Volume 247, number 2, 242–246 FEB 07032 April 1989

The primary structure of rat ribosomal protein L18a

Yuji Aoyama, Yuen-Ling Chan, Oded Meyuhas[+] and Ira G. Wool

Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL 60637, USA and [+]*The Developmental Biochemistry Research Unit, Institute of Biochemistry, The Hebrew University and Hadassah Medical School, Jerusalem, Israel*

Received 23 January 1989; revised version received 27 February 1989

The amino acid sequence of rat ribosomal protein L18a was deduced from the sequence of nucleotides in a recombinant cDNA. Ribosomal protein L18a contains 175 amino acids and has a molecular mass of 20047 Da. Hybridization of the cDNA to digests of rat nuclear DNA and to a preparation of poly(A)$^+$ mRNA suggests that there are 8–11 copies of the L18a gene and that the mRNA for the protein is about 700 nucleotides in length. Rat L18a is related to *Schizosaccharomyces pombe* L17 and perhaps to *Halobacterium marismortui* L19.

Ribosomal protein L18a; Amino acid sequence; cDNA; Evolution; (Rat)

1. INTRODUCTION

We and others are engaged in an attempt to obtain a solution to the structure of eukaryotic ribosomes. The motivation for this undertaking is the belief that knowledge of the structure is essential for a rational, molecular account of the function of the organelle in protein synthesis. No solution of the structure is possible without 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. The covalent structure of the four species of RNA and the sequence of amino acids in 26 of the 70 to 80 proteins have been determined either directly or they have been deduced from the sequence of nucleotides in recombinant cDNAs (cf. [1] for references). We report here the structure of rat ribosomal protein L18a which we have inferred from the sequence of nucleotides in a recombinant cDNA.

Correspondence address: I.G. Wool, Department of Biochemistry and Molecular Biology, The University of Chicago, 920 East 58th Street, Chicago, IL 60637, USA

The nucleotide sequence presented here has been submitted to the EMBL/GenBank data base under the accession no. Y07499

2. EXPERIMENTAL

The recombinant DNA procedures and the methods used to determine the sequence of nucleotides in the nucleic acid were either described or cited before [2].

3. RESULTS AND DISCUSSION

A recombinant cDNA was isolated from a library prepared from poly(A)$^+$ mRNA and provisionally identified by hybridization-selection and translation as encoding ribosomal protein L18 [3]. The sequence of nucleotides had an open reading frame but the location of the initiation codon could not be established and there was no termination codon; moreover, the deduced sequence in the cDNA insert did not correspond to the NH_2-terminal amino acid sequence of L18 (results not given). A reexamination of the identity of the translation product revealed that on one-dimensional SDS gel electrophoresis the protein co-migrated with rat ribosomal protein L18a and not L18 (fig.1); in two-dimensional urea gels it is hard to distinguish these two proteins. The tentative conclusion at the time was that the incomplete cDNA encoded part of the rat L18a sequence, not L18; however, we were unable to confirm this suspicion immediately because we

Published by Elsevier Science Publishers B.V. (Biomedical Division)

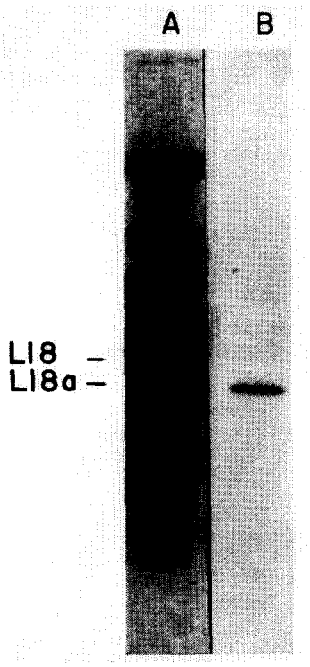

Fig.1. Analysis of the product of the translation of the hybrid-selected mRNA. A reticulocyte lysate (10 μl) was incubated with hybrid-selected mRNA (selected with the DNA insert in pL18a-2) and extracted with 67% acetic acid and the protein precipitated with 90% acetone. Electrophoresis was in SDS polyacrylamide gels: (A) Coomassie brilliant blue stain of all the proteins of the 60 S ribosomal subunit; (B) fluorography of the [^{35}S]methionine labeled product of the translation of the hybrid-selected mRNA. The bands that contain L18 and L18a are designated on the left.

