Vol. 185, No. 1, 1992 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
May 29, 1992 Pages 356-362

THE PRIMARY STRUCTURE OF RAT RIBOSOMAL PROTEIN L11

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

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

Received April 17, 1992

SUMMARY: The amino acid sequence of the rat 60S ribosomal subunit protein L11 was deduced from the sequence of nucleotides in a recombinant cDNA. Ribosomal protein L11 has 178 amino acids and a molecular weight of 20,239. Hybridization of the cDNA to digests of nuclear DNA suggests that there are 6-8 copies of the L11 gene. The mRNA for the protein is about 800 nucleotides in length. Rat L11 is homologous to ribosomal proteins from other eukaryotes and is related to the L5 family of proteins from eubacterial and archaebacterial ribosomes.

A coherent molecular account of the function of mammalian ribosomes in protein synthesis requires comprehension of the structure of the particles. A prerequisite to the solution of the structure is the sequence of nucleotides and of amino acids in the constituent nucleic acids and proteins. A commitment has been made to the acquisition of this data for mammalian (rat) ribosomes (1). We report here the structure of rat ribosomal protein L11 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 the nucleic acids were either described or cited before (2,3). Oligodeoxynucleotide probes for the cDNA encoding rat ribosomal protein L11 were synthesized predicated on sequences of amino acids present in a peptide from the protein. The NH_2 terminus of L11 is blocked but a peptide formed by acid hydrolysis at an aspartic acid-proline bond (between positions 120 and 121) was isolated and sequenced by Edman degradation in an automated gas phase sequencer. Probe I, based on the sequence PGFSIADKK, was a mixture of 288 oligodeoxynucleotides, each 26 bases long; probe II was based on the sequence KEEAMRXFQ and contained 192 oligodeoxynucleotides, each 26 bases in length. The oligodeoxynucleotides were synthesized on a solid support by the methoxyphosphoramidite method using an Applied Biosystems, 380B, DNA synthesizer (4). The probes were made radioactive by labeling the 5' end with [^{32}P]ATP using T4 polynucleotide kinase (2,3).

*To whom correspondence should be addressed at Department of Biochemistry and Molecular Biology, The University of Chicago, 920 E. 58th St., Chicago, IL 60637.

0006-291X/92 $4.00

RESULTS AND DISCUSSION

The Sequence of Nucleotides in Recombinant cDNAs Encoding Rat Ribosomal Protein L11

