Structural Characterization and Alternate Splicing of the Gene Encoding the Preadipocyte EGF-like Protein Pref-1†

Cynthia M. Smas, Daniel Green, and Hei Sook Sul*

Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts 02115

Received January 14, 1994; Revised Manuscript Received May 5, 1994®

ABSTRACT: Preadipocyte factor 1 (pref-1), a member of the EGF-like protein family, is a transmembrane protein with six tandem EGF-like repeats in the putative extracellular domain. Expression of pref-1 is abolished during the in vitro differentiation of 3T3-L1 preadipocytes to adipocytes, and constitutive expression of pref-1 in preadipocytes inhibits their differentiation [Smas, C. M., & Sul, H. S. (1993) *Cell 73*, 725–734]. In the present studies, we have isolated and characterized genomic clones for pref-1 and have identified multiple pref-1 transcripts generated by alternate splicing. The pref-1 gene consists of five exons and four introns spanning approximately 7.3 kb. By primer extension analysis, the transcription start site was determined to be 169 bp upstream from the translation initiation codon. We have identified functional promoter sequences by transient transfection using a 2.1 kb fragment of the pref-1 5′ flanking region linked to a luciferase gene; the pref-1–luciferase fusion gene construct gave 20-fold higher promoter activity as compared to the promoterless vector. Analysis of exon–intron junctions reveals that unlike the majority of the mammalian EGF-like genes, EGF-like repeats of pref-1 are not encoded by discrete exons. Through RT–PCR and the isolation and analysis of multiple pref-1 cDNA clones, we have identified, in addition to full-length pref-1, five alternately spliced forms with various in-frame deletions of all or a part of the sixth EGF-like repeat, juxtamembrane, and predicted transmembrane domains. We conclude, by comparing cDNA and genomic sequences, that all of the multiple forms of pref-1 transcript have in-frame deletions generated by alternate splicing within exon 5.

Preadipocyte factor 1 (pref-1)[1] is a newly described member of the epidermal growth factor-like family of proteins that we have cloned from 3T3-L1 preadipocytes (Smas & Sul, 1993). Analysis of pref-1 cDNA sequence predicts a transmembrane protein containing six tandem epidermal growth factor-like (EGF) repeats in the putative extracellular region. This 35–40 amino acid motif, first identified in EGF and characterized by 6 cysteines that form 3 disulfide loops (Carpenter & Cohen, 1990), is present in a variety of proteins that function in cell growth, cell adhesion, and differentiation. These include the diffusible growth factors EGF, TGF-α (Massague, 1990), amphiregulin (Plowman et al., 1990), and HB-EGF (Higashiyama et al., 1991). Each of these molecules exhibits high-affinity binding to the EGF receptor, produces mitogenic responses in EGF-sensitive cells, and in addition to the characteristic cysteine spacing in the EGF-like domain has other conserved amino acid residues involved in receptor interaction. In addition to those proteins that act through the EGF receptor, EGF-like domains are also present in a number of molecules that function in protein–protein interaction including extracellular matrix proteins and cell adhesion molecules (Siegelman et al., 1990; Krusius et al., 1987), and the *Drosophila* proteins notch and delta that interact via their EGF-like domains to determine cell fate choice in the neurogenic ectoderm (Rebay et al., 1991; Fehon et al., 1990). On the basis of the predicted structure of the EGF-like repeats of the pref-1 protein, it is unlikely it functions through the EGF receptor; rather, the cysteine spacing and amino acid sequence within the individual EGF-like domains of pref-1 are most similar to those of the *Drosophila* protein delta (Smas & Sul, 1993).

We have previously addressed pref-1 function utilizing 3T3-L1 cells (Green & Kehinde, 1974), a clonal murine adipogenic cell line in which we initially identified the pref-1 cDNA. In vitro adipocyte conversion in these cells is accompanied by significant alterations in the expression of adipocyte-specific proteins (Paulauskis & Sul, 1988; Ntambi et al., 1988; de Herreros & Birnbaum, 1988; Rubin et al., 1978) as well as structural and extracellular matrix components (Aratani & Kitagawa, 1988; Calvo et al., 1991). A variety of growth and hormonal factors (Green & Kehinde, 1975; Smith et al., 1988), some that function as negative regulators (Ignotz & Massague, 1985), affect adipocyte conversion. However, the precise combination of factors required for the initiation of and progression through the adipogenic program and the gene(s) that function as regulatory switches in the process have not been identified. The expression pattern of pref-1 mRNA during adipocyte differentiation is unique. The single 1.7 kb mRNA is abundant in preadipocytes, and it is the only gene whose expression is completely abolished during the differentiation of 3T3-L1 preadipocytes to adipocytes. Likewise, the pref-1 protein, present as multiple discrete forms ranging in size from 45 to 60 kDa, is down-regulated during 3T3-L1 differentiation. Constitutive expression of pref-1 in preadipocytes, thus in effect blocking its down-regulation, drastically inhibits adipocyte conversion (Smas & Sul, 1993) and demonstrates that pref-1 mediates the adipocyte differentiation process. In vivo, pref-1 is detected in a variety of embryonic tissues, but in the adult has only been detected

† Supported in part by National Institutes of Health and American Cancer Society grants.

