Vol. 179, No. 1, 1991 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
August 30, 1991 Pages 314-318

THE PRIMARY STRUCTURE OF RAT RIBOSOMAL PROTEIN S28

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

Department of Biochemistry and Molecular Biology,
The University of Chicago, Chicago, Illinois 60637

Received June 24, 1991

SUMMARY: The amino acid sequence of the rat 40S ribosomal subunit protein S28 was deduced from the sequence of nucleotides in a recombinant cDNA. Ribosomal protein S28 has 69 amino acids and has a molecular weight of 7,836. Hybridization of the cDNA to digests of nuclear DNA suggests that there are 8-10 copies of the S28 gene. The mRNA for S28 is about 450 nucleotides in length. Rat S28 is homologous to *Saccharomyces cerevisiae* S33.

Ribosomes are complex ribonucleoprotein organelles that catalyze peptide bond formation and protein synthesis in all organisms in the biosphere; those in eukaryotic cells have 70 to 80 proteins and 4 species of RNA (1). An effort is being made to determine the primary structure of all of these molecules for a single mammalian species, the rat. The purpose is to establish a data base that will assist in solving the structure of the particles. 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 S28 which we have inferred from the sequence of nucleotides in a recombinant cDNA.

MATERIAL AND METHODS

The recombinant DNA procedures and the methods used to determine the sequence of nucleotides in nucleic acids have been described or cited (2, 3). An oligodeoxynucleotide probe for the cDNA encoding rat ribosomal protein S28, based on the sequence of amino acids in a cyanogen bromide peptide prepared from the protein, was synthesized. The probe was a mixture of 288 different oligodeoxynucleotides, each 26 bases in length, complementary to the sequence encoding IIRNVKGPV (residues 42-50 in S28). The oligodeoxynucleotides were synthesized on

* To whom correspondence should be addressed at Department of Biochemistry and Molecular Biology, University of Chicago, 920 East 58th Street, Chicago, IL 60637. Fax: 312-702-0439.

0006-291X/91 $1.50

a solid support by the methoxyphosphoramidite method using an Applied Biosystems, 380B, DNA synthesizer (4). The sequence of amino acids in a cyanogen bromide fragment from S28 was determined by Edman degradation in an automated gas phase sequencer.

RESULTS AND DISCUSSION

The Sequence of Nucleotides in Recombinant cDNAs Encoding Rat Ribosomal Protein S28

A random selection of 20,000 cells from two cDNA libraries of 20,000 and 30,000 independent transformants that had been constructed from regenerating rat liver poly(A)$^+$mRNA (2, 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 mRNA for rat ribosomal protein S28. Three clones gave a positive hybridization signal with the probe. The DNA from the plasmids of the 3 transformants was isolated and digested with restriction endonucleases. These clones had inserts that ranged in length from 350 to 400 nucleotides. One of the clones was selected, pRS28-1, and the sequences of nucleotides from both strands of the cDNA and overlapping sequences for each restriction site were obtained.

The cDNA insert in pRS28-1 is 298 nucleotides long and has a 5' noncoding sequence of 11 bases, a single open reading frame of 210, a 3' noncoding sequence of 77 plus a long 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 (TGA) at position 208; it encodes 69 amino

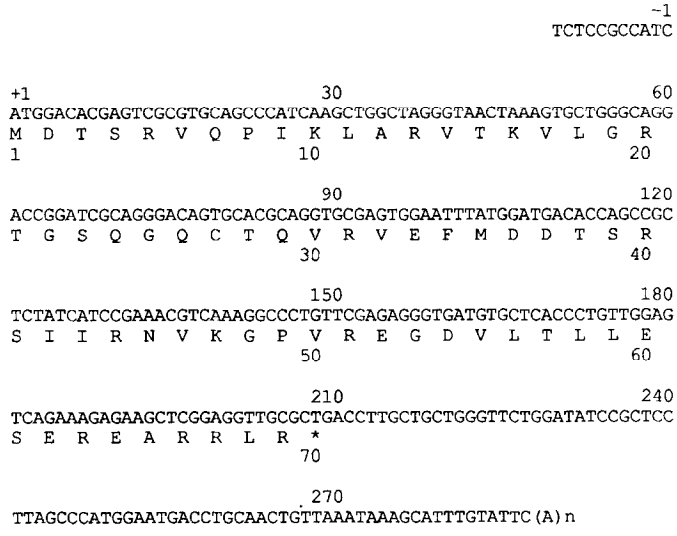

Fig. 1. *The sequence of nucleotides in the cDNA insert in plasmid pRS28-1 and the amino acid sequence encoded in the open reading frame.* The positions of nucleotides in the cDNA inserts are given above the residue; the positions of amino acids in protein S28 are designated below the residue.

acids (Fig. 1). The initiation codon occurs in the context AUC*AUG*G which deviates little from the optimum ACC*AUG*G (5). The hexamer AATAAA that directs post-transcriptional cleavage-polyadenylation of the 3' end of the precursor of the mRNA (6) is at position 270-275, 12 nucleotides upstream of the start of the poly(A) stretch. The first 5 nucleotides in the 5' noncoding sequence are pyrimidines, i.e. TCTCC (Fig. 1). There are consecutive pyrimidines at the 5' end of most, if not all, mammalian ribosomal protein mRNAs (cf. (1) for references and discussion) and the sequences may play a role in the regulation of translation (7). The pyrimidine stretch usually defines the 5' end of a ribosomal protein mRNA and its presence there implies that the cDNA is full length; if that holds for the S28 mRNA the 5' noncoding sequence is exceptionally short, only 11 nucleotides (Fig. 1).

