Vol. 178, No. 1, 1991 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
July 15, 1991 Pages 322-328

THE PRIMARY STRUCTURE OF RAT RIBOSOMAL PROTEIN L17

Katsuyuki Suzuki and Ira G. Wool*

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

Received June 3, 1991

SUMMARY: The amino acid sequence of the rat 60S ribosomal subunit protein L17 was deduced from the sequence of nucleotides in two recombinant cDNAs. Ribosomal protein L17 has 184 amino acids and has a molecular weight of 21,383. Hybridization of the cDNA to digests of nuclear DNA suggests that there are 17-19 copies of the L17 gene. The mRNA for the protein is about 720 nucleotides in length. Rat L17 is homologous to human L17 and related to *Saccharomyces cerevisiae* YL17, *Halobacterium marismortui* L23, *Halobacterium halobium* L22e, *Escherichia coli* L22 and other members of the prokaryotic L22 family.

Ribosomes are complex ribonucleoprotein organelles 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 species 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 particles. 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 L17 which we have inferred from the sequence of nucleotides in recombinant cDNAs.

MATERIAL 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). Two oligodeoxynucleotide probes for the cDNA encoding rat ribosomal protein L17 were synthesized; the sequences were based on that in a human cDNA (4) which encodes a protein that is related to *Halobacterium marismortui* ribosomal protein L23. Probe 1 was a mixture of 288 different oligodeoxynucleotides, each 23 bases in length, complementary to the sequence encoding

* To whom correspondence should be addressed. Fax: 312-702-0439.

0006-291X/91 $1.50

MVRYSLDP (residues 1-8); probe 2 was a mixture of 384 different oligodeoxynucleotides, each 24 bases in length, complementary to the sequence encoding KQKLMARE (residues 177-184). The oligodeoxynucleotides were synthesized on a solid support by the methoxyphosphoramidite method using an Applied Biosystems, 380B, DNA synthesizer (5).

The *in vitro* transcription and translation procedures and the methods used to identify the protein encoded in the open reading frame in pL17-11 were also described or cited before (6).

RESULTS AND DISCUSSION

The Sequence of Nucleotides in Recombinant cDNAs Encoding Rat Ribosomal Protein L17

The sequence of amino acids encoded in the open reading frame of a cDNA from a human peripheral blood library was found to be related to the *Halobacterium marismortui* ribosomal protein L23 (4) and, as we discovered from a search of a library of more than 800 ribosomal protein amino acid sequences, to *Saccharomyces cerevisiae* YL17 (7). It was likely then that the human cDNA encoded a ribosomal protein and we undertook to isolate the rat homologue and to determine its identity. 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 two oligodeoxynucleotide probes that were synthesized to be complementary to the sequence of nucleotides predicted to be present in the mRNA for the homologous rat ribosomal protein. Four clones gave a positive hybridization signal with the probe. The DNA from the plasmids of the 4 transformants was isolated and digested with restriction endonucleases. These clones had inserts that ranged in length from 330 to 620 nucleotides. Two of the clones were selected, pL17-11 and pL17-12, and the sequences of nucleotides from both strands of the cDNAs and overlapping sequences for each restriction site were obtained.

The cDNA insert in pL17-11 is 620 nucleotides long and has a 5' noncoding sequence of 22 bases, a single open reading frame of 555, a 3' noncoding sequence of 43 including a poly(A) stretch of 15 (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 553; it encodes 184 amino acids (Fig. 1). The initiation codon occurs in the context AAG*AUG*G which deviates from the optimum ACC*AUG*G (8). The hexamer AATAAA that directs post-transcriptional cleavage-polyadenylation of the 3' end of the precursor of the mRNA (9) is at position 566-571, 12 nucleotides upstream of the start of the poly(A) stretch.

The sequence of nucleotides in pL17-12 is identical to that in pL17-11 except that it contains 8 additional bases in the 5' noncoding region and lacks a good portion of the 3' part of

