Volume 14 Number 19 1986 Nucleic Acids Research

Sequences coding for the ribosomal protein L14 in *Xenopus laevis* and *Xenopus tropicalis*; homologies in the 5' untranslated region are shared with other r-protein mRNAs

Elena Beccari, Paola Mazzetti, AnnaMaria Mileo, Irene Bozzoni, Paola Pierandrei-Amaldi[1] and Francesco Amaldi[2]

Centro Acidi Nucleici CNR, Dipartimento di Genetica e Biologia Molecolare, Università di Roma I, P.le Aldo Moro 5, 00185 Roma, [1]Istituto di Biologia Cellulare CNR, Via Romagnosi 18A, 00196 Roma and [2]Dipartimento di Biologia, II Università di Roma 'Tor Vergata', Via Orazio Raimondo, 00173 Roma, Italia

Received 5 June 1986; Revised and Accepted 10 September 1986

ABSTRACT

In the haploid genome of Xenopus laevis there are two genes coding for the r-protein L14. It is not known if they are located on the same chromosome. cDNA clones deriving from the transcripts of the two genes have been isolated from an oocyte messenger cDNA bank showing that they are both expressed. We have studied the structure of one of the L14 genes by Electron Microscopy, restriction mapping and sequencing. An allelic form of the L14 gene was also isolated. It contains a large deletion covering the 5'end region up to the middle of the third intron. The 5' end of the X.laevis L14 gene was compared to that of the corresponding gene in the closely related species X.tropicalis and found to be highly conserved. The L14 gene has multiple initiation sites, but the large majority of the transcripts start in the middle of a pyrimidine tract not preceded by a canonical TATA box as in other eukaryotic housekeeping genes. The X.laevis L1 and L14 genes have a common decanucleotide in the first exon in the same position with regard to the initiator ATG which just precedes the first intron. The decanucleotide shows homology with the X.laevis 18S rRNA.

INTRODUCTION

The ribosomal protein genes are a group of coordinately expressed sequences active in all cell types, belonging to the so called "housekeeping" category.

Regulation of this set of genes has been extensively investigated in E.coli where specific r-proteins are the effectors of an "autogenous" regulation operating at the level of translation (1).

In eukaryotes the r-protein genes are not clustered in the genome and more than one gene codes for each r-protein. A great deal of knowledge about the structure is accumulating as a result of cloning and sequencing of r-protein genes in yeast (2-3), Drosophila (4-6) , Xenopus (7-8), mouse (9-10).

In this study we have characterized a gene coding for L14 in X.laevis and in X.tropicalis. Sequence comparison between the L14 and the L1 r-protein genes in X.laevis has shown some common structural features. Homologous sequences are located in the 5' flanking region in corresponding position. Both genes initiate transcription in a pyrimidine tract not preceded by a canonical TATA box. The decanucleotide GAGAAGCAGC is located in the first exon of both genes, in the same position with respect to the initiator ATG which immediately precedes the first intron. These homologies, found also in another r-protein sequence, would support the idea that specific sequences, different from the described "consensus" of the polII transcribed genes, are important for the expression of this set of genes.

METHODS

Construction of the X.laevis and X.tropicalis libraries

The DNA was purified by standard methods from the nucleated erythrocytes of a single individual in both species. EcoRI partial digests of about 20 kb were ligated to EcoRI digested Charon 4 arms and packaged (11).

Isolation and characterization of X.laevis and X.tropicalis L14 clones

The libraries were plated on the C3000 bacterial strain and screened by hybridization with the nick-translated L14 cDNA insert of the recombinant pXom92 plasmid. Phage DNA preparation,restriction digestions, Southern and Northern blotting, nick-translation and subcloning were all carried out according to standard procedures (12).

Electron Microscopy

Heteroduplexes were formed by denaturing the mixed phage DNAs with 0.1N NaOH,20mM EDTA for 10 min at room temperature, than neutralysing with 0.2 M Tris-HCl pH 7.2. The reassociation was for 2 hours at r.t. after adding an equal volume of formamide.

R-loops with oocyte polyA+ RNA were formed in 70% formamide, R-loop buffer (11) by incubating at 48°C for 3-4 hours. Spreadings and preparation of the grids were as described (11).

Oocyte injection

pBR322 subclones of the L14 genes were injected as described (13).

DNA sequence analysis

DNA sequencing was carried out according to Maxam and Gilbert (14) with the addition of a T specific reaction (15). Fragments were end labelled with T4 polynucleotide kinase and strand separated. The chemical reaction products were electrophoresed on urea-polyacrylamide 20% and 6% gels

Poly(A)+ RNA preparation

Total poly(A)+ RNA was prepared from stage 2-4 Xenopus laevis oocytes as described (16).