* To whom correspondence should be addressed at the Department of Nutrition, Harvard School of Public Health, 665 Huntington Ave., Boston, MA 02115. Phone: (617) 432-0938. FAX: (617) 432-2435.

® Abstract published in *Advance ACS Abstracts,* July 1, 1994.

[1] Abbreviations: EGF, epidermal growth factor; TGF, transforming growth factor; HB-EGF, heparin binding epidermal growth factor; pref-1, preadipocyte factor 1; SDS, sodium dodecyl sulfate; RT–PCR, reverse transcription–polymerase chain reaction; SV, simian virus; CMV, cytomegalovirus; β-GAL, β-galactosidase; LUC, luciferase.

0006-2960/94/0433-9257$04.50/0

in adrenal (Smas & Sul, 1993). The expression of the mRNAs for pGE2 (Helman et al., 1990) in neuroendocrine tumors and dlk (Laborda et al., 1993) in small cell lung carcinoma, that are apparently identical to pref-1, suggests that pref-1 may have a broader role in cell growth and differentiation.

We report here the isolation and characterization of the murine gene for pref-1. We determine the exon–intron organization of the gene and compare the domain structure of the gene to those of other members of the EGF-like family. We identify the transcription start site and show that the 5′ flanking region acts as a functional promoter in preadipocytes. We demonstrate that the gene undergoes alternate splicing of the extracellular juxtamembrane region to result in six forms of the pref-1 transcript.

EXPERIMENTAL PROCEDURES

Genomic Southern Blot. High molecular weight mouse genomic DNA was prepared as described from 3T3-L1 preadipocytes (Sambrook et al., 1989). DNA was digested to completion with a panel of restriction enzymes, run on 0.7% agarose gels, and transferred to Hybond (Amersham). Filters were prehybridized at 65 °C for 4 h in 6× SSC, 5× Denhardt's reagent, 0.5% SDS, and 100 μg/mL herring sperm DNA and hybridized for 16–20 h under identical conditions with the addition of EDTA to 0.01 M and [α-^{32}P]dATP (DuPont, New England Nuclear) labeled random-primed full-length mouse pref-1 cDNA as probe. Final washes were at 65 °C for 1 h in 0.1× SSC/0.1% SDS.

Isolation of Pref-1 Genomic Clones. A total of 6.4 × 10^5 plaques of a mouse genomic library purchased from Stratagene and prepared from NIH-3T3 cells in the vector λ FIX II was screened by hybridization of duplicate filters to random-primed, [α-^{32}P]dATP-labeled full-length pref-1 cDNA probe. Following 4 h prehybridization at 65 °C in 5× SSC, 50 mM phosphate buffer, pH 7.5, 4× Denhardt's reagent, 0.1% SDS, and 50 μg/mL herring sperm DNA, probe was added and hybridization continued under the above conditions for 16–20 h. Filters were washed for 1 h in 0.1% SDS/0.1× SSC at 65 °C, and exposed to X-ray film (Fuji) for 18 h at −80 °C with an intensifying screen. Two positive clones were detected. Phage DNA from both clones was purified according to established procedures (Sambrook et al., 1989), and subjected to restriction digest followed by Southern hybridization. Two independent probes, prepared from pref-1 cDNA restriction fragments and specific to either the 5′ or the 3′ portion of the pref-1 cDNA, were used. One of the two genomic clones, gPref-1, was determined to encompass the complete structural gene. Fragments generated by *Bgl*II restriction digest of the genomic clone gPref-1 were subcloned into the pBSSKII(+) vector (Sratagene) for further characterization.

DNA Sequencing. Dideoxy sequencing was performed on double-stranded DNA using Sequenase version 2.0 (United States Biochemical) as recommended by the manufacturer. Exons were localized by sequencing the pref-1 gene utilizing oligonucleotide primers based on the pref-1 cDNA sequence. Sequences at the exon–intron junctions were determined by direct sequencing across the junctions.

RNA Preparation and Northern Blot Analysis. 3T3-L1 cell monolayers were rinsed twice with PBS and harvested by scraping in guanidinium isothiocyanate (GTC) followed by centrifugation over CsCl (Chirgwin et al., 1979). For tissues, tissue was excised, frozen in liquid nitrogen, and homogenized in GTC with a polytron. For Northern analysis, RNA was size-fractionated on 1% agarose gels in 2.2 M formaldehyde, 20 mM MOPS, and 1 mM EDTA, stained with ethidium bromide, and transferred to Hybond (Amersham). Following prehybridization for at least 4 h in 50% formamide, 5× SSC, 5× Denhardt's, 0.5% SDS, and 50 μg/mL herring sperm DNA at 42 °C, filters were hybridized for 16–20 h under identical conditions with the addition of at least 1 × 10^6 cpm/mL of ^{32}P random-primed pref-1 cDNA as probe. Final washes were for 1 h in 0.1× SSC/0.1% SDS at 65 °C. Filters were exposed with an intensifying screen to Fuji RX film.

