Nucleic Acids Research, Vol. 19, No. 18 4983–4990

Complex pattern of alternative splicing generates unusual diversity in the leader sequence of the chicken link protein mRNA

Ferenc Deák[1], Endre Barta[1,2], Silvija Mestric[1+], Markus Biesold[1] and Ibolya Kiss[1*]
[1]Institute of Biochemistry, Biological Research Center of the Hungarian Academy of Sciences, H-6701 Szeged, PO Box 521 and [2]Agricultural Biotechnology Center, H-2101 Gödölló, Hungary

Received June 5, 1991; Revised and Accepted August 27, 1991 GenBank accession no. M35040

ABSTRACT

We report here the isolation of the 5′ end and the promoter region of the gene for chicken cartilage link protein, and demonstrate extensive heterogeneity of the leader sequence arising from differential utilization of multiple splice sites within the 5′-most exon. The 500-base pairs (bp) exon 1 consists of solely untranslated sequence and is followed by an intron >33 kilobase pairs (kb). Together, the five exons predict a gene size longer than 100 kb. Multiple transcription initiation sites were mapped 34, 46, 56, 66 and 76 bp downstream of a TATA-like motif. Sequence analysis revealed that in addition to the non-spliced variant, multiple mRNA species were generated by alternative splicing resulting in the exclusion of 92, 166, 170, 174 and 263 nucleotides (nt), respectively, from exon 1. Polymerase chain reaction confirmed the existence of various splice forms, and showed cell type- and developmental stage-specific expression for one group of them. Secondary structure predictions indicated that the leaders of the splice forms could form stable hairpin structures with different free energies of formation (up to $\Delta G = -110$ kcal/mol), suggesting translational control. The splice variant detected in the largest amount had the least stable predicted hairpin ($\Delta G = -31.7$ kcal/mol).

INTRODUCTION

Link protein (LP), an abundant glycoprotein of the cartilagenous extracellular matrix, stabilizes the proteoglycan aggregates via simultaneous binding to hyaluronic acid and chondroitin sulfate proteoglycan (1–5). Two or three forms of LP found in several species (6–8), seem to differ only in proteolytic cleavage of a short peptide and in the degree of glycosylation (7, 9).

The precursor of chicken LP consists of 355 amino acids including a secretory signal peptide (10). The mature protein is divided into three structural and functional domains stabilized by disulfide bridges (11, 12). The immunoglobulin-like domain interacts with proteoglycan, while the two tandemly repeated domains bind to hyaluronate (13, 14). Link proteins may have yet other functions as they are also found in tissues other than cartilage (15–18). Furthermore, an alternatively spliced exon encoding an extra segment between the signal peptide and the first domain of rat chondrosarcoma LP has been reported (19).

In chicken cartilage, mRNA size classes of 5.8–6.0 kb and 3.0 kb, differing in their 3′ untranslated region (3′-UTR), were detected (10). The LP mRNA species are encoded by a single gene per haploid genome (12). The four protein coding exons, which evolved by duplication and exon shuffling, are scattered over a genomic region longer than 70 kb in chicken (12). The 5′ end of the gene, however, has not been isolated previously.

In the present study, we characterize the first exon and the 5′-flanking region of the chicken LP gene. A comparison of the nucleotide sequences of genomic and cDNA clones, and polymerase chain reaction (PCR) products revealed that alternative splicing within the 5′-UTR gave rise to the formation of multiple LP mRNA species in chondrocytes. The potential role of the heterogeneous 5′-UTR in the regulation of gene expression is discussed.

MATERIALS AND METHODS

Primer extension, cDNA and genomic cloning

All positions are given from the first nucleotide of the translation start codon of the LP precursor (10, 12). Oligonucleotides (Table I) were synthesized by phosphoramidite chemistry (20) and labeled with T4 polynucleotide kinase (21).

RNA was prepared from sterna of 14-day-old chicken embryos, and from sterna and articular cartilage of 6-week-old (juvenile) chicks as described (22). For analytical primer extension, 0.1 pmoles of end-labeled primer was annealed with 4 μg of poly(A)$^+$ RNA immediately after denaturation with 10 mM methyl-mercuric hydroxide, and elongated with 40 units of M-MLV reverse transcriptase (BRL) in 50 μl of 50 mM Tris-HCl

* To whom correspondence should be addressed

+ Present address: Pliva Research Institute, 41000 Zagreb, I.L. Ribara 89, Yugoslavia

(pH=8.3), 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT, supplemented with 0.5 mM dNTP, 40 μg/ml actinomycin D and 20 units of placental RNase inhibitor.

For the construction of the primer extension library, 0.25 pmoles of P3r primer and 5 μg poly(A)$^+$ RNA were used. The cDNA-RNA hybrids were tailed using dGTP and terminal transferase and annealed with 1.15 pmoles of oligo(dC)-tailed *Bam*HI linker in the presence of 1.5 units of RNase H. The second cDNA strand was synthesized with 10 units of DNA polymerase I. The ends were polished with T4 DNA polymerase and ligated to phosphorylated *Sal*I linker. After *Bam*HI and *Sal*I digestion, the cDNA was size fractionated by Sepharose CL-4B column chromatography and inserted into pUC12 vector. Recombinant plasmids were introduced into *Escherichia coli* DH5 cells (23). Colonies were screened by hybridization to end-labeled P2r and P3r as described (10).

A chicken genomic library (12) was screened using the 277-bp *Eco*RII-*Sac*I fragment (from position −312 to −35) of pLP7G12.