could not isolate, and therefore could not sequence, L18a. A search of our library revealed that the amino acid sequence encoded in the putative L18a cDNA insert was related to an NH_2-terminal fragment of *Schizosaccharomyces pombe* L17 [4]. What is important here is that the amino acid composition of *S. pombe* L17 [4] was similar to that of rat L18a [5]. It appeared then that the cDNA, now designated pL18a-1, encoded ribosomal protein L18a (not L18) and that it was related to *S. pombe* L17 (see later). This encouraged us to synthesize an oligodeoxynucleotide [6], a 20-mer, that reproduced a portion of the original cDNA sequence, and to use it to screen a new cDNA library of 30000 independent transformants constructed from poly(A)$^+$ mRNA prepared from regenerating rat liver [2]. Eight colonies gave a positive hybridization signal. The DNA from these plasmids was isolated, digested with restriction endonucleases, and analyzed by gel electrophoresis. Southern blot hybridization with the oligonucleotide probe confirmed that the inserts contained cDNA for L18a. One of the clones, designated pL18a-2, was selected and the sequences of nucleotides in both strands of the cDNA, and overlapping sequences for each restriction site, were obtained.

The cDNA insert in pL18a-2 contains 580 nucleotides and includes a 5′-noncoding sequence of 27 nucleotides, a single open reading frame, and a 3′-noncoding sequence of 25 nucleotides (fig.2). The open reading frame of 528 nucleotides begins at an ATG codon at a position that we designate +1 and ends with a termination codon (TAG) at position 526 (fig.2). The initiation codon occurs in the context GCUAUGA which differs from the optimum ACCAUGG [7]. The 3′-noncoding sequence has the hexamer AATAAA (positions 546–551), presumed to be the recognition sequence directing post-transcriptional cleavage-polyadenylation of the 3′-end of pre-mRNA [8].

The reading frame in pL18a-2 specifies a protein of 175 amino acids (fig.2). This protein was identified as rat ribosomal protein L18a in the following way: translation of mRNA selected by hybridization to pL18a-1 led to the identification, by gel electrophoresis, of the product as encoding L18a (fig.1); and by the congruence of the amino acid composition inferred from the cDNA to that previously derived [5] from a hydrolysate of purified L18a (table 1).

The molecular mass of rat L18a is 20047 Da close to that of 21300 Da estimated from the migration of the purified protein in SDS gels [5]. We do not know if the NH_2-terminal methionine encoded in the L18a mRNA is removed after translation. However, the residue next to the initiator methionine is lysyl which has been reported [9] to protect against NH_2-terminal processing; hence, we predict that the methionine is retained in the protein.

The cDNA insert in pL18a-2 was made radioactive and used to probe digests made with restriction endonucleases (*Bam*HI, *Eco*RI, or *Hin*dIII) from rat liver nuclear DNA [10]. The number of hybridization bands suggest that there are 8–11 copies of the L18a gene (fig.3A). Other mammalian ribosomal protein genes have been found to