Primer extension

The 35 bp PstI-BamHI restriction fragment from the L14 cDNA clone was end labelled with T4 polynucleotide kinase and strand separated. This primer was annealed to 5 µg X.laevis oocyte poly(A)+ RNA in 80% formamide, 0.4 M NaCl and 40 mM MOPS pH 6.5 at 46°C for 5 hours. The mixture was ethanol precipitated and resuspended in 20 µl 50 mM Tris-HCl pH 8.3, 5 mM MgCl2, 40 mM KCl, 2 mM DTT, 1mM dNTP, 20 RNAsine units and 15 AMV reverse transcriptase units and incubated at 42°C for two hours. The extended DNA was phenol extracted, precipitated and run in a 6% sequencing gel.

S1 mapping

The 178 bp HinfI-AvaI fragment in the 5' end region of the genomic L14 clone was end labelled and strand separated. The appropriate strand was incubated with 6 µg poly(A)+ RNA in 80% formamide, 0.4 M NaCl and 40 mM MOPS pH 6.5 at 37°C o.n. The mixture was diluted 15-fold with S1 buffer and incubated at 37°C for 30 min with 100 S1 units. After phenol extractions and precipitation the nucleic acids were run on a 10% urea-acrylamide sequencing gel in order to see DNA fragments as short as 2 bases. The conditions of hybridization were not stringent in order to maintain a DNA-RNA hybrid 12-13 bp long.

The S1 mapping was also carried out using the 168 bp EcoRI-HindIII 5' end fragment of the full length cDNA clone. The procedure was the same, but the conditions of hybridization were more stringent, at 49 °C.

RESULTS AND DISCUSSION

Isolation of L14 gene copies from the genomic library

Two phages, λXlrp10 and λXlrp84, both carrying sequences for the L14 ribosomal protein gene, were isolated from a genomic library made with the DNA of a single individual. The EcoRI digestion pattern of the two recombinant phages was compared and they were found to share three bands of flanking regions, but differ in the fragments complementary to the cDNA clone. One phage carried L14 sequences in two fragments 4.4 and 6.4 kb long, the other in only one 7 kb band (fig 1a).

When the heteroduplexes between the two phages were examined at the E.M.

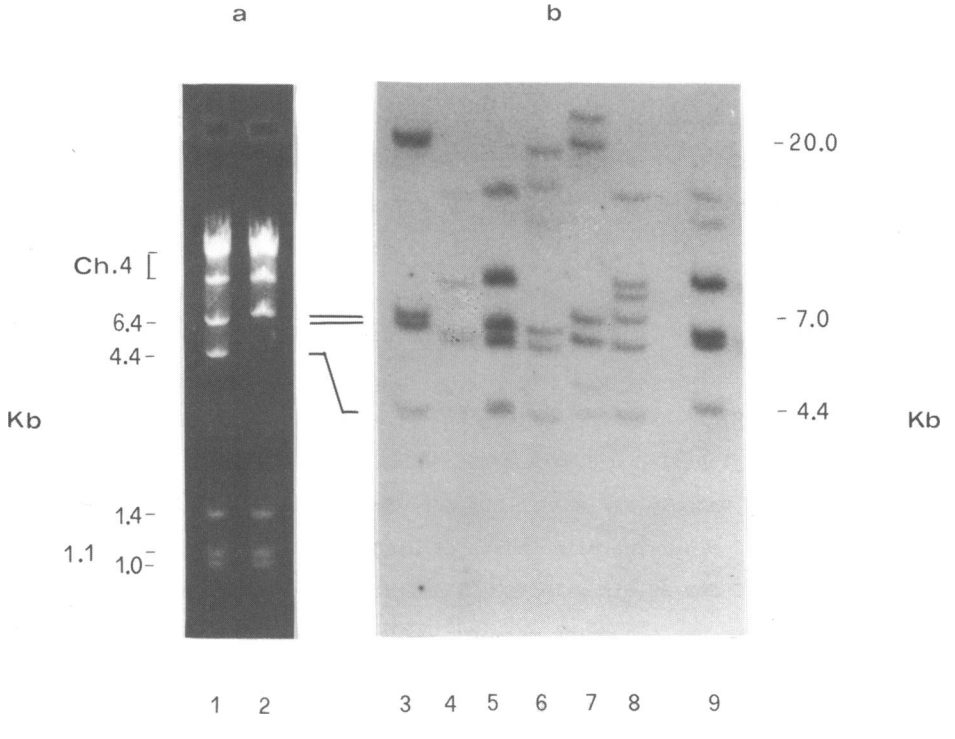

Fig.1- a) Agarose gel electrophoresis of EcoRI digests of λXlrp10 (1) and λXlrp84 (2); CH4 indicates Charon4 arms. b) Southern blot of EcoRI digested X.laevis DNA from single individuals (3-9), hybridized to the L14 cDNA. In lane 3 the DNA from the frog used to construct the library. The bands cloned in the recombinant phages are indicated. The DNA bands carrying L14 sequences appear polymorphic in the population. Band length is indicated in Kb.