Table I. Oligonucleotide primer sequences

Name	Sequence	Position
	Forward primers	
PA	5′-CAAGTGGTCAGAAGTATCACGT-3′	−605 to −584
PB1	5′-TAGTTCGGGACTGGTGTGCG-3′	−488 to −469
PB2	5′-TTAGTTCGGGACTGGTGTGCGGTGCAGAGC-3′	−489 to −460
PC	5′-ACTTGGGAGCTCCACACAA-3′	−46 to −28
	Reverse primers	
PDr	5′-GTCTCTTTGTGGCTCTGGGTGGCAGAGGAC-3′	−159 to −188
PEr	5′-GGTGGCAGAGGACCTAAAAAAGCTCTGCAC-3′	−176 to −188
		and −452 to −468
PFr	5′-CTGCCCCAGCCTCCGTGCCCCTCAGCTTCT-3′	−404 to −433
PGr	5′-GGTGGCAGAGGACTCACACACAGAGTGTCC-3′	−176 to −188
		and −359 to −375
P2r	5′-CTTGTCATCTTCACAGTCAC-3′	8 to −12
P3r	5′-ACCACAAGTAGGCGGGGTCC-3′	128 to 109

Reverse transcription coupled with PCR (RT-PCR)

The procedure was similar to published methods (24, 25) and was reproducible in repeated experiments. cDNA was synthesized by extension of primer P3r using 2 μg poly(A)$^+$ RNA as template. Following alkaline hydrolysis of RNA and ethanol precipitation, the cDNA was amplified, as indicated, using 2.5 units of *Taq* polymerase (NEBL) and 100 pmoles of primers in 67 mM Tris-HCl (pH=8.8), 16.6 mM $(NH_4)_2SO_4$, 10 mM 2-mercaptoethanol, 6.7 mM $MgCl_2$, 0.2 mM of each dNTP, 10% (v/v) DMSO in a final volume of 100 μl. The last five cycles were carried out by adding 10 μCi of radioactive reverse primer for analysis on sequencing gels.

In other experiments primer PDr was annealed to 2 μg poly(A)$^+$ or 20 μg total RNA at 64°C and reverse transcribed in a final volume of 50 μl. After denaturation for 5 min at 100°C, 4 μl of this reaction mixture was added to 10 mM Tris-HCl (pH=7.9), 2 mM $MgCl_2$, 10% Triton X-100, 1 μM of each primer, 100 μg/ml bovine serum albumin, and 200 μM of each

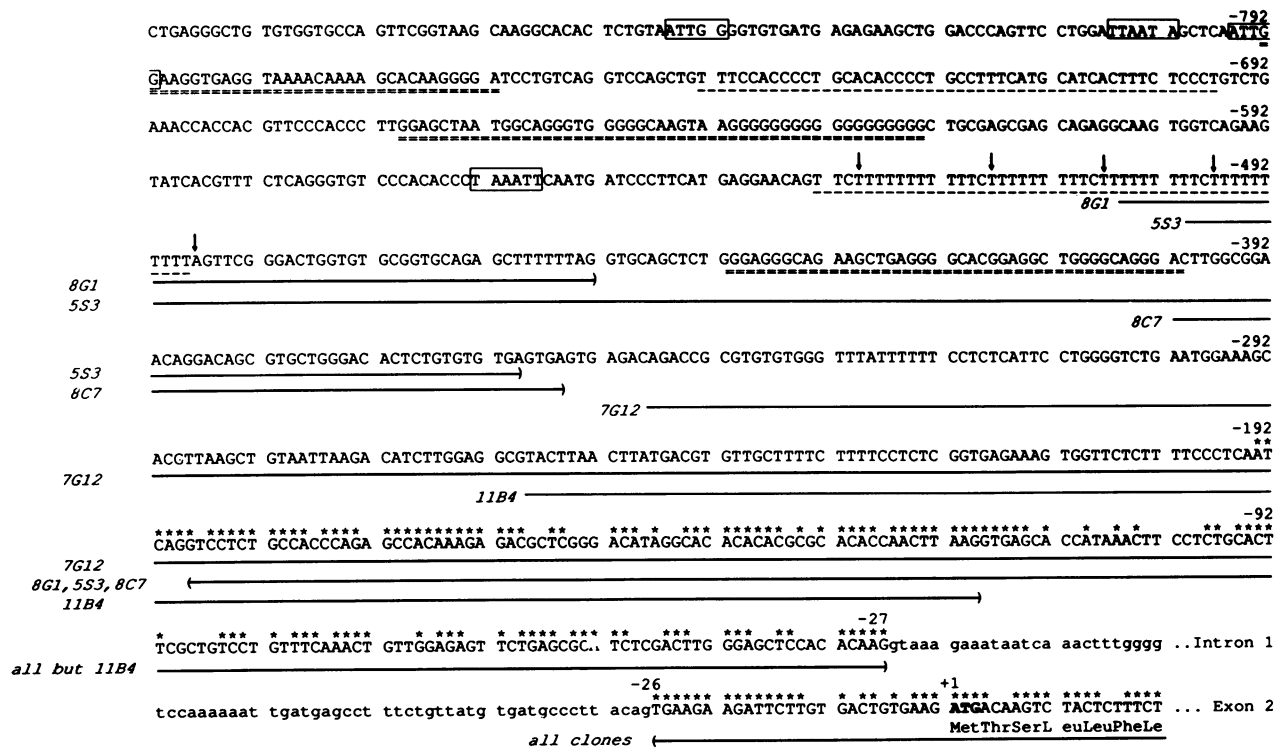

Fig. 1. Nucleotide sequence of the 5′ end of the chicken LP gene. Positions are given from the translation start site within exon 2, not including intron 1 (nucleotides in lower case letters). Nucleotide sequences represented by cDNA clones are indicated by solid lines. All sequences were determined from both directions. TATA and CCAAT motifs are boxed. Homopurine and homopyrimidine tracts longer than 30 nt are denoted by double and single broken lines, respectively. Arrows point to the major transcription start sites. Nucleotides identical with the human sequence (31) are marked with asterisks above the sequence. Physical map of cDNA (**A**) and genomic (**B**) clones spanning the 5′ end of the LP gene. Solid bars represent the vectors. Hybridization to primers P2r (−) and P3r (=) is indicated below the map of each cDNA. A more detailed restriction map of the 1.7-kb *Sal*I-*Eco*RI genomic fragment is shown in an expanded format. Open box indicates the position of exon 1. Horizontal arrows denote the sequencing strategy. Abbreviations of restriction enzymes: A, *Ava*I; B, *Bam*HI; C, *Sac*I; E, *Eco*RI; R, *Eco*RII; S, *Sal*I; V, *Ava*II.

dNTP in a final volume of 50 μl. Amplification was performed with 1.25 units of *Taq* polymerase using a Coy TempCycler.

DNA analysis and secondary structure prediction

The nucleotide sequence was determined by the method of Sanger et al. (26) using M13 subclones constructed as described (10). Transcription start points were mapped with T4 DNA polymerase on cloned genomic DNA as suggested by Hu and Davidson (27). Fragments separated by polyacrylamide gel electrophoresis were transferred to nitrocellulose filters after denaturation by boiling for 10 min in 1 M ammonium-acetate, 20 mM NaOH. DNA fragments were labeled by random primer extension (28).

