Volume 262, number 2, 253–255 FEBS 08254 March 1990

The primary structure of rat ribosomal protein S24

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 5 February 1990

The amino acid sequence of rat ribosomal protein S24 was deduced from the sequence of nucleotides in a recombinant cDNA. S24 contains 133 amino acids and has a molecular mass of 15 413. Hybridization of the cDNA to digests of nuclear DNA suggests that there are 12–16 copies of the S24 gene. The mRNA for the protein is about 600 nucleotides in length. Rat S24 is homologous to *Xenopus laevis* S19 and related to *Halobacterium morismortui* ribosomal protein S15.

Ribosomal protein S24; Amino acid sequence; cDNA; (Rat)

1. INTRODUCTION

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 it 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 S24. Mammalian ribosomal protein S24 can be crosslinked to initiation factor eIF-3 [2] which is necessary for the binding of mRNA to ribosomes and, hence, it is located at the site where the factor binds and where initiation of translation of mRNA occurs.

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 [3]. The NH_2-terminal amino acid sequence of a mixture of rat ribosomal proteins S23 and S24 was determined by Edman degradation using an Applied Biosystems, Model 470A, automated gas phase protein sequencer. Radioactive rat ribosomal protein S24 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 S24 and other ribosomal proteins. The scoring matrix was Dayhoff's MDM '78 [6].

3. RESULTS AND DISCUSSION

Two cDNA libraries of 30 000 and of 20 000 independent transformants were contructed from poly(A)$^+$ mRNA prepared from regenerating rat liver [3]. A random selection of 25 000 cells from each library was screened for clones that hybridized to a cDNA (generously provided by F. Amaldi) that encoded *Xenopus laevis* ribosomal protein S19 [7]. Related rat and *Xenopus* ribosomal proteins have similar sequences of amino acids [1]. Three colonies gave a positive hybridization signal with the probe. DNA from the plasmids of the 3 transformants was isolated, digested with restriction endonucleases, and analyzed by gel electrophoresis. One of the clones, designated pS24-5, had an insert approximately 450 nucleotides in length and Southern blot hybridization with the oligonucleotide probe confirmed that it contained cDNA for a protein related to *X laevis* S19; it was later identified as rat ribosomal protein S24. The length of the rat S24 coding sequence was anticipated from the molecular weight of the protein [8] to be about 400 nucleotides. The sequences of nucleotides were determined in both strands of the cDNA insert in pS24-5 and in overlapping sequences for each restriction site.

The cDNA insert in pS24-5 contains 466 nucleotides and includes a 5′ noncoding sequence of 32 nucleotides, an open reading frame of 402 nucleotides, and a 3′ noncoding sequence of 32 nucleotides (fig.1). In the other two reading frames the sequence is interrupted by many termination codons. The open reading frame begins at an ATG codon at a position that we designate +1 and ends with a termination codon (TAG) at nucleotides 399–402; it encodes 133 amino acids (fig.1). The initiation codon occurs in the context AUCAUGA which deviates from the optimum ACCAUGG [9]. The 3′ noncoding sequence lacks the hexamer AATAAA which is the recognition sequence directing post-transcriptional cleavage-polyadenylation of the 3′ end of pre-mRNA [10]. The first nine nucleotides of the S24 cDNA (positions −32 through