(not shown), the clones appeared to match perfectly along the entire length, except for the looping out of a 3 kb single strand, belonging to λXlrp10. In order to find out if the deletion or the insertion of a region could be due to some cloning artifact, we made a Southern blot of the EcoRI digested DNA from different individuals of Xenopus laevis, including the one used to construct the library, analyzing the molecular weight of the bands hybridizing to the cDNA probe (fig 1b). The fragments 4.4, 6.4 and 7.0 kb are indeed present in the genome of the donor female. The genes appear polymorphic in the population,as shown by the different pattern of L14 sequences displayed; the number of bands suggests 4 genes for L14 in the diploid state.

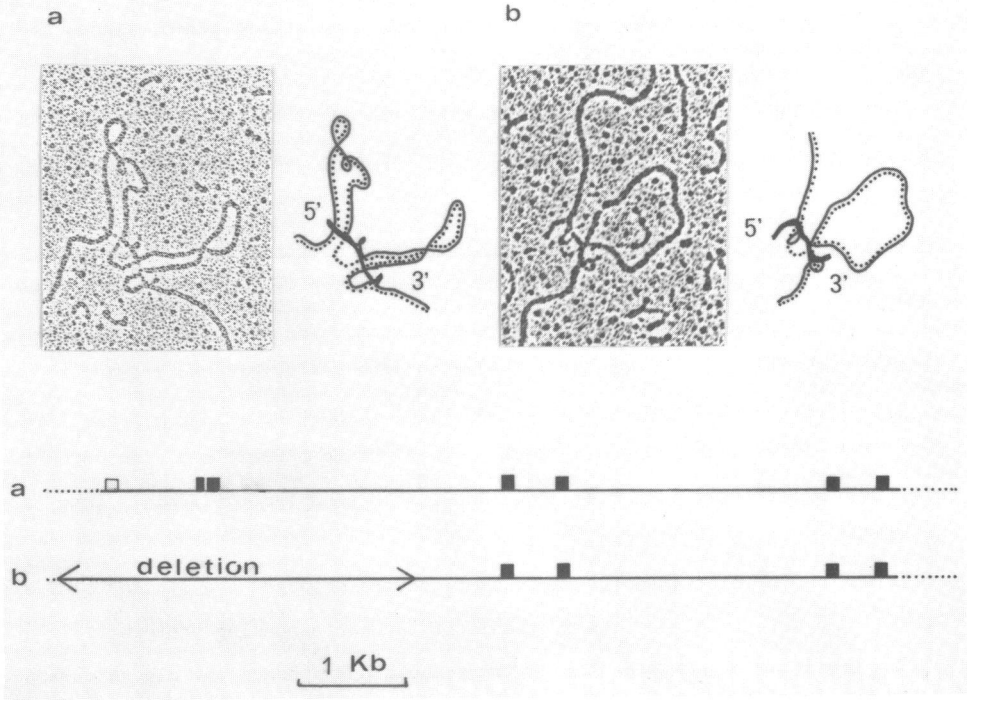

Fig.2- a) R-loop of λXlrp10 DNA hybridized to X.laevis oocyte poly(A)+ RNA. b) R-loop of λXlrp84 with the same RNA. One of the DNA strands is dotted in the drawings. RNA is the heavy line. The exon-intron pattern of a and b is shown on the bottom. The 5'exon not annealed to L14 mRNA in a) is drawn as an empty box. The DNA strands are reassociated in the intronic regions.

Electron Microscopic analysis of R-loops formed by the recombinant phages

As a first approach to the structure of the L14 gene we performed R-loop mapping to visualize the exon-intron arrangement in the DNA carried by the phages. Pictures of λXlrp10 hybridized to L14 mRNA showed 5 exons, 100-120 bp long, separated by 4 introns 2800, 300, 2500 and 400 bp long (fig.2a). At the 5' end of the messenger, a free tail was often visible, indicating that it failed to hybridize to the first exon of the L14 gene and therefore suggesting the presence of an additional 5' intron, which was in fact detected later by sequencing.

In λXlrp84 the pattern of introns differed from the previous clone due to the absence of the first 2800 bp intron, while the others were conserved (fig.2b).

The two clones therefore appear to carry two allelic L14 genes, since the regions flanking the genes are homologous. One is missing about 3 kb of genomic sequence. We did not sequence this gene, but by restriction mapping we could deduce that the deletion involves the first exon and the sequences up to the middle of the third intron.

Expression of the L14 gene after injection into X.laevis oocytes

A good test for the functionality of a gene is its expression in a system as similar as possible to the natural environment. The injection of X.laevis cloned genes into the nucleus of X.laevis oocytes meets this requirement.

