Volume 249, number 1, 119-122 FEB 07157 May 1989

The primary structure of rat ribosomal protein L34

Yuji Aoyama, Yuen-Ling Chan and Ira G. Wool

Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL 60637, USA

Recieved 8 March 1989

The amino acid sequence of rat ribosomal protein L34 was deduced from the nucleotide sequence in a recombinant cDNA and confirmed from the NH_2-terminal amino acid sequence of the protein. Ribosomal protein L34 contains 117 amino acids (the NH_2-terminal methionine is removed after translation of the mRNA) and has an M_r of 13498. Hybridization of the cDNA to digests of nuclear DNA suggests that there are 7–9 copies of the L34 gene. The mRNA for the protein is about 630 nucleotides in length.

Ribosomal protein L34; Amino acid sequence; cDNA; (Rat)

1. INTRODUCTION

An attempt is being made to accumulate a set of data which it is hoped will eventually encompass the sequences of amino acids in all the proteins in the ribosomes of a mammalian species, the rat. The motivation for this compilation is the value it is perceived to 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 L34 which was deduced from the nucleotide sequence of a recombinant cDNA and confirmed from the NH_2-terminal sequence of the protein.

2. EXPERIMENTAL

The recombinant DNA procedures and the methods used to determine the sequence of nucleotides in the nucleic acids were either described or cited before [1]. A probe for the cDNA encoding rat ribosomal protein L34, based on the sequence of 8 amino acids near the NH_2 terminus (residues 12–19) of the protein, was contained in a mixture of 768 different oligodeoxynucleotides, 23 nucleotides in length, that was synthesized on a solid support by the methoxyphosphoramidite method using an Applied Biosystems, model 380B, DNA synthesizer [2]; the oligonucleotides were purified by polyacrylamide gel electrophoresis.

Radioactive rat ribosomal protein L34 cDNA was hybridized to restriction enzyme digests of genomic DNA [3], and to a preparation of rat liver poly(A)$^+$ mRNA [4].

The computer programs, RELATE and ALIGN [5], were used to assess possible evolutionary relationships between rat L34 and other ribosomal proteins. The scoring matrix was Dayhoff's MDM '78 [5].

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

3. RESULTS AND DISCUSSION

Two cDNA libraries of 30000 and 20000 independent transformants were constructed from poly(A)$^+$ mRNA prepared from regenerating rat liver [1]. A random selection of 14000 colonies from each library was screened for clones that hybridized to an oligodeoxynucleotide probe that was synthesized to be complementary to the nucleotide sequence predicted to be present in the portion of the mRNA that encoded the 8 amino acids (YNTASNKT) near the NH_2 terminus of rat ribosomal protein L34. Four colonies gave a positive hybridization signal with the probe. The DNA from the plasmids of the 4 transformants was isolated, digested with restriction endonucleases, and analyzed by gel electrophoresis. One of the clones, designated pL34-1, had an insert of approximately 450 nucleotides in length and Southern blot

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/89/$3.50

hybridization with the oligonucleotide probe confirmed that it contained cDNA for L34. The anticipated length of the L34 coding sequence, calculated from the molecular mass of the protein [6], is 430 nucleotides. The sequences of nucleotides from both strands of the cDNA insert in pL34-1, and overlapping sequences for each restriction site, were obtained.

The cDNA insert in pL34-1 contains 448 nucleotides and includes a 5′-noncoding sequence of 30 nucleotides, a single open reading frame, a 3′-noncoding sequence of 64 nucleotides, and a terminal poly(A) stretch (fig.1). In the other two reading frames the sequence is interrupted by many termination codons. The open reading frame of 354 nucleotides begins at an ATG codon at a position that we designate +1 and ends with a termination codon (TAA) at position 352; it encodes 117 amino acids (fig.1). The initiation codon occurs in the context AGGAUGG which does not deviate greatly from the optimum ACCAUGG [7]. The 3′-noncoding sequence has the hexamer AATAAA (position 374–379) which is presumed to be the recognition sequence directing post-transcriptional cleavage-polyadenylation of the 3′-end of pre-mRNA [8].

The first three nucleotides of the L34 cDNA (positions −30 through −28 in fig.1) are pyrimidines, i.e. TCC. This may be the end of a longer pyrimidine stretch; such sequences have been reported to be present in the 5′-untranslated region of many eukaryotic ribosomal protein mRNAs [9] and may play a role in the regulation of their translation.