RNA secondary structure was predicted with the FOLD program of Zuker (29) included in the Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin (30).

RESULTS

Alternative splicing in the 5'-UTR of the chicken LP gene

Genomic clones for chicken LP isolated previously carried the entire protein coding region and 3'-UTR, but not the first exon of the gene (12). In order to isolate the 5' end of the gene, we constructed a cDNA library using poly(A)$^+$ RNA purified from sterna of chicken embryos as template, and a specific primer P3r (Table I), complementary to the LP mRNA from position 109 to 128 relative to the translation start site. The library was screened with a mixture of end-labeled P2r and P3r oligonucleotides (Table I), specific to exons 2 and 3, respectively. Five of the positive cDNA clones with the longest inserts were sequenced.

The 277-bp *Eco*RII-*Sac*I fragment of the most 5'-end clone pLP7G12 hybridized to RNA species of the same mobility as described for LP mRNA (10) (data not shown). The same fragment was used as a hybridization probe to screen a chicken genomic library. One positive clone, λgLP100 was isolated and characterized. Restriction mapping and Southern hybridization revealed that the clone carried exon 1 alongwith 800 bp of upstream sequence and 19.7 kb of the putative first intron (data not shown). Even though the previously described genomic clone λgLP12.1 (12) extended 13.6 kb upstream of the 5'-most translated exon, named exon 2 in this paper, the restriction maps of the two clones did not overlap, nor did the two clones cross hybridize. Therefore, we concluded that the first intron is larger than 33 kb. Since the protein coding exons including the 3'-UTR cover a genomic region of 70 kb or longer (12), hence this indicates a primary transcript over 100 kb in length.

The nucleotide sequence of the first exon and flanking regions was determined (Fig. 1). Alignment of the genomic and cDNA sequences revealed complete identity along exon 2; however, only one of the new cDNA clones, pLP7G12 overlapped continuously with the corresponding genomic sequence. Shorter or longer internal segments bordered by consensus donor and acceptor splice sites (Table II) were absent from the 5'-UTR of the other cDNA clones. This observation strongly indicates that the templates of the cDNA molecules were generated by alternative splicing events within exon 1, defining at least five different splice variants for LP mRNA. A homology search revealed that in the leader region, two out of four potential acceptor sites were utilized, while out of the seven putative donor sites, six have contributed to the generation of observed diversity (Table II).

Two short open reading frames were found in the 5'-UTR from −300 to −276 and from −248 to −122. Both of these upstream initiation codons are removed by excision of the alternative introns from the 8C7-, 5S3- and 8G1-type messages. Polymorphism was

Table II. Putative splice sites at the 5' end of the LP gene

donor site[a]	pos[b]	acceptor site[c] pos[b]	cDNA clone
		ttttttttagTTC −485	
		gctttttagGTG −451	
TAGgtgcag	−451	cctcaatcagGTC −188	pLP 8G1
TGTgtgagt	−362	cctcaatcagGTC −188	pLP D48
TGAgtgagt	−358	cctcaatcagGTC −188	pLP 5S3
TGAgtgaga	−354	cctcaatcagGTC −188	pLP 8C7
TCGgtgaga	−220		
AAGgtgagc	−118	gcccttacagTGA intron1	pLP 11B4
AAGgtaaag	−26	gcccttacagTGA intron1	all pLP clones except 11B4
C/A AGgtragt		yyyyyynyagG	consensus[d]

[a]motifs containing AGgtr or gtrag sequence. r: a or g.
[b]position of the splice site in the sequence as in Fig. 1.
[c]motifs differing by not more than two nucleotides from the sequence yyyyyy*nyag*G and only at positions not underlined. y: c or t.
[d]Shapiro and Senapathy (32).

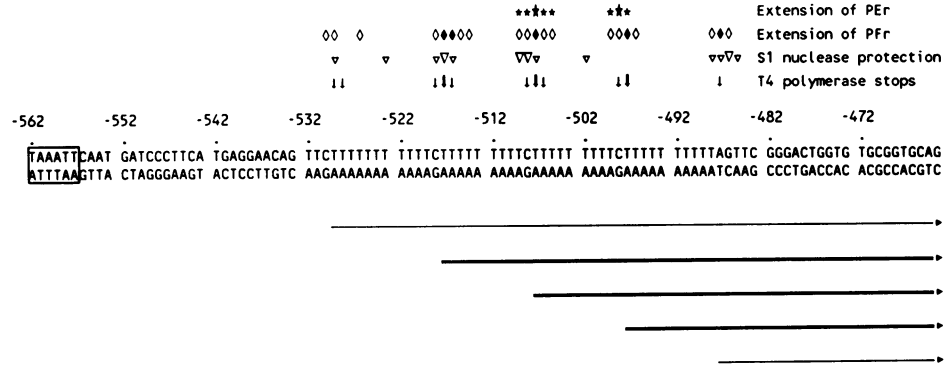

Fig. 2. Composite representation of the transcription initiation sites determined by various techniques. The TATA-like sequence is boxed. Thickness of the horizontal arrows below the sequence reflects the relative frequency of initiation. Nuclease S1 mapping of the 5' end of the chicken LP gene. Uniformly labeled, single stranded probes, complementary to the mRNA from positions indicated by solid black bars in the schematic diagram, were made by extension of the sequencing primer on M13 subclones. The probes were hybridized to 60 μg of total RNA and treated as described (43). The length of the protected fragments, represented by thin lines, are given in nucleotides. A, B and C refer to respective probes. Lanes G and T, dideoxy sequencing ladder; lane 1, undigested probe; lanes 2 and 3, probe/RNA hybrid treated with nuclease S1 at 30°C or 37°C, respectively; lane 4, probe digested without RNA added.

detected between −370 and −366, where some of the cDNA clones lacked one of the repeated CT dinucleotides. Sequence elements of potential functional significance were identified in the 5′-flanking region of the gene. Two TATA-like (at −806 and −562) and two CCAAT motifs (at −845 and −795) were found (Fig. 1). The sequence from −387 to −370 matches the consensus cleavage site for chicken topoisomerase (33). Purine and pyrimidine rich segments alternate throughout the sequence. The nucleotide sequence from position −193 to −27 has 69% identity with the corresponding human sequence (31).