The L14 gene sequences present in the EcoRI fragments 4.4 and 6.4 kb long from λXlrp10 were introduced as EcoRI partials into pBr322 in a recombinant plasmid called pl14 and the unique EcoRI 7 kb band from λXlrp84 into a pBr322 recombinant called p84.

Injection of pl14 into the nucleus of oocytes produced the synthesis of a mature mRNA able to direct the synthesis of the corresponding protein (13).

The injection of p84, on the contrary, did not produce any discrete mRNA species, showing that the gene was not functional (not shown).

Mapping of the L14 capsite

We sequenced by the Maxam and Gilbert procedure (14) the pl14 region containing the 5' end of the gene. The cDNA used to screen the library contained only 80% the mRNA sequence starting from the poly(A) tail (7). In order to

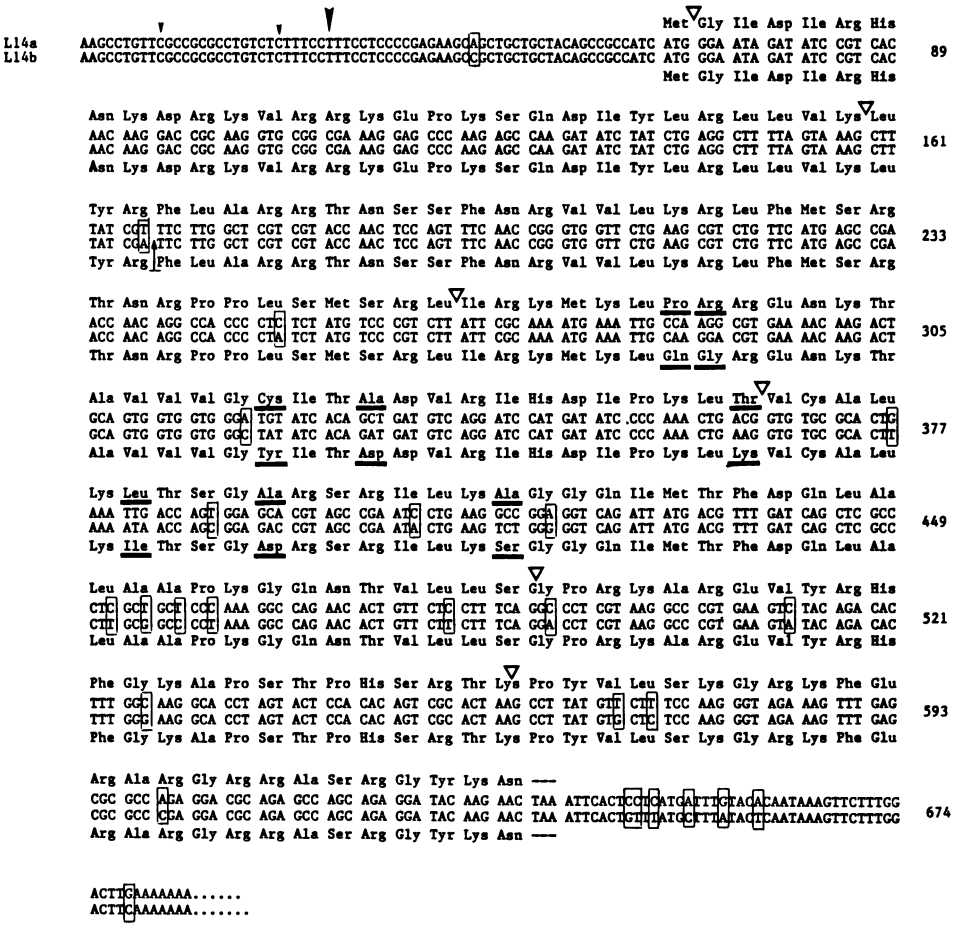

Fig.3- Sequences and translated proteins of cDNAs originated from the L14a and L14b genes. The L14b incomplete cDNA starts from position 167 (indicated by an arrow): the preceding region is obtained from the genomic clone coding for L14b. Major and minor capsites are indicated by arrow tips. Triangles indicate the position of introns. The 8 aminoacid differences are underlined. Neutral nucleotide substitutions are boxed.

define the exon borders in the 5' sequence we needed a full-length cDNA that we screened out of a cDNA library of poly(A)+ X.laevis oocyte RNA (kindly provided by D.Melton). Fig.3 shows the sequence of the cDNA isolated.