Primer Extension. Three synthetic oligonucleotide primers were used in primer extension analyses. PR-1 (5′CGAGATGATCGCGACCGGAGCCCTC3′), corresponding to nucleotides +166 to +190 of the pref-1 cDNA (with +1 the transcription start site, see later), PR-2 (5′AGCAGCCCCTGCACCGCCTCCGCTC3′), corresponding to nucleotides +116 to +141, and PR-3 (5′GCACGGGCGTCCAGCGGGAAGAAAG3′), corresponding to nucleotides +50 to +74. Primers were labeled at the 5′ end with [γ-^{32}P]ATP (>6000 Ci/mmol; DuPont, New England Nuclear) and T4 polynucleotide kinase (New England Biolabs). Following purification on Sephadex G-25 (Sigma Chemical), 1.5 × 10^6 cpm of probe was ethanol-precipitated with 20 μg of either 3T3-L1 preadipocyte or mouse liver total RNA. This pellet was resuspended in 20 μL of hybridization solution (0.25 M KCl, 10 mM Tris-HCl, pH 8.3, and 1 mM EDTA) and annealed for 45 min at 65 °C, followed by 45 min at 45 °C. After slow-cooling to room temperature, the reaction mixture was adjusted to 70 mM KCl, 50 mM Tris-HCl, pH 8.3, 10 mM $MgCl_2$, 10 mM dithiothreitol, 500 μM each dNTP, 50 units of human placental ribonuclease inhibitor (Amersham), and 35 units of avian myeloblastosis virus reverse transcriptase in a final volume of 80 μL. The reactions were incubated for 5 min at 37 °C followed by 1 h at 42 °C and ethanol-precipitated, and the products were analyzed on 6% sequencing gels. The identical primer was used in parallel sequencing reactions to determine the exact position of the transcription start site.

Pref-1 Promoter Construct and Transient Transfections. An *Alw*NI restriction fragment encompassing from +120 in the pref-1 cDNA through approximately −2100 of the 5′ flanking region was subcloned into the *Bgl*II site of the promoterless luciferase reporter vector pGL2-basic (Promega) to construct pPREF-LUC. For analysis of pref-1 promoter activity in 3T3-L1 preadipocytes, cells were cultured in Dulbecco's modified Eagle's medium containing 10% fetal bovine serum. Confluent cells in 60 mm dishes were transfected by the calcium phosphate method using 10 μg of DNA of either the promoterless pGL2-basic vector, the SV40 promoter construct pGL2-SV40, or the pPREF-LUC construct. To normalize for the efficiency of transfection, 1 μg of the β-galactosidase expression plasmid pCMV β-GAL was cotransfected with the test plasmids. Cells were harvested 48 h posttransfection and lysed in 200 μL of buffer (1% Triton X-100, 25 mM glycylglycine, pH 7.8, 15 mM $MgSO_4$, 4 mM EGTA, and 1 mM dithiothreitol), and enzyme activities in cytosolic extracts were assayed spectrophotometrically for luciferase (Braiser et al., 1989) using a luminometer (Berthold, Nashua, NH) and for β-galactosidase (Sambrook et al., 1989). The mean and the standard error of the mean were calculated on the basis of triplicate transfections of each construct, and Student's *t*-test was applied to test the statistical significance of the data.

RT–PCR Analysis. For analysis of alternate forms of pref-1 by RT–PCR, first-strand cDNA was synthesized using the primer pcrb (5′TGCTTAGATCTCCTCATCACCAG3′) that