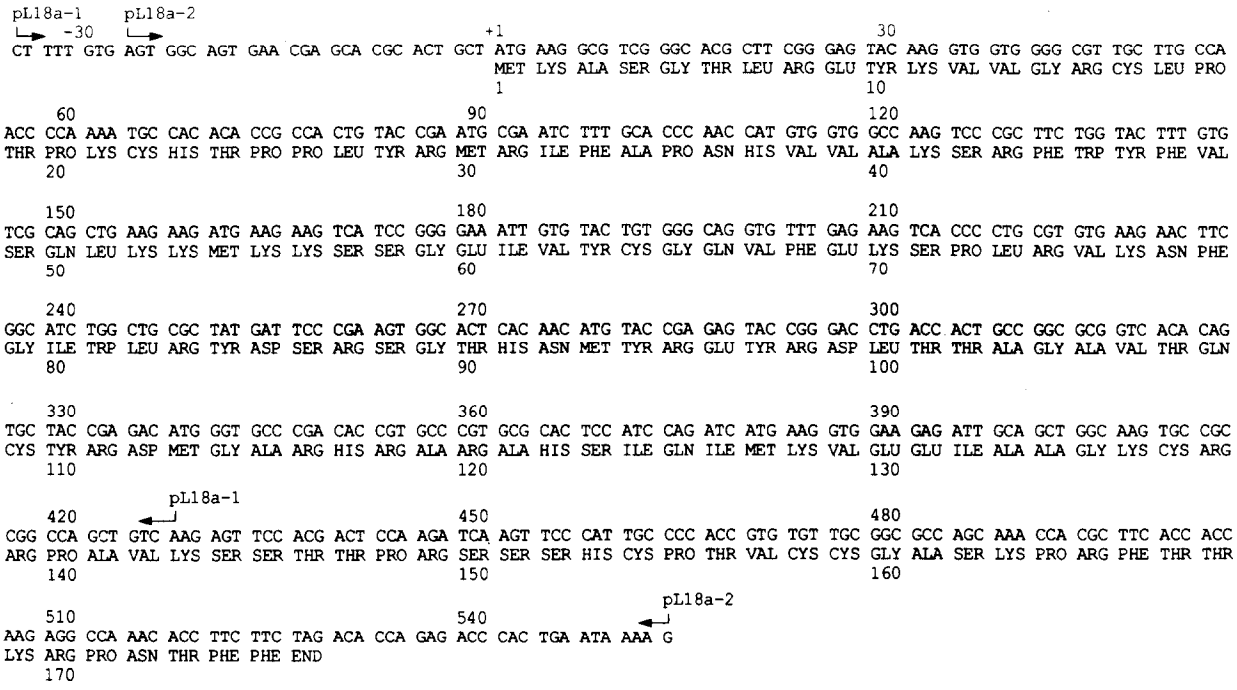

Fig.2. The sequence of nucleotides in the cDNA inserts in plasmids pL18a-1 and pL18a-2 and the amino acid sequence encoded in the open reading frame. The position of the nucleotides in the cDNA inserts is given above the residue; the position of amino acids in protein L18a is designated below the residue. The start and the termination sites of pL18a-1 and pL18a-2 are designated by arrows.

Table 1

Amino acid composition of L18a

Amino acids	A	B
Alanine	16	12
Arginine	15	19
Aspartic acid and asparagine	9	3+4
Cysteine	n.d.	8
Glutamic acid and glutamine	11	6+4
Glycine	17	10
Histidine	7	6
Isoleucine	6	6
Leucine	8	7
Lysine	16	15
Methionine	4	6
Phenylalanine	8	8
Proline	13	11
Serine	12	15
Threonine	10	13
Tryptophan	n.d.	2
Tyrosine	5	8
Valine	12	12
Residues		175

The amino acid composition (in numbers of residues) of rat ribosomal protein L18a was determined either (A) from an analysis of a hydrolysate of the purified protein [5] or inferred (B) from the sequence of nucleotides in a recombinant cDNA

be present in multiple copies [11]. However, in no instance has it been shown that more than one of the genes is functional [12–14]. The presumption is that for each ribosomal protein the genome contains only one gene that is expressed and that the other copies are non-functional pseudogenes.

To determine the size of mRNA coding for L18a, poly(A)$^+$ mRNA from rat liver was separated by electrophoresis and screened for hybridization bands using radioactive pL18a-2 cDNA. One distinct band was detected, the size being about 700 bases in length (fig.3B).

The order of amino acids in L18a was compared, using the computer program RELATE [15], to more than 330 other ribosomal protein sequences in a library that we have compiled. The comparison that yielded the highest score was with an NH_2-terminal fragment of 45 residues from *S. pombe* L17 (11.9 SD units) [4]. An alignment of rat L18a with the NH_2-terminus of *S. pombe* L17 reveals 22 identities of 45 possible matches for a score of 16.14 SD units (fig.4). Moreover, there are four positions at which the amino acid changes are conservative. It is most likely that the two proteins