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

```
                                                                -30
                                                                TCCTCCTTTACAGCTCGGGCACCGTAGCCATC

+1                              30                              60                              90
ATGAATGACACGGTAACCATCCGGACCAGGAAGTTCATGACAAACCGTCTGCTTCAGAGGAAACAGATGGTCATTGATGTCCTTCATCCT
M  N  D  T  V  T  I  R  T  R  K  F  M  T  N  R  L  L  Q  R  K  Q  M  V  I  D  V  L  H  P
                               10                              20                              30

                               120                             150                             180
GGGAAGGCCACAGTACCAAAGACAGAAATTCGGGAAAAGCTGGCCAAGATGTACAAAACCACACCAGATGTCATCTTTGTATTTGGATTC
G  K  A  T  V  P  K  T  E  I  R  E  K  L  A  K  M  Y  K  T  T  P  D  V  I  F  V  F  G  F
                               40                              50                              60

                               210                             240                             270
AGAACCCACTTTGGTGGAGGCAAGACAACTGGCTTTGGCATGATCTATGATTCTTTAGATTATGCAAAGAAGAATGAGCCTAAACACAGA
R  T  H  F  G  G  G  K  T  T  G  F  G  M  I  Y  D  S  L  D  Y  A  K  K  N  E  P  K  H  R
                               70                              80                              90

                               300                             330                             360
CTTGCGAGACATGGCCTTTATGAGAAGAAAAAGACCTCCCGCAAACAGCGAAAAGAACGAAAGAACAGAATGAAGAAGGTCAGGGGGACT
L  A  R  H  G  L  Y  E  K  K  K  T  S  R  K  Q  R  K  E  R  K  N  R  M  K  K  V  R  G  T
                              100                             110                             120

                               390                             420
GCAAAGGCCAATGTTGGTGCTGGCAAAAAGCCAAAGGAGTAGTTCTACGGTGACTTGATGTTCTGCTGTGATAT
A  K  A  N  V  G  A  G  K  K  P  K  E  *
                              130
```

Fig.1. The sequence of nucleotides in the cDNA insert in plasmid pS24-5 and the amino acid sequence encoded in the open reading frame. The positions of the nucleotides in the cDNA insert are given above the residue; the positions of amino acids in protein S24 are designated below the residue.

−24 in fig.1) are pyrimidines, i.e. TCCTCCTTT. Pyrimidine sequences have been reported to be present in the 5′ untranslated region of many eukaryotic ribosomal protein mRNAs [1] and may play a role in the regulation of their translation.

The rat ribosomal protein specified by pS24-5 was identified as S24 by hybridization of the plasmid DNA to poly(A)$^+$ mRNA and selection of the complimentary mRNA. The latter was translated in a reticulocyte lysate and the radioactive product identified as S24 by its migration on two-dimensional polyacrylamide gels (fig.2). Ribosomal protein S24 has not been resolved from S23 either by conventional ion-exchange chromatography or by electrophoresis in polyacrylamide gels containing sodium dodecyl sulfate [8]; albeit in the latter procedure the band is diffuse and gives the impression of two proteins with slightly different molecular weights [8]. 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 clearly indicates that they are separate proteins (cf. fig.3). The translation product obtained from the

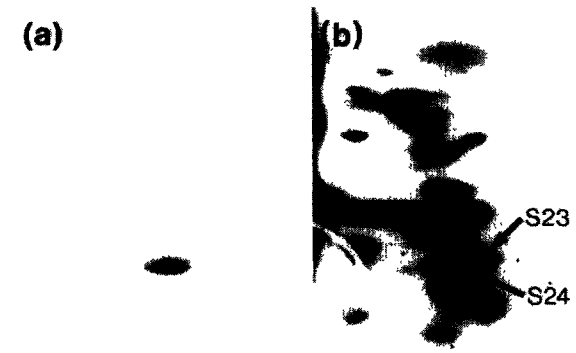

Fig.2. Two-dimensional electrophoresis of the product of the translation of the hybrid-selected mRNA for S24. A reticulocyte lysate (10 μl) was incubated with the mRNA selected with pS24-5 from a preparation of rat liver poly(A)$^+$ mRNA and a sample (3 μl) was extracted with 67% acetic acid and the protein precipitated with 90% acetone. Electrophoresis of the radioactive translation product (labeled with [^{35}S]methionine) was with 100 μg of carrier protein from 40 S 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) Fluorography of the gel; (b) Coomassie brilliant blue stain of the gel.