Mapping of transcription initiation sites of the LP gene

Three different techniques were used to define the transcription start site and the results are summarized in Fig. 2. S1 nuclease mapping using probes of three different lengths (data not shown) indicated multiple initiations in a region from position −529 to −485. A strong protection was observed at the 3′ end of a poly(T) tract around −486 and two additional frequent initiation sites were mapped at −517 ± 1 and −508 ± 1 by each probe (Fig. 2).

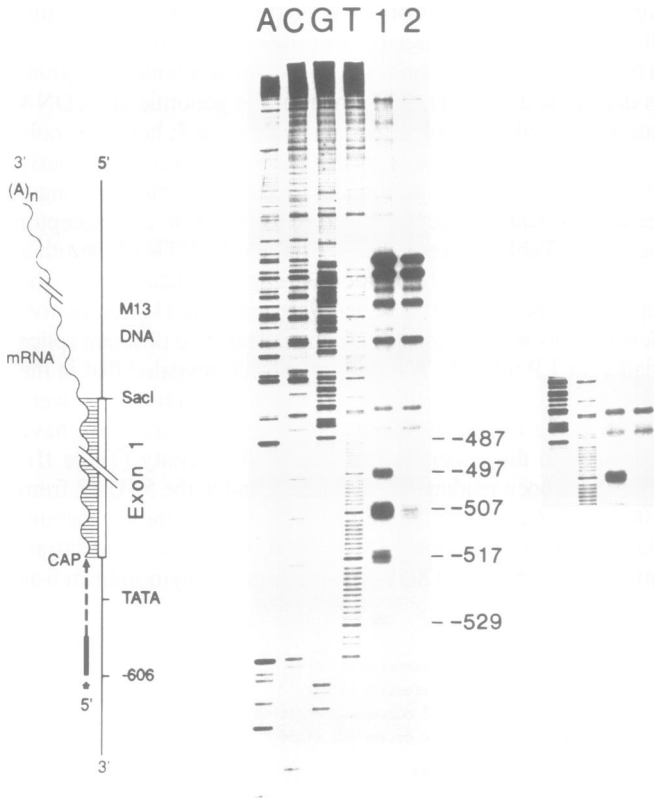

Fig. 3. Mapping of transcription start sites by T4 DNA polymerase. The diagram to the left depicts the experimental strategy. An M13 DNA carrying the RNA-complementary strand of exon 1 and the promoter region was used as template. Solid bar with asterisk represents the ^{32}P end-labeled PA primer, broken arrow indicates the primer extension products. Lanes A, C, G and T, dideoxy sequencing ladders of the same template DNA from unlabeled PA primer in the presence of [^{35}S]dATP. Lane 1, 4 μg poly(A)$^+$ RNA was hybridized to the template before primer extension. Lane 2, primer extension in the absence of hybridized mRNA. The numbers at the right margin of the autoradiogram denote the nucleotide positions. That part of the autoradiogram which carries the extension product at position −487 is shown after prolonged exposure, in an inlet to the right. Analysis of PCR products on a denaturing polyacrylamide gel. **D**, PC/P3r primer pair was used in RT-PCR under conditions described for Fig. 6. A and B. Lane 1, products of RT-PCR. Lane 3, PCR performed from 1 ng insert of pLP8G1 as template. M, pUC12 *Hae*III digest used as size marker.

Performing S1 protection assays with other probes, we could not, however, detect any transcript originating from the TATA motif at −806 (data not shown). These results indicated that the TATA-like motif at −562 may serve as a promoter for the LP gene.

S1 nuclease sensitivity of the poly(U) stretch at the 5′ end of the LP mRNA may have contributed to the S1 pattern observed. Therefore, we also mapped the 5′ end using T4 DNA polymerase (Fig. 3). A specific primer, PA, was annealed to a single stranded antisense DNA spanning the 5′ end of the LP gene, and the primer was extended with T4 DNA polymerase in the presence of hybridized RNA. Multiple start sites were determined within the poly(T) stretch from −529 to −487. The major sites were located at −497, −507 ± 1 and −517 ± 1, and two minor sites at −529 and at −487. The extension product representing the last site was faint but visible in the original autoradiogram.

In primer extension experiments (Fig. 4) the primer PEr, complementary to the splice form represented by pLP8G1, resulted in two major clusters of extension products 60 ± 1 and 69 ± 2 nt in length, thus defining transcription initiation at −498 ± 1 and −507 ± 2, respectively. In contrary, when PFr, a primer specific to all but 8G1-type splice variants was used, the extension products indicated multiple start sites within a wider region, between −529 and −487 (Fig. 2). The most abundant transcripts (113 nt) initiated around −517. The difference between the primer extension results in repeated experiments suggests that the LP splice forms may differ in the relative frequency of utilization of the multiple start sites. However, due to the complex nature and high stability of the leader (see below)

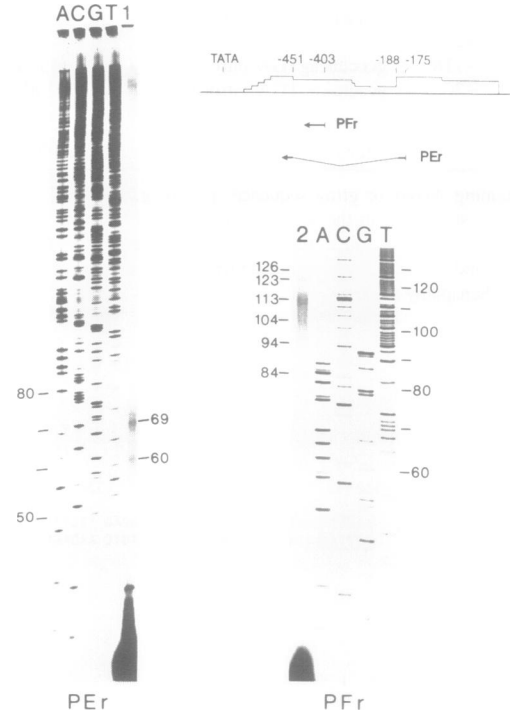

Fig. 4. Primer extension analysis of the 5′ end of the LP gene. 4 μg poly(A)$^+$ RNA was hybridized at 60°C to 5′ ^{32}P end-labeled oligonucleotides, complementary to exon 1 at positions shown in the schematic diagram. Exon 1 is boxed. cDNA was synthesized at 42°C for 1 h with M-MLV reverse transcriptase. The extension products from primers PEr (lane 1) and PFr (lane 2) were analysed on sequencing gels. Lanes A, C, G and T, dideoxy sequencing ladders. Numbers at the margins of the autoradiograms denote the length in nucleotides.

which impeded the design and extension of primers specific for the distinct splice forms, we could not test this hypothesis.