A random selection of 60,000 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 synthesized to be complementary to the sequence of nucleotides predicted to be present in the mRNA for the rat ribosomal protein L11. Two clones gave a positive hybridization signal with the probes. The DNA from the plasmids of the 2 transformants was isolated and digested with restriction endonucleases. These clones had inserts that were about 500 nucleotides in length. Preliminary analysis of the sequences of nucleotides in these clones indicated that they encoded rat ribosomal protein L11 but lacked the 3' end of the open reading frame and the untranslated region. A second rat cDNA library (kindly provided by M. Brownstein, NIH) was screened using a cDNA from one of the clones described above. Three clones hybridized to this probe. The cDNA inserts in these 3 clones were about 600 bases in length; one clone was selected, pcD-L11-14, and the sequence of nucleotides from both strands of the cDNA was obtained.

```
                                                                    -1
                                                          TCTTCCTGTTCTCCACC

+1                              30                              60                              90
ATGGCGCAAGATCAGGGTGAAAAGGAGAACCCCATGCGGGAACTTCGTATCCGCAAGCTCTGCCTCAATATCTGCGTGGGGGAGAGTGGA
M  A  Q  D  Q  G  E  K  E  N  P  M  R  E  L  R  I  R  K  L  C  L  N  I  C  V  G  E  S  G
                          10                              20                              30

                               120                             150                             180
GACAGACTGACCCGAGCAGCCAAGGTGTTGGAGCAGCTCACAGGCCAGACCCCTGTGTTTTCCAAAGCTAGATACACCGTCAGGTCCTTT
D  R  L  T  R  A  A  K  V  L  E  Q  L  T  G  Q  T  P  V  F  S  K  A  R  Y  T  V  R  S  F
                          40                              50                              60

                               210                             240                             270
GGCATCCGGCGAAATGAGAAAATCGCTGTTCACTGCACTGTCCGTGGAGCCAAGGCAGAGGAGATTCTGGAGAAAGGCCTGAAGGTGCGA
G  I  R  R  N  E  K  I  A  V  H  C  T  V  R  G  A  K  A  E  E  I  L  E  K  G  L  K  V  R
                          70                              80                              90

                               300                             330                             360
GAGTACGAATTACGGAAGAATAACTTCTCAGATACCGGCAACTTTGGCTTTGGGATTCAGGAACACATTGACCTAGGCATCAAATATGAC
E  Y  E  L  R  K  N  N  F  S  D  T  G  N  F  G  F  G  I  Q  E  H  I  D  L  G  I  K  Y  D
                         100                             110                             120

                               390                             420                             450
CCAAGCATTGGCATCTATGGCCTGGATTTCTATGTGGTGCTGGGTAGGCCAGGTTTCAGCATCGCAGACAAGAAGCGCAGAACAGGCTGC
P  S  I  G  I  Y  G  L  D  F  Y  V  V  L  G  R  P  G  F  S  I  A  D  K  K  R  R  T  G  C
                         130                             140                             150

                               480                             510                             540
ATTGGGGCCAAACACAGAATCAGCAAAGAGGAGGCCATGCGCTGGTTCCAGCAGAAGTATGATGGCATCATCCTTCCTGGCAAATAAACT
I  G  A  K  H  R  I  S  K  E  E  A  M  R  W  F  Q  Q  K  Y  D  G  I  I  L  P  G  K  *
                         160                             170

                               570
TTATCCAAAGGCCAATAAAGTTTTCTC(A)n
```

Fig. 1. *The sequence of nucleotides in the cDNA insert in plasmid pcD-L11-14 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 amino acids in protein L11 are designated below the residue.

The cDNA insert in pcD-L11-14 is 584 nucleotides long and has a 5' noncoding sequence of 17 bases, a single open reading frame of 537, a 3' noncoding sequence of 30 and 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 AUG codon at a position that we designate +1 and ends with a termination codon (TAA) at position 535; it encodes 178 amino acids (Fig. 1). The initiation codon occurs in the context ACC*AUG*G which conforms to the consensus (5). The hexamer AATAAA that directs post-transcriptional cleavage-polyadenylation of the 3' end of the precursor of the mRNA (6) is at position 554-559, 8 nucleotides upstream of the start of the poly(A) stretch.

The Primary Structure of Rat Ribosomal Protein L11

The rat ribosomal protein encoded in the open reading frame in pcD-L11-14 was identified as L11 from the coincidence of the sequence of amino acids deduced from the sequence of nucleotides with that of 48 of 50 residues in a peptide from Lll (positions 121-170) determined by Edman degradation in an automated gas phase sequencer; the exceptions were two amino acids in the peptide that could not be identified.

The molecular weight of rat ribosomal protein L11, calculated from the sequence of amino acids deduced from pcD-L11-14, is 20,239; close to the 21,300 estimated before (7) from SDS-PAGE of the purified protein. We do not know whether the NH_2-terminal methionine encoded in the L11 mRNA is removed after translation. However, the residue next to the initial methionyl in L11 is alanyl which has been reported (8) to favor NH_2-terminal processing.

Protein L11 has an excess of basic residues (17 arginyl, 16 lysyl, and 3 histidyl) over acidic ones (8 aspartyl and 14 glutamyl) (Table I). The basic residues tend to be clustered; for example, 5 of the 12 residues at positions 8-19; 4 of the 4 between positions 144 and 147; and 5 of the 11 at positions 154-164. L11 has a number of hydrophilic regions; for example, 9 of the 16 residues at positions 4-19, 19 of the 45 at positions 52-96, and 12 of the 22 at positions 143-164 are charged.