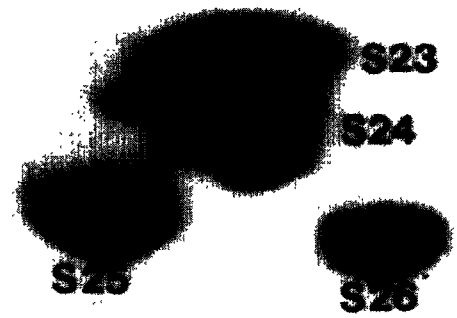

Fig.3. Two-dimensional polyacrylamide gel showing the partial resolution of S23 and S24. This is a region of a polyacrylamide gel stained with Coomassie brilliant blue after electrophoresis in two dimensions of a mixture of proteins from rat 40 S ribosomal subunits showing partial separation of S23 and S24.

mRNA selected with pS24-5 (fig.2) aligns with the lower portion of the S23/S24 spot that has been designated S24. A mixture of rat 40 S ribosomal subunit proteins was resolved by high-performance liquid chromatography and a sample was isolated that contained some S23 but was primarily S24. The predominant sequence obtained from this preparation by Edman degradation using a gas phase automated sequencer was XXDTVTIXTRKFMTNRL which corresponds to the sequence of amino acids deduced from the sequence of nucleotides in pS24-5.

The molecular weight of rat ribosomal protein S24, calculated from the sequence of amino acids, is 15 413. Protein S24 lacks tryptophan and cysteine and has a large excess of basic residues (13 arginyl, 23 lysyl, and 4 histidyl) over acidic ones (5 aspartyl and 6 glutamyl) (table 1). Many of the basic residues are clustered; for example, 15 of the 26 residues between positions 93 and 118 are basic. Indeed, there are two hydrophilic regions of the protein; positions 31–49 (8 of 18 residues are charged); positions 80–118 (24 of 39 residues are charged).

The cDNA insert in pS24-5 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 12–16 copies of the S24 gene (data not shown). There are multiple copies of many other mammalian ribosomal protein genes (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 genome contains only one ribosomal protein gene that is expressed and that the other copies are nonfunctional pseudogenes.

To determine the size of the mRNA for S24, glyoxylated total poly(A)$^+$ mRNA from rat liver was separated by electrophoresis and screened for hybridization bands using radioactive pS24-5 cDNA. One band of about 600 bases was detected (data not shown).

The sequence of amino acids in rat ribosomal protein S24 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 S24 is homologous to *X. laevis* S19; the RELATE score is 52.1 SD units. In an alignment of the amino acid sequences there are 129 identities out of 132 possible matches (the ALIGN score is 57.3). The differences are two conservative changes, an arginine in *Xenopus* to a lysine in rat at position 93 and a glutamic acid to an aspartic acid at position 133, and the insertion of a proline at position 131 in the rat sequence. Rat S24 is also related to *Halobacterium morismortui* S15 – the RELATE score is 10.7 and in the alignment there are 25 identities in 102 possible matches (the ALIGN score is 15.3). The sequence of amino acids in rat S24 was searched for internal repeats but none were found.

Table 1

The amino acid composition (in numbers of residues) of S24 determined from the sequence of nucleotides in a recombinant cDNA

Alanine	7	Leucine	7
Arginine	13	Lysine	23
Aspartic acid	5	Methionine	6
Asparagine	5	Phenylalanine	6
Cysteine	0	Proline	5
Glutamic acid	6	Serine	2
Glutamine	3	Threonine	13
Glycine	11	Tryptophan	0
Histidine	4	Tyrosine	4
Isoleucine	5	Valine	8
		Residues	133

The determination of the sequence of amino acids in rat S24 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 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 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/Gen Bank/DDBJ Nucleotide Sequence Databases under the accession number X51538

REFERENCES

[1] Wool, I.G., Endo, Y., Chan, Y.L. and Glück, A. (1990) in: The Structure, Function and Evolution of Ribosomes (Hill, W.E. ed) Am. Soc. Microbiol., Washington, DC, in press.
[2] Tolan, D.R., Hershey, J.W. and Traut, R.T. (1983) Biochemie 65, 427–436.
[3] Chan, Y.L., Lin, A., McNally, J. and Wool, I.G. (1987) J. Biol. Chem. 262, 12879–12886.
[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, vol. 5, suppl. 3, pp. 1–8, National Biomedical Research Foundation, Washington, DC.
[7] Amaldi, F., Beccari, E., Bozzoni, I., Luo, Z.X. and Pierandrei-Amaldi, P. (1982) Gene 17, 311–316.
[8] Collatz, E., Wool, I.G., Lin, A. and Stöffler, G. (1976) J. Biol. Chem. 251, 4666–4672.
[9] Kozak, M. (1986) Cell 44, 283–292.
[10] Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211–214.