Volume 251, number 1,2, 89–93 FEB 07368 July 1989

The primary structure of rat ribosomal protein L26

Veronica Paz, Joe Olvera, Yuen-Ling Chan and Ira G. Wool

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

Received 2 May 1989

The amino acid sequence of rat ribosomal protein L26 was deduced from the sequence of nucleotides in a recombinant cDNA and confirmed from the NH_2-terminal amino acid sequence of the protein. Rat L26 contains 145 amino acids and has a molecular mass of 17 266 Da. Hybridization of the cDNA to digests of nuclear DNA suggests that there are 8–16 copies of the L26 gene. The mRNA for the protein is about 650 nucleotides in length. Protein L26 has a sequence of 9 residues that may be repeated in three places.

Ribosomal protein L26; Amino acid sequence; cDNA; Elongation factor-2 binding domain; (Rat)

1. INTRODUCTION

Importance has been attached 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. The covalent structure of the four species of RNA has been established and there are rational proposals for their secondary structure; in addition, eighty-two proteins have been isolated from the particles and the sequences of amino acids in more than 30 have been determined either directly or they have been deduced from the sequence of nucleotides in recombinant cDNAs ([1] and references therein). We report here the structure of rat ribosomal protein L26 which we have inferred from the sequence of nucleotides in a recombinant cDNA and which we have confirmed by sequencing portions of the protein. Protein L26 can be crosslinked to elongation factor 2 [2], an observation that implies that the protein forms part of the domain concerned with binding of the factor to the 60 S ribosomal subunit prior to translocation of peptidyl-tRNA from the A to the P site during each of the reiterative cycles of peptide bond formation.

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

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 [3]. A probe for the cDNA encoding rat ribosomal protein L26, based on the sequence of 7 amino acids at the NH_2 terminus of the protein, was contained in a mixture of 64 different oligodeoxynucleotides, 20 nucleotides in length, that was synthesized on a solid support by the methoxyphosphoramidite method using an Applied Biosystems, model 380B, DNA synthesizer [4]; the oligonucleotides were purified by polyacrylamide gel electrophoresis.

Radioactive rat ribosomal protein L26 cDNA was hybridized to genomic DNA that had been digested with restriction enzymes [5], and to a preparation of rat liver poly(A)$^+$ mRNA [6].

The computer programs, RELATE and ALIGN [7], were used to assess possible evolutionary relationships between rat L26 and other ribosomal proteins and to search for repeated sequences in the protein. The scoring matrix was Dayhoff's MDM '78 [7].

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

3. RESULTS AND DISCUSSION

A random selection of 20000 colonies taken from two cDNA libraries of 30000 and of 20000 independent transformants, which had been constructed from poly(A)$^+$ mRNA prepared from regenerating rat liver [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 portion of the mRNA that encoded 7 amino acids (MKFNPFV) near the NH_2 terminus of rat ribosomal protein L26. Thirteen colonies gave a positive hybridization signal with the probe. The DNA from the plasmids of the 13 transformants was isolated, digested with restriction endonucleases, and analyzed by gel electrophoresis. One of the clones, designated pL26-5, had an insert of approximately 520 base pairs and Southern blot hybridization with the oligonucleotide probe confirmed that it contained cDNA for L26. The sequences of nucleotides in both strands of the cDNA insert in pL26-5, and in overlapping sequences for each restriction site, were determined.

The cDNA insert in pL26-5 contains more than 508 nucleotides and includes a 5′-noncoding sequence of 36 nucleotides, a single open reading frame, a 3′-noncoding sequence of more than 34 nucleotides, and a terminal poly(A) stretch whose length could not be determined (fig.1). In the other two reading frames the sequence is interrupted by many termination codons. The open reading frame of 438 nucleotides begins at an ATG codon at a position that we designate +1 and ends with a termination codon (TAG) at position 436; it encodes 145 amino acids (fig.1). The initiation codon occurs in the context AAAAUGA which differs from the optimum ACCAUGG [8]. The 3′-noncoding sequence has the hexamer AATAAA (position 463–468), presumed to be the recognition sequence directing post-transcriptional cleavage-polyadenylation of the 3′-end of pre-mRNA [9].