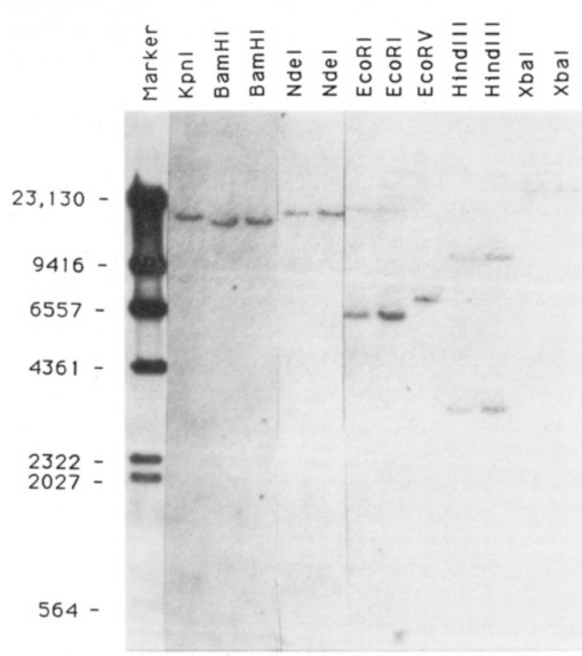

FIGURE 1: Pref-1 genomic Southern analysis. Southern blot hybridization of mouse genomic DNA. DNA was isolated from 3T3-L1 preadipocytes, and 10 μg was digested with each of the indicated restriction enzymes, followed by electrophoresis on a 0.7% agarose gel. The DNA was transferred to a nylon membrane (Hybond, Amersham) and hybridized with radiolabeled pref-1 cDNA probe. Size markers are shown at the left in base pairs and represent *Hin*dIII-digested λ DNA. Following exposure of the membrane to the pref-1 signal, it was rehybridized to radiolabeled λ DNA probe for autoradiographic visualization of the markers.

corresponds to nucleotides 1306–1328 of the pref-1 cDNA, with the transcription start site at +1. For reverse transcription, a 25 μL reaction containing 1 μg of total 3T3-L1 preadipocyte RNA, 500 μM each dNTP, 6.5 μM pcrb primer, 50 mM Tris-HCl, pH 8.3, 10 mM dithiothreitol, 10 mM $MgCl_2$, and 17 units of avian myeloblastosis virus reverse transcriptase was incubated for 1 h at 42 °C. For PCR of the first-strand cDNA, 5 μL of the reverse transcription reaction was brought to a final volume of 50 μL containing 500 μM each dNTP, 50 mM KCl, 10 mM Tris-HCl, pH 9.0, 0.1% Triton X-100, 1.5 mM $MgCl_2$, 6.5 μM each of the oligonucleotide primers pcrb (above) and pcra (5′CCAGCTG-GATTCGTCGACAAGACCT3′, nucleotides 763–787 of the pref-1 cDNA), and 1 unit of TAQ DNA polymerase (Promega). The reaction was subjected to 30 cycles of PCR with denaturation at 96 °C (1 min), annealing at 55 °C (1 min), and extension at 72 °C (2 min) with the final extension reaction for 5 min. The negative control for RT–PCR consisted of parallel reactions minus the reverse transcriptase. Products were fractionated on 1.2% agarose gels, stained with ethidium bromide, transferred to Hybond (Amersham), and subjected to Southern hybridization using a probe corresponding to the cDNA sequence flanked by the pcra and pcrb primer pair. To determine the exact structure of the alternate forms of pref-1, PCR reaction products were cloned into the pCR vector (Invitrogen), according to manufacturer's instructions, and recombinants were analyzed by DNA sequencing.

RESULTS

Mouse Genomic Southern Blot. Hybridization of *Kpn*I-, *Bam*HI-, *Nde*I-, *Eco*RI-, *Eco*RV-, *Hin*dIII-, and *Xba*I-digested mouse genomic DNA with the complete pref-1 cDNA as probe reveals a simple pattern of hybridization (Figure 1).

Table 1: Exon–Intron Organization of the Murine Pref-1 Gene[a]

Consensus Sequences at Splice Sites

```
       a                       tt  tttttt
AG  gt agt.................... tt        ncag   G
       g                       cc  cccccc
```

Exon (length)	5′ Splice Donor		Intron (approx. length)	3′ Acceptor Site	
I (235)	ACC-TAT-G Thr-Phe-G	gtgagc	A (1300)	.....tttctcttcctcag	GG-GCT-GAA ly-Ala-Glu
II (64)	GTC-TGC-AG Val-Cys-Ar	gtaatg	B (500)	.....ctttgtccacacag	G-TGC-CAT g-Cys-His
III (131)	GAA-ATA-G Glu-Ile-A	gtgggc	C (2500)	.....cccccatttcccag	AC-GTT-CGG sp-Val-Arg
IV (142)	ATC-AAT-GG Ile-Asn-Gl	gtaaat	D (1400)	.....tctttgtgttacag	T-TCT-CCC y-Ser-Pro
V (1032)					

[a] Exon sequences are in capital letters, and intron sequences are in lower case with the length of each in base pairs indicated in parentheses. The amino acids interrupted at splice sites are shown below the nucleotide sequence. Splice site consensus sequences are from Shapiro and Senapathy (1987).

This suggests that pref-1 is probably encoded by a single-copy gene. The fact that a single *Eco*RV band of approximately 8 kb hybridizes with the complete cDNA probe further indicates a relatively small size for the pref-1 gene.

