Vol. 162, No. 1, 1989 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
July 14, 1989 Pages 364-370

THE PRIMARY STRUCTURE OF RAT RIBOSOMAL PROTEIN L21

K.R.G. Devi, Yuen-Ling Chan, and Ira G. Wool*

Department of Biochemistry and Molecular Biology,

The University of Chicago, Chicago, Illinois 60637

Received June 5, 1989

SUMMARY: The covalent structure of rat ribosomal protein L21 was deduced from the sequence of nucleotides in a recombinant cDNA and confirmed from the NH_2-terminal amino acid sequence of the protein. Ribosomal protein L21 contains 159 amino acids (the NH_2-terminal methionine is removed after translation of the mRNA) and has a molecular weight of 18,322. Hybridization of the cDNA to digests of nuclear DNA suggests that there are 16-23 copies of the L21 gene. The mRNA for the protein is about 680 nucleotides in length.

A solution of the structure of eukaryotic ribosomes is deemed important since it is believed, with cause, to be essential for a rational, molecular account of the function of the organelle in protein synthesis. A prime requirement for solving the structure is knowledge of the sequence of nucleotides and amino acids in the constituent nucleic acids and proteins. A commitment has been made to the acquisition of this data for mammalian (rat) ribosomes. We report here the structure of rat ribosomal protein L21 which we have inferred from the sequence of nucleotides in a recombinant cDNA and which we have confirmed by sequencing portions of the protein.

MATERIALS AND METHODS

The recombinant DNA procedures and the methods used to determine the sequence of nucleotides in the nucleic acid were either described or cited before (1, 2). The strategy that was used to design the probes for the cDNA encoding rat ribosomal

*To whom correspondence should be addressed at Department of Biochemistry and Molecular Biology, The University of Chicago, 920 East 58th Street, Chicago, Illinois 60637.

0006-291X/89 $1.50

protein L21 based on a sequence of 12 amino acids in the protein has also been reported (1). Probe 1, a mixture of 96 different oligodeoxynucleotides each 20 nucleotides in length based on the sequence IYKKGDI, and probe 2, a mixture of 288 different oligodeoxynucleotides each 20 nucleotides in length based on the overlapping sequence DIVDIKG, were synthesized on a solid support by the methoxyphosphoramidite method using an Applied Biosystems, Model 380B, DNA synthesizer (3), and the oligonucleotides were purified by polyacrylamide gel electrophoresis.

Radioactive rat ribosomal protein L21 cDNA was hybridized to restriction enzyme digests of genomic DNA (2), and to a preparation of rat liver poly(A)$^+$mRNA.

The computer programs RELATE and ALIGN were used to assess possible evolutionary relationships between rat L21 and other ribosomal proteins. The scoring matrix was Dayhoff's MDM '78 (4).

RESULTS AND DISCUSSION

The Sequence of Nucleotides in a Recombinant cDNA Encoding Rat Ribosomal Protein L21.

A random selection of 24,000 colonies from two cDNA libraries of 30,000 and of 20,000 independent transformants were screened for clones hybridizing to two oligonucleotide probes complementary to the nucleotides predicted to be present in the portion of the mRNA that encodes 12 amino acids (IYKKGDIVDIKG) near the amino terminus of rat ribosomal protein L21. Three colonies gave a positive hybridization signal with the probe. The DNA from the plasmids of the three transformants was isolated and digested with restriction endonucleases. These clones had inserts of 0.6 to 0.7 kb and Southern blot hybridization with the probe confirmed that all contained cDNA for L21. The anticipated length of the L21 coding sequence, calculated from the molecular weight of the protein (5), is 530 nucleotides. The clone with the largest insert, designated pL21-2, was selected and the sequences of nucleotides from both strands of the cDNA and overlapping sequences for each restriction site were obtained.

The cDNA insert in pL21-2 contains 554 nucleotides and includes a 5' noncoding sequence of 39 nucleotides, a single open reading frame, 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 of 483 nucleotides begins at an ATG codon at a position that we designate +1 and ends with a termination codon TAA at position 481; it encodes 160 amino acids (Fig. 1). The initiation codon occurs in the context AAAATGA which differs from the optimum ACCATGG (6). The hexamer AATAAA, presumed to be the recognition sequence directing post transcriptional cleavage-polyadenylation of the 3' end of pre-mRNA (7), is located at position 497-502, 13 nucleotides upstream of the start of a poly(A) stretch of 12 nucleotides.