To sum up, the results obtained by three different strategies are generally consistent with the conclusion that the LP gene is transcribed from multiple start points located within the poly(T) strech 34 ± 1, 46 ± 1, 56 ± 1, 66 ± 1 and 76 ± 1 bp downstream of the TATA-like motif (Fig. 2). Nucleotide sequences at the initiation sites match the YYCAYYYY consensus sequence (34) described for other eukaryotic genes except for the lack of A at position 4. The initiation sites at −507, −517 and −497 are used more frequently than the others. The major discrepancies in S1 mapping (two minor sites and increased frequency of cleavage around −486) can be explained by the S1 sensitivity of the 5′ end of the hybrid. Indeed, more accurate data were obtained with T4 DNA polymerase as compared to the other two techniques, due to the fact that this analysis was neither influenced by the base composition nor the secondary structure of the RNA.

PCR analysis of the multiple LP mRNA species

The expected fragment sizes for various combinations of forward and reverse primers in RT-PCR analysis are depicted in Fig. 5. Using two different primer pairs (Fig. 6A and B), splice variant represented by cDNA clone 8G1 was found to be the most abundant among the LP mRNA species in embryonic sternal cartilage, followed by the 8C7/5S3-type variants. Bands corresponding in size to 7G12 and 11B4 splice forms were not detectable in PCR even upon prolonged overexposure of the autoradiograms; however, several shorter products were found, corresponding in size to the hypothetical messages Hyp2−4 (Fig. 6A, lane 2).

In order to test if additional exons hidden in the long first and second introns have contributed to the diversity observed, another RT-PCR was carried out with the PC/P3r pair of primers. A single product of 174 bp was monitored, identical in size to that synthesized on the cDNA clone pLP8G1 (not shown). Longer PCR products were not observed even after long exposure. Therefore, we concluded that no alternative exons located between exon 1 and exon 3 were detectable in chicken cartilage.

The PCR products and the cDNA clones were further correlated by designing the PB2/PDr pair of primers, each 30 nt in length, thus decreasing the complexity and minimizing the nonspecific background in RT-PCR. Forty cycles produced DNA fragments visible in ethidium-bromide stained gels irrespective of whether poly(A)$^+$ or total RNA was used as template (Fig. 7A). All the bands corresponding to the various cDNA clones were observed, except for the 7G12-type PCR fragment. In fact, hybridization to a specific probe clearly indicated the presence of the 7G12-type PCR product (Fig. 7B, lane 2) appearing as fragments of ever-increasing length. The explanation for this phenomenon is the presence of direct repeats, which led to the out-of-register hybridization of competing PCR products at subsequent cycles of amplification. In addition to several shorter repeats, a tandem repeat of 23 nt containing four mismatches was found at −338 and −243. Similar artefacts were detected by others when DNA template containing tandem repeats was used in PCR (35).

Surprisingly, apart from the expected PCR products detected in denaturing gels (Fig. 6C), two additional bands (marked X

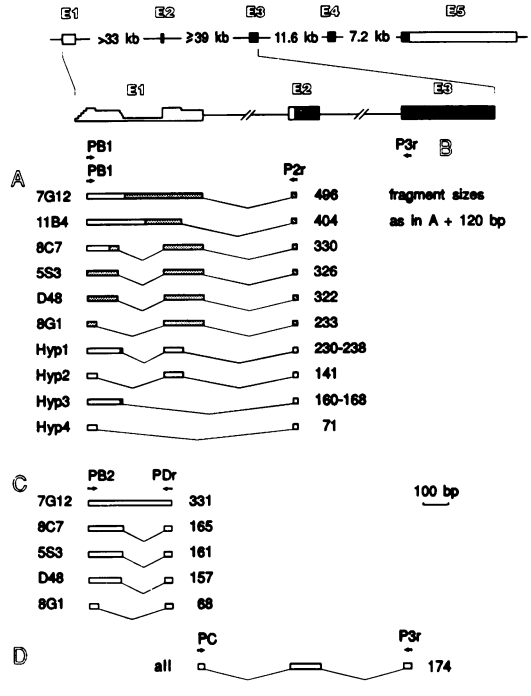

Fig. 5. Diagram of alternative splicing in the LP leader region. The structure of the gene is shown at the top of the figure. Exons (E1-E5) are numbered from the 5′ end of the gene (note the difference from ref. 12). Open boxes represent the 5′- and 3′-UTRs and black boxes the protein coding exons. Arrows indicate the positions of the primers used in PCR. The PCR fragments assumed to be formed with various primer pairs, based on the structure of the cDNA clones, are depicted. The regions harbored by the cDNA clones are shown as dotted bars for comparison. Primer pairs PB1/P2r (A), PB1/P3r (B), PB2/PDr (C) and PC/P3r (D) were chosen. Fragment sizes expected for the various splice forms are shown in base pairs at the right side of the diagrams.

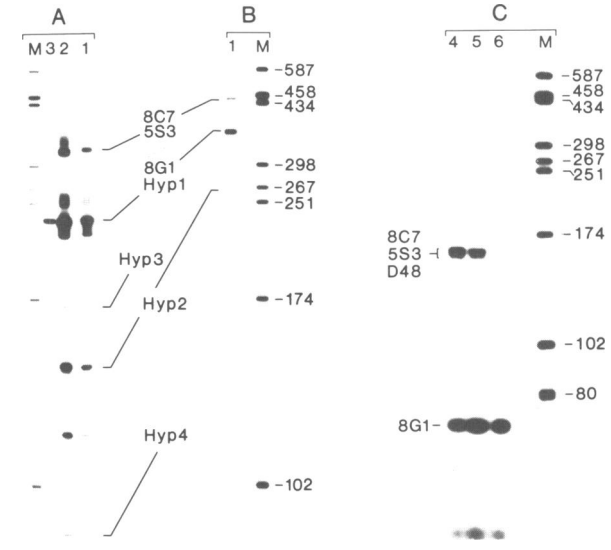

Fig. 6. Analysis of the PCR products on denaturing polyacrylamide gels. A, B and C represent primer combinations as in Fig. 5. **A** and **B**, cDNA sythesized from poly(A)$^+$ RNA of chick embryo sterna was subjected to 30 cycles of amplification at 94°C for 1.5 min, 48°C for 2.5 min and 75°C for 3.5 min in the presence of 10% DMSO. Lane 1, Products of RT-PCR. Lane 2, as lane 1, but 10-fold amount was loaded. Lane 3, PCR performed from 1 ng insert of pLP8G1 as template. M, pUC12 *Hae*III digest used as size marker. **C**, Total RNA from embryonic sternal (lane 4), juvenile sternal (lane 5) and juvenile articular (lane 6) chondrocytes was used in RT-PCR employing the PB2/PDr primer pair. 30 cycles of amplification were performed at 94°C for 1 min, 65°C for 2 min and 75°C for 3 min. Numbers to the right indicate the fragment sizes in nucleotides.