The first 7 nucleotides of the L26 cDNA (positions −36 to −30 in fig.1) are pyrimidines, i.e. TCCTTTT. Pyrimidine stretches have been reported to be present at the start of the 5′-untranslated region of many eukaryotic ribosomal protein mRNAs [10–13] and may play a role in the regulation of their translation.

The reading frame in pL26-5 is flanked by initiation and termination codons and specifies a sequence of 145 amino acids (fig.1) which was identified as rat ribosomal protein L26 in the following manner: the recombinant cDNA clone (pL26-5) was selected using an oligodeoxynucleotide probe that was complementary to the codons for a sequence of 7 amino acids at the NH_2-terminus of L26. The amino acid composition (table 1) inferred from the cDNA is close to that previously derived from a hydrolysate of purified L26 [14]. But most importantly, the sequence of amino acids deduced from the sequence of nucleotides in pL26-5 corresponds precisely to the NH_2-terminal 13 residues (MKFNPFVTSDRSK) determined by Edman degradation of protein L26 using an automated gas phase sequencer.

The NH_2-terminal methionine encoded in the L26 mRNA is not removed after translation since it is found in the amino acid sequence derived from the protein. The residue next to the initial methionyl is lysyl which has been reported [15] to protect against NH_2-terminal processing. The molecular mass of rat ribosomal protein L26, calculated from the sequence of amino acids, is 17266 Da which approximates that of 18600 Da estimated from the migration of the purified protein in SDS gels [14].

Protein L26 has 45 basic residues (17 arginyl, 24 lysyl, and 4 histidyl) and 15 acidic ones (5 aspartyl and 10 glutamyl). As has been noted before for ribosomal proteins, the basic residues tend to be clustered; for example, 6 of 8 residues at positions 11–18 and 13 of 27 at positions 110–136. The carboxyl-terminus of L26 is remarkably hydrophilic; 23 of 38 residues (positions 108–148) are charged. The protein is unusual in that it lacks both cysteine and tryptophan.

The cDNA insert in pL26-5 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 [5]. The number of hybridization bands suggest that there are 8–16 copies of the L26 gene (fig.2A). Other mammalian ribosomal protein genes have been found to be present in multiple copies [11,12,16]. However, in no instance has it been shown that more than one of the genes is functional [11,12,16]. The presumption is that for each ribosomal protein the genome contains only one gene that is expressed, that the other copies are

```
      -30                                        +1                                        30
TCC TTT TGC GGC CAT CGC TGG ATT GCA GCC GCC AAA ATG AAG TTC AAT CCC TTC GTG ACT TCT GAC CGA AGC AAG AAC CGC AAA CGG CAT
                                                MET LYS PHE ASN PRO PHE VAL THR SER ASP ARG SER LYS ASN ARG LYS ARG HIS
                                                  1                                       10

     60                                         90                                       120
TTC AAT GCA CCA TCT CAC ATT CGG AGG AAG ATC ATG TCT TCT CCG CTT TCC AAA GAA CTG AGA CAG AAG TAT AAT GTT CGG TCT ATG CCC
PHE ASN ALA PRO SER HIS ILE ARG ARG LYS ILE MET SER SER PRO LEU SER LYS GLU LEU ARG GLN LYS TYR ASN VAL ARG SER MET PRO
     20                                         30                                       40

    150                                        180                                      210
ATT CGA AAG GAC GAC GAA GTT CAG GTT GTT CGA GGA CAC TAC AAA GGC CAG CAG ATT GGC AAA GTG GTC CAA GTG TAC AGG AAG AAA TAC
ILE ARG LYS ASP ASP GLU VAL GLN VAL VAL ARG GLY HIS TYR LYS GLY GLN GLN ILE GLY LYS VAL VAL GLN VAL TYR ARG LYS LYS TYR
     50                                         60                                       70

    240                                        270                                      300
GTC ATC TAC ATT GAG CGA GTC CAG CGA GAG AAG GCT AAC GGC ACA ACT GTC CAT GTG GGC ATC CGC CCC AGC AAG GTG GTT ATC ACC AGG
VAL ILE TYR ILE GLU ARG VAL GLN ARG GLU LYS ALA ASN GLY THR THR VAL HIS VAL GLY ILE ARG PRO SER LYS VAL VAL ILE THR ARG
     80                                         90                                       100

    330                                        360                                      390
CTA AAG CTG GAC AAG GAC CGC AAG AAG ATC CTG GAG AGG AAA GCC AAG TCC CGG CAA GTA GGA AAG GAG AAG GGC AAA TAC AAG GAA GAA
LEU LYS LEU ASP LYS ASP ARG LYS LYS ILE LEU GLU ARG LYS ALA LYS SER ARG GLN VAL GLY LYS GLU LYS GLY LYS TYR LYS GLU GLU
    110                                        120                                      130

    420                                        450
ACT ATC GAG AAG ATG CAG GAG TAG AAA AAT AAT GCC CAG TTT TCA TTA AAT AAA GCT T
THR ILE GLU LYS MET GLN GLU END
    140
```

