Glycoconjugate Journal (1994) **11**:353–362

Galectin-3: differential accumulation of distinct mRNAs in serum-stimulated mouse 3T3 fibroblasts

PATRICIA G. VOSS, YEOU-GUANG TSAY and JOHN L. WANG*
Department of Biochemistry, Michigan State University, East Lansing, MI 48824, USA

Received 19 April 1994, revised 17 June 1994

The murine *Galectin-3* gene spans ~12 kb of DNA and contains six exons, with the translation initiation codon located in exon II. On the basis of restriction mapping and sequence analysis of the DNA upstream of exon II, primer extension assays, rapid amplification of cDNA ends, and ribonuclease protection assays were designed and carried out to determine the initiation site of transcription and the sequence of exon I. The results revealed at least two transcription initiation sites (α and δ), each of which appears to be specifically associated with the use of alternative donor splice sites, resulting in distinct mRNA species. Type I message initiates at transcription start site δ, uses splice donor site No. 2, retaining a 27 bp sequence, whereas type II message initiates at transcription start site α, uses splice donor site No. 1, resulting in the loss of the 27 bp sequence. Primer extension assays carried out with mRNA isolated from 3T3 fibroblasts at various times after serum stimulation indicate that while the type II message varies in level only a little over the first 20 h, there is dramatic accumulation of the type I message, which peaks at 16 h post mitogen addition.

Keywords: lectin; carbohydrate-binding protein; gene structure; alternative splicing; mitogen stimulation

Introduction

Galectin-3 [1] is the new name for the galactose/lactose-specific lectin previously known under a number of different designations, including Carbohydrate Binding Protein 35 (CBP35) [2], Mac-2 [3], IgE-binding protein [4], CBP30 [5], L-29 [6], and L-34 [7]. Because the present paper represents the first communication from this laboratory using the newly adopted nomenclature, we will use the designation *Galectin-3* when we refer to the gene/protein in general, assuming that studies carried out on the gene/protein under any one of the above names is applicable to all of them. There are instances, however, in which the specific molecule used by one laboratory is slightly (but of significance) different from the corresponding molecule of another laboratory (e.g. the cDNAs reported for murine CBP35, Mac-2, and L-34 are of different lengths). In this case, we will use the designation, old name/Galectin-3 (e.g. CBP35/Galectin-3), to highlight the specific source of the molecule.

In mouse 3T3 fibroblasts, the predominant portion of Galectin-3 was found to be intracellular by immunofluorescence staining of fixed and permeabilized cells [8]. There was prominent labelling of the cell nucleus and variable staining of the cytoplasm. Subcellular fractionation studies indicated that the protein was associated with ribonucleoprotein complexes involved in the processing of pre-mRNA molecules [9]. The distribution of the lectin between the nucleus and the cytoplasm was dependent on the proliferation state of the cells: in quiescent cultures of 3T3 fibroblasts, galectin-3 was found primarily in the cytoplasm, whereas in proliferating cultures, it was located predominantly in the nucleus [10].

Galectin-3 is an immediate-early gene in mouse 3T3 cells [11]. When serum-starved, quiescent cultures of 3T3 cells were stimulated by the addition of serum, there was an early rise in the transcription rate of the *Galectin-3* gene. This increase in transcription was not dependent on *de novo* protein synthesis, inasmuch as it was not inhibited by cycloheximide. On Northern blots, there was an increase in the level of the Galectin-3 mRNA, which exhibited a peak at about 3 h. This was followed by a transient decrease and then a second rise, leading up to a ~six-fold elevation over the next 10 h. Such a bimodal kinetic pattern suggests that there may be two regulatory events governing the expression of the gene following mitogen addition.

Although a single band (~1.3 kb) was detected on Northern blots using the cDNA for Galectin-3 as a probe [3, 11–13], a number of lines of evidence suggest that there may be more than one mRNA species. First, two types of cDNA, that differed by the presence (type I) or absence (type II) of 27 bp near the 5′-end, were identified during the initial cloning of Mac-2/Galectin-3 [3]. S1 nuclease analysis

* To whom correspondence should be addressed.

indicated that both types I and II cDNAs represent complementary copies of authentic cytoplasmic messages. Second, sequence analyses of genomic clones of Galectin-3 revealed that the basis for the two types of mRNA was alternative splicing and that the 27 bp insert (in type I) represented the 3′ portion of exon I between two alternative splice donor sites [14, 15]. Finally, reverse transcription of poly A-containing RNA from a number of cell lines and tissues, followed by polymerase chain reaction (PCR) amplification, revealed two DNA products derived from the use of the alternative donor splice sites [14]. We now show that during the stimulation of quiescent 3T3 fibroblasts, the bimodal pattern observed for the increase in Galectin-3 mRNA is also reflected by the differential accumulation of distinct mRNA species, as revealed by primer extension analysis. The use of alternative sites for the initiation of transcription, coupled with the use of alternative donor splice sites, contribute to the regulation of the types and levels of Galectin-3 mRNA species.

Materials and methods

Screening of the genomic libraries

The cDNA for CBP35/Galectin-3 [2] was labelled with [^{32}P]dCTP using the random priming kit (Boehringer, Mannheim) and the method of Feinberg and Vogelstein [16]. The specific activity of the labelled cDNA was $\sim 10^8$ cpm per μg DNA. This probe was used to screen a λ EMBL4 library derived from mouse liver chromosomal DNA (Clonetech Laboratory) using procedures described previously [17]. Three positive clones were identified. Since these three clones appeared to have the same inserts, only one clone, designated as λG_3, was characterized by restriction mapping and DNA sequence analysis (see Fig. 1A).

A 1.7 kb *Pst* I-*Sal* I fragment at the left end of the λG_3 clone (Fig. 1A) was isolated and labelled with [^{32}P]dCTP as described above. This was then used as a hybridization probe to screen a λ FIXII library, derived from chromosomal DNA of NIH 3T3 cells (Stratagene). The λGH clone was identified and isolated from this screening. The same λ FIXII NIH 3T3 cell genomic library was sequentially screened with three labelled probes: (a) a 1.3 kb *Hind* III-*Hind* III fragment from the λGH clone (Fig. 1A); (b) the 4.6 kb *Not* I-*Hind* III fragment at the 5′-end of the λGH clone (Fig. 1A); and (c) the WJ20 oligonucleotide (Fig. 1B). One clone, λ15A, was positive for all three hybridizations (Fig. 1A).

