implies that L12 forms part of the domain necessary for EF-1 dependent aminoacyl-tRNA binding to the A site and for EF-2 dependent translocation of peptidyl-tRNA.

The determination of the sequence of amino acids in rat L12 is a contribution to a definition of the structure of a component of an important ribosomal domain as well as to a set of data 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 of this data is its perceived 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 Dr. Juan P.G. Ballesta of the Universidad Autónoma de Madrid for providing us with the amino acid sequence of *S. cerevisiae* L15 before publication, to Yuen-Ling Chan for advice and for fruitful discussions, and to Arlene Timosciek for aid in the preparation of the manuscript. The sequence data are in the EMBL/Gen Bank/DDBJ Nucleotide Sequence Databases under the accession number X53504.

REFERENCES

1. Wool, I.G., Endo, Y., Chan, Y.L. and Glück, A. (1990) in Structure, Function and Evolution of Ribosomes (Hill, W.E., ed.), Amer. Soc. Microbiol., Washington, D.C., in press.
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. Beaucage, S.L. and Caruthers, M.H. (1981) Tetrahedron Lett. 22, 1859-1862.
5. Kozak, M. (1986) Cell 44, 283-292.
6. Tsurugi, K., Collatz, E., Wool, I.G. and Lin, A. (1976) J. Biol. Chem. 251, 7940-7946.
7. Landon, M. (1977) Meth. Enzymol. 47, 145-149.

8. Pucciarelli, G., Remacha, M., Vilella, M.D. and Ballesta, J.P.G. (1990) Nucleic Acids Res., in press.
9. Nieuwint, R.T.M., Molenaar, C.M.T., van Bommel, J.H., van Raamsdonk-Duin, M.M.C., Mager, W.H. and Planta, R.J. (1985) Curr. Genet. 10, 1-5.
10. Rogers, J.H. (1985) Int. Rev. Cytol. 93, 187-279.
11. 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.
12. Shimmin, L.C., Newton, C.H., Ramirez, C., Yee, J., Downing, W., Louie, A., Matheson, A.T. and Dennis, P.P. (1989) Can J. Microbiol. 35, 164-170.
13. Shimmin, L.C. and Dennis, P.P. (1989) EMBO J. 8, 1225-1235.
14. Wittmann-Liebold, B. (1984) Adv. Protein Chem. 35, 56-78.
15. Sor, F. and Nomura, M. (1987) Mol. Gen. Genet. 210, 52-59.
16. El-Baradi, T.A.L., de Regt, V.C.H.F., Einerhand, S.W.C., Teixido, J., Planta, R.J., Ballesta, J.P.G. and Raué, H.A. (1987) J. Mol. Biol. 195, 909-917.
17. Uchiumi, T. and Ogata, K. (1986) J. Biol. Chem. 261, 9668-9671.
18. Uchiumi, T., Kikuchi, M., Terao, K., Iwasaki, K. and Ogata, K. (1986) Eur. J. Biochem. 156, 37-48.