Fig.1. The sequence of nucleotides in the cDNA insert in plasmid pL26-5 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; numbering begins with the A of the AUG codon as position +1; nucleotides 5′ to the initiation codon are assigned negative numbers. The position of amino acids in protein L26 is designated below the residue.

non-functional pseudogenes; however, this presumption derives from the analysis of only a limited number of families.

Table 1

The amino acid composition (in numbers of residues) was determined either (A) from an analysis of a hydrolysate of purified L26 [14] or inferred (B) from the sequence of nucleotides in a recombinant cDNA

Amino acids	A	B
Alanine	7	3
Arginine	15	17
Aspartic acid and asparagine	11	5+5
Cysteine	n.d.	0
Glutamic acid and glutamine	15	10+8
Glycine	10	7
Histidine	4	4
Isoleucine	9	10
Leucine	6	5
Lysine	21	24
Methionine	3	4
Phenylalanine	5	3
Proline	6	5
Serine	6	9
Threonine	5	5
Tryptophan	n.d.	0
Tyrosine	7	6
Valine	15	15
Residues		145

To determine the size of the mRNA for L26, glyoxylated total poly(A)$^+$ mRNA from rat liver was separated by electrophoresis and screened for hybridization bands using radioactive pL26-5 cDNA. One band of about 650 bases was detected (fig.2B).

The sequence of amino acids in rat ribosomal protein L26 was compared, using the computer program RELATE [7], to the sequences of amino acids in more than 400 other ribosomal proteins contained in a library we have compiled. The only comparison to yield a possibly significant score (4.8 SD units) was with a fragment from *Schizosaccharomyces pombe* L27 [17]; an alignment reveals 8 identities out of 28 possible matches – 10 if conservative changes (isoleucine/valine and lysine/arginine) are scored. However this alignment is between the NH_2-terminus of *S. pombe* L27 and the carboxyl-terminus of rat L26. Moreover, the *S. pombe* NH_2-terminal fragment of L27 is more closely related to the NH_2-terminus of rat L37 than was reported previously [17]; the RELATE score is 10.9 SD units and there are 17 identities and two conservative substitutions (serine/threonine and isoleucine/valine) in 28 possible matches of residues (the alignment score is 10.12 SD units).

