Vol. 202, No. 1, 1994 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
July 15, 1994 Pages 314-320

THE PRIMARY STRUCTURE OF RAT RIBOSOMAL PROTEIN S23

Yoshihisa Kitaoka, Joe Olvera, and Ira G. Wool*

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

Received June 1, 1994

SUMMARY: The amino acid sequence of the rat 40S ribosomal subunit protein S23 was deduced from the sequence of nucleotides in a recombinant cDNA. Ribosomal protein S23 has 142 amino acids, the NH_2-terminal methionine is removed after translation of the mRNA, and a molecular weight of 15,666. Hybridization of the cDNA to digests of nuclear DNA suggests that there are 6 to 13 copies of the S23 gene. The mRNA for the protein is about 650 nucleotides in length. Rat S23 is identical to a human ribosomal protein and is also related to *Saccharomyces cerevisiae* S28, to *Tetrahymena thermophila* S12, and to the prokaryotic S12 family of ribosomal proteins.

An attempt is underway to determine the structure of eukaryotic ribosomes. The endeavor is motivated by a desire to provide a rational account of the biochemistry underlying the function of the organelle in protein synthesis. It is by no means certain, perhaps not even likely, that knowledge of the structure will lead *pari passu* to an understanding of the function of ribosomes. However, it is difficult to imagine being able to describe function in molecular detail without knowledge of the structure of the particle. At any rate, if one has undertaken to attempt to solve the structure it is obvious that a requisite is the sequences of nucleotides and of amino acids in the constituent nucleic acids and proteins. A commitment has been made both to the larger enterprise, i.e. to the solution of the structure of mammalian (rat) ribosomes, and to the acquisition of the chemical data (1). As a part of this undertaking, we report here the amino acid sequence of rat ribosomal protein S23.

MATERIAL 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 (2-4). Two oligodeoxynucleotide probes were prepared based on amino acid sequences that are identical in *Saccharomyces*

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

0006-291X/94 $5.00

cerevisiae S28 (5) and in *Tetrahymena thermophila* S12 (6); the sequences are KQPNSAIRKC and GDIPGVRFK. The first probe was a mixture of 256 different oligodeoxynucleotides, each 29 nucleotides in length; the second was a mixture of 192, each 26 bases long. The oligodeoxynucleotides were synthesized on a solid support by the methoxyphosphoramidite method using an Applied Biosystems, Model 380B, DNA synthesizer.

RESULTS AND DISCUSSION

The Sequence of Nucleotides in a Recombinant cDNA Encoding Rat Ribosomal Protein S23

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 (2, 3) was screened for clones that hybridized to two oligodeoxynucleotide probes that were related to amino acid sequences predicted to occur in rat ribosomal protein S23. Four clones gave a positive hybridization signal with the probes. The sequence of nucleotides in both strands of the cDNA insert in one of these clones, pS23-14, was determined. The cDNA insert is 501 nucleotides long, has a 5' noncoding sequence of 28 bases (if the 12 additional bases in a second cDNA, pS23-18, are included), a single open reading frame of 432 nucleotides, and a 3' noncoding sequence of 41 bases and a poly(A) stretch (Fig. 1).

The open reading frame in pS23-14 begins at an AUG codon at a position that we designate +1 and ends with a termination codon (TAA) at position 430; it encodes 143 amino acids (Fig. 1). The initiation codon occurs in the context AAG*ATG*G; the consensus sequence is ACC*ATG*G (7). The hexamer AATAAA that directs posttranscriptional cleavage-polyadenylation of the 3' end of the precursor of the mRNA (8) is at position 456-461, 12 nucleotides upstream of the start of the poly(A) stretch. The first 9 nucleotides of the 5' noncoding sequence are pyrimidines, i.e. CTTTCTCTC (Fig. 1). The 5' end of most, if not all, eukaryotic ribosomal protein mRNAs have pyrimidine sequences (1) that are presumed to have a role in the regulation of their translation (9).