Isolation and Domain Structure of the Murine Pref-1 Gene. Screening of a λ FIX II mouse genomic library with the complete pref-1 cDNA identified two λ genomic clones. Southern analysis using pref-1 cDNA restriction fragments specific for both the 5′ and 3′ portions of the cDNA revealed that one clone, gPREF-1, hybridized to both probes. Subsequent restriction and sequence analysis confirmed that clone gPREF-1 contained the total 7.3 kb structural gene, approximately 8 kb of 5′ flanking sequence, and 1.5 kb of 3′ flanking sequence. The structural gene consists of five exons ranging in size from 64 to 1033 bp. Sequences at the exon–intron junctions (Table 1) conform to consensus splice donor and acceptor sites (Shapiro & Senapathy, 1987). Unlike the majority of vertebrate EGF-like genes where individual EGF-like repeats are often encoded by separate exons, the structure of the pref-1 gene reveals only an approximate correlation between exon–intron junctions and the predicted functional domains of the pref-1 protein (Figure 2). The entire 5′ untranslated region, the translation start site, and the presumed signal sequence through to the start of the first EGF-like repeat are encoded by exon 1. Exon 2 contains the first two disulfide loops of EGF-like repeat 1. Exon 3 codes for the remaining disulfide loop of the first repeat, and the complete second EGF-like repeat. Exon 4 comprises all of the third EGF-like repeat and the initial cysteine of the first disulfide loop of EGF-like repeat 4. The remainder of the molecule, including the second cysteine of the first disulfide loop and the second and third disulfide loops of EGF-like repeat 4, all of repeats 5 and 6, and the juxtamembrane, transmembrane, and cytoplasmic domains, is encoded by exon 5. Only the placement of an intron between exons 1 and 2, separating the signal sequence from the start of the EGF-like repeats, suggests an exact correlation between exon boundaries and protein domains. The remainder of the gene is not modular. Introns split EGF-like repeats and exon 5 encodes several EGF-like repeats as well as the juxtamembrane, transmembrane, and cytoplasmic regions.

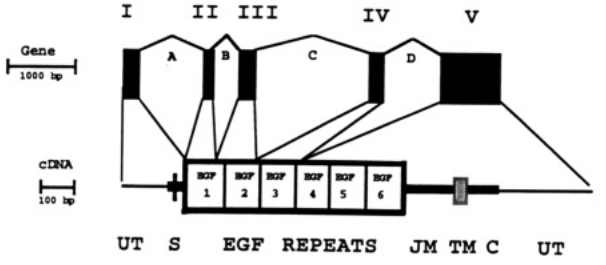

FIGURE 2: Organization of the murine pref-1 gene and alignment of protein domains with the exon–intron boundaries. The structural organization of the gene is shown at the top with the corresponding protein domains shown below. The respective size scales are at the left. Exons are indicated by solid boxes and are numbered accordingly (I–V), and positions of the four introns (A–D) are indicated. The domains of the pref-1 cDNA and protein are designated as follows: UT, untranslated region; S, signal sequence; EGF REPEATS, EGF-like repeats; JM, juxtamembrane; TM, transmembrane; C, cytoplasmic domain.

Analysis of Transcription Start Site. To identify the 5′ end of the pref-1 transcript, we performed primer extension studies. Each of three oligonucleotide primers PR-1, PR-2, and PR-3, when used in primer extension, resulted in a single extension product that stopped at the same nucleotide. This nucleotide is indicated by arrows in Figure 3A. This site was identified when 3T3-L1 preadipocyte RNA was used as template, and is located 169 bases preceding the AUG translation start site. No product was detected when mouse liver RNA was employed. This agrees with the fact that pref-1 mRNA is not detected in liver by Northern analysis (Smas & Sul, 1993). The exact location of transcription initiation was determined by comparison with a DNA sequencing reaction that used the indicated primers. A pref-1 genomic subclone that encompassed 3000 bp of 5′ flanking sequence through exon 3 was used as template. All references to the cDNA sequence in this paper are numbered with the start site, identified herein, at +1. A consensus cap site, CAGCTC (Bucher & Trifonov, 1986), correctly predicts the transcription start site as the A at +1, the same nucleotide demonstrated by primer extension.

Since sequences proximal to the transcription start site did not show traditional TATA or CAAT boxes, we addressed whether the pref-1 5′ flanking region sequence acts as a functional promoter in 3T3-L1 preadipocytes. We transfected into 3T3-L1 preadipocytes either the plasmid pPREF-LUC, which contained 2.1 kb of pref-1 5′ flanking region linked to the luciferase reporter gene, or as controls the promoterless pGL2-basic vector, or pGL2-SV40, an SV-40 promoter-driven luciferase positive control plasmid. To normalize for transfection efficiency, a β-galactosidase expression construct was cotransfected with each of the above three plasmids. Background activity in our luciferase assay averaged 300–400 relative light units. The normalized mean luciferase activities of triplicate dishes transfected with pGL2-basic, pPREF-LUC, and pGL2-SV40 were 9150 ± 975, 195 587 ± 5494*, and 502 598 ± 37 835*, respectively (±SEM, asterisk indicates activity significantly different than that of pGL2-basic, $p < 0.01$). This demonstrates that the pref-1 promoter construct pPREF-LUC induces a level of luciferase activity more than 20 times higher than that of the promoterless pGL2-basic vector, and approximately 40% that of the strong viral promoter construct pGL2-SV40. Thus, we have identified the 5′ end of the structural gene for pref-1 and the 5′ flanking sequence that contains a functional promoter highly active in 3T3-L1 preadipocytes.