The sequence of amino acids in L26 was

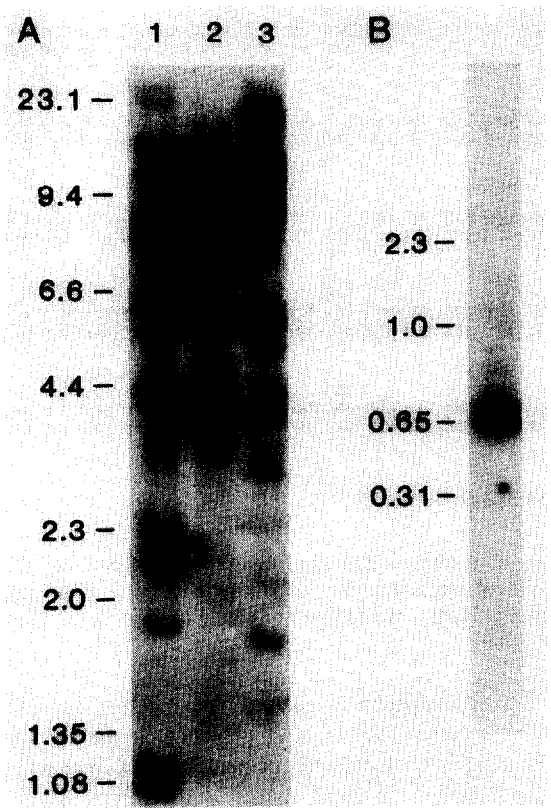

Fig.2. Hybridization of ribosomal protein L26 cDNA to rat genomic DNA and to poly(A)$^+$ mRNA. (A) Rat nuclear DNA (10 μg) was digested with restriction enzymes: *Bam*HI (lane 1); *Eco*RI (lane 2); *Hin*dIII (lane 3). The digests were resolved by electrophoresis in 0.7% agarose gels and transferred to MSI nylon filters. Uniformly labeled radioactive cDNA from pL26-5 was hybridized to the immobilized genomic DNA. The position to which DNA standards of the size designated (in kilobases) migrate is shown to the left. (B) The poly(A)$^+$ mRNA (1 μg) prepared from rat liver was treated as described before [6] and hybridized to radioactive L26 cDNAs as above. The size of the mRNA was estimated by comparison to the mobility in the same gel of single-stranded DNA restriction fragments that had been calibrated by comparison with the mobility of 18 S and 28 S rRNAs.

searched for internal duplications. There is a potential repeat of 9 residues that occurs three times at positions 55, 71, and 126 (fig.3). The consensus sequence, a residue that occurs at the same position in at least two of the three putative repeats, is VQV-R-KYK. It is not possible, unfortunately, to test the statistical significance of these repeats; i.e. to determine if they are authentic duplications or merely fortuitous. However, we

$_{55}$V Q V V R G H Y K
$_{71}$V Q V Y R K K Y V
$_{126}$R Q V G K E K G K
Consensus V Q V – R – K Y K

Fig.3. Possible repeat sequences in ribosomal protein L26. The number on the left designates the position in the L26 sequence of the first amino acid in the putative repeat. The consensus sequence, a residue that occurs at the same position in at least two of the three repeats, is given at the bottom.

have found multiple, related, basic, tandem and nontandem duplications in several ribosomal proteins before ([6] and references therein); this has suggested to us that they have functional significance, although we have as yet no indication what this might be. Possibilities are a role in an interaction with RNA (ribosomal, transfer, or messenger) or in directing the protein to the nucleolus for assembly of ribosomes.

The determination of the sequence of amino acids in rat L26 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 this 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 L26, 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/DDBJ Nucleotide Sequence Databases under the accession number X14671 rat rpl26.

REFERENCES

[1] Devi, K.R.G., Chan, Y.L. and Wool, I.G. (1989) Biochim. Biophys. Acta, in press.
[2] Nygard, O., Nilsson, L. and Westermann, P. (1987) Biochim. Biophys. Acta 910, 245–253.

[3] Chan, Y.L., Lin, A., McNally, J. and Wool, I.G. (1987) J. Biol. Chem. 262, 12879–12886.
[4] Beaucage, S.L. and Caruthers, M.H. (1981) Tetrahedron Lett. 22, 1859–1862.
[5] Chan, Y.L. and Wool, I.G. (1988) J. Biol. Chem. 263, 2891–2896.
[6] Glück, A., Chan, Y.L., Lin, A. and Wool, I.G. (1989) Eur. J. Biochem., in press.
[7] 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.
[8] Kozak, M. (1986) Cell 44, 283–292.
[9] Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211–214.
[10] Lin, A., Chan, Y.L., McNally, J., Peleg, D., Meyuhas, O. and Wool, I.G. (1987) J. Biol. Chem. 262, 12665–12671.
[11] Wiedemann, L.M. and Perry, R.P. (1984) Mol. Cell. Biol. 4, 2518–2528.
[12] Dudov, K.P. and Perry, R.P. (1984) Cell 37, 457–468.
[13] Wagner, M. and Perry, R.P. (1985) Mol. Cell. Biol. 5, 3560–3576.
[14] Tsurugi, K., Collatz, E., Wool, I.G. and Lin, A. (1976) J. Biol. Chem. 251, 7940–7946.
[15] Flinta, C., Persson, B., Jornvall, H. and Von Heijne, G. (1986) Eur. J. Biochem. 154, 193–196.
[16] Klein, A. and Meyuhas, O. (1984) Nucleic Acids Res. 12, 3763–3776.
[17] Otaka, E., Higo, K. and Itoh, T. (1983) Mol. Gen. Genet. 191, 519–524.