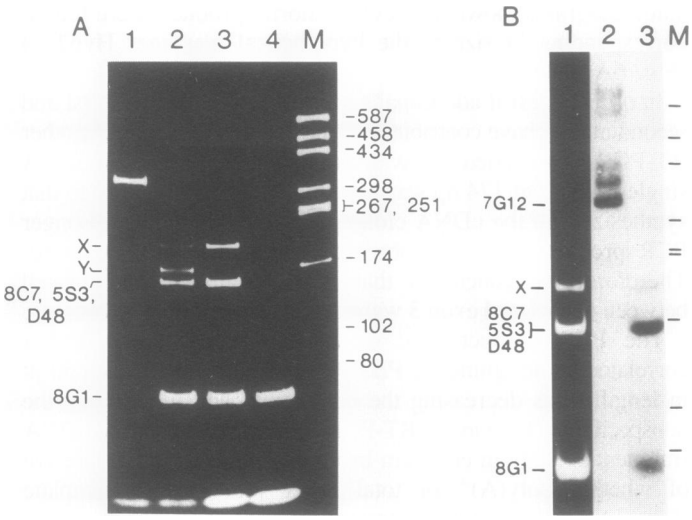

Fig. 7. Analysis of the PCR products on nondenaturing gels. 40 cycles of amplification were performed in RT-PCR under conditions as in Fig. 6C. **A**, UV-fluorescent picture of the ethidium-bromide stained gels. Total RNA obtained from embryonic sternal (lane 2), juvenile sternal (lane 3) and juvenile articular (lane 4) chondrocytes was used. Control reaction was performed employing a cloned genomic DNA fragment carrying exon 1 (lane 1). **B**, Southern hybridization. RT-PCR products obtained using RNA isolated from juvenile sternal chondrocytes were separated on a 5% polyacrylamide gel (lane 1), blotted and hybridized to a continuously labeled 7G12-specific probe (lane 2) extending from position −347 to −188, and to the 5′ end-labeled oligonucleotide PGr (Table I), specific for the 5S3 form (lane 3). M, positions of the pUC12 *Hae*III fragments as in A.

and Y in Fig. 7) were visualized in non-denaturing gels. The apparent mobilities of these bands decreased by increasing the temperature during electrophoresis, and these bands hybridized to the 5S3-specific oligonucleotide (Fig. 7B, lane 3). To confirm the identity of the PCR fragments, all the bands were excised from the non-denaturing gels, cloned and sequenced. Sequencing of the 160 bp fragment confirmed the presence of the 5S3 and 8C7 types, and in addition, revealed the existence of a new splice variant (D48), which utilized the donor site located at position −362. Sequence analysis also proved that band X represented the heteroduplexes of 5S3/D48 and 8C7/5S3 PCR products, differing in length by 4 nt. On the other hand, due to the polymorphism of the embryonic population, the 2 nt difference resulted in the heteroduplex band Y. This band was not observed when RNA from sternal or articular cartilage of a homozygotic animal was used in RT-PCR (Fig. 7A, lanes 3 and 4). The formation of heteroduplexes can be explained by competition of the accumulating PCR products with the oligonucleotide primers for annealing to template DNA.

To sum up, in spite of the difficulties due to the complexity of the 5′-UTR, PCR confirmed the diversity of the LP leader sequence and revealed differences in the relative amounts of the various splice forms. The 68-bp fragment, representing the 8G1 splice variant, was the most abundant products after various amplification regimens. The 8C7/5S3/D48 splice forms, however, showed cell type-specific expression, as being detectable in significantly lower amounts in juvenile articular compared to embryonic and juvenile sternal cartilage, in repeated experiments (Fig. 6C and 7A). Finally, the 7G12-type PCR product was found in very small amounts when either RNA source was used.

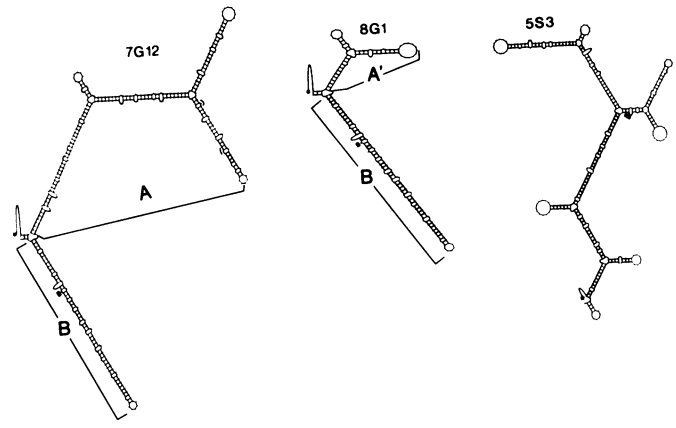

Fig. 8. Predicted secondary structures of the LP leaders. The program FOLD (29) was used to predict the most stable stem-loop structure in the 5′-UTR of LP mRNAs as shown from position −529 to +35 for the 7G12- and 8G1-type variants, and from −529 to +128 for the 5S3-type. The overall free energy of formation (ΔG) values are −150.3, −71.8 and −140.4 kcal/mol for splice forms 7G12, 8G1 and 5S3, respectively. A, B and A' represent different domains with predicted free energies of −110.2 kcal/mol, −38.8 kcal/mol and −31.7 kcal/mol, respectively. Asterisks mark the cap of the longest transcript, solid bars and thin lines show the positions of the initiating and upstream AUGs, respectively.