To make sure that the cDNA clone contained the L14 capsite and to locate it on the genomic clone, we made a primer extension analysis using a 35 bp

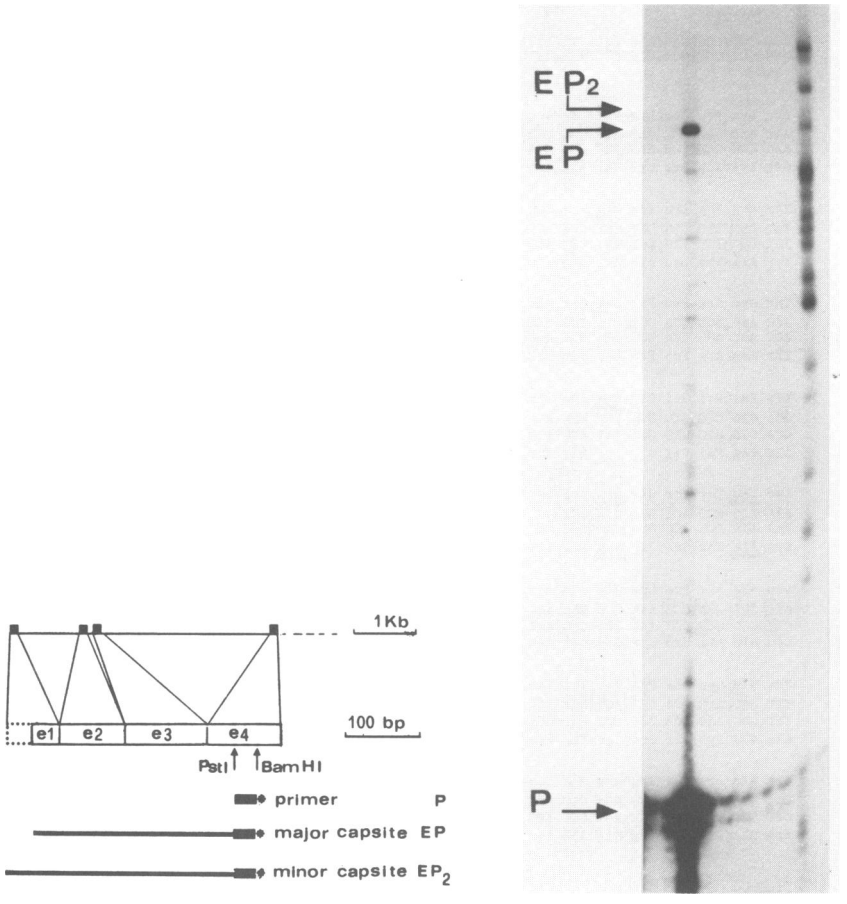

Fig.4- Primer extension. A PstI-BamHI fragment of the cDNA clone was 5' end-labelled, hybridized to oocyte poly(A)+ RNA and extended with AMV reverse transcriptase. P=primer, EP=extended primer. EP2 indicates a faint band 30 bp longer than EP. M.W.markers: HpaII digested pBr322.

PstI-BamHI fragment of the cDNA clone (fig.4). This fragment was terminally labeled and hybridized to total oocyte poly(A)+ RNA. The primer was extended by AMV reverse transcriptase up to 320 nucleotides, but a longer faint band was also visible at the position corresponding to 350 nucleotides (fig.4). The strong band indicated a stop of the reverse transcriptase about 30 bp downstream of the 5' end of the cDNA clone. The faint band corresponded to the length of the cDNA. The major band indicated a start of transcription at