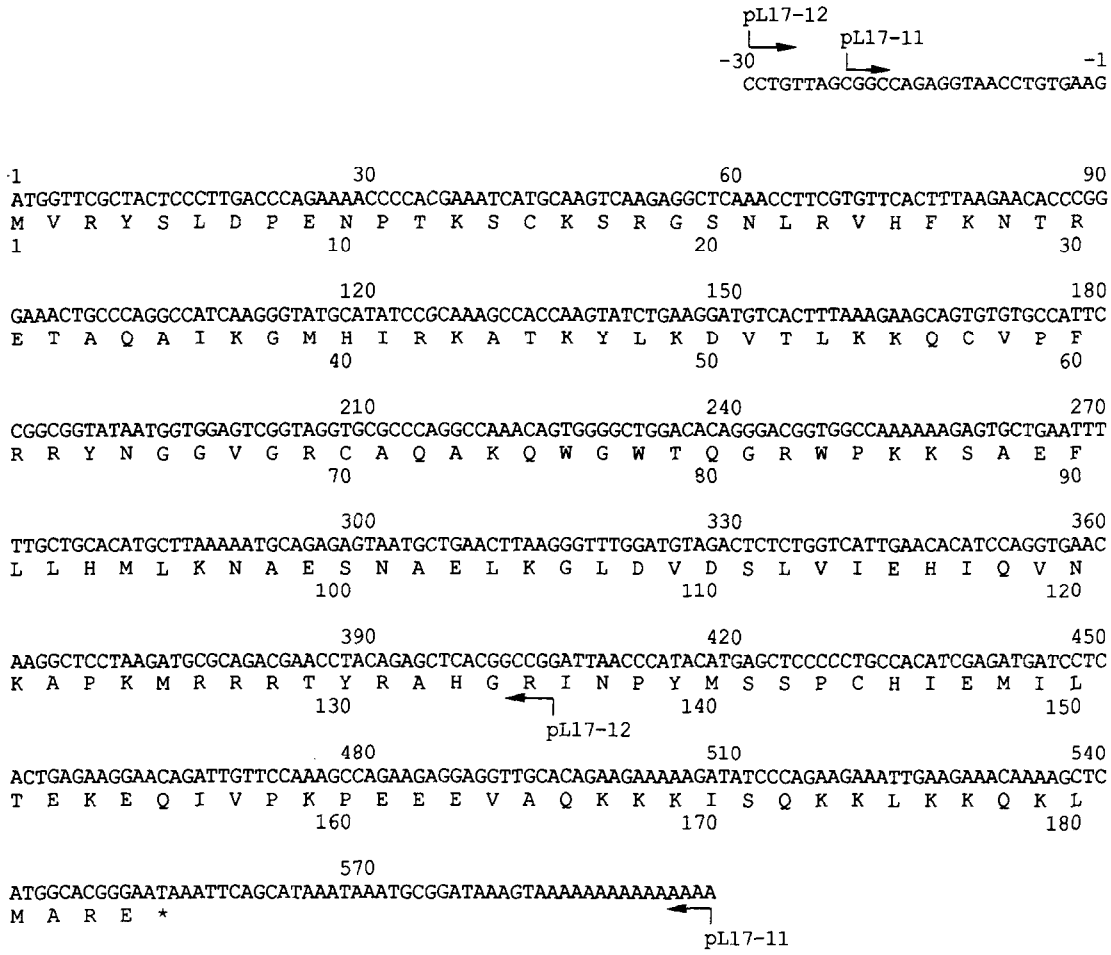

Fig. 1. *The sequences of nucleotides in the cDNA inserts in plasmids pL17-11 and pL17-12 and the amino acid sequence encoded in the open reading frame.* The positions of the nucleotides in the cDNA inserts are given above the residue; the positions of amino acids in protein L17 are designated below the residue; the start and stop sites for pL17-11 and pL17-12 are indicated by the verticals of the bent arrows.

the open reading frame (Fig. 1). Since the parts of pL17-11 and pL17-12 that coincide have identical nucleotide sequences they are likely to be derived from the same gene.

The Primary Structure of Rat Ribosomal Protein L17

The rat ribosomal protein encoded in the open reading frame in pL17-11 was identified as L17 by transcription of the pL17-11 cDNA, translation of the RNA transcript in a nuclease-treated reticulocyte lysate, and identification of the radioactive product from its migration on two-dimensional polyacrylamide gels (Fig. 2). In addition, the amino acid composition inferred from the cDNA is very close to that obtained before (10) from an hydrolysate of purified L17 (Table I).