Alternate Splicing Generates Multiple Pref-1 Transcripts. During our previous isolation and characterization of pref-1 cDNA clones, one clone isolated had an in-frame deletion of 225 bp in the extracellular juxtamembrane region. The presence of a deleted form of the pref-1 cDNA raised the possibility that alternate splicing is used as a mechanism for the generation of various pref-1 transcripts. To investigate whether this deleted pref-1 transcript is present in vivo, and to determine whether other alternate forms exist, we searched for alternate forms of pref-1 using RT–PCR.

RT–PCR amplifications were performed with a primer pair flanking the sequence missing from the originally identified deleted cDNA clone. Two major products result (Figure 4A), both of which hybridize to pref-1-specific probe (Figure 4B). The longer product corresponds in size to the 566 bp band which would be generated from the undeleted, full-length pref-1 and the shorter one to the 341 bp band which would result from a 225 bp deletion in the pref-1 transcript. An identical pattern of products was detected by RT–PCR of adult mouse adrenal tissue, the only adult tissue known to express pref-1 (unpublished results). No product was detected when reverse transcriptase was omitted from the reactions. To determine the exact structure of these forms and to identify other forms of pref-1 transcript that may exist in 3T3-L1 preadipocytes, RT–PCR reaction products were subcloned and sequenced.

In addition to full-length pref-1, we have identified five alternate forms of pref-1 transcript, all generated within exon 5, that contain various in-frame deletions of the extracellular juxtamembrane region (Figure 5). Three of these differ significantly in regard to the size of the deletion. Full-length undeleted pref-1 is referred to here as form A. Form B is the result of a 153 bp deletion of a portion of the sixth EGF-like repeat and part of the juxtamembrane region. Form C has the same 5′ splice donor point as form B, but has a further downstream 3′ acceptor site to result in a 219 bp deletion that encompasses a portion of the sixth EGF-like repeat, all of the juxtamembrane region, and several amino acids of the predicted transmembrane domain. Form D has the same 3′ acceptor point as form C but has an even more upstream 5′ donor site to result in a 279 bp deletion that includes all the region deleted in form C and, in addition, the complete sixth EGF-like repeat. Aside from these major differences among the alternate forms, minor differences were also detected. Form C2 is identical to form C except that the 3′ acceptor site is 6 bp after that of form C. Analogously, in form D2 the acceptor site is 6 bp after that of form D, and this is the same acceptor site as that used in form C2. The sequences at the 5′ and 3′ ends of the deleted regions have consensus splice signals; they begin with GT and end with AG, and a polypyrimidine tract is present just before the 3′ end of the deletion (Shapiro & Senapathy, 1987).

To rule out the possibility that the alternate products detected were artifacts generated by RT–PCR and to confirm that they represent actual forms of the pref-1 transcript present in vivo, we screened approximately 420 000 phage plaques of the 3T3-L1 cDNA library from which pref-1 was originally isolated. Approximately 0.05% of the phage plaques hybridized to pref-1 probe, and 15 of these were characterized by restriction analysis and sequencing in regard to alternately spliced products. Of these cDNA clones, 10 corresponded to the undeleted pref-1 (form A), 2 to form B, 1 to form C, and 2 to form C2. The detection of forms B, C, and C2 by two