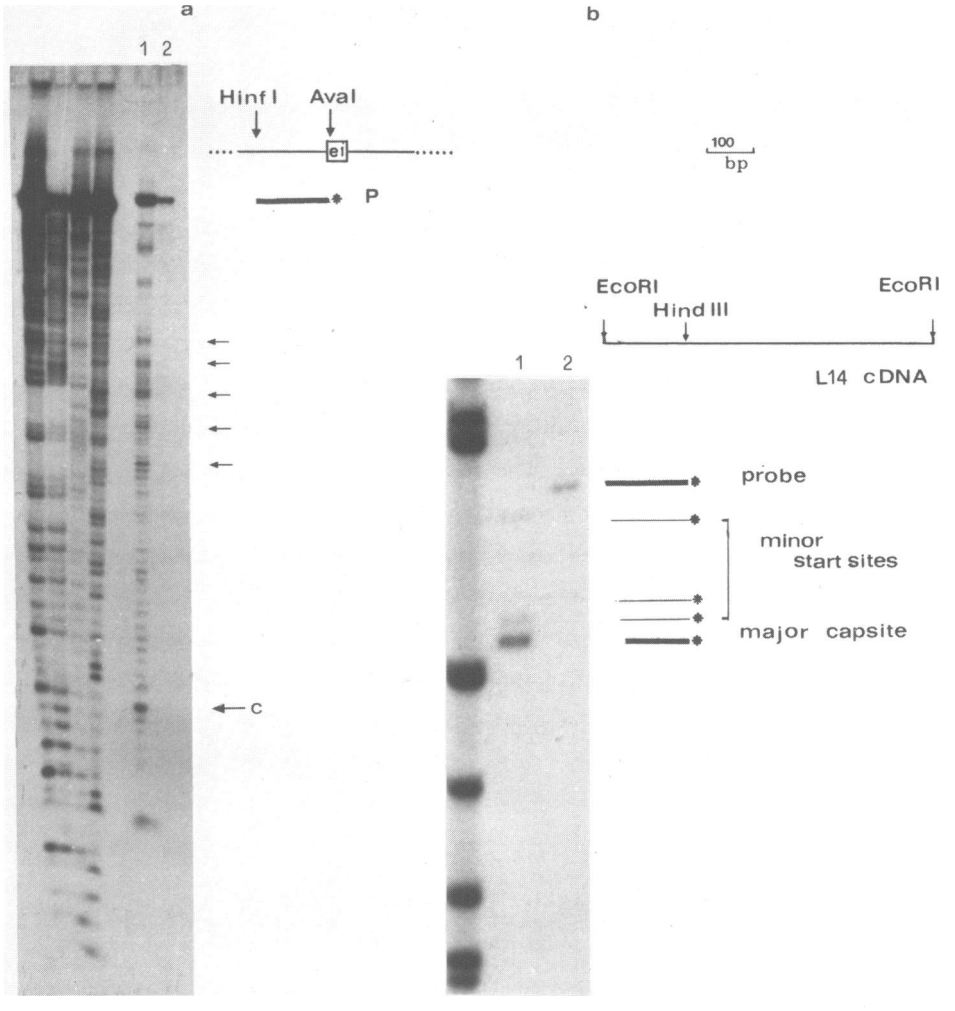

Fig.5- S1 mapping to determine the sites for transcription initiation. a) P is a fragment of the genomic clone labelled at the AvaI site of the first exon. Arrows point to the fragments protected from S1 digestion after hybridization to oocyte poly(A)+ RNA. c=major capsite. The sequence ladder is from an unrelated fragment. In lane 2, P was hybridized to a procaryotic total RNA. b) The probe is a fragment of the L14a cDNA ,which extends 5' to the major capsite. 1) probe hybridized to poly(A)+ X.laevis RNA 2) to procaryotic RNA. Samples are treated as described in Methods.

the level of the pyrimidine stretch which also surrounds the capsite of some other r-protein genes so far sequenced, like the L1 in Xenopus laevis and the L30 and L32 in mouse (9,10). In all these cases the capsite had been deter-

mined by primer extension. The presence of a longer band and the finding of a cloned cDNA containing additional nucleotides indicated that other capsites could be present or, alternatively, that the pyrimidine tract could be a strong stop for the reverse transcriptase, while the real capsite was located upstream.

To clarify this point, we performed a conventional S1 mapping on a fragment on the genomic clone, taking advantage of a AvaI site located 13 bp downstream of the presumptive major capsite. In this case no artifacts due to the reverse transcriptase could be imagined to operate. Fig.5a shows that a protection band in fact appeared at the level of the expected major capsite, together with additional bands. The same analysis was performed on the cDNA clone (fig.5b). Again a major protection band was observed ,corresponding to the startsite in the pyrimidine tract, plus additional faint bands indicating the presence of RNA molecules transcribed 5' to that point.

In conclusion most of the L14 mRNA molecules are transcribed from a capsite located in a pyrimidine stretch, like in other eukaryotic r-protein genes. A minority of the transcripts start upstream of that point. It must be noted that the major capsite is not preceded by a canonical TATA box.

The L14 protein

The aminoacid sequence of the L14 protein, deduced from the full-length cDNA sequence, is shown in fig.3. The protein is made by 188 aminoacids. Basic aminoacids, like Lysine, Arginine and Histidine constitute about 28% of the total. This was indeed expected from the physiological and electrophoretic properties of the L14 protein.

We compared the L14 with the aminoterminal sequences determined by Wittmann-Liebold et al (17) on the rat liver ribosomal proteins. The homology with the rat L18 r-protein is very high (fig.6). Over the first 19 aminoacids 13 are identical. The similarity could possibly be higher since two of the differing rat aminoacids gave faint spots and were not confirmed by a second method of analysis. For this reason they are given in brackets by the authors.

Comparison between the transcripts of the two L14 genes

In fig.3 the nucleotide and aminoacid sequences of the two L14 cDNAs are compared. L14b is the incomplete cDNA and the 5' region is taken from the

Rat	L18	Gly - Val - Asp - Ile -(Arg)- His - Asn - Lys - Glu -(Ser)-
X.laevis	L14	Gly - Ile - Asp - Ile - Arg - His - Asn - Lys - Asp - Arg -
		- Lys - Val -(Ser)-(Arg)- Lys - Leu - Pro - Lys - Leu -.....
		- Lys - Val - Arg - Arg - Lys - Glu - Pro - Lys - Ser -....