The Primary Structure of Rat Ribosomal Protein S28

The rat ribosomal protein encoded in the open reading frame in pRS28-1 was identified as S28 by the exact congruence of the order of amino acids deduced from the sequence of nucleotides in the cDNA with the 33 residues (positions 36-68) in a cyanogen bromide peptide from the protein determined by Edman degradation using an automated gas phase sequencer. In addition, the amino acid composition inferred from the cDNA is very close to that obtained (8) from an hydrolysate of purified S28 (Table I).

The molecular weight of rat ribosomal protein S28, calculated from the sequence of amino acids deduced from pRS28-1, is 7,836. Since the NH_2 terminus of S28 is blocked we do not know whether the initial methionine encoded in the S28 mRNA is removed after translation. However, the residue next to the first methionyl in S28 is aspartyl and it has been reported (9) that when the NH_2-terminal methionine is acetylated (as may be the case in S28) the predominant penultimate amino acid is aspartic.

Protein S28 has an excess of basic residues (11 arginyl and 3 lysyl) over acidic ones (4 aspartyl and 5 glutamyl) (Table I). The protein is quite hydrophilic (23 of 69 residues are charged) especially the carboxyl-terminal half (where 17 of 39 residues are charged). S28 has no histidine, tryptophan, nor tyrosine.

The Number of Copies of the S28 Gene

The cDNA insert in pRS28-1 was made radioactive and used to probe digests of rat liver DNA made with restriction endonucleases *Bam*HI, *Eco*RI, or *Hin*dIII (3). The number of hybridization bands suggests that there are 8-10 copies of the S28 gene (data not shown). There are multiple copies of many mammalian ribosomal protein genes; however, where an analysis has been done only one of the genes has been found to be functional and the presumption is that the other copies are retroposon pseudogenes (cf. (1) for references and discussion).

TABLE I. Amino acid composition of rat ribosomal protein S28

Amino Acid	A	B
Alanine	3	2
Arginine	10	11
Aspartic acid and asparagine	6	4 + 1
Cysteine	n.d.	1
Glutamic acid and glutamine	10	5 + 4
Glycine	6	5
Histidine	0	0
Isoleucine	2	3
Leucine	6	6
Lysine	3	3
Methionine	n.d.	2
Phenylalanine	1	1
Proline	2	2
Serine	5	5
Threonine	6	6
Tryptophan	n.d.	0
Tyrosine	0	0
Valine	7	8
Residues		69

The amino acid composition (in numbers of residues) determined either (A) from an hydrolysate of S28 (8) or inferred (B) from the sequence of nucleotides in a recombinant cDNA.

The Size of the mRNA Encoding Rat Ribosomal Protein S28

To determine the size of the mRNA coding for S28 poly(A)$^+$mRNA from rat liver was separated by electrophoresis and screened for hybridization bands using radioactive pRS28-1 cDNA. One distinct band of about 450 nucleotides was detected (data not shown).

Comparison of the Sequence of Amino Acids in Rat S28 with Ribosomal Proteins from Other Species

The sequence of amino acids in rat S28 was compared, using the computer programs RELATE and ALIGN (10), to those in more than 850 other ribosomal proteins contained in a library that we have compiled. Rat ribosomal protein S28 is homologous to *Saccharomyces cerevisiae* S33 (11); the RELATE score is 22.5 SD units and in an alignment of the amino acid sequences there are 49 identities in 67 possible matches (73%); the ALIGN score is 33.7 SD units.

The sequence of amino acids in S28 was searched for internal duplications and one possible repeat was found: MDTSR (at positions 1-5) and MDDTSR (at positions 35-40).

The determination of the sequence of amino acids in rat S28 is a contribution to a data set 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 is its anticipated use 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.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health Grant GM 21769. We are grateful to our colleague Anton Glück for assistance with the computer analyses, to Veronica Paz for technical assistance, and to Arlene Timosciek for aid in the preparation of the manuscript. The sequences are in the EMBL/Gen Bank/DDBJ Nucleotide Sequence Databases under the accession number X59277.

REFERENCES

1. Wool, I.G., Endo, Y., Chan, Y.L. and Glück, A. (1990) in The Ribosome: Structure, Function and Evolution (Hill, W.E., *et al.* ed.), pp. 203-214, Amer. Soc. Microbiol., Washington, D.C.
2. Chan, Y.L., Lin, A., McNally, J. and Wool, I.G. (1987) J. Biol. Chem. 262, 12879-12886.
3. Chan, Y.L. and Wool, I.G. (1988) J. Biol. Chem. 263, 2891-2896.
4. Beaucage, S.L. and Caruthers, M.H. (1981) Tetrahedron Lett. 22, 1859-1862.
5. Kozak, M. (1986) Cell 44, 283-292.
6. Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211-214.
7. Levy, S., Avni, D., Hariharan, N., Perry, R.P. and Meyuhas, O. (1991) Proc. Natl. Acad. Sci. USA 88, 3319-3323.
8. Goerl, M., Welfle, H. and Bielka, H. (1978) Biochim. Biophys. Acta 519, 418-427.
9. Flinta, C., Persson, B., Jörnvall, H. and von Heijne, G. (1986) Eur. J. Biochem. 154, 193-196.
10. 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, D.C.
11. Leer, R.J., van Raamsdonk-Duin, M.M.C., Schoppink, P.J., Cornelissen, M.T.E., Cohen, L.H., Mager, W.H. and Planta, R.J. (1983) Nucleic Acids Res. 11, 7759-7768.