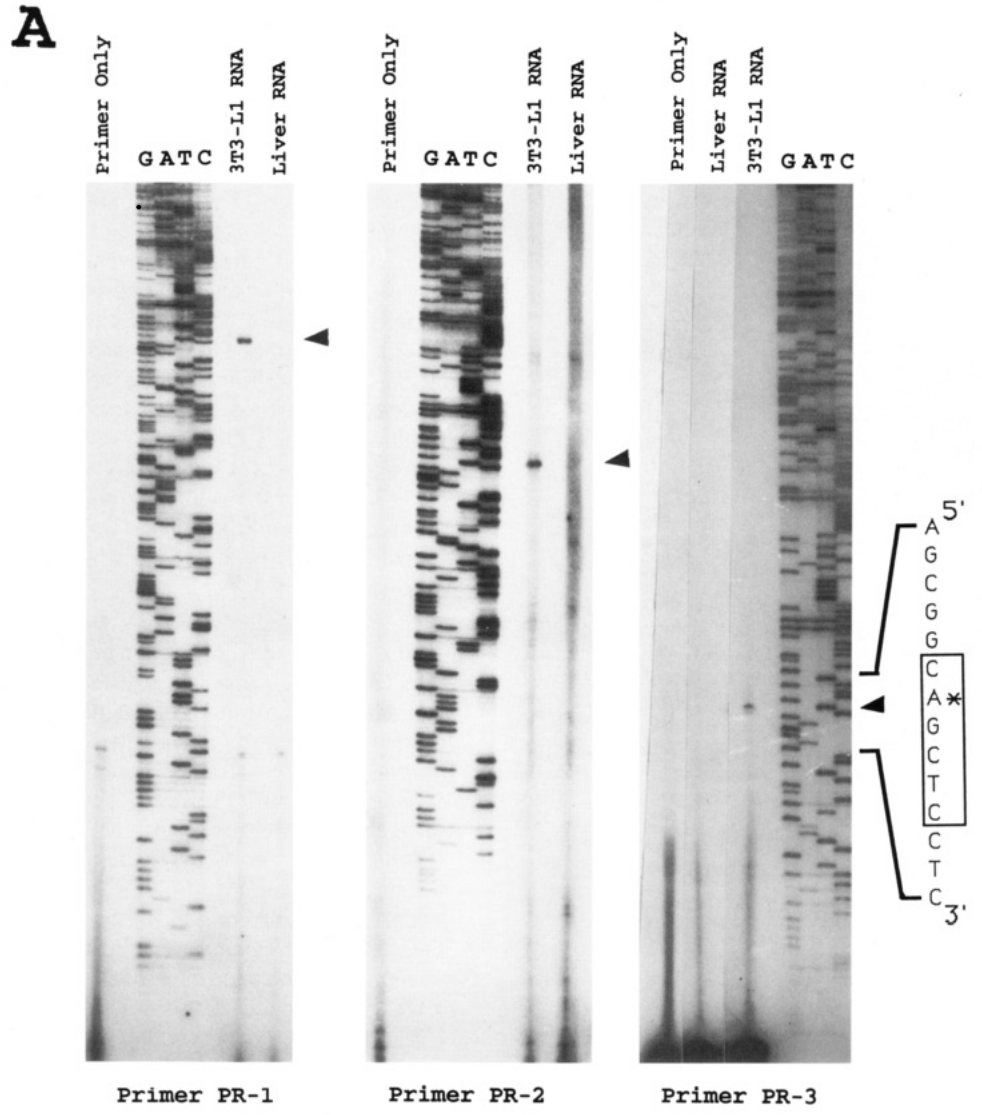

FIGURE 3: Identification of the transcription initiation site of the pref-1 gene. (A) Primer extension analysis. Total RNA (20 μg) prepared from mouse liver or from 3T3-L1 preadipocytes was hybridized to the indicated ^{32}P end-labeled primer. Products from each primer extension reaction, as well as probe alone, were analyzed on 6% sequencing gels. A sequencing ladder (lanes marked G, A, T, and C) was generated by using the indicated primer with a pref-1 genomic DNA subclone as template in dideoxy sequencing reactions and was run in parallel with the primer extension products. The specific primer extension products that result for PR-1, PR-2, and PR-3 are shown by arrows. The sequence at the transcription start site is indicated at the right and is the reverse complement of that shown on the sequencing ladder. The identified transcription initiation site is indicated by an asterisk, and the consensus cap site is boxed. (B) The locations of the three primers used in primer extension analysis are underlined and labeled. The cap site sequence is underlined and the transcription start site indicated by a forward arrow.

different methods, RT–PCR and cDNA clone analysis, confirms that they represent multiple pref-1 transcripts present in vivo. Failure to detect cDNA clones for forms D and D2 likely reflects their low frequency in the mRNA population. Under the assumption that the number of each of the various forms of the pref-1 cDNAs isolated reflects their relative frequency in the 3T3-L1 preadipocyte mRNA population, it is likely that the discrete transcript of approximately 1.7 kb detected by Northern analysis of 3T3-L1 preadipocyte RNA (Figure 6) corresponds to form A, the longest pref-1 transcript.

