Vol. 201, No. 1, 1994 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
May 30, 1994 Pages 102-107

THE PRIMARY STRUCTURE OF RAT RIBOSOMAL PROTEIN L13

Joe Olvera and Ira G. Wool*

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

Received April 7, 1994

SUMMARY: The amino acid sequence of the rat 60S ribosomal subunit protein L13 was deduced from the sequence of nucleotides in two recombinant cDNAs. Ribosomal protein L13 has 210 amino acids, the NH_2-terminal methionine is removed after translation of the mRNA and has a molecular weight of 24,094. Hybridization of the cDNA to digests of nuclear DNA suggests that there are 8 to 10 copies of the L13 gene. The mRNA for the protein is about 870 nucleotides in length. Rat L13 is related to ribosomal proteins from other eukaryotes.

Importance attaches to obtaining a solution of the structure of eukaryotic ribosomes since knowledge of the structure is believed, with cause, to be essential for a rational, molecular account of the function of the organelle in protein synthesis. It is hard to imagine solving the structure without knowing the sequence of nucleotides and amino acids in the constituent nucleic acids and proteins. A commitment has been made to the acquisition of this data for mammalian (rat) ribosomes (1). As a part of this undertaking, we report here the amino acid sequence of rat ribosomal protein L13.

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 L13 was isolated from a fraction of all the 60S subunit proteins (5) by high performance liquid chromatography in a reverse phase column (Bio-Rad, Hi-Pore C4) with a 0 to 60% acetonitrile gradient containing 0.1% trifluoroacetic acid at a flow rate of 0.8 ml/min. The sequence of the NH_2-terminal 13 amino acids of L13, and the sequence of 116 additional amino acids in peptides prepared from L13 with cyanogen bromide or by acid hydrolysis, were determined by Edman degradation in an Applied Biosystems, Model 470A, automated gas phase sequencer. Two

*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

probes were prepared to identify cDNAs encoding L13: the first was a mixture of 192 different oligodeoxynucleotides, each 23 nucleotides in length based on the sequence NVYKKEKA (residues 159-166); the second was a mixture of 256 different oligodeoxynucleotides, each 29 nucleotides long based on the sequence EEEKNFKAFA (residues 171-180). 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 Recombinant cDNAs Encoding Rat Ribosomal Protein L13

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 (2, 3) was screened for clones that hybridized to two oligodeoxynucleotide probes that were related to amino acid sequences in rat ribosomal protein L13. Three clones gave a positive hybridization signal with the probes. The DNA from the plasmids of two of the transformants, those with largest inserts, was isolated and digested with restriction endonucleases. The sequence of nucleotides in both strands of the cDNA inserts in these two clones, pL13-2 and pL13-3, was determined. In pL13-2 the cDNA insert is 596 nucleotides long, has a 5' noncoding sequence of 39 bases, a single open reading frame of 557 nucleotides; but no termination codon and no 3' noncoding sequence (Fig. 1). The cDNA insert in pL13-3 is 446 nucleotides long, with a coding sequence of 408 nucleotides, and a 3' noncoding sequence of 38 bases and a poly(A) stretch; it lacks a 5' noncoding sequence and an initiation codon (Fig. 1). The overlapping sequences in pL13-2 and pL13-3 are identical, hence, they are likely to be derived from the same gene. The two cDNAs together encode all of L13 and for convenience we shall refer to them as pL13-2,3.

The cDNA in pL13-2,3 is 713 nucleotides long, has 5' and 3' noncoding sequences of 39 and 38 bases, and an open reading frame of 636 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 634; it encodes 211 amino acids (Fig. 1). The initiation codon occurs in the context GCC*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 656-661, 13 nucleotides upstream of the start of the poly(A) stretch. The first 8 nucleotides of the 5' noncoding sequence are pyrimidines, i.e. TCCTTTCC (Fig. 1). The 5' end of most, if not all, eukaryotic ribosomal protein mRNAs have pyrimidine sequences (1) that are presumed to have a role in the regulation of their translation (8).

