Volume 263, number 1, 85–88 FEBS 08299 April 1990

The primary structure of rat ribosomal protein S16

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

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

Received 22 January 1990; revised version received 14 February 1990

The amino acid sequence of rat ribosomal protein S16 was deduced from the sequence of nucleotides in a recombinant cDNA and confirmed from the NH_2-terminal amino acid sequence of the protein. S16 contains 145 amino acids (the NH_2-terminal methionine is removed after translation of the mRNA) and has a molecular mass of 16304. Hybridization of the cDNA to digests of nuclear DNA suggests that there are 11–13 copies of the S16 gene. The mRNA for the protein is about 700 nucleotides in length. Rat S16 is homologous to mouse S16 (there are 2 amino acid changes and a residue is deleted) and related to *Halobacterium morismortui* ribosomal protein S3 and to *Escherichia coli* S9.

Ribosomal protein S16; Amino acid sequence; cDNA; (Rat)

1. INTRODUCTION

A commitment has been made to the determination of the sequences of amino acids in all of the proteins in the ribosomes of a single mammalian species, the rat. The motivation for this compilation is the value it is perceived that it will 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 sequence of amino acids in rat ribosomal protein S16.

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 S16, based on the sequence of 10 amino acids (residues 25–34) of the mouse ribosomal protein S16 [2], was contained in a mixture of 32 different oligodeoxynucleotides, 29 nucleotides in length, that was synthesized on a solid support by the methoxyphosphoramidite method using an Applied Biosystems, model 380B, DNA synthesizer [3]; the oligonucleotides were purified by polyacrylamide gel electrophoresis. The NH_2-terminal amino acid sequence of rat ribosomal protein S16 was determined by Edman degradation using an Applied Biosystems, model 470A, automated gas phase protein sequencer.

Radioactive rat ribosomal protein S16 cDNA was hybridized to restriction enzyme digests of genomic DNA [4], and to a preparation of rat liver poly(A)$^+$ mRNA [5].

The computer programs, RELATE and ALIGN [6], were used to assess possible evolutionary relationships between rat S16 and other ribosomal proteins. The scoring matrix was Dayhoff's MDM '78 [6].

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 of 20000 independent transformants were constructed from poly(A)$^+$ mRNA prepared from regenerating rat liver [1]. A random selection of 25000 cells from each library was screened for clones that hybridized to an oligodeoxynucleotide probe that encoded a sequence of 10 amino acids (residues 25–34) of the mouse ribosomal protein S16 [2]. Homologous rat and mouse ribosomal proteins have the same or very similar sequences of amino acids [7]. Five colonies gave a positive hybridization signal with the probe. DNA from the plasmids of the 5 transformants was isolated, digested with restriction endonucleases, and analyzed by gel electrophoresis. One of the clones, designated pS16-8, had an insert approximately 540 nucleotides in length and Southern blot hybridization with the oligonucleotide probe indicated that it might contain cDNA for S16. The anticipated length of the S16 coding sequence, calculated from the molecular weight of the protein [8], is 465 nucleotides. The sequences of nucleotides from both strands of the cDNA insert in pS16-8, and overlapping sequences for each restriction site, were obtained.

The cDNA insert in pS16-8 contains 544 nucleotides and includes a 5′ noncoding sequence of 52 nucleotides, an open reading frame of 441 nucleotides, and a 3′ noncoding sequence of 51 nucleotides (Fig. 1). Long stretches of 2 of the 3 possible reading frames are open; the third is interrupted by termination codons. Of the two open reading frames only one starts with a methionine codon and encodes the NH_2-terminal sequence of S16 as determined directly from the protein (but see later). This open reading frame begins at an ATG codon at a position that we designate +1 and