The amino acid composition of L11 (Table I) derived from the nucleotide sequence of pcD-L11-14 closely approximates that obtained from an hydrolysate of a puried protein originally designated L12. In the report (7) of the amino acid composition of ribosomal proteins L11 and L12 the values for the two were transposed as was shown earlier (9). The protein from which the amino acid sequence was obtained (referred to above) was shown by two-dimensional PAGE to be L11.

The Number of Copies of the L11 Gene

The cDNA insert in pcD-L11-14 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

TABLE I. Amino acid composition of rat ribosomal protein L11

Amino Acid	A	B
Alanine	13	10
Arginine	18	17
Aspartic acid and asparagine	16	8 + 6
Cysteine	n.d.	4
Glutamic acid and glutamine	25	14 + 7
Glycine	21	19
Histidine	3	3
Isoleucine	16	15
Leucine	15	13
Lysine	16	16
Methionine	2	3
Phenylalanine	9	8
Proline	6	5
Serine	7	7
Threonine	7	7
Tryptophan	n.d.	1
Tyrosine	6	6
Valine	12	9
Residues		178

The amino acid composition (in numbers of residues) was determined either (A) from an hydrolysate of purified L11 but designated L12 (7) or inferred (B) from the sequence of nucleotides in pcD-L11-14.

hybridization bands suggest that there are 6 to 8 copies of the L11 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 L11

To determine the size of the mRNA coding for L11 poly(A)$^+$mRNA from rat liver was separated by electrophoresis and screened for hybridization bands using radioactive pcD-L11-14 cDNA. One distinct band of about 800 nucleotides was detected (data not shown).

```
                1                                                   50
RL11            ......... ....MAQDQG EKENPMRELR IRKLCLNICV GESGDR...L
ScL16           ......... .......MST KAQNPMRDLK IEKLVLNISV GESGDR...L
TtL21           ......... .......MTD KKENKMREVK IAKLVINCCV GESGDK...L
DdV18           ......... ......... .....MRGLR IEKLVLNICV GESGDR...L
MvL5            ......... ....MSFQEV WEKEPMKKPR IQKVTVNFGV GEAGDR...L
HmL5e           ......... ...SSESESG GDFHEMREPR IEKVVVHMGI GHGG...RDL
BstL5           MNRLKEKYVK EVVPALMSKF NYKSIMQVPK IEKIVINMGV GDAVQNPKAL
EcL5            .AKLHDYYKD EVVKKLMTEF NYNSVMQVPR VEKITLNMGV GEAIADKKLL
Consensus                     ......  .....Mr..r i.Kl.ln..v Ge.gdr   L