Cloning and sequencing

The 6.4 kb *Sal* I-*Eco* RI fragment of λG_3 clone (Fig. 1A) was first cloned into the PUC-18 plasmid. After digestion with *Pst* I and *Hind III*, the individual fragments were subcloned into the sequencing vectors M13mp18 and M13mp19 [18]. The 1.7 kb *Sal*-I-*Pst* I fragment and parts of the 2.5 kb *Pst* I-*Hind* III fragment were sequenced using the dideoxy method [19] (Sequenase Kit, United States Biochemical Co.).

The 6.4 kb *Eco* RI-*Eco* RI fragment and the 6.7 kb *Eco* RI-*Sal* I fragment from the λGH clone (Fig. 1A) were cloned into the PUC-18 plasmid. After *Hind* III digestion, these fragments were further subcloned into M13mp18 and M13mp19 vectors [18]. Both single-stranded and double-stranded DNA sequencing analyses were performed with the Sequenase Kit. Eight oligonucleotide primers, synthesized on the basis of the CBP35/Galectin-3 cDNA sequence [2], were used in carrying out the sequencing reactions.

Southern blot analysis of restriction digests of the λ15A clone yielded a 2.3 kb *Eco* RI-*Eco* RI fragment (Fig. 1A) that hybridized to the oligonucleotide probes WJ20, WJ17, and WJ26 (Fig. 1B). This fragment was subcloned into the pBluescript-KS$^+$ vector (Stratagene) and nested deletion mutants were made at the 3′-end of the insert using exonuclease III and S1 nuclease [20]. Sequencing was performed using the dideoxy method [19]. One of these deletion mutants, designated NB-3i, spanned the region from the *Eco* RI restriction site ($\sim$2 kb upstream of exon I) through all of exon I and terminated at the T residue highlighted by an asterisk just inside intron I (Fig. 1A and 1B). This deletion mutant was chosen as the template for riboprobe synthesis to be used in the ribonuclease (RNase) protection assays (see below).

Oligonucleotide synthesis

All oligonucleotides were synthesized on an Applied Biosystems DNA Synthesizer (Model 380B) in the Macromolecular Structure Facility at Michigan State University. Besides the sequencing primers, the following oligonucleotides were synthesized and used for screening genomic fragments containing exon I, for primer extension analysis and rapid amplification of cDNA ends (RACE) procedures: (a) WJ14 (22-mer, 5′ CGAAAAGCTGTCTGCCATTTTC 3′) is complementary to the sequence of exon II (Fig. 1B) of murine *Galectin-3*. (b) WJ17 (24-mer, 5′ GATTAGTGC-TCCACCCGCCGCTCC 3′) is complementary to the sequence of exon I in the 5′-untranslated region (Fig. 1B) of murine *Galectin-3*, as reported in the cDNA sequence [3, 7] and in the genomic sequence [14, 15]. (c) WJ20 (22-mer, 5′ CCGCCGCTCCGCTAGCTGTCAG 3′) is complementary to the sequence of exon I in the 5′-untranslated region [Fig. 1B], as reported for the Mac-2/Galectin-3 cDNA [3] and the murine *Galectin-3* genomic sequence [14, 15]. These sequences were identical to each other over this region but were different from that reported for the L-34/Galectin-3 cDNA [7], which contained an extra base, G. (d) WJ26 (24-mer, 5′ GTAGTGCTCTCTGTGCCG-CTCACC 3′) is complementary to the sequence of the 27 bp insert in exon I in the 5′-untranslated region

