Vol. 193, No. 1, 1993 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
May 28, 1993 Pages 106-112

THE PRIMARY STRUCTURE OF RAT RIBOSOMAL PROTEIN S9

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

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

Received March 18, 1993

SUMMARY: The amino acid sequence of the rat 40S ribosomal subunit protein S9 was deduced from the sequence of nucleotides in a recombinant cDNA. Ribosomal protein S9 has 193 amino acids, the NH_2-terminal methionine is removed after translation of the mRNA, and has a molecular weight of 22,360. Hybridization of the cDNA to digests of nuclear DNA suggests that there are 14 to 16 copies of the S9 gene. The mRNA for the protein is about 1,000 nucleotides in length in part because of an especially long 5' noncoding region (103 nucleotides). Rat S9 is related to ribosomal proteins from other eukaryotes, *Saccharomyces cerevisiae* YS11 and *Dictyostelium discoideum* rp 1024, and to the eubacterial, archaebacterial, and chloroplast family of S4 ribosomal proteins. We have identified the product of the *Trypanosoma brucei* gene U as the homolog of rat ribosomal protein S9.

An effort is being made to determine the sequences of amino acids in all of the proteins in the ribosomes of a single mammalian species, the rat (1). These data are necessary, although they may not be sufficient, for a solution of the structure of the organelle and for 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 S9. Rat S9 is related to *Escherichia coli* ribosomal protein S4 which latter is essential for the initiation of the assembly of 30S ribosomal subunits (2) presumably by establishing the mature conformation of 16S rRNA (3). *E. coli* strains bearing mutations in S4 (designated *ram* for ribosome assembly mutation) have a very high translational error rate (4).

MATERIALS AND METHODS

The recombinant DNA procedures and the methods used to determine the sequence of nucleotides in the nucleic acids have been described or cited (5-7). Rat ribosomal protein S9 was isolated from a fraction of all the 40S subunit proteins (8) by high performance liquid

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

0006-291X/93 $4.00

chromatography in a reverse phase column (Bio-Rad, Hi-Pore C4) with a 0 to 60% acetonitrile gradient containing 0.1% trifluoroacetic acid at a flow rate of 0.8 ml/min. The sequence of the NH_2-terminal 27 amino acids was determined by Edman degradation in an Applied Biosystems, Model 470A, automated gas phase sequencer. In addition, a cyanogen bromide fragment of S9 was isolated and the amino acid sequence of 30 residues (positions 93-122) were determined. These two amino acid sequences were used to design primers for the polymerase chain reaction (9). The DNA generated in this way was 320 nucleotides in length and was made radioactive by the random primer method (10).

RESULTS AND DISCUSSION

The Sequence of Nucleotides in a Recombinant cDNA Encoding Rat Ribosomal Protein S9

A random selection of 24,000 cells from two cDNA libraries of 20,000 and 30,000 independent transformants that had been constructed from regenerating rat liver poly(A)$^+$mRNA (5, 6) was screened for clones that hybridized to the oligodeoxynucleotide probe that was generated in the polymerase chain reaction and that was related to amino acid sequences in rat ribosomal protein S9. Two clones gave a positive hybridization signal with the probe. The DNA from the plasmids of the 2 transformants was isolated and digested with restriction endonucleases. These clones had cDNA inserts of about 300 bases and preliminary analysis of the sequences of nucleotides indicated that they encoded only a part of S9. The cDNA insert in one of these plasmids, pS9-1, was made radioactive and used to screen a third rat cDNA library (kindly provided by M. Brownstein, NIH). Seven clones hybridized to this probe. One clone, pcD-S9-30, was selected and the sequence of nucleotides in both strands of the cDNA was determined.

The cDNA insert in pcD-S9-30 has a 5' noncoding sequence of 26 bases, a single open reading frame of 585, a 3' noncoding sequence of 76 and a long poly(A) stretch (Fig. 1). In the other two reading frames the sequence is interrupted by termination codons. Primer extension experiments established that the 5' noncoding sequence is actually 103 nucleotides long (results not shown), longer than in pcD-S9-30, indeed, amongst the longest of any rat ribosomal protein mRNA. The open reading frame begins at an AUG 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 AAC*ATGC*; the consensus is ACC*ATG*G (11). The hexamer AATAAA that directs posttranscriptional cleavage-polyadenylation of the 3' end of the precursor of the mRNA (12) is at position 640-645, 16 nucleotides upstream of the start of the poly(A) stretch.

