The primary structure of rat ribosomal protein S7

Katsuyuki Suzuki, Joe Olvera and Ira G. Wool

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

Received 12 July 1990

The amino acid sequence of the rat 40S ribosomal subunit protein S7 was deduced from the sequence of nucleotides in two recombinant cDNAs and confirmed from the amino acid sequence of a cyanogen bromide peptide obtained from the protein. Ribosomal protein S7 has 194 amino acids and has a molecular mass of 22 113. Hybridization of the cDNA to digest of nuclear DNA suggests that there are 14–16 copies of the S7 gene. The mRNA for the protein is about 725 nucleotides in length. Rat S7 is homologous with *Xenopus laevis* S8. The protein contains a possible internal duplication of 10 residues.

Ribosomal protein S7 (rat); Amino acid sequence; cDNA

1. INTRODUCTION

An effort is underway to determine the sequences of amino acids in all of the proteins in the ribosomes of a single mammalian species, the rat [1]. Knowledge of the structure of ribosomes is presumed to be essential (albeit perhaps not in itself sufficient) for a rational, molecular account of the function of the organelle. As a part of this endeavor we report here the covalent structure of rat ribosomal protein S7 which we have inferred from the sequences of nucleotides in recombinant cDNAs.

2. EXPERIMENTAL

The recombinant DNA procedures and the methods used to determine the sequence of nucleotides in nucleic acids were either described or cited before [2,3].

3. RESULTS AND DISCUSSION

A random selection of 24000 cells from two cDNA libraries of 20000 and 30000 independent transformants that had been constructed from regenerating rat liver poly(A)$^+$ mRNA [3] was screened for clones that hybridized to a cDNA (generously provided by F. Amaldi) encoding *Xenopus laevis* ribosomal protein S8 [4]. Related rat and *X. laevis* ribosomal proteins have similar sequences of amino acids [1]. Five clones gave a positive signal on hybridization with the probe. The DNA from the plasmids of the 5 transformants was isolated and digested with restriction endonucleases. These clones had inserts that were either approx. 430 or 550 nucleotides in length and Southern blot hybridization with the probe indicated that the inserts might contain cDNA for a protein related to *X. laevis* S8. The sequences of nucleotides were determined in both strands of the cDNA inserts in the plasmids designated pS7-1 and pS7-3 and in overlapping sequences for each restriction site. The open reading frame in pS7-1 lacks the codons for the 69 carboxyl-terminal amino acids and pS7-3 lacks the NH_2-terminal 43 residues; together they encode the entire protein. For convenience we refer to them as pS7-1,3.

The cDNA insert in pS7-1,3 contains 670 nucleotides; it includes a 5′ noncoding sequence of 35 nucleotides, a single open reading frame of 585 nucleotides, and a 3′ noncoding sequence of 50 nucleotides with a terminal poly(A) stretch (Fig. 1). In the other two reading frames the sequence is interrupted by termination codons. The open reading frame begins at an ATG codon at a position that we designate +1 and ends with a termination codon (TAA) at position 583; it encodes 194 amino acids (Fig. 1). The initiation codon occurs in the context GCCAUGU which deviates from the optimum ACCAUGG [5]. The 3′ noncoding sequence has the hexamer AATAAA (positions 598–603) which is the recognition sequence directing post-transcriptional cleavage and polyadenylation of the 3′ end of pre-mRNA [6].

The first eight nucleotides of the S7 cDNA (positions −28 through −35 in Fig. 1) are pyrimidines, i.e. CTCTTTCT. Pyrimidine sequences are present at the 5′ end of many eukaryotic ribosomal protein mRNAs [1] and may play a role in the regulation of their translation.