The reading frame in pL34-1 is flanked by initiation and termination codons and specifies a protein of 117 amino acids (fig.1). This protein was identified as rat ribosomal protein L34 in the following manner: the recombinant cDNA clone (pL34-1) was selected using an oligodeoxynucleotide probe that was complementary to the codons for a sequence of 8 amino acids near the NH_2 terminus of L34. The amino acid composition inferred from the cDNA is close to that previously derived from an hydrolysate of purified L34 [6], although, there are appreciable discrepancies in the numbers of glycines, alanines, and histidines (table 1). The sequence of amino acids deduced from the sequence of nucleotides in pL34-1 corresponds precisely to the NH_2-terminal 33 residues determined by Edman degradation of protein L34 using an automated gas-phase sequencer.

The NH_2-terminal methionine encoded in the L34 mRNA is removed after translation since it is not found in the amino acid sequence derived from the protein. The residue next to the initial methionyl is valyl which has been reported [10] to favor NH_2-terminal processing. The M_r of rat ribosomal protein L34, calculated from the sequence of amino acids (but without the NH_2-terminal methionine), is 13 367 close to that of 15 800 estimated from the migration of the purified protein in SDS gels [6].

Protein L34 has a large excess of basic residues

```
-30                                    +1                                  30                                  60
  TCC GGG GAC GTT GCT GTA GAG GCA CTC AGG ATG GTC CAG CGT TTG ACA TAC CGT CGT AGG CTT TCC TAC AAT ACA GCC TCT AAC AAA ACT
                                        MET VAL GLN ARG LEU THR TYR ARG ARG ARG LEU SER TYR ASN THR ALA SER ASN LYS THR
                                        1                                   10                                  20

                                     90                                  120                                 150
  AGG CTG TCT CGA ACT CCT GGC AAC AGG ATC GTT TAC CTT TAC ACC AAG AAG GTC GGG AAA GCA CCT AAA TCC GCA TGT GGC GTG TTG CCA
  ARG LEU SER ARG THR PRO GLY ASN ARG ILU VAL TYR LEU TYR THR LYS LYS VAL GLY LYS ALA PRO LYS SER ALA CYS GLY VAL LEU PRO
                                     30                                  40                                  50

                                    180                                 210                                 240
  GGC AGA CTG CGA GGG GTC GTT GCT GTG AGA CCC AAA GTC CTC ATG AGA TTG TCC AAG ACA AAG AAG CAC GTT CAG CAG GGC CTA TGG TGG
  GLY ARG LEU ARG GLY VAL VAL ALA VAL ARG PRO LYS VAL LEU MET ARG LEU SER LYS THR LYS LYS HIS VAL GLN GLN GLY LEU TRP TRP
                                    60                                  70                                  80

                                    270                                 300                                 330
  CTC CAT GTG CGC CAA GTG TGT CCG GAC AGG ATC AAG CGG GCT TTC CTT ATC GAG GAG CAG AAA ATC GTT GTG AAA GTG TTG AAG GCA CAA
  LEU HIS VAL ARG GLN VAL CYS PRO ASP ARG ILU LYS ARG ALA PHE LEU ILU GLU GLU GLN LYS ILU VAL VAL LYS VAL LEU LYS ALA GLN
                                    90                                  100                                 110

                                    360                                 390
  GCA CAG AGT CAG AAA GCA AAA TAA ATG GGC AGC TTT TTT AAG TAA TAA AAA TCA AGC TCT CTC TAA AAA AAA AAA AAA AAA AAA AAA A
  ALA GLN SER GLN LYS ALA LYS END
```

Fig.1. The nucleotide sequence of the cDNA insert in plasmid pL34-1 and the amino acid sequence encoded in the open reading frame. The position of the nucleotides in the cDNA insert is given above the residue; the position of amino acids in protein L34 is designated below the residue.

Table 1

The amino acid composition (in numbers of residues) was determined either (A) from an analysis of an hydrolysate of purified L34 [6] or inferred (B) from the sequence of nucleotides in a recombinant cDNA

Amino acids	A	B
Alanine	15	8
Arginine	14	14
Aspartic acid and asparagine	5	1 + 3
Cysteine	n.d.	2
Glutamic acid and glutamine	7	2 + 8
Glycine	14	6
Histidine	7	2
Isoleucine	5	4
Leucine	11	12
Lysine	14	15
Methionine	2	2[a]
Phenylalanine	1	1
Proline	5	5
Serine	10	6
Threonine	6	9
Tryptophan	n.d.	2
Tyrosine	3	4
Valine	17	14
Residues		117

[a] The NH_2-terminal methionine is removed after translation of the mRNA

(14 arginyl, 15 lysyl, and 2 histidyl) over acidic ones (1 aspartyl and 2 glutamyl). Protein L34 is quite hydrophobic (43 of 116 residues) and the hydrophobic amino acids tend to occur in clusters.