In the 5' untranslated region, at positions -39 to -31, there is a sequence (CTTTCCGCC) in which 8 of 9 residues are pyrimidines (Fig. 1). Sequences of pyrimidines have been reported to be present at the end of the 5' untranslated region

```
          -30                                    +1                                    30
CTT TCC GCC GGG GCA GCC ATC TTC CAG TAA CTC GCC AAA ATG ACG AAC ACA AAA GGA AAG AGG AGG GGT ACT CGG TAT ATG TTC TCT AGA
                                                   MET THR ASN THR LYS GLY LYS ARG ARG GLY THR ARG TYR MET PHE SER ARG
                                                   1                                     10

        60                                     90                                    120
CCT TTT AGG AAA CAT GGA GTC GTT CCT TTG GCC ACA TAC ATG CGA ATC TAC AAG AAG GGT GAT ATT GTA GAC ATC AAG GGA ATG GGC ACT
PRO PHE ARG LYS HIS GLY VAL VAL PRO LEU ALA THR TYR MET ARG ILE TYR LYS LYS GLY ASP ILE VAL ASP ILE LYS GLY MET GLY THR
        20                                     30                                    40

       150                                    180                                   210
GTT CAA AAA GGA ATG CCC CAT AAG TGT TAC CAT GGC AAA ACC GGA AGA GTC TAC AAT GTC ACC CAG CAT GCC GTG GGC ATC ATT GTG AAC
VAL GLN LYS GLY MET PRO HIS LYS CYS TYR HIS GLY LYS THR GLY ARG VAL TYR ASN VAL THR GLN HIS ALA VAL GLY ILE ILE VAL ASN
        50                                     60                                    70

       240                                    270                                   300
AAG CAA GTT AAA GGC AAG ATT CTG GCC AAG AGG ATC AAT GTG CGG ATT GAG CAC ATC AAG CAC TCA AAG AGC AGA GAC AGC TTC CTG AAG
LYS GLN VAL LYS GLY LYS ILE LEU ALA LYS ARG ILE ASN VAL ARG ILE GLU HIS ILE LYS HIS SER LYS SER ARG ASP SER PHE LEU LYS
        80                                     90                                    100

       330                                    360                                   390
CGG GTG AAG GAG AAC GAT CAG AAG AAA AAG GAA GCC AAA GAG AAG GGC ACC TGG GTT CAG CTT AAC GGC CAG CCT GCG CCA CCC AGA GAA
ARG VAL LYS GLU ASN ASP GLN LYS LYS LYS GLU ALA LYS GLU LYS GLY THR TRP VAL GLN LEU ASN GLY GLN PRO ALA PRO PRO ARG GLU
       110                                    120                                   130

       420                                    450                                   480
GCC CAC TTT GTG AGG ACT AAC GGA AAG GAG CCT GAG CTG CTG GAG CCC ATT CCA TAC GAA TTC ATG GCC TAA TGT ATA CAA AGG AAA TAA
ALA HIS PHE VAL ARG THR ASN GLY LYS GLU PRO GLU LEU LEU GLU PRO ILE PRO TYR GLU PHE MET ALA END
       140                                    150                                   160

       510
AGA ACC CGA ACT GC
```

Fig. 1. The sequence of nucleotides in the cDNA insert in plasmid pL21-2 and the amino acid sequence encoded in the open reading frame. The position of the nucleotides in the cDNA insert is given above the residue; the position of amino acids in the protein (derived from the nucleotide sequence) is designated below the residue.

of many eukaryotic ribosomal protein mRNAs (8-13) and may play a role in the regulation of their translation.

The Primary Structure of Rat Ribosomal Protein L21

The reading frame in pL21-2 is flanked by initiation and termination codons and specifies a protein of 160 amino acids (Fig. 1). This protein was identified as rat ribosomal protein L21 in the following manner: The recombinant cDNA clone pL21-2 was selected using two oligodeoxynucleotide probes that were complementary to the codons for a sequence of 12 amino acids near the NH_2 terminus of L21. The amino acid composition inferred from the cDNA is very close to that previously derived (5) from an hydrolysate of purified L21 (Table I). The sequence of amino acids deduced from the sequence of nucleotides in pL21-2 corresponds to the NH_2-terminal 45 residues determined directly from protein L21 (data not shown).