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

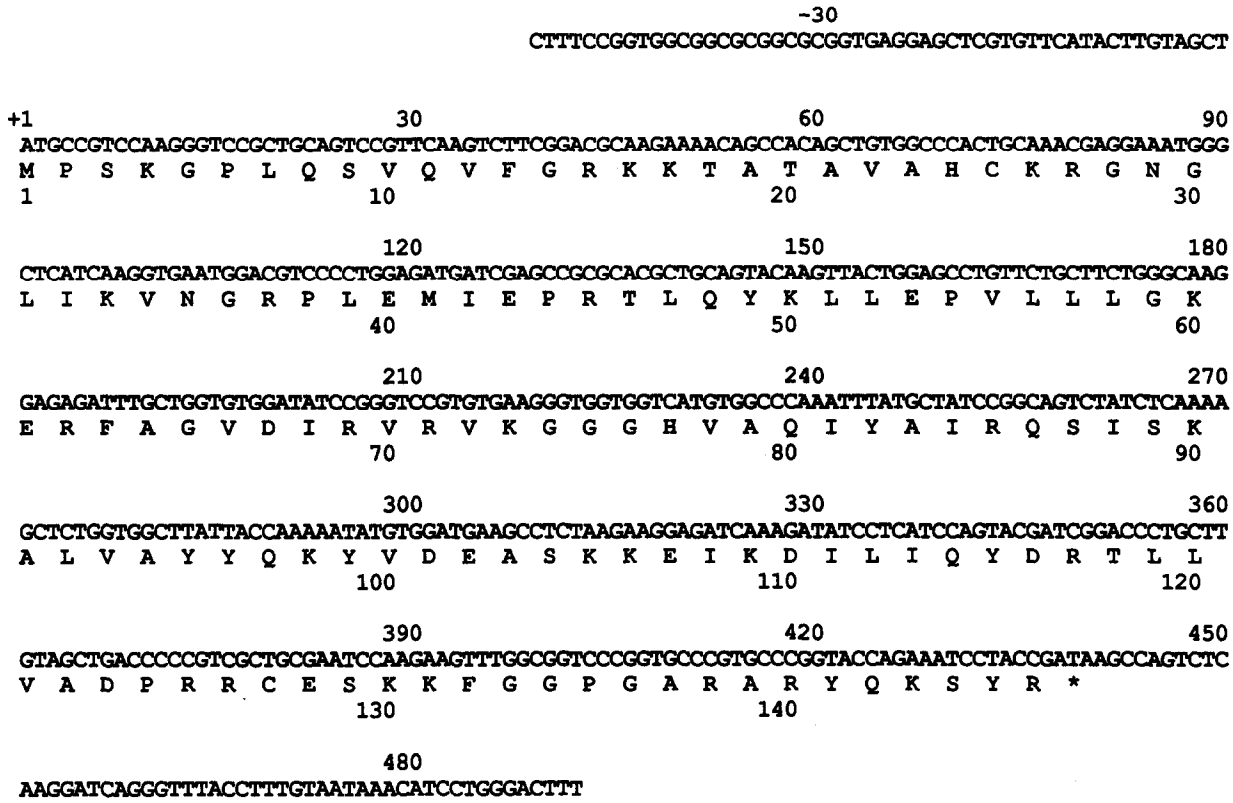


Fig. 1. The sequence of nucleotides in the cDNA insert in plasmid pS16-8 and the amino acid sequence encoded in the open reading frame. The position of the nucleotides in the cDNA insert is given above the residue; the position of amino acids in protein S16 is designated below the residue.

ends with a termination codon (TAA) at position 439; it encodes 146 amino acids (Fig. 1). The initiation codon occurs in the context GCUAUGC which deviates from the optimum ACCAUGG [9]. The 3′ noncoding sequence has the hexamer AATAAA (position 473–478) which is the recognition sequence directing post-transcriptional cleavage-polyadenylation of the 3′ end of pre-mRNA [10].

The first 6 nucleotides of the S16 cDNA (positions −52 through −47 in Fig. 1) are pyrimidines, i.e. CTTTCC. Pyrimidine sequences have been reported to be present in the 5′ untranslated region of many eukaryotic ribosomal protein mRNAs [2] and may play a role in the regulation of their translation.

The rat ribosomal protein specified by pS16-8 was identified as S16 in the following manner: the amino acid composition (Table I) inferred from the cDNA very closely approximates that previously derived from a hydrolysate of purified S16 [8]. The sequence of amino acids deduced from the sequence of nucleotides in pS16-8 corresponds precisely to the NH_2-terminal 18 residues determined by Edman degradation of protein S16 (see later).