Fig.6- Comparison between the aminoacid sequences of rat L18 and of X.laevis L14. The aminoacid differences in X.laevis are underlined. The rat aminoacids in brackets formed weak spots in thin-layer chromatography and could not be confirmed by HPLC (18).

genomic sequence. 38 nucleotides differ and 8 aminoacids. The nucleotide changes resulting in aminoacid differences are clustered in two central exons of the gene, suggesting a selective pressure against changes in the aminoacid composition of the amino and carboxy-proximal ends of the protein.

Evolutionary conservation of the 5' end of the L14 gene

It was of interest to examine if the stuctural features of the 5' end of the X.laevis L14 gene had been conserved in the closely related species X.tropicalis. Therefore we screened a X.tropicalis library, using as a probe the X.laevis cDNA. A recombinant phage was isolated. We found by hybridization that the 5' end of the X.tropicalis gene was carried by a 1200 bp PstI-EcoRI fragment that we sequenced. The comparison of the sequences of the 5' end of the L14 genes in the two species is shown in fig.7. The conservation

```
-165  GGACTCGGTAAACAAAAACAACTTCGCATTGTGCAACTGATTTATCCTGCCCTTGCCACAACATTTCAGTCCTCCGCCCACCACCGAGTGT
 -74  TCAGGCTAGAGCGGCCACCACAAACTTCCGGTTATCAGGTGTTCCCATAAGCCTGTTCGCCGCGCCCCTGTCTCTTTCCTTTCCTCCCCGAG
 +14  AAGCCGCTGCTGCTACAGCCGCCATCATGGTGAGAAACCAGACTACTTTATAATCCGCTGCAATCCCGCTTTCTTCTCCAGCGCGTACTCA
+105  CAATCAGGGGATGGCTGGCTCCTAGGACTGTGGGCTACCAGTACCGGGCCAGTGTTGGGAGTGAATGTGTGCGAGAAGTGTTGTTAAATAA
+196  AGGATGTGTAACTCTTCCTCCGTTCAGCTTCGGGCCTGGGACTGGGCGCACACACTGTTCTATGGCTGCTGGTTAAGTAAGCGGACTTGGC
+287  ATGTTTCCAGTGGCCGAGACATGGGATTAAAGAACCTTTTCAGAACCAGAGCGGGAAGCCAGACATTAAAATCACTGTCACTCCCCTCCTG
+378  AACCTGAACTGGTTAAACACATGATACCACTGGCGACCCTAATAGGTGAGTTCGTCCATCGCTACCGCTTAACCAGATCGCCTTGCCTAGA
+469  GCAGCCCTTGCGAT.....
```

Fig.7- 5' end region of the X.laevis L14 gene compared to the corresponding X.tropicalis sequence. Dots indicate homology in the X.tropicalis gene. The first exon, containing the non-translated part of the messenger and the first ATG, is boxed.

is very high. The genes are 90% homologous in the 100 bp preceding the capsite and identical, except one base, in the first exon. The homology between the two Xenopus species extends to about 500 nucleotides of the first intron and after that point the two genes diverge completely. The full homology of the first exon, which contains the untranslated part of the messenger, indicates a strong pressure for the maintenance of the sequence, which must have a functional role in the expression of the gene.

Structural features common to the L1 and L14 genes of X.laevis

We compared the structure of the L14 gene with that of another X.laevis r-protein gene fully sequenced, L1 (8). Common features are the absence of a canonical TATA box, a capsite located in a pyrimidine tract and the first exon made by the untranslated part of the messenger. The initiator ATG is separated from the rest of the mRNA sequence by a large intron in both genes.

Fig.8 shows the homologies in the 5' end regions of L1 and L14. Three short similar sequences are located upstream of the capsite. In the first exon a common sequence, the decanucleotide GAGAAGCAGC (boxA), is located in the same position relative to the ATG preceding the first intron. One of the two L14 genes has a C in place of the fourth A of the decanucleotide and this