The Primary Structure of Rat Ribosomal Protein L13

The rat ribosomal protein encoded in the open reading frame in pL13-2,3 was identified by the correspondence of the amino acid sequence encoded in the plasmid DNA with 128 of 129

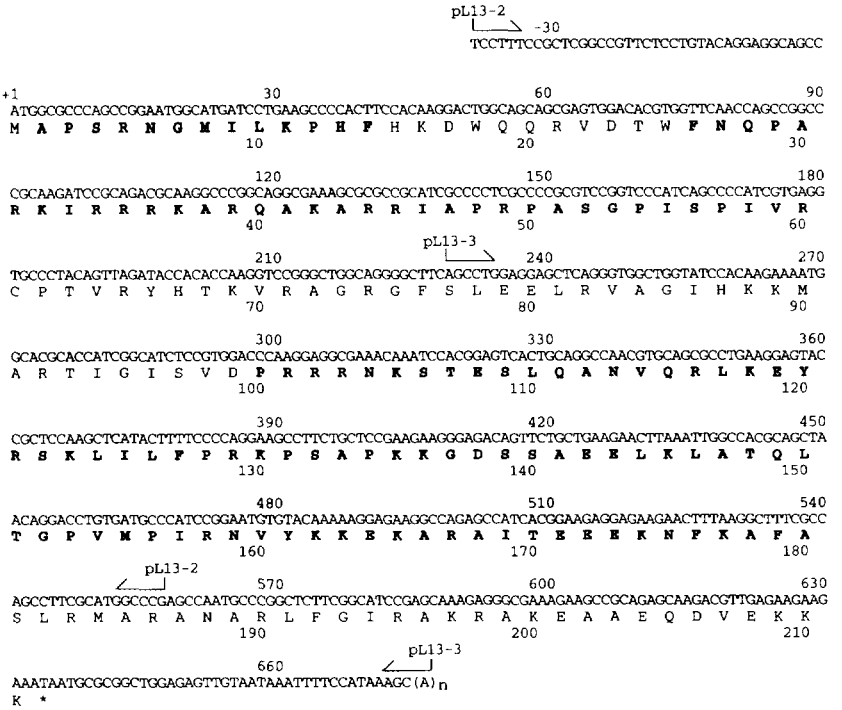

Fig. 1. *The sequence of nucleotides in the cDNA insert in pL13-2,3 and the amino acid sequence encoded in the open reading frame.* The positions of the nucleotides in the cDNA are given above the residue; the positions of the amino acids in protein L13 are designated below the residue. The initial and terminal nucleotides in pL13-2 and pL13-3 are designated by the vertical of the bent arrows. The amino acid residues in bold letters were determined also by Edman degradation of L13.

amino acids in L13 determined directly by Edman degradation (Fig. 1). We discuss the one exception, a serine/arginine ambiguity later. In addition, the amino acid composition inferred from the sequence of nucleotides in pL13-2,3 and that determined from an hydrolysate of the protein (9) are in close agreement (Table I).

The molecular weight of rat ribosomal protein L13, calculated from the sequence of amino acids deduced from pL13-2,3 is 24,225. However, the NH_2-terminal methionine encoded in the L13 mRNA is removed after translation, i.e. it is not in the NH_2-terminal sequence determined directly from the protein. The residue next to the initial methionine in L13 is alanine which favors NH_2-terminal processing (10). Thus, the mature processed L13 has 210 residues and the molecular weight is 24,094, close to the 26,300 estimated before (9) from SDS-PAGE of the purified protein.

Protein L13 has a large excess of basic residues (29 arginyl, 25 lysyl, and 4 histidyl) over acidic ones (5 aspartyl and 13 glutamyl) (Table I). There are several clusters of basic residues;

TABLE I. Amino acid composition of rat ribosomal protein L13

Amino Acid	A	B
Alanine	25	25
Arginine	30	29
Aspartic acid and asparagine	12	5 + 7
Cysteine	n.d.	1
Glutamic acid and glutamine	20	13 + 8
Glycine	9	9
Histidine	4	4
Isoleucine	12	12
Leucine	12	12
Lysine	21	25
Methionine	4	5*
Phenylalanine	7	7
Proline	13	14
Serine	8	12
Threonine	7	8
Tryptophan	n.d.	2
Tyrosine	4	3
Valine	11	10
Residues		211*

The amino acid composition (in numbers of residues) determined either (A) from a hydrolysate of purified L13 (9) or inferred (B) from the sequence of nucleotides in pL13-2,3.
*The NH_2-terminal methionine is removed after translation of the mRNA.
n.d., not determined.

for example, 11 of 19 amino acids at positions 31-49 (Fig. 1). There are highly charged regions in L13 at positions 31-49, at 99-108, at 129-145, at 162-177, and at 195-211.