The Primary Structure of Rat Ribosomal Protein S9

The rat ribosomal protein encoded in the open reading frame in pcD-S9-30 was identified by the correspondence of the amino acid sequence encoded in the plasmid DNA with two amino

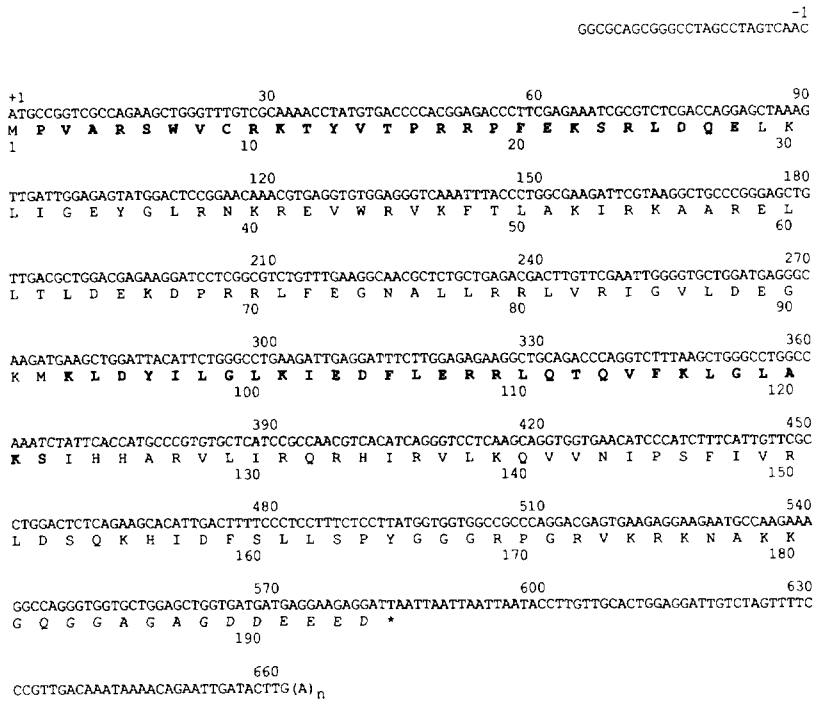

Fig. 1. *The sequence of nucleotides in the cDNA insert in pcD-S9-30 and the amino acid sequence encoded in the open reading frame.* The positions of the nucleotides in the cDNA are given above the residue; the positions of the amino acids in protein S9 are designated below the residue. The amino acid residues in bold letters were determined also by Edman degradation of S9.

acid sequences in S9 determined by Edman degradation: the NH_2-terminal 27 residues and 30 in a cyanogen bromide fragment (positions 93-122), (cf. Fig. 1). In addition, the amino acid composition inferred from the sequence of nucleotides in pcD-S9-30 and that determined from an hydrolysate of the protein (13) are in close agreement (Table I).

The molecular weight of rat ribosomal protein S9, calculated from the sequence of amino acids deduced from pcD-S9-30, is 22,491; close to the 21,200 estimated before (13) from SDS-PAGE of the purified protein. The NH_2-terminal methionine encoded in the S9 mRNA is removed after translation. Thus, the mature processed S9 has 193 residues and the molecular weight is 22,360.

Protein S9 has an excess of basic residues (25 arginyl, 19 lysyl, and 4 histidyl) over acidic ones (11 aspartyl and 13 glutamyl) (Table I). The basic residues, as is common for ribosomal proteins, tend to be clustered; for example, 7 of the 12 residues at positions 169-180. The carboxyl-terminal 6 residues are acidic. Indeed, there are a number of highly charged regions in S9 and one notable hydrophobic district (positions 140-149).

TABLE I. Amino acid composition of rat ribosomal protein S9

Amino Acid	A	B
Alanine	11	10
Arginine	28	25
Aspartic acid and asparagine	15	11 + 4
Cysteine	n.d.	1
Glutamic acid and glutamine	19	13 + 7
Glycine	17	16
Histidine	4	4
Isoleucine	11	11
Leucine	24	25
Lysine	19	19
Methionine	1	2*
Phenylalanine	7	7
Proline	7	7
Serine	7	7
Threonine	5	5
Tryptophan	n.d.	2
Tyrosine	5	4
Valine	14	14
Residues		194*

The amino acid composition (in numbers of residues) was determined either (A) from an hydrolysate of purified S9 (13) or inferred (B) from the sequence of nucleotides in pcD-S9-30.

*The NH_2-terminal methionine is removed after translation of the mRNA, thus the mature protein has 193 residues.