                51                                                  100
RL11            TRAAKVLEQL TGQTPVFSKA RYTVRSFGIR RNEKIAVHCT VRGAKAEEIL
ScL16           TRASKVLEQL SGQTPVQSKA RYTVRTFGIR RNEKIAVHVT VRGPKAEEIL
TtL21           TKAAKVLKDL SGQEPVFSRA RYTIRSFGIK RNEKMAVHVT IRGDKARDIL
DdV18           VRAAKVLEQL TGQTPVYSKA RYTVRSFNIR RNEQIAAHVT VRGEKAAEIL
MvL5            TIGAKVIETL TGQAPVRTLA KQTNPAFGIR KKLPIGLKVT LRGKNAEEFL
HmL5e           ANAEDILGEI TGQMPVRTKA KRTVGEFDIR EGDPIGAKVT LRDEMAEEFL
BstL5           DSAVEELTLI AGQRPVVTRA KKSIAGFRLR QGMPIGAKVT LRGERMYEFL
EcL5            DNAAADLAAI SGQKPLITKA RKSVAGFKIR QGYPIGCKVT LRGERMWEFF
Consensus       ..a..vl..l tGQ.Pv.tkA r.tv..F.Ir ....i..kvT lRg.ka.e.l

                101                                                 150
RL11            EK.....GLK VREY.ELRKN NFSDTGNFGF GIQEHIDL.G IKYDPSIGIY
ScL16           ER.....GLK VKEY.QLRDR NFASTGNFGF GIDEHIDL.G IKYDPSIGIF
TtL21           TR.....GLK VKE.MELRKK NFSNTGNFGF GIQEHIDL.G MKYDPSTGIF
DdV18           E.....IGLN VR*..... ......... ......... .........
MvL5            ENAFVAFKVS GK...VLYAS SFDKVGNFSF GVPEHIDFPG QKYDPTVGIY
HmL5e           QT.....ALP L...AELATS QFDDTGNFSF GVEEHTEFPS QEYDPSIGIY
BstL5           DKLISVSLPR VRDFRGVSKK AFDGRGNYTL GIKEQLIFPE IDYDKVNKVR
EcL5            ERLITIAVPR IRDFRGLSAK SFDGRGNYSM GVREQIIFPE IDYDKVDRVR
Consensus       er     .l. vre  .l... .F...GNf.f Gi.Ehid. .  ..YDps.gi.

                151                                                 200
RL11            GLDFYVVLGR PGFSIADKKR RTGCIGAKHR ISKEEAMRWF QQKYDGIILP
ScL16           GMDFYVVMNR PGARATRRKR CKGTVGNSHK TTKEDTVSWF KQKYDADVLD
TtL21           GMDFYVVLER PGTRVARRRR ATSRVGNNQM ISKEECINWF KTEFEGNVY.
DdV18           ......... ......... ......... ......... .........
MvL5            GMDICVTFEK PGYRVKSRKL KRSHIPAKHL VKKEEAIEYI EKKFGAEVVM
HmL5e           GLDVTVRLVR PGYRVAKRDK ASRSIPTKHR LNPADAVAFI ESTYDVEVSE
BstL5           GMDIVIV... ......... ....TTANTD EEARELLALL GMPFQK*...
EcL5            GLDI..... ......... .TITTTAK.. .SDEEGRALL ....AAFDFP
Consensus       G.D..v...r PG.rv..rkr ....i..... ...ee.v... .......v..

                201
RL11            GK..
ScL16           K*..
TtL21           ....
DdV18           ....
MvL5            E*..
HmL5e           *...
BstL5           ....
EcL5            FRK*
Consensus
```