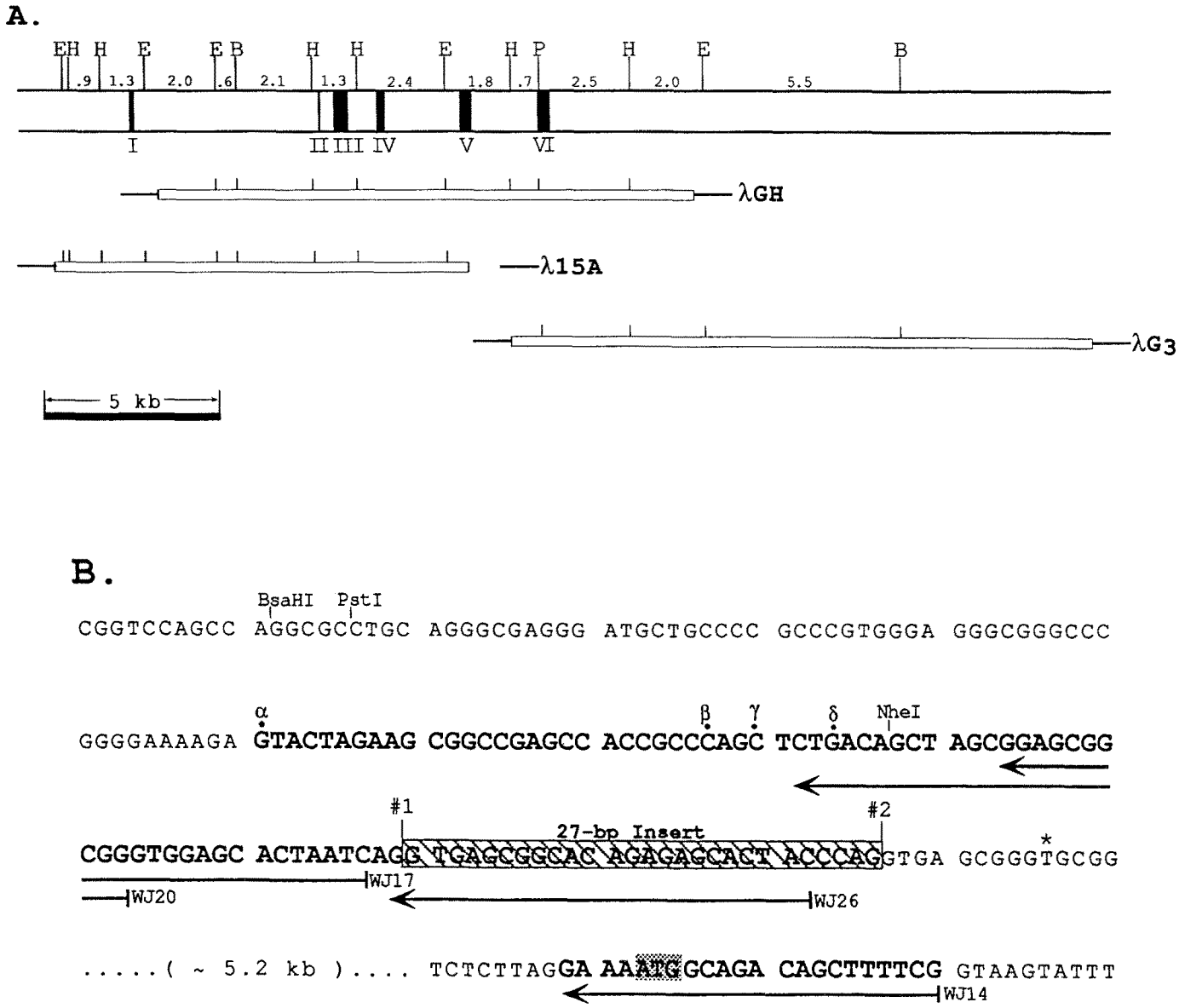

Figure 1. (A) Organization and restriction map of the murine *Galectin-3* gene. Exons are indicated by dark lines/boxes and are numbered I–VI. Restriction enzyme sites are indicated by the letters: E, *Eco* RI; H, *Hind* III; B, *Bam* HI; and P, *Pst I*. The numbers between restriction enzyme sites indicate the size of the restriction fragment. λGH, λ15A, and λG_3 are the genomic clones from which the information on the organization and restriction map of the gene are derived. (B) Detailed structure of and relationship between exons I and II. The bases in exons are in bold lettering. Initiation sites of transcription deduced from previously published [3, 15] and the present data are indicated by α, β, γ, and δ above the respective bases. Two alternative splice donor sites are labelled #1 and #2. The hatched rectangle highlights a 27 bp sequence that is either retained or spliced out, depending on the use of splice donor sites #2 or #1, respectively. The stippled ATG highlights the translation initiation codon in exon II. *Bsa* HI, *Pst* I, and *Nhe* I indicate the position of the restriction enzyme cleavage sites. The riboprobe used in the ribonuclease protection assay spans the genomic DNA from the *Bsa* HI site to the T residue highlighted by an asterisk just inside intron I. The length and 5′ → 3′ direction of the oligonucleotide probes WJ14, WJ17, WJ20, and WJ26 are indicated by the arrows.

(Fig. 1B), as reported for Mac-2/Galectin-3 cDNA [3] and the murine *Galectin-3* genomic sequence [14, 15]. (e) dT-adapter (35-mer, 5′ GACTCGAGTCGACAT-CGA$(T)_{17}$ 3′) and a 17-mer adapter (5′ GACTCGAGTCG-ACATCG 3′), both of which were used for the RACE procedure.

Primer extension analysis

The source and culture conditions for Swiss mouse 3T3 fibroblasts have been described previously [11]. Non-synchronized populations of 3T3 cells were seeded at 5×10^3 cells per cm^2 and harvested after 48 h of culture. Cultures used for synchronization were seeded at 5×10^3

cells per cm^2. After 24 h, the medium was changed to Dulbecco modified Eagle's medium containing 0.2% calf serum for 48 h. Cells were stimulated by the readdition of serum (10%) and harvested at various time points (1, 6, 16, and 21 h).

Cytoplasmic RNA was isolated following the protocol described [20], except that water-saturated phenol heated to 65 °C was used to extract the RNA. The concentration of RNA was determined by absorbance readings at 260 and 280 nm. The intactness of the RNA preparations was checked by ethidium bromide staining of the RNA separated on a formaldehyde-0.8% agarose gel.

The primer oligonucleotide, WJ14, was labelled at the 5′-end using T4 polynucleotide kinase and γ-[^{32}P]ATP (specific activity, $\sim 2.2 \times 10^7$ cpm per µg). An aliquot of this primer (10^5 cpm) was hybridized to 40 µg of cytoplasmic RNA in 15 µl of hybridization buffer (0.15 M NaCl, 0.01 M Tris, pH 8.3, 1 mM EDTA) for 1.5 h at 65 °C. After cooling to room temperature, 30 U of avian myeloblastosis virus reverse transcriptase (Life Sciences, Inc.) in 30 µl of reaction buffer (20 mM Tris, pH 8.3, 10 mM $MgCl_2$, 6 mM dithiothreitol, 150 µg per ml actinomycin D, and 0.3 mM each of the dNTPs) were added. The reaction was allowed to proceed for 1 h at 42 °C, after which it was diluted with 105 µl of 10 mM Tris, pH 8, 1 mM EDTA, and then treated with RNase A (2 mg ml^{-1}; Sigma) at 37 °C for 15 min. The samples were then phenol extracted, ethanol precipitated, and the products of the primer extension reaction were analysed on a 9% acrylamide-7 M urea gel. After autoradiography, the intensity of the radioactive bands was quantitated on a BioImage Analyzer. The lengths of the primer-extended fragments were estimated by comparison with an unrelated DNA sequencing reaction and a ladder generated by ^{32}P-end labelling fragments from an *Msp* I digest of pBr322 DNA.

Rapid amplification of cDNA ends (RACE)

The protocol of Frohman [21] was used to amplify the 5′-end of the Galectin-3 mRNA. Poly A-containing RNA was isolated from non-synchronized 3T3 cells using the Promega Poly A-Tract 1000 isolation kit. Using 1.5 µg of this poly A-containing RNA as template, the oligonucleotide, WJ14, complementary to exon II (Fig. 1B), was used to prime cDNA synthesis, extending to the 5′-end of the mRNA. Using terminal transferase (10 U; United States Biochemical Co.), a poly A-tail was added to the 3′-end of the first cDNA strand. A first round of amplification of the extended cDNA was carried out using the dT-adapter oligonucleotide, which was complementary to the poly A-tail and which contained sequences for restriction sites *Xho* I, *Sal* I and *Cla* I, as the 5′-end primer and the WJ14 oligonucleotide (same primer used to produce the cDNA) as the 3′-end primer. The conditions for amplification were: 25 cycles, melting at 95 °C for 40 s, annealing at 55 °C for 1 min, and extending at 72 °C for 3 min. A second round of amplification was carried out with 1–5 µl of the first amplification per reaction, using nested primers: the 17-mer adapter oligonucleotide as the 5′-end primer and either WJ17 or WJ20 (Fig. 1B) as the 3′-end primer. Amplification was carried out for 40 cycles under the same melting, annealing, and extension conditions as above. The PCR products were resolved on 1.5% agarose gels and revealed by ethidium bromide staining. The gels were then subjected to Southern blotting (22) with a ^{32}P-labelled genomic fragment bounded by *Pst* I and *Nhe* I sites (Fig. 1B).

Ribonuclease protection assay

A riboprobe was synthesized from NB-3i, described above as one of the nested deletion mutants generated from the 2.3 kb *Eco* RI-*Eco* RI fragment of the λ15A genomic clone (Fig. 1A). The template was linearized by digestion with *Bsa* HI, and the α-[^{32}P]CTP-labelled riboprobe was transcribed using the T7 RNA polymerase promoter. The resulting riboprobe starts at the *Bsa* HI restriction site, covering all of exon I, and terminates at the T residue (highlighted by asterisk in Fig. 1B) just inside intron I. Poly A-containing RNA (2 µg) from non-synchronized 3T3 cells was hybridized with the riboprobe (10^6 cpm) at 35 °C for 14 h. The reaction mixture was then digested with 1 U of RNase ONE (Promega) at 37 °C for 60 min. The size of the protected fragment(s) was analysed by electrophoresis on a 6% denaturing polyacrylamide gel.