Potential secondary structure of the LP leaders

Fig. 8 shows computer models of the most stable secondary structures predicted for the 5′ end of three LP splice forms. The 7G12-type variant appears to possess the most extensive secondary structure, which can be divided into two regions: 1/ domain A from position −477 to −126 contains most of the leader including both upstream open reading frames and is predicted to form a very stable stem-loop structure with $\Delta G = -110.2$ kcal/mol; and 2/ domain B from −120 to +30 includes the translation start codon for the LP precursor and has a predicted $\Delta G = -38.8$ kcal/mol. The computer generated hairpin structure of the 8G1 splice form also involves domain B but domain A is significantly truncated with a predicted $\Delta G = -31.7$ kcal/mol. In contrary, the secondary structures predicted for the 5S3, 8C7 and D48 variants are entirely different from that of 7G12-type. It suggests that deletion of the leader sequence from position −358 to −189 may influence the folding pattern around the translation start site.

DISCUSSION

The present studies demonstrate that the transcription unit for chicken LP has a complex structure. Even though the gene comprises of only 5 exons, the size of the transcription unit is larger than 100 kb, due to the presence of long introns and unusually long UTRs (529 bp for 5′-UTR and 4.8 kb for 3′-UTR). Long leaders are, however, known for mRNAs of important regulatory proteins involved in growth and development (36−40).

The LP gene is transcribed from multiple initiation sites located between 34 and 76 bp downstream of a TATA-like motif. Although this motif differs from the canonical sequence of RNA-polymerase II-dependent promoters (41), preliminary experiments indicate that a 140 bp fragment carrying this motif is indeed able to work as a promoter in transient expression assays. The occurrence of multiple start sites and the high GC content of the

LP promoter resembles to the features of housekeeping genes (42), but differs from those in the lack of the classical Sp1 binding sites. GC-rich sequences were also found in the 5'-flanking region of other cartilage-specific genes (43, 44).

One of the most striking features of the LP gene is the complex structure of the 5'-UTR (Fig. 5) arising from alternative splicing within exon 1, via differential utilization of six donor and two acceptor splice sites. The generation of six different splice forms confirmed by sequencing of cDNA clones and PCR fragments, however, provides only a minimum estimate. Additional splice variants may exist as suggested by the appearance of PCR fragments Hyp1-Hyp4 (Fig. 5A and 6A). The reason why these forms are not represented by the cDNA clones, probably is that the cDNA clones, harboring short inserts, have not been selected for further analysis. The occurrence of multiple start sites, and the utilization of distinct polyadenylation sites (10) may also increase the heterogeneity. Furthermore, it is possible that alternative splicing in the translated region, although not detected in chicken cartilage, may contribute to the diversity of LP mRNA in other species or tissues, as observed recently in rat chondrosarcoma (19).

Complex structure of the 5' leader has been described in a few cases (45–48). Alternative splicing within the 5'-UTR is often combined with alternative promoter usage (47, 48). In case of the LP gene, where the multiple LP mRNA species arise from the same promoter, the gene expression may further be controlled by regulated splicing by providing specific splicing factors. It is possible that some of the LP splice forms (7G12, 11B4) may in fact represent processing intermediates, which are kept nonfunctional by temporarily retaining the intron. Examples of regulated splicing have been reviewed recently (36).

The fact that heterogeneity is confined to the highly conserved 5'-UTR of the LP mRNA which shows cell type- and stage-specific expression, opens the possibility that the various splice forms are subjected to some kind of translational control. The differences in the predicted secondary structure of the various splice forms seem to support this assumption. The splice form which had the least stable predicted secondary structure (8G1), was relatively more abundant in chondrocytes of different origin compared to the form with potentially more stable leader structure (7G12; $\Delta G = -110$ kcal/mol). The hypothesis of extensive secondary structure formation in the LP leader is substantiated considering the difficulties in primer extension and RT-PCR analysis of the mRNA, and also in sequence analysis.

The various LP mRNA species may have different stabilities. Alternatively, they may be translated at different rates, since hairpin structures with predicted free energies greater than −50 kcal/mol, located upstream of the initiator AUG, have been proposed to inhibit the translation of the mRNA by interfering with the migration of the ribosomal initiation complex (49, 50). It has been confirmed recently for the mRNAs of ornithine decarboxylase and several proto-oncogenes that long leaders, with extensive secondary structure and upstream AUG codons, impair translation (36, 39, 40, 51, 52). Binding of regulatory proteins, as demonstrated for the leaders of ferritin mRNA and poliovirus RNA (53–56), may also be involved in translational control of alternatively spliced LP mRNAs.

Short open reading frames located upstream of the actual translation start site may also influence the translation efficiency of an mRNA (50, 52). Two of the LP splice variants (7G12 and 11B4) carry upstream open reading frames in retained introns. Since all three AUG codons, including the initiator codon for LP, occur in unfavorable context for initiating translation (50, 57), the upstream AUG codons may also have a regulatory role.

Data presented here provide a basis for further analysis of the role of the 5'-flanking regions and various leader sequences in the transcriptional and translational regulation of the expression of LP in different cell-types and during development.

ACKNOWLEDGEMENTS

We thank F.Solymosi for valuable discussion, L.Szilák for the synthesis of the oligonucleotides, I.Fekete and A.Simon for excellent technical assistance, and A.Borka for the artwork. This work was supported by Grant OKKFT Tt 1986/296 from the National Foundation of Technical Development of Hungary and Grant OTKA 168 and 896 (to I.K.) and a special fund from the Biological Research Center (to F.D.).