Fig. 2. *An alignment of proteins related to rat L11.* The reiterative multiple sequence alignments were made with the programs GAP and PRETTY from the GCG Sequence Analysis Software Package (24). In the *consensus* invariant amino acids are in capital letters; conservative changes (R, K, and H; I, V, and L; D and E; and T and S) and residues present in 4 of 5, 5 of 6, 5 of 7, or 6 of 8 sequences are in small letters; a consensus was derived only for positions occupied by at least 5 residues. The abbreviations for the ribosomal proteins are: RL11, rat L11, ScL16, *Saccharomyces cerevisiae* L16; TtL21, *Tetrahymena thermophila* L21; DdV18, *Dictyostelium discoideum* vegetative specific gene V18 product; MvL5, *Methanococcus vannielii* L5; HmL5e, *Halobacterium marismortui* L5; BstL5, *Bacillus stearothermophilus* L5; EcL5, *Escherichia coli* L5.

Comparison of the Sequence of Amino Acids in Rat L11 with Ribosomal Proteins from Other Species

The sequence of amino acids in rat L11 was compared, using the computer programs RELATE and ALIGN (10), to those in more than 1,000 other ribosomal proteins contained in a library that we have compiled. Rat L11 is homologous to *Saccharomyces cerevisiae* L16 (11); the RELATE score is 45.5 S.D. units. In an alignment of the amino acid sequences there are 120 identities out of 174 possible matches (the ALIGN score is 76.8 S.D. units). Rat L11 is also related to *Tetrahymena thermophila* L21 (12) - the RELATE score is 45.8 and the ALIGN score is 79.1 with 110 identities in 172 possible matches; and to *Dictyostelium discoideum* vegetative specific gene V18 (13) - the RELATE score is 28.6 and the ALIGN score is 39.2 with 66 identities in 79 possible matches. In addition, rat L11 is related to various eubacterial and archaebacterial members of the ribosomal protein L5 family: to *Methanococcus vannielii* L5 (14) - the RELATE and ALIGN scores are 18.8 and 44.3; to *Halobacterium marismortui* L5 (15) - 22.0 and 47.4; to *Bacillus stearothermophilus* L5 (16) - 4.7 and 9.8; and to *Escherichia coli* L5 (17) - 4.2 and 8.3. A simultaneous alignment of the proteins related to rat L11 is in Fig. 2. Some of these relationships have been noted before (12, 18). Thus there is conservation of this ribosomal protein across the three kingdoms (cf. Fig. 2) suggesting it was a component of an early ribonucleoprotein ribosome.

In eubacteria three proteins bind to 5S rRNA; in *E. coli* they are L5, L18, and L25 (19, 20); in eukaryotes a single protein, L5, binds to the nucleic acid (21-23). The eubacterial ribosomal proteins L5 are related to eukaryotic L11 which latter is not known to associate with 5S rRNA; moreover, the amino acid sequence of eubacterial L5 is not related to rat L5 (2). We cannot account for these anomalies.

The sequence of amino acids in L11 was searched for internal duplications; none were found.

The determination of the sequence of amino acids in rat L11 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 Herta Trampitsch who began this study, to Anton Glück and Katsuyuki Suzuki for

computer analyses, 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 X62146.

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. Tsurugi, K., Collatz, E., Wool, I.G., and Lin, A. (1976) J. Biol. Chem. 251, 7940-7946.
8. Flinta, C., Persson, B., Jörnvall, H., and von Heijne, G. (1986) Eur. J. Biochem. 154, 193-196.
9. Suzuki, K., Olvera, J., and Wool, I.G. (1990) Biochem. Biophys. Res. Comm. 172, 35-41.
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. Teem, J., Abovich, N., Kaufer, N., Schwindinger, W., Warner, J., Levy, A., Woolford, J., Leer, R., van Raamsdonk-Duin, M., Mager, W., Planta, R., Schultz, L., Freisen, J., Fried, H., and Rosbash, M. (1984) Nucleic Acids Res. 12, 8295-8312.
12. Rosendahl, G., Andreasen, P.H., and Kristiansen, K. (1991) Gene 98, 161-167.
13. Singleton, C.K., Manning, S.S., and Ken, R. (1989) Nucleic Acids Res. 17, 9679-9692.
14. Auer, J., Spicker, G., and Boeck, A. (1989) J. Mol. Biol. 209, 21-36.
15. Hatakeyama, T., and Hatakeyama, T. (1990) Biochem. Biophys. Acta 1039, 343-347.
16. Kimura, R., and Kimura, J. (1987) FEBS Lett. 210, 85-90.
17. Ceretti, D.P., Dean, D., Davis, G.R., Bedwell, D.M., and Nomura, M. (1983) Nucleic Acids Res. 11, 2599-2616.
18. Kenmochi, N., Maeda, N., and Tanaka, T. (1991) Biochim. Biophys. Acta 1088, 445-447.
19. Yu, R.S.T., and Wittmann, H.G. (1973) Biochim. Biophys. Acta 324, 375-385.
20. Huber, P.W., and Wool, I.G. (1984) Proc. Natl. Acad. Sci. U.S.A. 81, 322-326.
21. Blobel, G. (1971) Proc. Natl. Acad. Sci. U.S.A. 68, 1881-1885.
22. Terao, K., Takahashi, Y., and Ogata, K. (1975) Biochim. Biophys. Acta 402, 230-237.
23. Huber, P.W., and Wool, I.G. (1986) J. Biol. Chem. 261, 3002-3005.
24. Devereux, J., Haeberli, P., and Smithies, O. (1984) Nucleic Acids Res. 12, 387-395.