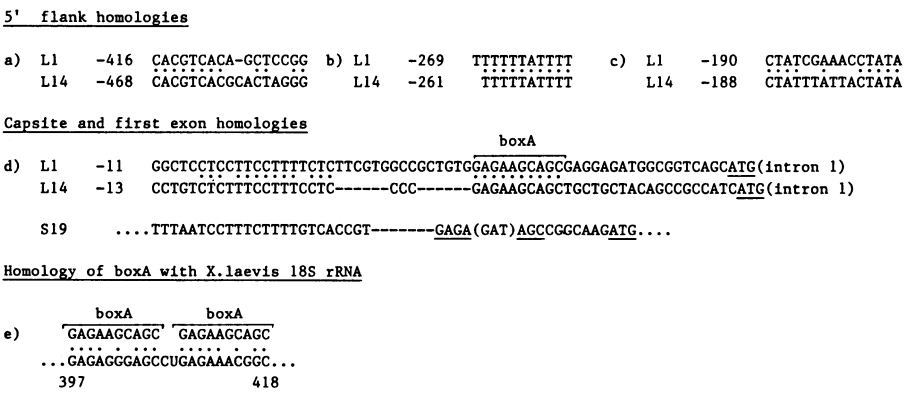

Fig.8- Homologies of 5'end sequences of L1, L14 and S19 r-protein genes in Xenopus laevis. a,b,c) genomic sequences of L1 and L14. d) the same plus S19 sequences from a cDNA clone: here capsite and position of the first intron are not determined. e) the decanucleotide called box A, common to L1 and L14, is 80% homologous to two consecutive regions of X.laevis 18S rRNA in position 397 to 418.

is the only difference between the two genes in the first exon. By computer homology search we found that the decanucleotide is 80% homologous to a region of the X.laevis 18S rRNA at position 397 to 418 (18). The sequence GAGAGGGAGC CU GAGAAACGGC of the 18S contains two consecutive 80% homology boxes separated by two nucleotides. The significance of the conserved decanucleotide is difficult to assess until more r-protein gene sequences are analyzed in X.laevis. A very similar sequence, GAGAGATAGCCGGC, is present at the 5' end of the cDNA of S19, another X.laevis r-protein (7). In this case the sequence precedes the initiator ATG of the open reading frame of three nucleotides only.

ACKNOWLEDGEMENTS

We thank Mr.G.Pisaneschi and Mr.A.Di Francesco for excellent technical assistance.This work was supported by grants from "Progetto finalizzato Ingegneria Genetica e Basi Molecolari delle Malattie Ereditarie, CNR" and from "Istituto Pasteur - Fondazione Cenci Bolognetti".

REFERENCES

1. Nomura,M.,Gourse,R.and Baughman,G.(1984) Ann.Rev.Biochem.53,75-118.
2. Fried,H.M.and Warner,J.R.(1984) p.169-192 in G.Stein and J.Stein (ed) Recombinant DNA and Cell proliferation. Academic Press,inc.,New York
3. Teem,J.L.,Abovich,N.,Kaufer,N.F.,Schwindinger,W.F.,Warner,J.R.,Lery,A., Woolford,J.Leer,R.J.,van Raamsdonk-Duin,M.M.C.,Mager,W.H.,Planta,R.J. Schultz,L.,Friesen,J.D.,Fried,H.and Rosbash,M.(1984)
4. Vaslet,C.A.,O'Connell,P.,Izquierdo,M.and Rosbash,M. (1980). Nature 285, 674-675
5. Fabijanski,S.and Pellegrini,M. (1982) Gene 18,267-276
6. Burns,D.K.,Stark,B.C.,Macklin,M.D. and Chooi,W.Y.(1984) Mol.Cell Biol. 12,2643-2652
7. Amaldi,F.,Beccari, E.,Bozzoni,I.,Luo,Z.L.and Pierandrei-Amaldi,P. (1982) Gene 17, 311-316
8. Loreni,F.,Ruberti,I.,Bozzoni,I.,Pierandrei-Amaldi,P. and Amaldi,F.(1985) EMBO J. 4, 3483-3488
9. Wiedemann,L.M.and Perry,R.P. (1984) Molec.Cell Biol.4, 2518-2528
10. Dudov,K.P.and Perry,R.P.(1984) Cell 37 , 457-468
11. Bozzoni,I.,Tognoni,A.,Pierandrei-Amaldi,P.,Beccari,E.,Buongiorno-Nardelli M. and Amaldi,F. (1982)J.Mol.Biol.161,353-371
12. Maniatis,T.,Fritsch,E.F.and Sambrook,J.(1982) Molecular cloning. A laboratory manual. Cold Spring Harbor
13. Bozzoni,I.,Fragapane,P.,Annesi,F.,Pierandrei-Amaldi,P.,Amaldi,F.and Beccari,E.(1984)J.Mol.Biol.180,987-1005
14. Maxam A.M. and Gilbert W.(1980) Methods in Enzymology 65,499-560

15. Rubin,C.M.and Schmid,C.W.(1980) Nucl.Acids Res.8,4613-4619
16. Pierandrei-Amaldi,P.and Beccari,E.(1980)Eur.J.Biochem. 106,603-611
17. Wittmann-Liebold,B.,Geissler,A.W.,Linn,A. and Wool,Ira(1979) J.Supramol. Struct. 12,425-433
18. Salim,M. and Maden,B.E.H.(1981) Nature 291, 205-208