Genomic Southern blot analyses

Mouse liver nuclei were isolated from a homogenate of 15 g of liver tissue by centrifugation at 500 × **g** for 5 min. The nuclei was suspended in 10 mM Tris (pH 8.0) containing 0.14 M NaCl, 1.5 mM $MgCl_2$, 0.5% sodium dodecyl sulfate and 100 µg ml^{-1} proteinase K at 37 °C for 16 h. DNA was isolated by ethanol precipitation and purified by CsCl gradient centrifugation. The DNA was then digested with *Eco* RI, *Hind* III and *Bam* HI. The initial digestion used 20 U enzyme per µg of DNA and was carried out for 20 h. An additional 10 U of enzyme per µg of DNA was added and incubated for another 4 h. The digested DNA was separated by agarose gel electrophoresis (0.8%), subjected to Southern blotting [22] using ^{32}P-labelled CBP35/Galectin-3 cDNA (10^8 cpm per µg) as a hybridization probe.

Results and discussion

Organization of the Galectin-3 *gene*

A λ EMBL4 library derived from mouse liver chromosomal DNA was screened using the cDNA for CBP35/Galectin-3 [2] as a hybridization probe. Of the 10^6 plaques screened, three gave reproducible positive signals. One of these, designated as λG_3, was further cloned by lower density rescreening and plaque purification. Digestion of λG_3 with *Pst* I and sequence analysis in both directions revealed the

single *Pst* I site (Fig. 1A) in the nucleotide sequence of the cDNA [2]. Comparison of the genomic sequence with the cDNA sequence defined exon VI, corresponding to the 3′-end of the cDNA clone. The 1.7 kb fragment, derived from the 'left' end of λG_3 by *Pst* I digestion (Fig. 1A), was used as a hybridization probe to search for other genomic clones from the λ FIXII library of NIH 3T3 cell chromosomal DNA. Clone λGH was selected from such a screening. Finally, clone λ15A was obtained by screening the same λ FIXII library using fragments derived from the 5′-end of λGH and using the oligonucleotide probe WJ20 (Fig. 1B).

Restriction mapping and nucleotide sequence analysis of these genomic clones defined the exons and exon-intron boundaries of the murine *Galectin-3* gene. With the exception of minor differences and points of clarification to be detailed below, the organization of this gene, as depicted in Fig. 1A, agrees with those reported by Gritzmacher *et al.* [14] and by Rosenberg *et al.* [15]. The gene spans ~12 kb of genomic DNA and contains six exons. The Met initiation codon is located in exon II (Fig. 1B), characterized by the consensus initiation sequence A/G NN-ATG-G [23]. Comparison of the nucleotide sequence of the cDNA clone for CBP35/Galectin-3 [2] with the genomic sequence revealed that the CBP35/Galectin-3 cDNA clone coded for the entire polypeptide chain except for the NH_2-terminal Met residue.

Such a comparison also showed that certain amino acid residues derived from the genomic sequence differed from the corresponding residues derived from the cDNA sequence [2]. These differences were also found in a comparison of the sequences of the CBP35/Galectin-3 cDNA [2] with the cDNA clones for mouse L-34/Galectin-3 [7] and Mac-2/Galectin-3 [3]. Therefore, we have resequenced the regions in the CBP35/Galectin-3 cDNA where there was a discrepancy, with the following results. (In this description, the numbering system for the amino acid residues starts with translation initiation methionine as residue 1; on this basis, the first residue reported for the CBP35/Galectin-3 cDNA becomes amino acid residue 2.) First, there was an error due to sequencing in CBP35/Galectin-3 cDNA for residues 110–112, and the sequence for these positions should be Ser-Ala-Pro, consistent with the reported sequences of L-34/Mac-2/Galectin-3 [3, 7] and with the murine genomic sequences [14, 15]. The consequence of this revision is that the polypeptide contains only one cysteine (at residue 186). Second, resequencing of the CBP35/Galectin-3 cDNA has confirmed the amino acid assignments previously reported [2]: Arg at residue 2, Gln at residue 92, and Pro at residue 93. Therefore, the differences with the other sequences at the corresponding positions remain as confirmed differences, which might be due to cloning artefacts or polymorphism. The significance of these differences is exemplified in predicting the isoelectric point(s) of recombinant polypeptides produced from expression vectors containing the CBP35/Galectin-3 cDNA (e.g. Arg instead of Ala at residue 2) [24].

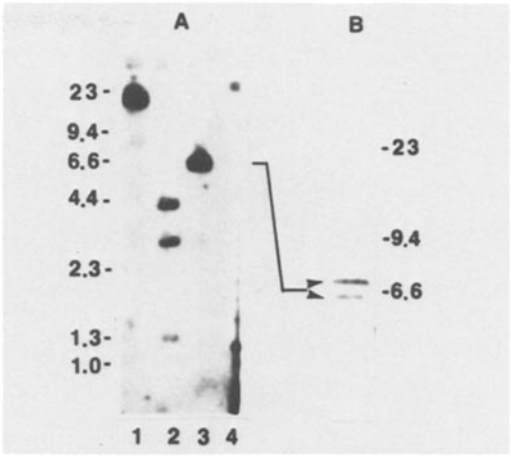

Figure 2. Southern blot hybridization of mouse genomic DNA with the cDNA clone for Galectin-3. Chromosomal DNA was isolated from mouse liver nuclei. The DNA was digested with restriction enzymes, separated on agarose gels (0.8%) and subjected to Southern blot analysis with ^{32}P-labelled CBP35/Galectin-3 cDNA (10^8 cpm per μg). Panel A: lane 1, *Bam* HI digest; lane 2, *Hind* III digest; lane 3, *Eco* RI digest; and lane 4, undigested DNA. In panel B, the same *Eco* RI digest sample that was used for lane 3 of panel A was electrophoresed for a longer time to better resolve the bands of a doublet. The numbers on the left and right indicate the positions of migration of size markers.