The cDNA insert in pL34-1 was made radioactive and used to probe digests made with the restriction endonucleases *Bam*HI, *Eco*RI, and *Hin*dIII from rat liver nuclear DNA [3]. The number of hybridization bands suggest that there are 7–9 copies of the L34 gene (fig.2A). Other mammalian ribosomal protein genes have been found to 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, that the other copies are non-functional pseudogenes; however, this presumption derives from the analysis of only a limited number of families.

To determine the size of the mRNA for L34, glyoxylated total poly(A)$^+$ mRNA from rat liver was separated by electrophoresis and screened for hybridization bands using radioactive pL34-1 cDNA. One band of about 630 bases was detected (fig.2B).

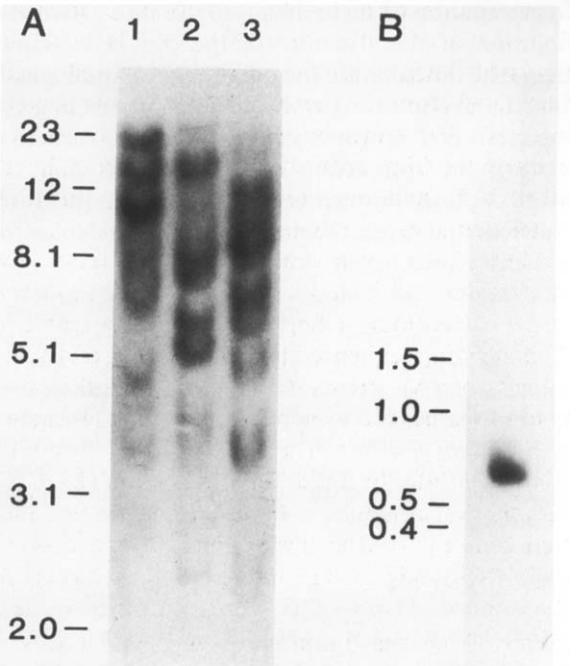

Fig.2. Hybridization of ribosomal protein L34 cDNA to rat genomic DNA and to poly(A)$^+$ mRNA. In A, rat nuclear DNA (10 μg) was digested with restriction enzymes: *Bam*HI (lane 1); *Eco*RI (lane 2); *Hin*dIII (lane 3). The digests were resolved by electrophoresis in 0.7% agarose gels and transferred to MSI nylon filters. Uniformly labeled radioactive cDNA from pL34-1 was hybridized to the immobilized genomic DNA. The position to which DNA standards of the size designated (in kilobase pairs) migrate is shown on the left. In B, the poly(A)$^+$ mRNA (1 μg) prepared from rat liver was treated as described before [4] and hybridized to radioactive L34 cDNAs as above. The size of the mRNA was estimated by comparison to the mobility in the same gel of DNA restriction fragments that had been calibrated by comparison with the mobility of 18 S and 28 S rRNAs.

The sequence of amino acids in rat ribosomal protein L34 was compared, using the computer program RELATE [5], to the sequence of amino acids in more than 400 other ribosomal proteins contained in a library we have compiled. We did not find any that are significantly similar to rat L34. The sequence of amino acids in L34 was also searched for internal duplications but none were found.

The determination of the sequence of amino acids in rat L34 is a contribution 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 these data is to use it to arrive 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 Alan Lin for the determination of the NH_2-terminal amino acid sequence of L34, to Anton Glück for help with the computer comparisons of amino acid sequences of ribosomal proteins, and to Arlene Timosciek for the preparation of the manuscript. The sequence data will appear in the EMBL/GenBank database under the accession number X14401.

REFERENCES

[1] Chan, Y.L., Lin, A., McNally, J. and Wool, I.G. (1987) J. Biol. Chem. 262, 12879–12886.
[2] Beaucage, S.L. and Caruthers, M.H. (1981) Tetrahedron Lett. 22, 1859–1862.
[3] Chan, Y.L. and Wool, I.G. (1988) J. Biol. Chem. 263, 2891–2896.
[4] Glück, A., Chan, Y.L., Lin, A. and Wool, I.G. (1989) Eur. J. Biochem., in press.
[5] 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, DC.
[6] Tsurugi, K., Collatz, E., Todokoro, K. and Wool, I.G. (1977) J. Biol. Chem. 252, 3961–3969.
[7] Kozak, M. (1986) Cell 44, 283–292.
[8] Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211–214.
[9] Wagner, M. and Perry, R.P. (1985) Mol. Cell. Biol. 5, 3560–3576.
[10] Flinta, C., Persson, B., Jornvall, H. and Von Heijne, G. (1986) Eur. J. Biochem. 154, 193–196.
[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] Klein, A. and Meyuhas, O. (1984) Nucleic Acids Res. 12, 3763–3776.