The protein specified by the reading frame in pS7-1,3 was identified, in the first instance, from the amino

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

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

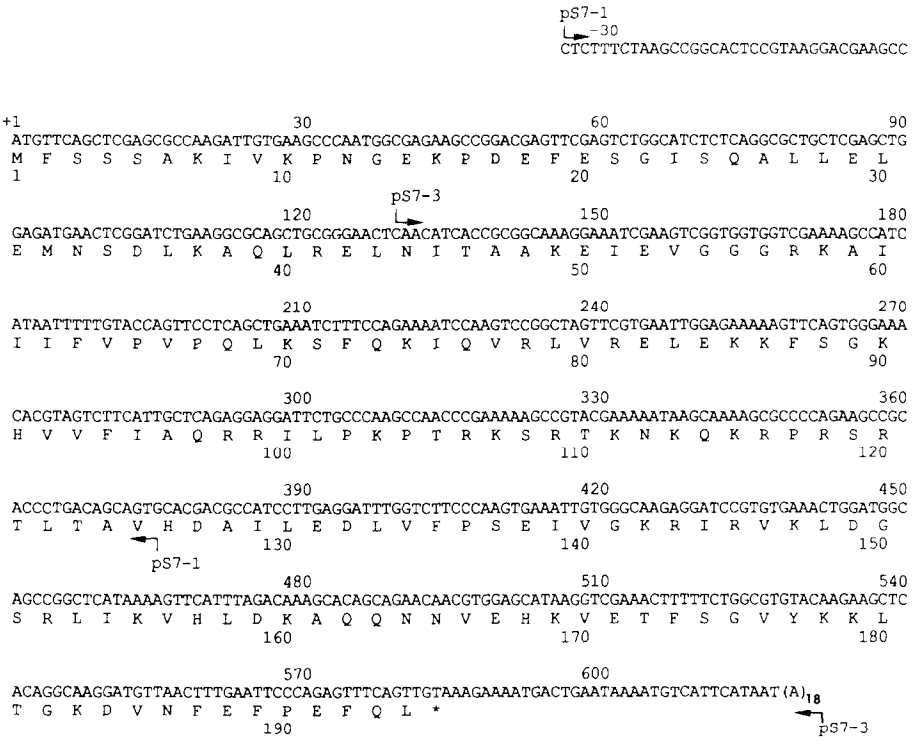

Fig. 1. The sequence of nucleotides in the cDNA inserts in plasmid pS7-1 and pS7-3 and the amino acid sequence encoded in the open reading frame. The positions of the nucleotides in the cDNA inserts are given above the residue; the positions of amino acids in protein S7 are designated below the residue; the start and stop sites for pS7-1 and pS7-3 are indicated by the vertical of the bent arrows.

acid composition (Table I) which closely approximates that derived before from a hydrolysate of rat ribosomal protein S7 [7]. The NH_2 terminus of S7 appears to be blocked since no reaction was obtained with dansyl reagent and no residues were obtained when the protein was analyzed in an automated gas phase sequencer. For this reason, S7 was cleaved with cyanogen bromide and a large peptide was prepared by high performance liquid chromatography. From the sequence of amino acids inferred from the cDNA two peptides were expected; the larger, derived from carboxyl terminus, should contain 162 residues. The sequence at the NH_2 terminus of this peptide was NSDLKAQLRELNITA which corresponds exactly to 15 residues (positions 33–47) arrived at from the cDNA.

The molecular mass of rat ribosomal protein S7, calculated from the sequence of amino acids deduced from pS7-1,3 is 22113, very close to that of 22200 estimated from the migration of the purified protein in sodium dodecyl sulfate gels [7]. Since the NH_2 terminus of S7 is blocked, we do not know whether the NH_2-terminal methionine encoded in the S7 mRNA is removed after translation.

Protein S7 lacks cysteine and tryptophan and has an excess of basic residues (14 arginyl, 24 lysyl, and 4 histidyl) over acidic ones (7 aspartyl and 16 glutamyl) (Table I). As has been noted before for ribosomal proteins, the basic residues tend to be clustered; for example, 12 of 23 residues at positions 98–120 are basic. There are two prominent hydrophilic regions in S7; residues 10–20 and 98–120.