REFERENCES

1. Hascall, V.C. and Sajdera, S.W. (1969) *J. Biol. Chem.* **244**, 2384–2396.
2. Heinegard, D. and Hascall, V.C. (1974) *J. Biol. Chem.* **249**, 4250–4256.
3. Oegema, T.R.,Jr., Brown, M. and Dziewiatkowski, D. D. (1977) *J. Biol. Chem.* **252**, 6470–6477.
4. Hardingham, T.E. (1979) *Biochem. J.* **177**, 237–247.
5. Thornton, D.J., Sheehan, J.K. and Nieduszynski, I.A. (1987). *Biochem. J.* **248**, 943–951.
6. De Luca S., Heinegard, D. and Hascall, V. C. (1977) *J. Biol. Chem.* **252**, 6600–6608.
7. Baker, J.R. and Caterson, B. (1979) *J. Biol. Chem.* **254**, 2387–2393.
8. Roughley P.J., Poole, A.R. and Mort, J.S. (1982) *J. Biol. Chem.* **257**, 11908–11914.
9. Le Gledic, S., Périn, J-P., Bonnet, F. and Jolles, P. (1983) *J. Biol. Chem.* **258**, 14759–14761.
10. Deák, F., Kiss, I., Sparks, K.J., Argraves, W.S., Hampikian, G. and Goetinck, P.F. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 3766–3770.
11. Neame, P.J., Christner, J.E. and Baker, J.R. (1986) *J. Biol. Chem.* **261**, 3519–3535.
12. Kiss, I., Deák, F., Mestric, S., Delius, H., Soós, J., Dékány, K., Argraves, W.S., Sparks, K.J. and Goetinck, P.F. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6399–6403.
13. Goetinck, P.F., Stirpe, N.S., Tsonis, P.A. and Carlone, D. (1987) *J. Cell Biol.* **105**, 2403–2408.
14. Périn, J.-P., Bonnet, F., Thurieau, C. and Jolles, P. (1987) *J. Biol. Chem.* **262**, 13269–13272.
15. Gardell, S., Baker, J., Caterson, B., Heinegard, D. and Roden, L. (1980) *Biochem. Biophys. Res. Commun.* **95**, 1823–1831.
16. Poole, A.R., Pidoux, I., Reiner, A., Cöster, L. and Hassell, J. R. (1982) *J. Cell Biol.* **93**, 910–920.
17. Fife, R.S., Caterson, B. and Myers, S.L. (1985) *J. Cell Biol.* **100**, 1050–1055.
18. Stirpe, N.S., Dickerson, K.T. and Goetinck, P. F. (1990) *Dev. Biol.* **137**, 419–424.
19. Rhodes, C., Doege, K., Sasaki, M. and Yamada, Y. (1988) *J. Biol. Chem.* **263**, 6063–6067.
20. Sinha, N.D., Biernat, J., McManus, J. and Köster, H. (1984) *Nucleic Acids Res.* **12**, 4539–4557.
21. Maxam, A. and Gilbert, W. (1980) *Methods. Enzymol.* **65**, 499–560.
22. Argraves, W.S., Deák, F., Sparks, K.J., Kiss, I. and Goetinck, P.F. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 464–468.
23. Hanahan, D. (1983) *J. Mol. Biol.* **166**, 557–580.
24. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) *Science* **239**, 487–491.
25. Lipson, K.E. and Baserga, R. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 9774–9777.
26. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
27. Hu, M. C-T. and Davidson, N. (1986) *Gene* **42**, 21–29.
28. Feinberg, A.P. and Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13.
29. Zuker, M. and Stiegler, P. (1981) *Nucleic Acids Res.* **9**, 133–148.
30. Devereux, J., Haeberli, P. and Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.

31. Dudhia, J. and Hardingham, T.E. (1990) *Nucleic Acids Res.* **18**, 1292.
32. Shapiro, M.B. and Senapathy, P. (1987) *Nucleic Acids Res.* **15**, 7155–7174.
33. Spitzner, J.R. and Muller, M.T. (1988) *Nucleic Acids Res.* **16**, 5533–5556.
34. Bucher, P. and Trifonov, E.N. (1986) *Nucleic Acids Res.* **14**, 10009–10026.
35. Jeffreys, A.J., Wilson, V., Neumann, R. and Keyte, J. (1988) *Nucleic Acids Res.* **16**, 10953–10971.
36. Kozak, M. (1988) *J. Cell Biol.* **107**, 1–7.
37. Kim, S-J., Glick, A., Sporn, M.B. and Roberts, A.B. (1989) *J. Biol. Chem.* **264**, 402–408.
38. Wen, L., Huang, J-K. and Blackshear, P.J. (1989) *J. Biol. Chem.* **264**, 9016–9021.
39. Propst, F., Rosenberg, M.P., Iyer, A., Kaul, K. and Vande Woude, G.F. (1987) *Mol. Cell. Biol.* **7**, 1629–1637.
40. Rao, C.D., Pech, M., Robbins, K.C. and Aaronson, S.A. (1988) *Mol. Cell. Biol.* **8**, 284–292.
41. Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P. (1980) *Science* **209**, 1406–1414.
42. Dynan, W.S. (1986) *TIG* **2**, 196–197.
43. Kiss, I., Deák, F., Holloway, R.G.,Jr., Delius, H., Mebust, K.A., Frimberger, E., Argraves, W.S., Tsonis, P.A., Winterbottom, N. and Goetinck, P.F. (1989) *J. Biol. Chem.* **264**, 8126–8134.
44. Kohno, K., Sullivan, M. and Yamada, Y. (1985) *J. Biol. Chem.* **260**, 4441–4447.
45. Reynolds, G.A., Goldstein, J.L. and Brown, M.S. (1985) *J. Biol. Chem.* **260**, 10369–10377.
46. Peralta, E.G., Winslow, J.W., Peterson, G.L., Smith, D.H., Ashkenazi, A., Ramachandran, J., Schimerlik, M.I. and Capon, D.J. (1987) *Science* **236**, 600–605.
47. Luo, X., Park, K., Lopez-Casillas, F. and Kim, K. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 4042–4046.
48. Nakajima, H., Yamasaki, T., Noguchi, T., Tanaka, T., Kono, N. and Tarui, S. (1990) *Biochem. Biophys. Res. Commun.* **166**, 637–641.
49. Kozak, M. (1989) *Mol. Cell. Biol.* **9**, 5134–5142.
50. Kozak, M. (1989) *J. Cell. Biol.* **108**, 229–241.
51. Grens, A. and Scheffler, I.E. (1990) *J. Biol. Chem.* **265**, 11810–11816.
52. Manzella, J.M. and Blackshear, P.J. (1990) *J. Biol. Chem.* **265**, 11817–11822.
53. Leibold, E.A. and Munro, H.N. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2171–2175.
54. Rouault, T.A., Hentze, M.W., Caughman, S.W., Harford, J.B. and Klausner, R.D. (1988) *Science* **241**, 1207–1210.
55. del Angel, R.M., Papavassiliou, A.G., Fernández-Tomás, C., Silverstein, S.J. and Racaniello, V.R. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 8299–8303.
56. Andino, R., Rieckhof, G.E. and Baltimore, D. (1990) *Cell* **63**, 369–380.
57. Kozak, M. (1987) *Nucleic Acids Res.* **15**, 8125–8148.