The Primary Structure of Rat Ribosomal Protein S23

The rat ribosomal protein encoded in the open reading frame in pS23-14 was identified as S23 by transcription of the plasmid DNA, translation of the RNA transcript in a rabbit reticulocyte lysate containing [^{3}H]leucine, and electrophoresis of the radioactive product in two dimensions in polyacrylamide gels (Fig. 2). Ribosomal protein S23 has not been resolved from S24 by conventional ion-exchange chromatography or by electrophoresis in polyacrylamide gels containing sodium dodecyl sulfate (10); albeit in the latter procedure the band is diffuse and gives the impression of two proteins with slightly different molecular weights (10). The two proteins are not completely resolved by electrophoresis in two dimensions in polyacrylamide gels containing urea either, although, in the best of circumstances the configuration of the single spot

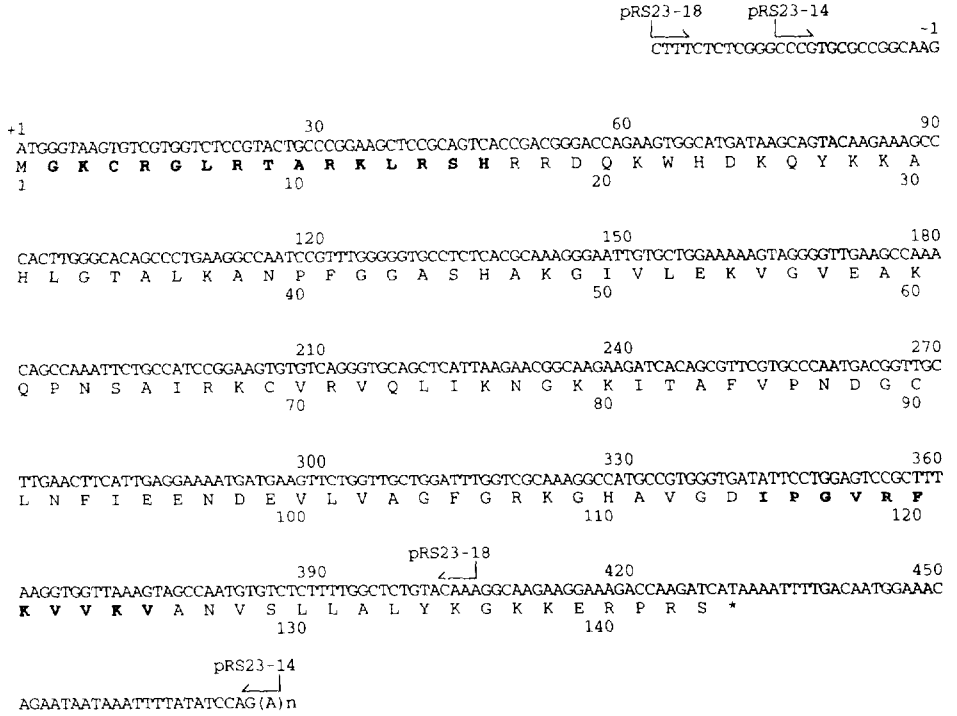

Fig. 1. *The sequence of nucleotides in the cDNA inserts in pS23-14 and pS23-18 and the amino acid sequence encoded in the open reading frame.* The positions of the nucleotides in the cDNA are given above the residue; the positions of the amino acids in protein S23 are designated below the residue. The initial and terminal nucleotides in pS23-14 and pS23-18 are designated by the vertical of the bent arrows. The amino acid residues in bold letters were determined also by Edman degradation of peptides from S23.

clearly indicates that they are separate proteins (but not in Fig. 2A). The translation product obtained from the pS23-14 transcript aligns with the S23/S24 spot (Fig. 2). The amino acid sequence of S24 was determined before (11) and it is different than the one encoded in pS23-14. In addition, a preparation that contained S23 and S24 was hydrolyzed with acid, the peptides separated, and the amino acid sequences determined by Edman degradation using an automated gas phase sequencer. The amino acid sequence in two of those peptides (residues 2-16 and 115-125; given in bold letters in Fig. 1) corresponded to amino acid sequences in pS23-14 and do not occur in S24 (11).

The molecular weight of rat ribosomal protein S23, calculated from the sequence of amino acids deduced from pS23-14, is 15,797. However, the NH_2-terminal methionine encoded in the S23 mRNA is removed after translation, i.e. it is not in the NH_2-terminal sequence determined directly from the protein. The residue next to the initial methionine in S23 is glycine which favors NH_2-terminal processing (12). Thus, the mature processed S23 has 142 residues and the