The NH_2-terminal methionine encoded in the S16 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 prolyl which has been reported [11] to favor NH_2-terminal processing. The molecular mass of rat ribosomal protein S16, calculated from the sequence of amino acids (but without the NH_2-terminal methionine), is 16304 close to that of 17100 estimated from the migration of the purified protein in sodium dodecyl sulfate gels [8].

Protein S16 has a large excess of basic residues (14 arginyl, 16 lysyl, and 2 histidyl) over acidic ones (5 aspartyl and 7 glutamyl); lacks tryptophan; and is quite hydrophobic (53 of 146 residues) – the hydrophobic amino acids tend to occur in clusters.

The cDNA insert in pS16-8 was made radioactive and used to probe digests made from rat liver nuclear DNA with the restriction endonucleases *Bam*HI, *Eco*RI and *Hin*dIII [4]. The number of hybridization bands suggest that there are 11–13 copies of the S16 gene (data not shown). There are multiple copies of many other mammalian ribosomal protein genes [12]. However, in no instance has it been shown that more than one of the genes is functional [13–15]. The presumption is that the genome contains only one ribosomal protein gene that is expressed and that the other copies are nonfunctional pseudogenes [16].

To determine the size of the mRNA for S16, glyoxylated total poly(A)$^+$ mRNA from rat liver was separated by electrophoresis and screened for hybridization bands using radioactive pS16-8 cDNA.

Table I

The amino acid composition (in numbers of residues) was determined either (A) from an analysis of a hydrolysate of purified S16 [8] or inferred (B) from the sequence of nucleotides in a recombinant cDNA

Amino acids	A	B
Alanine	13	12
Arginine	13	14
Aspartic acid and asparagine	9	2+5
Cysteine	n.d.	2
Glutamic acid and glutamine	15	8+7
Glycine	15	13
Histidine	2	2
Isoleucine	8	9
Leucine	13	13
Lysine	16	16
Methionine	2	2[a]
Phenylalanine	3	3
Proline	8	7
Serine	7	7
Threonine	5	4
Tryptophan	n.d.	0
Tyrosine	7	8
Valine	12	12
Residues		146

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

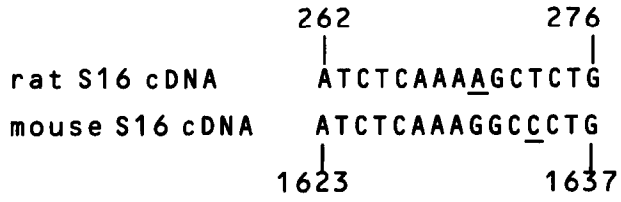
One band of about 700 bases was detected (data not shown).

The sequence of amino acids in rat ribosomal protein S16 was compared, using the computer program RELATE [6], to the sequence of amino acids in more than 500 other ribosomal proteins contained in a library we have compiled. Rat S16 is homologous to mouse S16 (but see below); the RELATE score is 47.6 SD units. Rat S16 is also related to *Escherichia coli* S9 – the RELATE score is 7.1 and in an alignment of the amino acid sequences there are 41 identities in 127 possible matches (the ALIGN score is 7.3); and to *Halobacterium morismortui* S3 – the RELATE score is 14.3 and in the alignment there are 53 identities in 130 possible matches (the ALIGN score is 30.1). The sequence of amino acids in rat S16 was searched for internal repeats but none were found.

When the sequence of amino acids in rat S16 deduced from the sequence of nucleotides as originally determined for pS16-8 was compared to that of the homologous mouse protein [2] there were differences at 15 positions. This was surprising since related mammalian ribosomal proteins diverge at a few residues if at all (cf. [7] for an example). For this reason we isolated rat S16 from a mixture of 40 S ribosomal subunit proteins by high-performance liquid chromatography and determined the NH_2-terminal 18 residues by Edman degradation. The amino acid sequence obtained from the protein (PSKGPLQSVQVFGRKKTA) did not correspond to the sequence deduced from either the mouse or the rat cDNAs, at least not after the sixth residue. However, it was noticed that the sequence of the next 12 amino acids did correspond to the deduced sequence if one shifted frames. Recall, 2 of the 3 reading frames in the S16 cDNA are substantially open.