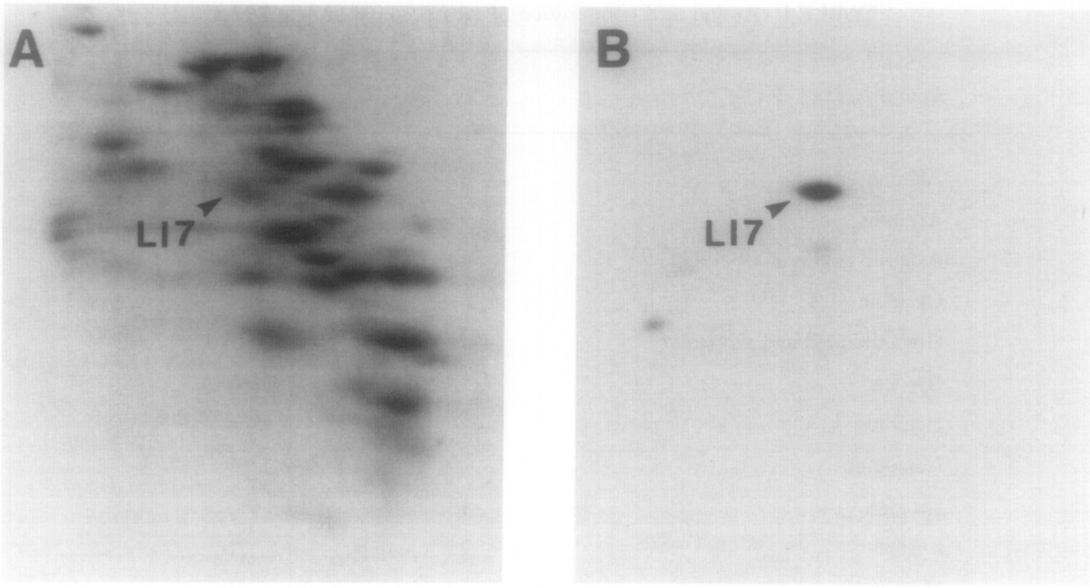

Fig. 2. *Two-dimensional electrophoresis of the product of the translation of the pL17-11 cDNA transcript.* A reticulocyte lysate (50 µl) was incubated with the RNA transcript of the pL17-11 cDNA (1 µg) and a sample (2.5 µl) of the product of the translation was extracted from the lysate with 67% acetic acid and the protein was precipitated with 90% acetone. Electrophoresis of the radioactive translation product (labeled with [^{35}S]-methionine) was with 100 µg of carrier protein from 60S ribosomal subunits. Electrophoresis in polyacrylamide gels containing urea was from left to right in the first dimension and from top to bottom in the second. Staining in *A* was with Coomassie brilliant blue; visualization of the gel in *B* was by fluorography.

The molecular weight of rat ribosomal protein L17, calculated from the sequence of amino acids deduced from pL17-11, is 21,383; close to the 22,100 estimated before (10) from SDS-PAGE of the purified protein. We do not know whether the NH_2-terminal methionine encoded in the L17 mRNA is removed after translation. However, the residue next to the initial methionyl in L17 is valyl which has been reported (11) to favor NH_2-terminal processing.

Protein L17 has a large excess of basic residues (15 arginyl, 26 lysyl, and 6 histidyl) over acidic ones (4 aspartyl and 13 glutamyl) (Table I). The basic residues tend to be clustered; for example, 8 of 15 residues at positions 121-135 and 8 of 13 between positions 167 and 179. L17 has a number of hydrophilic regions; for example, 17 of the 33 carboxyl-terminal residues are charged.

The Number of Copies of the L17 Gene

The cDNA insert in pL17-11 was made radioactive and used to probe digests of rat liver DNA made with restriction endonucleases *Bam*HI, *Eco*RI, or *Hind*III (3). The number of hybridization bands suggest that there are 17-19 copies of the L17 gene (data not shown). Many other mammalian ribosomal protein genes have been found to be present in multiple copies (cf.

TABLE I. Amino acid composition of rat ribosomal protein L17

Amino Acid	A	B
Alanine	12	12
Arginine	15	15
Aspartic acid and asparagine	14	4 + 8
Cysteine	n.d.	4
Glutamic acid and glutamine	22	13 + 10
Glycine	10	9
Histidine	6	6
Isoleucine	9	9
Leucine	13	13
Lysine	25	26
Methionine	4	7
Phenylalanine	4	3
Proline	9	9
Serine	9	10
Threonine	8	8
Tryptophan	n.d.	3
Tyrosine	6	5
Valine	10	10
Residues		184

The amino acid composition (in numbers of residues) determined either (A) from an hydrolysate of purified L17 (10) or inferred (B) from the sequence of nucleotides in a recombinant cDNA.

(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 L17

To determine the size of the mRNA coding for L17 poly(A)$^+$mRNA from rat liver was separated by electrophoresis and screened for hybridization bands using radioactive pL17-11 cDNA. One distinct band of about 720 nucleotides was detected (data not shown).

Comparison of the Sequence of Amino Acids in Rat L17 with Ribosomal Proteins from Other Species