The Number of Copies of the L13 Gene

The cDNA insert in pL13-2 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 8 to 10 copies of the L13 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 L13

To determine the size of the mRNA coding for L13, poly(A)$^+$mRNA from rat liver was separated by electrophoresis and screened for hybridization bands using radioactive pL13-2 cDNA. One distinct band of about 870 nucleotides was detected (data not shown).

Comparison of the Sequence of Amino Acids in Rat L13 with Ribosomal Proteins from Other Species

The sequence of amino acids in rat L13 was compared, using the computer programs RELATE and ALIGN (11), to those in more than 1,400 other ribosomal proteins contained in a library that we have compiled. This comparison did not yield a ribosomal protein with significant identity with L13.

However, a search of the EMBL/GenBank DNA database, using the program TFASTA (12), revealed several nucleotide sequences that are likely to encode plant and human homologs of L13. These nucleotide sequences, expressed sequence tags, were obtained by random sequencing and were not identified. The related proteins are from *Oryza sativa* (rice) - two partial sequences (accession numbers, D15755 and D15204) having about 60% amino acid identity with rat L13; a partial sequence from *Arabidopsis thaliana* (accession number Z17463) having 59% amino acid identity with rat L13; and a large number of human DNA sequences that encode portions of a protein that has amino acid identity with rat L13.

One of the human nucleotide sequences however, encodes the entire homolog of rat L13; the protein was designated "breast basic conserved protein" (13) since it was not recognized that it is a ribosomal protein. The amount of human L13 mRNA is decreased in malignant breast lesions compared to benign tumors; the mRNAs for two other ribosomal proteins, P2 (14) and the ubiquitin carboxyl extension protein S27a (15), are also decreased.

There is 96.2% amino acid identity in an alignment of rat and human L13 - 203 identities out of 211 possible matches; the ALIGN score is 94.3 and the RELATE score is 56.9 SD. Eight nonidentities is an unusually large number of differences for related mammalian ribosomal proteins and suggests an error in the nucleotide sequence.

There is a single discrepancy, at position 56, in the amino acid sequence obtained directly from rat L13 and that inferred from the cDNA. The residue from Edman degradation of rat L13 is Arg; the cDNA encodes a Ser. The human, the *O. sativa*, and the *A. thaliana* cDNAs all code for an Arg at this position. We suspect that the difference is either due to a fidelity error in the synthesis of the rat L13 cDNA or is a true microheterogeneity.

The determination of the sequence of amino acids in rat L13 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 fruitful discussions, to Anton Glück for assistance with the computer analyses, 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 X78327.

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. Głück, 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. Levy, S., Avni, D., Hariharan, N., Perry, R.P., and Meyuhas, O. (1991) Proc. Natl. Acad. Sci. USA 88, 3319-3323.
9. Tsurugi, K., Collatz, E., Wool, I.G., and Lin, A. (1976) J. Biol. Chem. 251, 7940-7946.
10. Flinta, C., Persson, B., Jörnvall, H. and von Heijne, G. (1986) Eur. J. Biochem. 154, 193-196.
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. Pearson, W.R., and Lipman, D.J. (1988) Proc. Natl. Acad. Sci. USA 85, 2444-2448.
13. Adams, S.M., Helps, N.R., Sharp, M.G.F., Brammar, W.J., Walker, R.A., and Varley, J.M. (1992) Human Molecular Genetics 1, 91-96.
14. Sharp, M.G.F., Adams, S.M., Elvin, P., Walker, R.A., Brammar, W.J., and Varley, J.M. (1990) Br. J. Cancer 61, 83-88.
15. Adams, S.M., Sharp, M.G.F., Walker, R.A., Brammar, W.J., and Varley, J.M. (1992) Br. J. Cancer 65, 65-71.