The Number of Copies of the S9 Gene

The cDNA insert in pcD-S9-30 was made radioactive and used to probe separate digests of rat liver DNA made with the restriction endonucleases *Bam*HI, or *Eco*RI, or *Hin*dIII (6). The number of hybridization bands suggest that there are 14 to 16 copies of the S9 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 retroposon pseudogenes.

The Size of the mRNA Encoding Rat Ribosomal Protein S9

To determine the size of the mRNA coding for S9, poly(A)$^+$mRNA from rat liver was separated by electrophoresis and screened for hybridization bands using radioactive pcD-S9-30 cDNA. One distinct band of about 1,000 nucleotides was detected (data not shown).

Comparison of the Sequence of Amino Acids in Rat S9 with Ribosomal Proteins from Other Species

The sequence of amino acids in rat S9 was compared, using the computer programs RELATE and ALIGN (14), to those in more than 1,200 other ribosomal proteins contained in a library that we have compiled; the program TFASTA (15) was used to search the GenBank DNA data base. Rat S9 is related to *Saccharomyces cerevisiae* YS11 (16) - the RELATE score is 46.2 S.D. units and in an alignment of the amino acid sequences there are 129 identities in 193 possible matches (67% identity; the ALIGN score is 70.5); and to *Dictyostelium discoideum* rp 1024 (17) - the RELATE score is 33.1 and in an alignment of the amino acid sequences there are 111 identities in 184 possible matches (60% identity; the ALIGN score is 57.0). In addition, rat ribosomal protein S9 is related to the eubacterial, archaebacterial, and organelle family of S4 proteins: to *Halobacterium marismortui* S4 (18) - the RELATE score is 12.2 and in an alignment of the amino acid sequences there are 59 identities in 171 possible matches (the ALIGN score is 24.8); to *E. coli* S4 (19) - the RELATE score is 3.3 and in an alignment there are 38 identities in 131 possible matches (the ALIGN score is 5.4); to *Zea mays* chloroplast S4 (20) - the RELATE score is 8.5 and in an alignment there are 36 identities out of 114 possible matches (the ALIGN score is 8.5).

It has been remarked before (1, 21) that the amino acid sequences of archaebacterial ribosomal proteins are closer to their eukaryotic than to their eubacterial homologs, although, the organization of archaebacterial genes mimics that of the eubacteria. This pattern holds for rat S9; it is more closely related to *H. marismortui* S4 (35% identity) than to *E. coli* S4 (29% identity) and *H. marismortui* S4 is more closely related to rat S9 (35% identity) than to *E. coli* S4 (23% identity). The relationship is even more striking when one inspects the alignments of the amino acid sequence of *H. marismortui* S4 with rat S9 and with *E. coli* S4: there are three regions of close identity between rat S9 and *H. marismortui* S4 that together span most of the two molecules and few gaps are needed for the alignment; on the other hand, to align *H. marismortui* S4 (171 amino acids) with *E. coli* S4 (203 amino acids) a number of long gaps are required and there are only two shorter regions of extended identity (comparison not shown). Indeed, the inference that rat S9 is related to *E. coli* S4 would be tenuous at best without the amino acid sequence of *H. marismortui* S4 to link them. In this respect the amino acid sequence of the archaebacterial ribosomal protein serves as a kind of Rosetta stone that allows one to identify related eukaryotic and eubacterial ribosomal proteins.

It should be noted that *E. coli* S4 is a primary 16S rRNA binding protein (2); indeed, many of the *E. coli* ribosomal proteins that can be correlated with a eukaryotic protein are primary rRNA binding proteins (unpublished observation of the authors). The implication is that rat S9 may serve a function similar to that of *E. coli* S4; S9 may help to organize 18S rRNA in

preparation for the association of other ribosomal proteins. We cannot, however, identify a sequence of amino acids that is conserved in rat S9 and in the eubacterial and archaebacterial S4 family that would support even a provisional designation as the rRNA binding module. Finally, rat S9 might play a role in translational fidelity just as its eubacterial homolog S4 does (4).

In *Trypanosoma brucei* there are two fructose bisphosphate aldolase genes that are close to each other in the genome but amidst a number of unrelated genes (22). One of these has been designated U, however, the protein encoded by U was not identified (22). A search of the GenBank DNA data base has revealed that U encodes the *T. brucei* homolog of rat ribosomal protein S9 - the RELATE score is 35.0 and in an alignment of the amino acid sequences there are 115 identities out of 186 possible matches (62% identity; the ALIGN score is 66.1).