DNA isolated from nuclei of mouse liver was digested with several restriction enzymes, separated by gel electrophoresis and then subjected to Southern blot analysis with the cDNA for CBP35/Galectin-3 [2], which contained sequences corresponding to Exons II-VI (Fig. 1A). A single fragment (19–20 kb) was observed after *Bam* HI digestion (Fig. 2A, lane 1). *Hind* III digestion yielded fragments of 4.2, 3.2 and 1.3 kb (Fig. 2A, lane 2). Although the *Eco* RI-digested material yielded a broad band in the experiment shown in Fig. 2A (lane 3), it actually contained two fragments, which could be resolved into a doublet upon longer electrophoresis of the same material (Fig. 2B). The positions of migration of the bands corresponded to DNA molecules of ~6.4 kb and ~7.0 kb. The lengths of each of these genomic DNA fragments that hybridized with the cDNA probe were in good agreement with those derived from the genomic clones isolated above (Fig. 1A). No other bands could be observed. These results indicate that the *Galectin-3* gene is a single gene in the normal mouse genome.

Identification of the transcription initiation site by primer extension assay

A 22-mer oligonucleotide complementary to the entire exon II, designated as WJ14 (Fig. 1B), was synthesized, labelled with ^{32}P, and was used as the primer for the reverse transcription of mRNA isolated from non-synchronized

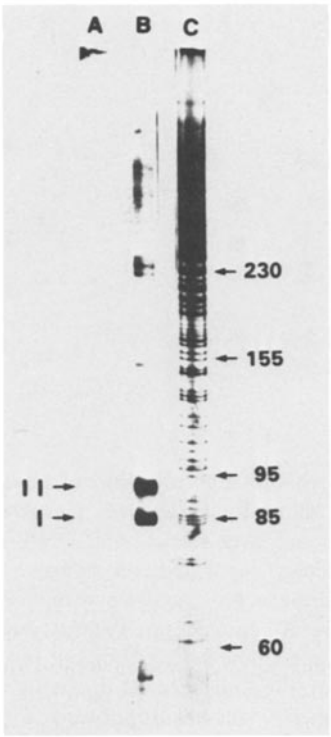

Figure 3. Primer extension assay of RNA isolated from non-synchronized 3T3 cells. The primer, WJ14, is complementary to the sequence of exon II (see Fig. 1B). ^{32}P-Labelled primer (10^5 cpm) was hybridized with cytoplasmic RNA (40 μg), followed by the addition of 30 U of avian myeloblastosis virus reverse transcriptase. The products of the primer extension reaction were analysed on a 9% acrylamide-7 M urea gel. Lane A: Product of the extension reaction using tRNA as a control. Lane B: Product of the extension reaction using cytoplasmic RNA. Lane C: An unrelated DNA sequencing reaction used to count nucleotide residues (numbers shown on the right). The numbers I and II on the left denote the positions of migration of the two observed primer extension products.

3T3 cells. The results showed that the products of the extension reaction yielded two major bands: band I, 85 nucleotides, and band II, 91 nucleotides (Fig. 3, lane B). Control tRNA, carried out in parallel, yielded no extension product (Fig. 3, lane A). We shall discuss these results in light of the available sequence of the murine *Galectin-3* gene [14, 15] and previous studies on PCR amplification of cDNA derived from reverse transcription of cytoplasmic mRNA [3]. This discussion is facilitated using the schematic diagram shown in Fig. 1B.

Band II, 91 nucleotides, is consistent with initiation of transcription at the position labelled α in Fig. 1B. In this message, splice donor site No. 1 (Fig. 1B) is used and the 27 bp insert is spliced out during processing. As a result, the length of the primer-extended product would consist of 69 nucleotides upstream of the 27 bp insert plus 22 nucleotides of the primer oligonucleotide, for a total length of 91 nucleotides. The use of transcription initiation site α (Fig. 1B) is consistent with the results of: (a) anchored PCR analysis of the Mac-2.9 cDNA clone, yielding pBC12, whose 5′-end starts with the sequence GTACTA... and which does not contain the 27 bp insert [3]; and (b) S1 nuclease mapping of the 5′-end of the *Mac-2/Galectin-3* gene after annealing to poly A-containing RNA from J774 macrophages [15]. It appears that the message corresponding to band II (Fig. 3) matches the type II cDNA reported by Rosenberg *et al.* [15].

Band I, 85 nucleotides, is consistent with initiation of transcription at the position labelled δ in Fig. 1B. In this message, splice donor site No. 2 (Fig. 1B) is used and the 27 bp insert is retained after processing. As a result, the length of the primer-extended product would consist of 36 nucleotides upstream of the insert, 27 nucleotides of the insert, plus 22 nucleotides of the primer, for a total length of 85 nucleotides. The use of transcription initiation site δ (Fig. 1B) is consistent with the results of anchored PCR analysis of the Mac-2.16 cDNA clone, yielding pBC7, whose 5′-end starts with the sequence GACAGC... and which contains the 27 bp insert (3). The message corresponding to band I (Fig. 3) matches, therefore, the type I cDNA of Rosenberg *et al.* [15].

There are other possible initiation sites of transcription, designated as positions β and γ in Fig. 1B. Both of these had been identified on the basis of S1 nuclease mapping [15]. If transcripts starting at these sites did not contain the 27 bp insert, the predicted primer extension products would be 62–65 nucleotides long (Fig. 1B). No such products could be observed in the experiment shown in Fig. 3. Therefore, transcripts starting at these sites are presumed to contain the 27 bp insert (type I message). In any case, it is clear that the type I and type II mRNAs differ in size by only a few nucleotides and, therefore, it is not suprising that previous Northern blots using the cDNA for Galectin-3 as a probe yielded a single band of ~1.3 kb [3, 11–13].

The primer extension assay shown in Fig. 3 was carried out with RNA derived from non-synchronized cultures. Although the data indicated that the levels of band I and band II are approximately equal, similar primer extension assays using RNA from synchronized 3T3 cell cultures suggest that there is differential accumulation of band I versus band II (see below), suggesting differential use of initiation site of transcription (site α versus site δ) and/or differential use of splice donor sites (site No. 1 versus site No. 2).