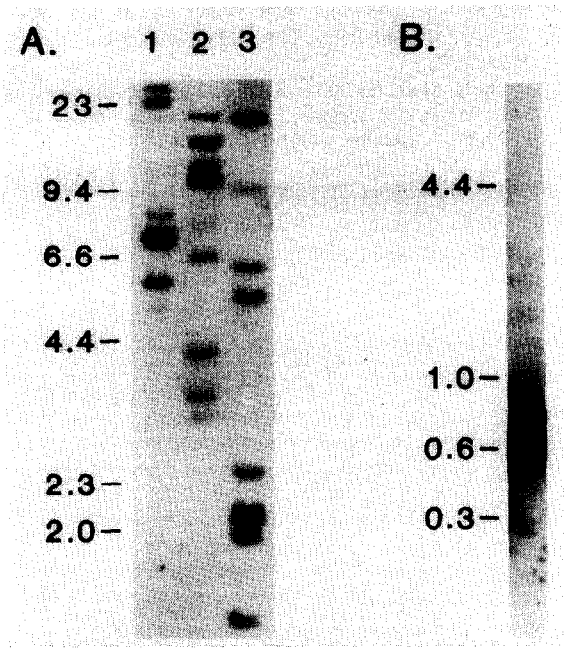

Fig.3. Hybridization of ribosomal protein L18a cDNA to rat genomic DNA and to poly(A)$^+$ mRNA. In A, rat nuclear DNA (10 μg) was digested with restriction enzymes: *Hin*dIII (lane 1); *Eco*RI (lane 2); or *Bam*HI (lane 3). The digests were resolved by electrophoresis in 0.7% agarose gels and transferred to MSI nylon filters. Uniformly labeled radioactive L18a cDNA insert from pL18a-2 was hybridized to the immobilized genomic DNA. The position to which DNA standards of the size designated (in kilobase pairs) migrate is shown to the left. In B, the poly(A)$^+$ mRNA (1 μg) prepared from rat liver was treated as described before [10] and hybridized to radioactive L18a cDNAs as above. The size of the mRNA was estimated by comparison to the mobility in the same gel of DNA restriction fragments.

are homologous, i.e. related to a common ancestral gene.

The other comparison that yields a significant score (5.64 SD units) is with a fragment of *Halobacterium marismortui* L19 [16]. In the alignment there are 6 identities out of 23 possible matches; the score is 5.46 SD units. The comparison does not allow a decision as to whether rat L18a and *H. marismortui* L19 are related; that will have to await completion of the amino acid sequence of the latter.

The determination of the sequence of amino acids in rat L18a is a contribution to a set of data which it is hoped will eventually encompass the structure of all the proteins in the ribosomes of this mammalian species. The primary motivation for the accumulation of these data is the value it must have 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.

Acknowledgements: This work was supported by a National Institutes of Health Grant GM 21769. We are grateful to Veronica Paz for technical assistance, to Arlene Timosciek for the preparation of the manuscript, and to Anton Glück for the computer comparisons.

REFERENCES

[1] Gallagher, M.J., Chan, Y.L., Lin, A. and Wool, I.G. (1988) DNA 7, 269–273.
[2] Chan, Y.L., Lin, A., McNally, J. and Wool, I.G. (1987) J. Biol. Chem. 262, 12879–12886.
[3] Meyuhas, O. and Perry, R.P. (1980) Gene 10, 113–129.
[4] Otaka, E., Higo, K. and Itoh, T. (1983) Mol. Gen. Genet. 191, 519–524.
[5] Tsurugi, K., Collatz, E., Todokoro, K., Ulbrich, N., Lightfoot, H.N. and Wool, I.G. (1978) J. Biol. Chem. 253, 946–955.
[6] Beaucage, S.L. and Caruthers, M.H. (1981) Tetrahedron Lett. 22, 1859–1862.
[7] Kozak, M. (1986) Cell 44, 283–292.
[8] Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211–214.

```
                1                   10                  20                  30                  40                  50
                '                   '                   '                   '                   '                   '
ERTL18A         M K A S G T L R E Y K V V G R C L P T P K C H T P P L Y R M R I F A P N H V V A K S R F W Y F V S Q
ESPL17          - - - - - A L K Z Y Z V V G R K V P T Z H Z P V P K L F R M R L F A P B Z S V A K S R Y G T F L K M
Common                    L .   Y   V V G R   . P T         P   L   R M R . F A P       V A K S R       F
```

Fig.4. A comparison of the NH_2-terminal amino acid sequences of rat ribosomal protein L18a and *S. pombe* protein L17. Identical residues are given in the line labeled common; the hyphens indicate that the first residue in the rat sequence is aligned with the sixth in the *S. pombe* protein; the dots designate conservative substitutions (arginine/lysine; isoleucine/leucine/valine).

[9] Flinta, C., Persson, B., Jornvall, H. and Von Heijne, G. (1986) Eur. J. Biochem. 154, 193–196.
[10] Chan, Y.L. and Wool, I.G. (1988) J. Biol. Chem. 263, 2891–2896.
[11] Monk, R.J., Meyuhas, O. and Perry, R.P. (1981) Cell 24, 301–306.
[12] Wiedemann, L.M. and Perry, R.P. (1984) Mol. Cell. Biol. 4, 2518–2528.
[13] Dudov, K.P. and Perry, R.P. (1984) Cell 37, 457–468.
[14] Wagner, M. and Perry, R.P. (1985) Mol. Cell. Biol. 5, 3560–3576.
[15] Dayhoff, M.O. (1978) in: Atlas of Protein Sequence and Structure, vol.5, suppl.3, pp.1–8, National Biomedical Research Foundation, Washington, DC.
[16] Walsh, M.J., McDougall, J. and Wittmann-Liebold, B. (1988) Biochemistry 27, 6867–6876.