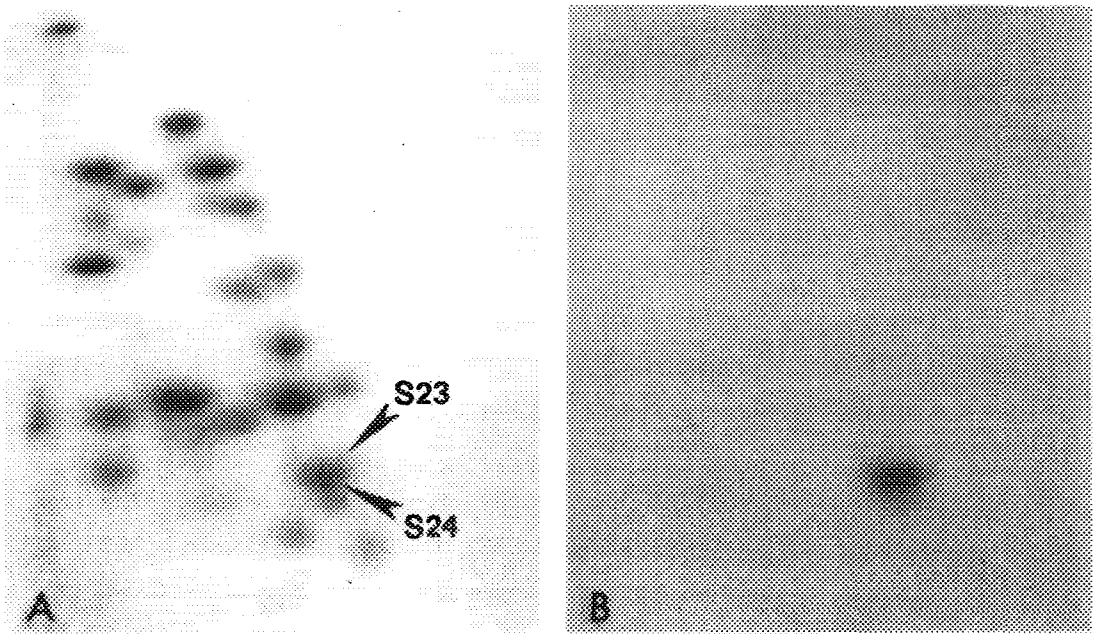

Fig. 2. *Two-dimensional electrophoresis of the product of the translation of the RNA obtained by transcription of the pS23-14 cDNA.* A reticulocyte lysate (50 µl) was incubated with the RNA and [^{3}H]leucine and a sample (10 µl) was extracted with 67% acetic acid and the protein precipitated with 90% acetone. Electrophoresis of the radioactive translation product was with 80 µg of carrier protein from 40S ribosomal subunits. Electrophoresis in polyacrylamide gels containing urea was from left to right in the first dimension and from top to bottom in the second. A, Coomassie brilliant blue stain of the gel; B, fluorography of the gel.

molecular weight is 15,666; the molecular weight estimated before (10) from SDS-PAGE of the purified protein was 18,800.

Protein S23 has a large excess of basic residues (12 arginyl, 20 lysyl, and 5 histidyl) over acidic ones (5 aspartyl and 6 glutamyl). The basic residues, as is common in ribosomal proteins, tend to be clustered; for example, 15 of 30 amino acids at positions 2-31 and 5 of 8 at positions 135-142 (Fig. 1). S23 has no methionine.

The Number of Copies of the S23 Gene

The cDNA insert in pS23-14 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 (3). The number of hybridization bands suggests that there are 6 to 13 copies of the S23 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 S23

To determine the size of the mRNA coding for S23, poly(A)$^+$mRNA from rat liver was separated by electrophoresis and screened for hybridization bands using radioactive pS23-14 cDNA. One distinct band of about 650 nucleotides was detected (data not shown).

Comparison of the Sequence of Amino Acids in Rat S23 with Ribosomal Proteins from Other Species

The sequence of amino acids in rat S23 was compared, using the computer programs RELATE and ALIGN (13), to those in more than 1,400 other ribosomal proteins contained in a library that we have compiled; the program TBLASTN (14) was used to search the GenBank DNA data base.

The comparison that yielded the closest identity was with a human protein that had been identified as a homolog of yeast ribosomal protein S28 (15); the RELATE score is 54.9 S.D. units. The amino acid sequences of the rat and human proteins are identical. Rat S23 is also related to *S. cerevisiae* S28 (5) - the RELATE score is 45.1 and in an alignment of the amino acid sequences there are 111 identities in 143 possible matches (78% identity); to *T. thermophila* S12 (6) - the RELATE score is 38.8 and in an alignment there are 97 identities in 140 possible matches (69% identity). S23 is also related to the members of the eubacterial, archaebacterial, and organelle S12 family of ribosomal proteins; representative examples are *Halobacterium halobium* S12 (16) - the RELATE score is 28.9 and in an alignment there are 77 identities in 141 possible matches (55% identity); and *Escherichia coli* S12 (17) - the RELATE score is 3.7 and in an alignment there are 33 identities in 112 possible matches (29% identity but see below). Thus there is conservation of this protein across the three kingdoms suggesting it was a component of an early ribonucleoprotein ribosome.