Confirmation by ribonuclease protection and RACE assays

The interpretation of our primer extension assay results suggests that the use of alternative sites of transcription initiation might be coupled to the alternative use of splice donor sites. For example, initiation at site α results in the use of splice donor site No. 1 and loss of the 27 bp insert from the transcript, whereas initiation at site δ results in the use of splice donor site No. 2 and retention of the 27 bp insert in the transcript. This notion was confirmed by a

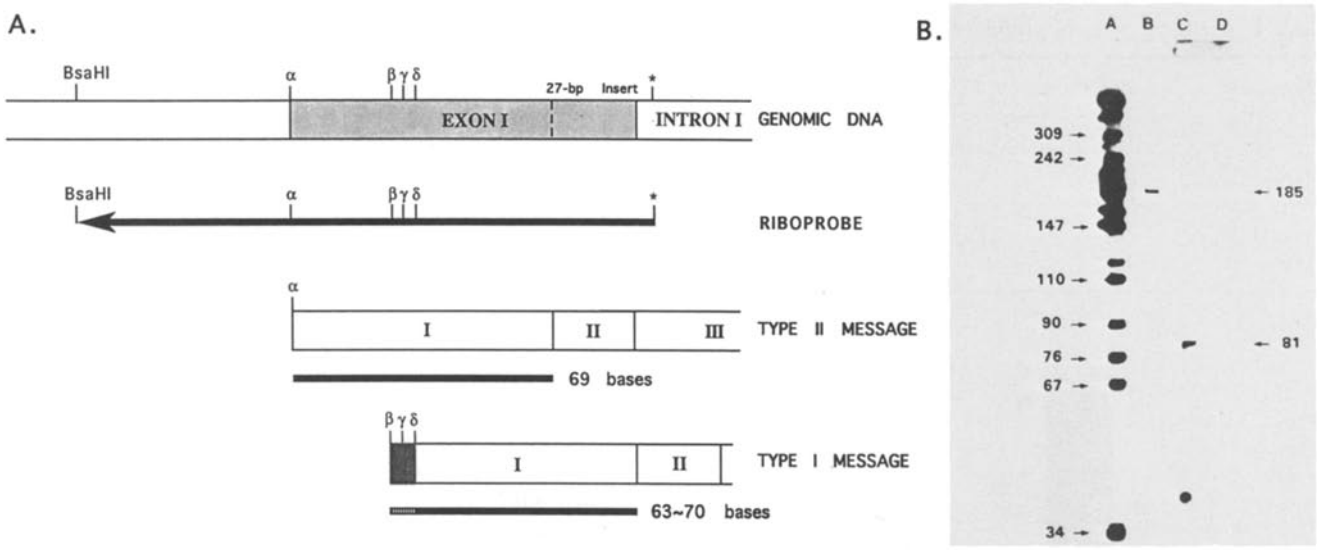

Figure 4. (A) Schematic diagram illustrating the design and predicted results of the ribonuclease protection assay. The relationship between the ends of the riboprobe (*Bsa* HI site to position indicated by asterisk) and the transcriptional initiation sites α, β, γ, and δ, as well as the 27 bp insert are shown. Type II message, which initiates at start site α and which does not contain the 27 bp insert, is expected to protect a fragment of the riboprobe containing 69 bases. Type I message, which initiates at δ (or β or γ) and which contains the 27 bp insert, is expected to protect a fragment of the riboprobe containing 63–70 nucleotides. (B) Ribonuclease protection assay with RNA from non-synchronized 3T3 cells. ^{32}P-Labelled riboprobe (10^6 cpm) was hybridized with poly A-containing RNA (2 µg), followed by digestion with RNase ONE. The protected fragments were analysed on a 6% acrylamide gel. Lane A: Size markers from an *Msp* I digest of pBr322 DNA. Lane B: Undigested riboprobe. Lane C: Protection of riboprobe by poly A-containing mRNA. Lane D: Control reaction containing riboprobe and tRNA. The numbers on the left indicate the sizes of the markers. The numbers on the right highlight the sizes of oligodeoxyribonucleotides whose positions of migration correspond to the undigested riboprobe and the mRNA protected fragment.

RNase protection assay using a riboprobe generated from a genomic fragment of the λ15A clone. This riboprobe starts at the *Bsa* HI restriction site at the 5′-end, includes the α, β, γ, and δ transcription initiation sites, the 27 bp insert, and terminates at the T residue highlighted by an asterisk (Fig. 1B, 4A). The labelled riboprobe was hybridized with poly A-containing mRNA from non-synchronized cultures of 3T3 cells, subjected to RNase digestion, and the protected fragment(s) was analysed on an acrylamide gel system (Fig. 4B).

In looking at the results of this experiment, it should be noted that oligodeoxyribonucleotides migrate faster (by ~10%) than oligoribonucleotides of the same size [25] and that the size markers in lane A of Fig. 4B are oligodeoxyribonucleotides derived from an *Msp* I digest of pBr322. Thus, the undigested riboprobe (~165 nucleotides) migrated to a position corresponding to an oligodeoxyribonucleotide containing ~185 bases (Fig. 4B, lane B). The portion of the riboprobe which was protected from RNase digestion by hybridization to mRNA migrated, apparently as a single band, to a position corresponding to an oligodeoxyribonucleotide containing ~81 bases (Fig. 4B, lane C). This would correspond to about 70 nucleotides in the riboprobe. In parallel controls, tRNA failed to protect any fragment of the riboprobe (Fig. 4B, lane D).

As illustrated in Fig. 4A, type II message (initiation at transcription start site α and using splice donor site No. 1, with resulting loss of the 27 bp insert) would yield a protected riboprobe fragment of 69 nucleotides. Alternatively, type I message (initiation at transcription start sites β, γ, or δ and using splice donor site No. 2, with retention of the 27 bp insert) would also yield a protected fragment of 63–70 nucleotides (Fig. 4A). Thus, the results (Fig. 4B, lane C) are consistent with either one or both of these possibilities. On the other hand, the results (Fig. 4B, lane C) would be inconsistent with: (a) initiation at transcription start site α, with retention of the 27 bp insert (predicted 96 nucleotides in protected fragment); and (b) initiation at transcription start sites β, γ, or δ, with loss of the 27 bp insert (predicted 36–43 nucleotides in the protected fragment).

Additional evidence consistent with the results and conclusions derived from the primer extension and RNase protection assays was obtained using the RACE procedure [21]. First, a cDNA was synthesized based on the Galectin-3 mRNA template and primed by WJ14, which is complementary to exon II (Fig. 1B). This cDNA was tailed with poly A so that a first round of amplification could be carried out using the dT-adapter as the 5′-end primer and WJ14 as the 3′-end primer. A second round of amplification was

Table 1. Predicted sizes of products from RACE procedures using WJ17 and WJ20 primers.