The molecular weight of rat ribosomal protein L21, calculated from the sequence of amino acids deduced from pL21-2, is 18,453. However the NH_2-terminal methionine encoded in the L21 mRNA is removed after translation since it is not found in the amino acid sequence derived from the protein. The residue next to the initiator methionine in L21 is threonyl which has been reported to favor NH_2-terminal processing (14). Thus, the number of residues in the mature protein is 159 and the

Table I Amino acid composition of rat ribosomal protein L21

Amino Acids	A	B
Alanine	8	7
Arginine	15	13
Aspartic acid and Asparagine	11	4 + 7
Cysteine	-	1
Glutamic acid and Glutamine	16	9 + 6
Glycine	15	14
Histidine	7	7
Isoleucine	10	10
Leucine	7	6
Lysine	21	23
Methionine	4	6[a]
Phenylalanine	5	5
Proline	9	9
Serine	5	4
Threonine	7	9
Tryptophan	-	1
Tyrosine	6	6
Valine	14	13
Residues		160

[a]The NH_2-terminal methionine is removed after translation of the mRNA.
The amino acid composition (in numbers of residues) was determined either (A) from an analysis of an hydrolysate of purified L21 (5) or inferred (B) from the sequence of nucleotides in a recombinant cDNA.

molecular weight is 18,322, close to that of 20,300 estimated from the migration of the purified protein in sodium dodecyl sulfate gels (5).

Protein L21 has an excess of basic over acidic residues; a total of 43 of the former and 13 of the latter (Table I). The basic residues tend to be clustered, occurring in groups of 3, or 3 of 4, or 4 of 5 or 6. This has been observed before for ribosomal proteins but its significance, if there is any, is not known. There is a striking hydrophilic region in L21 near the carboxyl terminus (positions 107-122). There are also several hydrophobic stretches with peaks near positions 25, 75 and 155. We note that 6 of the 9 prolyl residues occur near the carboxyl terminus (positions 132-155).

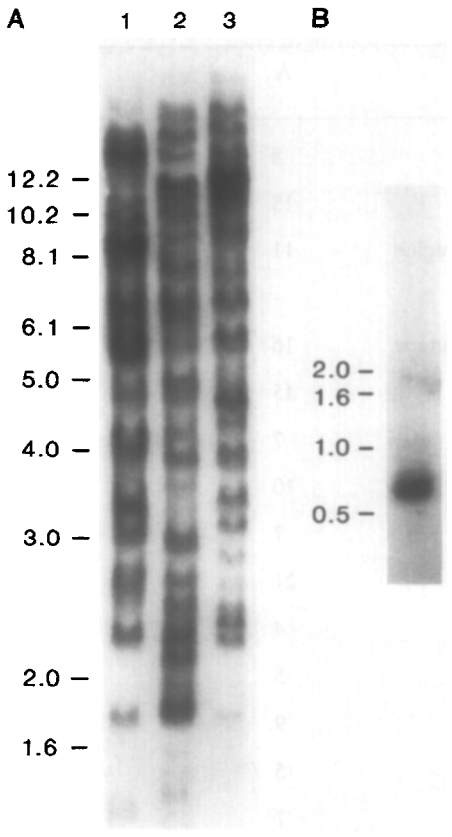

Fig. 2. Hybridization of ribosomal protein L21 cDNA to rat genomic DNA and to poly(A)$^+$mRNA. In A, rat nuclear DNA (10 μg) was digested with restriction enzymes: HindIII (lane I); EcoRI (lane II); or BamHI (lane III). The digests were resolved by electrophoresis in 0.7% agarose gels and transferred to GeneScreen Plus nylon filters. The uniformly labeled radioactive cDNA insert from pL21-2 was hybridized to the immobilized genomic DNA. The position to which DNA standards of the size designated (in kilobase pairs) migrate is shown to the left. In B, poly(A)$^+$mRNA (1 μg) prepared from rat liver was treated at 50 ^{o}C for 60 min with a solution of 50% dimethyl sulfoxide, 5.6% glyoxal, in 10 mM sodium phosphate (pH 7.0). The glyoxylated mRNA was separated by electrophoresis in a 1.2% agarose gel in 10 mM sodium phosphate buffer and transferred to GeneScreen Plus nylon filters by capillary pressure. The glyoxal was removed from the immobilized mRNA by mild alkaline treatment. Hybridization of radioactive pL21-2 cDNA was done in the same manner as to genomic DNA. The size of the mRNA was estimated by comparison to the mobility in the same gels of 18S and 28S rRNA and of DNA restriction fragments.