E. coli ribosomal protein S12, with S4 and S5, is essential for maintenance of accurate translation (18) perhaps by tuning the structure of 16S rRNA (19, 20). A mutation in the lysine at position 42 of *E. coli* S12, a residue that is in the streptomycin binding site, causes resistance to the antibiotic and leads to increased accuracy of translation by the ribosomes (21). It is noteworthy, that this lysine is conserved in rat S23 (it is at position 60) and is in a region where the amino acid sequences of the rat and *E. coli* proteins are closely related - between residues 60 and 78 of rat S23 there are 12 identities in 18 possible matches, i.e. 67% identity, in an alignment with residues 42 to 59 of *E. coli* S12 - there is one gap. There is one other conserved region in rat S23 and *E. coli* S12 - between residues 112 and 126 of rat S23 there are 8 amino acid identities and 2 conservative changes, i.e. 53% identity, or 67% similarity in an alignment with residues 86-100 in *E. coli* S12. These comparisons strengthen the inference that rat S23 and *E. coli* S12 are homologous and that they make similar contributions to the accuracy of

translation. Mutations in the *E. coli* ribosomal proteins S4 or S5 (*ram* mutations for *r*ibosomal *am*biguity) decrease the accuracy of translation (18, 22). The *S. cerevisiae* homologs of these ribosomal proteins function in a like manner to affect the accuracy of translation (23). It is quite likely that the mechanism of translational fidelity has been conserved during evolution (23) and that rat S23, which is related to *E. coli* S12 and *S. cerevisiae* S28, participates in the process.

The determination of the amino acid sequence of rat ribosomal protein S23 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 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 unraveling 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 Yuen-Ling Chan for advice and fruitful discussions, to Anton Glück for help with the computer analysis, and to Arlene Timosciek for the preparation of the manuscript. The sequence data will appear in the EMBL/GenBank/DDBJ Nucleotide Sequence Database under accession number X77398.

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. 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. Glück, A., Chan, Y.L., Lin, A.L., and Wool, I.G. (1989) Eur. J. Biochem. 182, 105-109.
5. Alksne, L.E., and Warner, J.R. (1993). *J. Biol. Chem.* 268, 10813-10819.
6. Nielsen, H., Andersen, P.H., Dreisig, H., Kristiansen, K., and Engberg, J. (1986). *EMBO J.* 5, 2711-2717.
7. Kozak, M. (1986) Cell 44, 283-292.
8. Proudfoot, N.J., and Brownlee, G.G. (1976) Nature 263, 211-214.
9. Levy, S., Avni, D., Hariharan, N., Perry, R.P., and Meyuhas, O. (1991) Proc. Natl. Acad. Sci. USA 88, 3319-3323.
10. Collatz, E., Wool, I.G., Lin, A., and Stöffler, G. (1976) J. Biol. Chem. 251, 4666-4672.
11. Chan, Y.L., Paz, V., Olvera, J., and Wool, I.G. (1990) FEBS Lett. 262, 253-255.
12. Flinta, C., Persson, B., Jörnvall, H. and von Heijne, G. (1986) Eur. J. Biochem. 154, 193-196.
13. 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.

14. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D. (1990). *J. Mol. Biol.* 215, 403-410.
15. Hori, N., Murakawa, K., Matoba, R., Fukushima, A., Okubo, K., and Matsubara, K. (1993). *Nucleic Acids Res.* 21, 4394.
16. Leffers, H., Gropp, F., Lottspeich, F., Zillig, W., and Garrett, R.A. (1989). *J. Mol. Biol.* 206, 1-17.
17. Funatsu, G., Yaguchi, M., and Wittmann-Liebold, B. (1977). *FEBS Lett.* 73, 12-17.
18. Gorini, L. (1974). In *Ribosomes* (Nomura, M., *et al.*, ed.), pp. 791-803, Cold Spring Harbor Laboratory, New York.
19. Stern, S., Powers, T., Changchien, L.M, and Noller, H.F. (1988). *J. Mol. Biol.* 201, 683-695.
20. Allen, P.N., and Noller, H.F. (1989). *J. Mol. Biol.* 208, 457-468.
21. Funatsu, G., and Wittmann, H.G. (1972). *J. Mol. Biol.* 68, 547-550.
22. Piepersberg, W., Böck, A., and Wittmann, H.G. (1975). *Mol. Gen. Genet.* 140, 91-100.
23. Alksne, L.E., Anthony, R.A., Liebman, S.W., and Warner, J.R. (1993). *Proc. Natl. Acad. Sci. USA* 90, 9538-9541.