Message	*WJ20-adapter PCR*				*WJ17-adapter PCR*			
	WJ20 primer	*Exon I*	*Poly A tail and adapter*	*Products predicted*	*WJ17 primer*	*Exon I*	*Poly A tail and adapter*	*Products predicted*
Type II	22	31	35	88	24	43	35	102
Type I	22	0	35	57	24	10	35	69

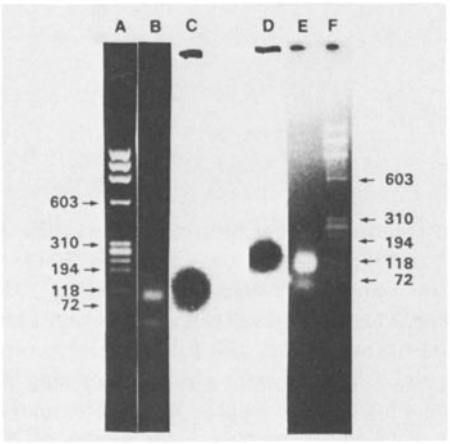

Figure 5. Rapid amplification of cDNA ends using RNA from non-synchronized 3T3 cells. Poly A-containing mRNA (1.5 μg) was reverse-transcribed using oligonucleotide WJ14 as a primer. A poly A tail was added to the 3′-end of the first cDNA, which was then subjected to two rounds of PCR amplification, using dT-adapter and WJ14 as primers in the first round and a 17-mer adapter and WJ17/WJ20 as nested primers in the second round. The PCR products were resolved on 1.5% agarose gels. Lanes A and F: ethidium bromide stained size markers derived from an *Hae* III digest of ϕX174. Lanes B and E: PCR products using the 17-mer adapter with WJ20 and WJ17, respectively, as the nested primers, stained with ethidium bromide. Lanes C and D: Autoradiography of the Southern blot of the gels in lanes B and E, respectively, probed with ^{32}P-labelled *Pst* I-*Nhe* I genomic fragment (Fig. 1B). The numbers on the left and right indicate the sizes of the markers.

performed using nested primers: a 17-mer adapter as the 5′-end primer and either WJ17 or WJ20 (Fig. 1B) as the 3′-end primer.

The results, shown in Fig. 5, are best discussed with the aid of Table 1. Using WJ20-adapter as the nested primers, transcription initiation at site α, use of splice donor site No. 1 and loss of the 27 bp insert (type II message) predicts a PCR product of 88 nucleotides. Transcription initiation at site δ, use of splice donor site No. 2 and retention of the 27 bp insert (type I message) predicts a PCR product of 57 nucleotides. The observed products of the RACE experiment, as revealed by ethidium bromide staining, are consistent with these predictions (Fig. 5, lane B). A *Pst* I-*Nhe* I genomic fragment (Fig. 1B) from clone 15A should have sufficient overlap with type II message to hybridize with the 88 nucleotide PCR product but should not hybridize with the 57 nucleotide product derived from type I message. This was found to be the case (Fig. 5, lane C).

Similar results and conclusions were obtained using the WJ17-adapter as the nested primers (Table 1). Type II message predicts a PCR product of 102 nucleotides, while type I message predicts a PCR of 69 nucleotides. Both of these were observed in the ethidium bromide stained gels (Fig. 5, lane E). Only the 102 nucleotide PCR product derived from type II message hybridized to the *Pst* I-*Nhe* I genomic fragment (Fig. 5, lane D), as would be expected on the basis of the fact that this genomic fragment overlaps with the 5′-end of the type I message by only 4–11 residues (Fig. 1B).

Differential accumulation of distinct transcripts after serum stimulation

Previous Northern blot analysis had shown that when quiescent cultures of 3T3 cells were stimulated by serum, there was an increase in the level of the Galectin-3 mRNA ($\sim$1.3 kb) within 30 min after mitogen addition [11]. The level of the mRNA reached a peak 1–2 h after stimulation and then decreased slightly. A more extended time course study revealed that the early rise ($<$3 h) was followed by a small decline before it increased some six-fold at 21 h. This prompted us to ask whether there was differential accumulation of type I versus type II mRNAs over the bimodal kinetic pattern. Using the oligonucleotide WJ14 as a primer, extension assays were carried out on RNA fractions derived from cells at various times after serum stimulation (Fig. 6). The results showed that while both the 91 nucleotide (type II) band and the 85-nucleotide (type I) band are present throughout the time course, it is the latter that most likely is responsible for the observed alterations in the Galectin-3 mRNA, as detected on Northern blots [11]. Densitometric scanning of the autoradiogram indicated that the intensity of the 91 nucleotide band changed little over the 20 h time course; at the most, there was about a 15% difference between the values at the various time points. In contrast, the 85 nucleotide band showed a slight decrease between 1 and 6 h (Fig. 6, lanes C and E). More dramatic, however, was the increase in the 85 nucleotide band at 16 and 21 h

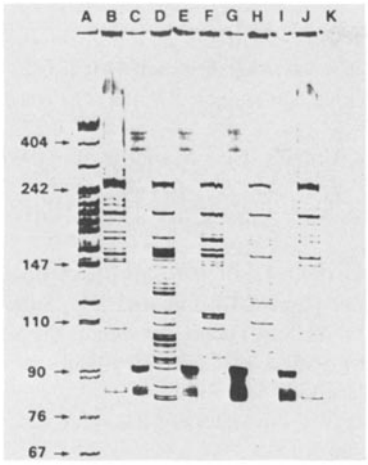

Figure 6. Primer extension assay of RNA isolated from serum-stimulated 3T3 cells. The assay was performed as in Fig. 3. Lane A: Size markers from an *Msp* I digest of pBr322 DNA. Lanes B, D, F, H, and J: Unrelated DNA sequencing reactions used to count nucleotide residues. Lanes C, E, G, and I: Products of the extension reaction using cytoplasmic RNA isolated from 3T3 cells at 1, 6, 16 and 21 h, respectively, post serum addition. Lane K: Product of the extension reaction using tRNA as a control. The numbers on the left indicate the sizes of the markers.

(Fig. 6, lanes G and I). At its peak, the level of the 85 nucleotide (type I) band was about three times that of the 91 nucleotide (type II) band (Fig. 6, lane G). These results indicate that the second peak in the bimodal kinetics of increase in Galectin-3 mRNA following serum stimulation is primarily associated with the accumulation of the 85 nucleotide (type I) message.

At present, the purpose of generating two (or more) distinct mRNA species for Galectin-3 is not clear. Since these messages differ only in exon I and since the ATG translation initiation codon is in exon II, the protein product derived from the two mRNAs is actually the same. The use of alternative initiation sites of transcription (α versus δ) may allow for using different promoters; one promoter may be stronger or promoter usage may be influenced by tissue/cell type, developmental stage or cell cycle phase. In any given cell type, such as the 3T3 fibroblast, both mRNAs are clearly found. A similar conclusion was also obtained in J774 macrophages [15], as well as a variety of other cell lines and tissues [14]. As demonstrated in the time course analysis following serum stimulation of 3T3 cells, the relative levels of the two mRNA species are clearly cell cycle-dependent.