The sequence of nucleotides in rat pS16-8 was determined again and two errors were uncovered. Inclusion of an extra C at position 19 had caused a reading frame shift, whereas omission of an A at position 270 had caused a shift back to the original frame. The corrected sequence (Fig. 1) encodes a protein whose composition is very close to that obtained from a hydrolysate of S16 (the amino acid composition derived from the original sequence of nucleotides had many discrepancies) and corresponds exactly to the first 18 residues determined directly from the protein.

However, the amino acid sequences of mouse and rat S16 were still quite different. In order to reconcile these differences we obtained the mouse ribosomal protein S16 cDNA and genomic DNA from R.P. Perry and determined the sequence of nucleotides. There are two errors in the published mouse nucleotide sequence just as there were at first in the rat sequence. The first error caused a reading frame shift because of the inclusion of an extra C at position 72 just as had been done with the rat cDNA at position 19 – these are the same sites in the sequence although the numbers are different. (Because this error leads to a frame shift, the probe, although based on an incorrect amino acid sequence (residues 35–34), still selected the correct clone.) The second error in the mouse sequence is the omission of a C at position 1634 which accomplishes substantially the same thing as the omission of the A at 270 in the rat cDNA, i.e. it shifts reading back to the original frame:



The corrected nucleotide sequences for the rat and mouse cDNAs encode proteins with 2 amino acid differences and 1 deletion. They are: the alanine at position 19 in the rat S16 sequence is deleted in mouse S16; residue 20 is threonyl in rat and leucyl in mouse; residue 46 is threonyl in rat and alanyl in mouse. Perry has communicated the corrections in the nucleotide sequence of the mouse S16 gene to the data bank (GenBank MusRPS16).

The lesson is how critical it is to have a partial amino acid sequence from the protein to authenticate the sequence derived from a gene or from a cDNA.

The determination of the sequence of amino acids in rat S16 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 them 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 Anton Glück for the computer comparisons of amino acid sequences of ribosomal proteins and to Arlene Timosciek for the preparation of the manuscript. The sequence data are in the EMBL/GenBank/DDBJ Nucleotide Sequence Databases under the accession number X17665.

REFERENCES

[1] Chan, Y.L., Lin, A., McNally, J. and Wool, I.G. (1987) J. Biol. Chem. 262, 12879–12886.
[2] Wagner, M. and Perry, R.P. (1985) Mol. Cell. Biol. 5, 3560–3576.
[3] Beaucage, S.L. and Caruthers, M.H. (1981) Tetrahedron Lett. 22, 1859–1862.
[4] Chan, Y.L. and Wool, I.G. (1988) J. Biol. Chem. 263, 2891–2896.
[5] Glück, A., Chan, Y.L., Lin, A. and Wool, I.G. (1989) Eur. J. Biochem. 182, 105–109.
[6] 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.
[7] Rajchel, A., Chan, Y.L. and Wool, I.G. (1988) Nucleic Acids Res. 16, 2347.
[8] Collatz, E., Ulbrich, N., Tsurugi, K., Lightfoot, H.N., MacKinley, W., Lin, A. and Wool, I.G. (1977) J. Biol. Chem. 252, 9071–9080.
[9] Kozak, M. (1986) Cell 44, 283–292.
[10] Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211–214.
[11] Flinta, C., Persson, B., Jornvall, H. and Von Heijne, G. (1986) Eur. J. Biochem. 154, 193–196.
[12] Monk, R.J., Meyuhas, O. and Perry, R.P. (1981) Cell 24, 301–306.
[13] Wiedemann, L.M. and Perry, R.P. (1984) Mol. Cell. Biol. 4, 2518–2528.
[14] Dudov, K.P. and Perry, R.P. (1984) Cell 37, 457–468.
[15] Klein, A. and Meyuhas, O. (1984) Nucleic Acids Res. 12, 3763–3776.
[16] Davies, B., Feo, S., Heard, E. and Fried, M. (1989) Proc. Natl. Acad. Sci. USA 86, 6691–6695.