Tripeptide Repeats in S9

There are five tripeptides in S9 whose amino acid sequences are repeated - LDE, RLD, PRR, RVL, and RRL (Table II). RRL occurs three times in S9, the others twice. The LDE tripeptide is unique to S9, i.e. it does not occur in any of the other 65 rat ribosomal proteins whose sequence has been determined; the other tripeptides occur in one or more additional rat ribosomal proteins (Table II). We cannot say whether the occurance of any of these tripeptide repeats is statistically significant and if they are what their contribution is (if there is any) to the structure or the function of the S9 ribosomal protein. It is worth noting that the PRR tripeptide has some resemblance to a nuclear localization signal (23).

The determination of the sequence of amino acids in rat S9 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

TABLE II. Tripeptide repeats in rat ribosomal protein S9

Tripeptide	Position in S9	Occurance in other rat ribosomal proteins
LDE	63-65; 87-89	
RLD	24-26; 150-152	L9
PRR	16-18; 68-70	L23a; S6*; S16; S8
RVL	127-129; 136-138	L3; L35; P0, S6; S11; S13; S18; S19
RRL	69-71; 79-82; 108-110	L5; L19; L29; L35; L37a; S6*; S15; S28; S14; L29; L34

*Indicates a ribosomal protein in which the tripeptide occurs twice.

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 Anton Glück and William Suzuki for assistance with the computer analyses and to Arlene Timosciek for aid in the preparation of the manuscript. The sequence is in the EMBL/GenBank/DDBJ Nucleotide Sequence Database under the accession number X66370.

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.*, eds.), pp. 203-214, Amer. Soc. Microbiol., Washington, D.C.
2. Nomura, M., and Held, W.A. (1974) in Ribosomes (Nomura, M. Tissieres, A., Lengyel, P., eds.), pp. 193-224, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
3. Stern, S., Powers, T., Changchien, L.M., Noller, H.F. (1989) Science 244, 783-790.
4. Davies, J., and Nomura, M. (1992) Annu. Rev. Genet. 6, 203-234.
5. Chan, Y.L., Lin, A., McNally, J., and Wool, I.G. (1987) J. Biol. Chem. 262, 12879-12886.
6. Chan, Y.L., and Wool, I.G. (1988) J. Biol. Chem. 263, 2891-2896.
7. Glück, A., Chan, Y.L., Lin, A.L., and Wool, I.G. (1989) Eur. J. Biochem. 182, 105-109.
8. Collatz, E., Lin, A., Stöffler, G., Tsurugi, K., and Wool, I.G. (1976) J. Biol. Chem. 251, 1808-1816.
9. Chan, Y.L., Devi, K.R.G., Olvera, J., and Wool, I.G. (1990) Arch. Biochem. Biophys. 283, 546-550.
10. Feinberg, A.P., and Vogelstein, B. (1984) Anal. Biochem. 137, 266-267.
11. Kozak, M. (1986) Cell 44, 283-292.
12. Proudfoot, N.J., and Brownlee, G.G. (1976) Nature 263, 211-214.
13. Collatz, E., Ulbrich, N., Tsurugi, K., Lightfoot, H.N., MacKinley, W., Lin, A., and Wool, I.G. (1977) J. Biol. Chem. 252, 9071-9080.
14. 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.
15. Pearson, W.R., and Lipman, D.J. (1988) Proc. Natl. Acad. Sci. USA 85, 2444-2448.
16. Mizuta, K., Hashimoto, T., Suzuki, K., and Otaka, E. (1991) Nucleic Acids Res. 19, 2603-2608.
17. Steel, L.F., Smyth, A., and Jacobson, A. (1987) Nucleic Acids Res. 15, 10285-10298.
18. Scholzen, T., and Arndt, E. (1992) J. Biol. Chem. 267, 12123-12130.
19. Wittmann-Liebold, B. (1984) Adv. Protein Chem. 36, 56-78.
20. Subramanian, A.R., Steinmetz, A., and Bogorad, L. (1983) Nucleic Acids Res. 11, 5277-5286.
21. Matheson, A.T., Auer, J., Ramirez, C., and Böck, A. (1990) in The Ribosome: Structure, Function, and Evolution (Hill, W.E., *et al.*, eds.), pp. 617-635, Amer. Soc. Microbiol., Washington, D.C.
22. Vijayasarathy, S., Ernest, I., Itzhaki, J.E., Sherman, D., Mowatt, M.R., Michels, P.A.M., and Clayton, C.E. (1990) Nucleic Acids Res. 18, 2967-2975.
23. Dingwall, C., and Laskey, R.A. (1991) Trends Biochem. Sci. 16, 478-481.