The Number of Copies of the L21 Gene

The cDNA insert in pL21-2 was made radioactive and used to probe digests prepared with restriction endonuclease (BamHI, EcoRI, or HindIII) from rat liver nuclear DNA (2). The number of hybridization bands suggest that there are 16-23 copies of the L21 gene (Fig. 2A). Many other mammalian ribosomal protein genes have been found to be present in multiple copies (15). However, in no instance has it been shown that more than one of the genes is functional (9, 10, 16). The presumption is that for each ribosomal protein the genome contains only one gene that is expressed,

that the other copies are non-functional pseudogenes. However, it needs to be emphasized that this presumption derives from the analysis of only a few families.

The Size of mRNA Encoding Rat Ribosomal Protein L21

To determine the size of mRNA coding for L21, total poly(A)$^+$mRNA from rat liver was separated by electrophoresis and screened for hybridization bands using radioactive pL21-2 cDNA (2). One distinct band of about 680 bases was detected (Fig. 2B).

Comparison of the Sequence of Amino Acids in Rat L21 with Ribosomal Proteins From Other Species

The sequence of amino acids in rat ribosomal protein L21 was compared, using the computer program RELATE (4), to the sequence of amino acids in more than 400 other ribosomal proteins contained in a library that we have compiled. We did not find any that are significantly similar to rat L21. However, it is noteworthy that marginal scores (the numbers in parentheses are the RELATE values in S.D. units) were obtained for the comparisons with rat (4.90), mouse (3.31) and human (3.51) ribosomal proteins L32. The sequence of amino acids in L21 was also searched for internal duplications without finding any.

The determination of the sequence of amino acids in rat L21 is a contribution to a set of data which it is hoped will eventually encompass 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.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health Grant GM 21769. We are grateful to our colleagues Alan Lin and Joe Olvera for assistance with the protein sequencing and to Arlene Timosciek for the preparation of the manuscript. The sequence data will appear in the EMBL/Gen Bank/DDBJ Nucleotide Sequence Database under the accession number X15216.

REFERENCES

1. Chan, Y.L., Lin, A., McNally, J., and Wool, I.G. (1987) J. Biol. Chem. 262, 12879-12886.
2. Chan, Y.L., and Wool, I.G. (1988) J. Biol. Chem. 263, 2891-2896.
3. Beaucage, S.L., and Caruthers, M.H. (1981) Tetrahedron Lett. 22, 1859-1862.
4. Dayhoff, M.O. (1978) In Atlas of Protein Sequence of Structure (M.O. Dayhoff, Ed.), Vol. 5, Suppl. 3, p. 1-8, National Biomedical Research Foundation, Washington, D.C.

5. Tsurugi, K., Collatz, E., Wool, I.G., and Lin, A. (1976) J. Biol. Chem. 251, 7940-7946.
6. Kozak, M. (1986) Cell 44, 283-292.
7. Proudfoot, N.J., and Brownlee, G.G. (1976) Nature 263, 211-214.
8. Chan, Y.L., Lin, A., McNally, J., Peleg, D., Meyuhas, O., and Wool, I.G. (1987) J. Biol. Chem. 262, 1111-1115.
9. Wiedemann, L.M., and Perry, R.P. (1984) Mol. Cell. Biol. 4, 2518-2528.
10. Dudov, K.P., and Perry, R.P. (1984) Cell 37, 457-468.
11. Wagner, M., and Perry, R.P. (1985) Mol. Cell. Biol. 5, 3560-3576.
12. Loreni, F., Ruberti, I., Bozzoni, I., Pierandrei-Amaldi, P., and Amaldi, F. (1985) EMBO J. 4, 3483-3488.
13. Devi, K.R.G., Chan, Y.L., and Wool, I.G. (1989) Biochim. Biophys. Acta, in the press.
14. Flinta, C., Persson, B., Jornvall, H., and von Heijne, G. (1986) Eur. J. Biochem. 154, 193-196.
15. Monk, R.J., Meyuhas, O., and Perry, R.P. (1981) Cell 24, 301-306.
16. Klein, A., and Meyuhas, O. (1984) Nucleic Acids Res. 12, 3763-3776.