Although the promoter region of the murine *Galectin-3* gene has yet to be defined by functional assays, sequence gazing reveals a GC-rich region containing a Sp1 transcription factor binding site [15]. No TATA-like box could be found. In this respect, the promoter region of the murine *Galectin-3* gene is similar to the human androgen receptor promoter, which is also TATA-less [26]. Two initiation sites of transcripion, located in a 13 bp region, have been identified and one of these is regulated by a single Sp1 binding sequence in a GC-box. It would be of interest to test whether the GC-rich region of the murine *Galectin-3* gene actually regulates transcription from start site α (and/or δ). It would also be of interest to determine whether the sequence between transcription start sites α and δ (33 bp) plays any promoter role in the initiation of transcription of type I messages.

The use of alternative sites of transcription initiation may have other consequences on parameters governing the level of expression. For example, type I and type II messages differ in sequence at two places: (a) type II mRNA has the 33 bp sequence between start site α and start site δ that is missing in type I mRNA; and (b) type II mRNA lacks the 27 bp insert present in type I mRNA. One or both of these may play a role in transcriptional pausing. Transcriptional pausing has been shown to occur downstream of the transcription initiation site for the *c-myc* gene [27, 28]. In this connection, it is interesting to cite several similarities between the *c-myc* and *Galectin-3* genes. First, both can use alternative initiation sites of transcription [15, 29]. Second, neither gene contains a coding sequence in exon I (ATG translation start codon in exon II) [14, 15, 29]. Third, both the *c-myc* and *Galectin-3* genes are immediate early genes in 3T3 fibroblasts. They are activated by serum stimulation and this activation is independent of *de novo* protein synthesis [11, 30, 31]. In both cases, protein synthesis inhibitors such as cycloheximide and anisomycin in fact 'super induce' the expression of the genes (mRNA level for a given gene is higher in the presence of the inhibitor than in its absence). Finally, both exhibit a bimodal kinetic pattern for the level of mRNA as a function of time after serum stimulation [11, 32, 33].

The differences in sequence in the 5′-untranslated regions of type I versus type II mRNA for Galectin-3 may also have important consequences on the stability of the respective messages. Therefore, the data on the differential accumulation of distinct mRNA species as a function of time after serum stimulation of quiescent 3T3 cells may reflect not only the activation of different transcriptional start sites but also the half lives of the transcripts. It remains to be demonstrated whether the sequences in the first exon can serve directly as stabilizing/destabilizing sequences (e.g. recognition by RNases) or as structural determinants that modify the recognition of destabilizing elements by RNases.

Acknowledgements

We thank Drs Shizhe Jia, Mark Sutton, and Quan Sun for screening, identification, and preliminary mapping of the genomic clones and for carrying out the genomic Southern hybridization experiments. We thank Mrs Linda Lang for her help in the preparation of the manuscript. This work was supported by grant GM38740 from the National Institutes of Health.

References

1. Barondes SH, Castronovo V, Cooper DNW, Cummings RD, Drickamer K, Feizi T, Gitt MA, Hirabayashi J, Hughes C, Kasai K, Leffler H, Liu F-T, Lotan R, Mercurio AM, Monsigny M, Pillai S, Porirer F, Raz A, Rigby PWJ, Rini JM, Wang, JL (1994) *Cell* **76**:597–8.
2. Jia S, Wang JL (1988) *J Biol Chem* **263**:6009–11.
3. Cherayil BJ, Weiner SJ, Pillai S (1989) *J Exp Med* **170**:1959–72.
4. Albrandt K, Orida NK, Liu F-T (1987) *Proc Natl Acad Sci USA* **84**:6859–63.
5. Sato S, Burdett I, Hughes RC (1993) *Exp Cell Res* **207**:8–18.
6. Leffler H, Masiarz FR, Barondes SH (1989) *Biochemistry* **28**:9222–9.
7. Raz A, Pazerini G, Carmi P (1989) *Cancer Res* **49**:3489–93.
8. Moutsatsos IK, Davis JM, Wang JL (1986) *J Cell Biol* **102**:477–83.
9. Wang JL, Werner EA, Laing JG, Patterson RJ (1992) *Trans Biochem Soc* **20**:269–74.
10. Moutsatsos IK, Wade M, Schindler M, Wang JL (1987) *Proc Natl Acad Sci USA* **84**:6452–6.
11. Agrwal N, Wang JL, Voss PG (1989) *J Biol Chem* **264**:17236–42.
12. Liu F-T, Albrandt K, Mendel E, Kulczycki A, Orida NK (1985) *Proc Natl Acad Sci USA* **82**:4100–4.
13. Raz A, Carmi P, Pazerini G (1988) *Cancer Res* **48**:645–9.
14. Gritzmacher CA, Mehl VS, Liu F-T (1992) *Biochemistry* **31**:9533–8.
15. Rosenberg IM, Iyer R, Cherayil B, Chiodino C, Pillai S (1993) *J Biol Chem* **268**:12393–400.
16. Feinberg AP, Vogelstein B (1983) *Anal Biochem* **132**:6–13.
17. Woo SLC (1979) *Methods Enzymol* **68**:389–401.
18. Messing J (1983) *Methods Enzymol* **101**:20–78.
19. Sanger F, Nicklen S, Coulson AR (1977) *Proc Natl Acad Sci USA* **74**:5463–7.
20. Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edition. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.
21. Frohman MA (1990) In *PCR Protocols, A Guide to Methods and Applications* (Innis MA, Gelfand DH, Sninsky JJ, White TJ, eds) pp. 28–38. San Diego: Academic Press.
22. Southern EM (1975) *J Mol Biol* **98**:503–8.
23. Kozak M (1986) *Cell* **44**:283–92.
24. Cowles EA, Agrwal N, Anderson RL, Wang JL (1990) *J Biol Chem* **265**:17706–12.
25. Ausebel FM, Brent R, Kinstron RE, Moore DD, Seidman JG, Smith JA, Struhl K (1993) *Current Protocols of Molecular Biology*. Supplement 21, pp. 4.7.1–4.7.8. New York: Greene Publishing Associates and John Wiley & Sons.
26. Faber PW, van Rooij HCJ, Schipper HJ, Brinkmann AO, Trapman J (1993) *J Biol Chem* **268**:9296–301.
27. Bentley DA, Groudine M (1988) *Cell* **53**:245–56.
28. Kerppola TK, Kane CM (1988) *Mol Cell Biol* **8**:4389–94.
29. Battey J, Moulding C, Taub R, Murphy W, Stewart T, Potter H, Lenoir G, Leder P (1983) *Cell* **34**:779–87.
30. Lau LF, Nathans D (1985) *EMBO J* **4**:3145–51.
31. Greenberg ME, Hermanowski AL, Ziff EB (1986) *Mol Cell Biol* **6**:1050–7.
32. Müller R, Bravo R, Burckhardt J, Curran T (1984) *Nature* **312**:716–20.
33. Lau LF, Nathans D (1987) *Proc Natl Acad Sci USA* **84**:1182–6.