The sequence of amino acids in rat L17 was compared, using the computer programs RELATE and ALIGN (12), to those in more than 800 other ribosomal proteins contained in a

library that we have compiled. Rat L17 is homologous to a human protein first reported to be related to *H. marismortui* L23 (4); the RELATE score is 77.9 S.D. units. In an alignment of the amino acid sequences there are 183 identities out of 184 possible matches (the ALIGN score is 81.5). The difference is at position 54 where there is a Q in the human protein and a K in the rat L17 sequence. Thus, the previously unidentified human protein is ribosomal protein L17. Rat L17 is related to *Halobacterium halobium* L22e (13); the RELATE score is 14.9; and the ALIGN score is 27.5 with 48 identities in 155 possible matches. Rat L17 is also related to *H. marismortui* L23 (14); the RELATE score is 11.0 and the ALIGN score is 29.5 with 51 identities in 154 possible matches. An NH_2-terminal fragment of 40 amino acids (34 residues were identified) from *S. cerevisiae* YL17 (7) also has a relationship to rat L17; the RELATE score is 14.5 and the ALIGN score is 17.6 with 21 identities in 40 possible matches. It is likely that when the entire sequence is determined yeast YL17 will prove to be homologous to rat L17. In addition, three ribosomal proteins designated L22 from eubacterial species (*Bacillus stearothermophilus* (15), *Escherichia coli* (16), and *Mycoplasma capricolum* (17)) and five from chloroplast ribosomes of various species of plants (*Nicotiana tabacum* (18), *Gracilaria tenuistipitata* (19), *Marchantia polymorpha* (20), *Spinacia oleracea,* (21) and *Euglena gracilis* (22)) are related to rat L17. The RELATE and ALIGN scores are low but significant; for example, for *E. coli* L22 the RELATE score is 4.6 and the ALIGN score is 8.7 with 23 identities in 110 possible matches.

A search of the EMBL/Gen Bank data base for nucleotide sequences related to the rat L17 cDNA led to an unexpected finding. The entire sequence of the rat ribosomal protein is encoded in what had been reported (23) to be an exceptionally long 3' noncoding region for a mouse hexokinase cDNA. In an alignment of the nucleotide sequence of the rat L17 cDNA and the 3' noncoding region of the mouse hexokinase cDNA there is 95% identity in 626 consecutive base pairs. The putative 3' region of the mouse hexokinase cDNA has in its middle a poly(A) stretch of 54 nucleotides that follows an open reading frame of 87 codons of which 60 are identical to a sequence in rat L17. Thus, it is most likely that the mouse ribosomal protein L17 cDNA was ligated to the 3' end of the hexokinase cDNA in the construction of the library. The structure of the mouse L17 ribosomal protein, however, cannot be derived from the mouse hexokinase cDNA sequence because it almost certainly has errors.

The sequence of amino acids in L17 was searched for internal duplications but none were found.

The determination of the sequence of amino acids in rat L17 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 fruitful discussions, to Joe Olvera for technical assistance, 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 X58389.

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. 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. Mager, D.L. and Freeman, J.D. (1990) Nucleic Acids Res. 18, 5301.
5. Beaucage, S.L. and Caruthers, M.H. (1981) Tetrahedron Lett. 22, 1859-1862.
6. Chan, Y.L., Devi, K.R.G., Olvera, J. and Wool, I.G. (1990) Arch. Biochem. Biophys. 283, 546-550.
7. Otaka, E., Higo, K. and Itoh, T. (1984) Mol. Gen. Genet. 195, 544-546.
8. Kozak, M. (1986) Cell 44, 283-292.
9. Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211-214.
10. Tsurugi, K., Collatz, E., Todokoro, K. and Wool, I.G. (1977) J. Biol. Chem. 252, 3961-3969.
11. Flinta, C., Persson, B., Jörnvall, H. and von Heijne, G. (1986) Eur. J. Biochem. 154, 193-196.
12. 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.
13. Mankin, A.S. (1989) FEBS Lett. 246, 13-16.
14. Arndt, E., Krömer, W. and Hatakeyama, T. (1990) J. Biol. Chem. 265, 3034-3039.
15. Krömer, W.J., Hatakeyama, T. and Kimura, M. (1990) Biol. Chem. Hoppe-Seyler 371, 631-636.
16. Zurawski, G. and Zurawski, S.M. (1985) Nucleic Acids Res. 13, 4521-4526.
17. Ohkubo, S., Muto, A., Kawauchi, Y., Yamao, F. and Osawa, S. (1987) Mol. Gen. Genet. 210, 314-322.
18. Tanaka, M., Wakasugi, T., Sugita, M., Shinozaki, K. and Sugiura, M. (1986) Proc. Natl. Acad. Sci. USA 83, 6030-6034.
19. Kao, J.S. and Wu, M. (1990) Nucleic Acids Res. 18, 3067.
20. Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H. and Ozeki, H. (1986) Nature 322, 572-574.
21. Zhou, D.X., Quigley, F., Massenet, O. and Mache, R. (1989) Mol. Gen. Genet. 216, 439-445.
22. Christopher, D.A., Cushman, J.C., Price, C.A. and Hallick, R.B. (1988) Curr. Genet. 14, 275-286.
23. Arora, K.K., Fanciulli, M. and Pedersen, P.L. (1990) J. Biol. Chem. 265, 6481-6488.