The cDNA inserts in pS7-1 and pS7-3 were made radioactive and used to probe digests made with restric-

Table I

Amino acid	A	B
Alanine	12	10
Arginine	13	14
Aspartic acid and asparagine	15	7+7
Cysteine	ND	0
Glutamic acid and glutamine	24	16+10
Glycine	14	10
Histidine	4	4
Isoleucine	13	14
Leucine	18	18
Lysine	22	24
Methionine	2	2
Phenylalanine	10	11
Proline	9	9
Serine	10	13
Threonine	8	7
Tryptophan	ND	0
Tyrosine	2	1
Valine	18	17
Residues		194

The amino acid composition (in numbers of residues) was determined either (A) from an analysis of a hydrolysate of purified S7 [7] or inferred (B) from the sequence of nucleotides in a recombinant cDNA

tion endonucleases (*Bam*HI, *Eco*RI, or *Hin*dIII) from rat liver DNA [3]. The number of hybridization bands suggest that there are 14–16 copies of the S7 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 pseudogenes.

To determine the size of the mRNA coding for S7, total poly(A)$^+$ mRNA from rat liver was separated by electrophoresis and screened for hybridization bands using radioactive pS7-1 and pS7-3 cDNAs. One distinct band of about 725 nucleotides was detected (data not shown).

The sequence of amino acids in rat ribosomal protein S7 was compared, using the computer programs RELATE and ALIGN [8], to the sequences of amino acids in more than 500 other ribosomal proteins contained in a library that we have compiled. The comparison that yielded the highest RELATE score (70.7 SD units) was with *X. laevis* S8 [9]. In an alignment of the amino acid sequences there are 189 identities in 194 possible matches (the ALIGN score is 82.1). The differences are at positions 4 (T in *X. laevis* S8 instead of the S in rat S7); at 55 (A instead of G); at 135 (Y instead of F); at 142 (R instead of K); and at 186 (V instead of N). The two proteins obviously are homologous, i.e. derived from a common ancestral gene.

The sequence of amino acids in S7 was searched for internal duplications. There is a possible repeat at positions 85–94 (KKFS GKHVVF) and 178–187 (KKLTGKDVNF); in these two stretches of 10 amino acids there are 6 identities (_) and one related (.) residue. It is difficult to know whether the similarity is significant and if it is what the function of these sequences are. We note, however, that a number of rat ribosomal proteins have similar duplications [1].

The determination of the sequence of amino acids in rat S7 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 National Institutes of Health Grant GM 21769. We are grateful to Yuen-Ling Chan for advice and fruitful discussions and to Arlene Timosciek for aid in the preparation of the manuscript. The sequence data are in the EMBL/GenBank/DDBJ Nucleotide Sequence Databases under the accession number X53377.

REFERENCES

[1] Wool, I.G., Endo, Y., Chan, Y.L. and Glück, A. (1990) in: Structure, Function, and Evolution of Ribosomes (Hill, W.E. ed.) Am. Soc. Microbiol., Washington, DC (in press).
[2] Chan, Y.L., Lin, A., McNally, J., Peleg, D., Meyuhas, O. and Wool, I.G. (1987) J. Biol. Chem. 262, 1111–1115.
[3] Chan, Y.L. and Wool, I.G. (1988) J. Biol. Chem. 263, 2891–2896.
[4] Amaldi, F., Beccari, E., Bozzoni, I., Luo, Z.X. and Pierandrei-Amaldi, P. (1982) Gene 17, 311–316.
[5] Kozak, M. (1986) Cell 44, 283–292.
[6] Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211–214.
[7] Collatz, E., Wool, I.G., Lin, A. and Stöffler, G. (1976) J. Biol. Chem. 251, 4666–4672.
[8] 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.
[9] Mariottini, P., Bagni, C., Annesi, F. and Amaldi, F. (1988) Gene 67, 69–74.