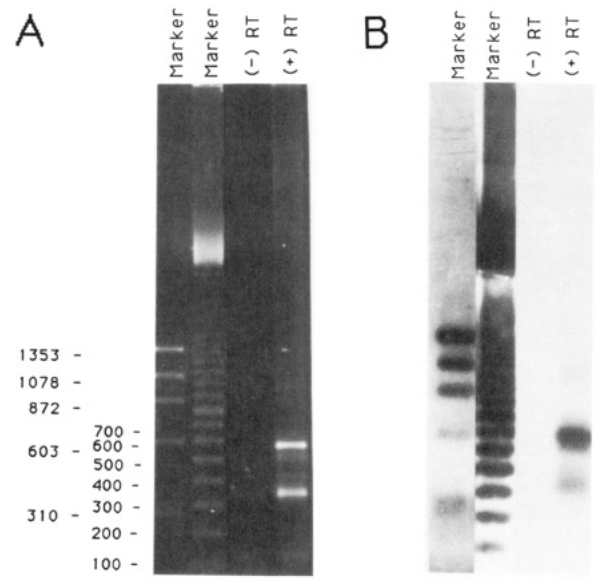

FIGURE 4: RT–PCR analysis of alternate splicing of the pref-1 gene. (A) Ethidium bromide stained gel of RT–PCR products. Total 3T3-L1 preadipocyte RNA was subjected to a RT–PCR reaction as described under Experimental Procedures, in either the absence (–) or the presence (+) of reverse transcriptase. The products were separated on a 1.2% agarose gel, visualized by ethidium bromide staining, transferred to a nylon membrane, and hybridized with a ^{32}P-labeled pref-1 probe that corresponds to the region between the two primers used in the RT–PCR reaction. The resulting Southern blot is shown in (B). Size markers are shown at the left and are a *Phi*X174/*Hae*III digest and a 100 bp ladder. These were visualized by hybridizing this portion of the membrane with radiolabeled markers, followed by autoradiography.

DISCUSSION

Exon–Intron Organization of the Pref-1 EGF-like Domains Resembles Invertebrate EGF-like Genes. Distinct structural and/or functional domains of proteins are often each encoded by single discrete exons. Additionally, the placement of the intron at the first, second, or third nucleotide of a codon, its phase, can reflect the evolutionary relatedness of these domains. Only exons that have introns of the same phase class at their 5′ and 3′ ends can be inserted, deleted, or duplicated by intron recombination without disruption of the reading frame (Patthy, 1987). On the basis of these features, the genes for the EGF-like family of proteins can be further classified into subfamilies. The first group includes those genes that encode growth factors whose products bind to the EGF receptor: EGF (Bell et al., 1986), amphiregulin (Plowman et al., 1990), HB-EGF (Fen et al., 1993), and TGF-α (Blasband et al., 1990). The growth factor-encoding portion of these genes is represented by two exons that are split by an intron at a conserved location between the fourth and fifth cysteine residues of the EGF-like domain. A second group of genes includes those encoding vertebrate and invertebrate EGF-like proteins that are modular in nature and that do not function through the EGF receptor. In these genes, each EGF-like domain is represented by a single exon that is flanked by symmetrical phase 1, or 1-1 class, introns (Patthy, 1988) and suggests that the overall arrangement of various domains in these proteins arose by exon shuffling (Patthy, 1987). These include the genes for LDL receptor (Sudhof et al., 1985), vitamin K dependent protein S (Ploos van Amstel et al., 1990), protein C (Foster et al., 1985), factor IX (Anson et al., 1984), murine homing receptor (Dowbenko et al., 1991), tissue plasminogen activator (Ny et al., 1984), and cartilage matrix proteins (Kiss et al., 1989) and the sea urchin gene uegf (Delgadillo-Reynoso et al., 1989).

Unlike the above genes, the EGF-like domains of pref-1 are not each encoded by a separate exon. Rather, individual exons of the pref-1 gene encode portions of a repeat domain or multiple domains. Additionally, the introns of the pref-1 gene are not symmetrical phase 1 introns as they are for other mammalian EGF-like genes, but are asymmetrical mixed phase introns. A third group would therefore include pref-1 and the EGF-like genes of *Drosophila* and *Caenorhabditis elegans*. EGF-like repeats of the notch, lin-12, and glp-1 proteins are not individually represented by exons, nor are the introns in the EGF-like coding regions of solely the 1-1 class (Kidd et al., 1986; Yochem et al., 1988; Yochem & Greenwald, 1989; Wharton et al., 1985), but are asymmetrical mixed phase introns, where exons may encode multiple or partial EGF-like domains. While not characterized in this regard, the vertebrate *notch* homologs *tan-1* (Ellisen et al., 1991) and *xotch* (Coffman et al., 1990) would presumably be in this third group.

Characteristics of the Pref-1 Promoter. Our analysis demonstrates that transcription of the pref-1 gene initiates at a single start site. Transient transfection analysis demonstrates that –2100 bp of 5′ flanking sequence is a functional promoter. While the cis-acting elements that control pref-1 gene transcription are currently under investigation, preliminary sequence analysis indicates that while the pref-1 gene does not contain consensus CAAT or TATA boxes at predicted distances from the +1 start site, other potential regulatory signals are present. Multiple GC boxes, often seen in TATA-less promoters, are present in the pref-1 5′ flanking region. This cis-acting element has been identified in the SV40 promoter region and in a wide variety of cellular genes where it binds the general transcription factor Sp1 (Gidoni et al., 1984). Also present are consensus AP-1 elements, which have been demonstrated to interact with jun/fos heterodimers (Mitchell et al., 1987), and binding sites for the AP-2 transcription factor that mediates response to phorbol ester and diacylglycerol (Imagawa et al., 1987). Future studies will address the role of these and other elements in the control of pref-1 gene transcription.

Alternate Splicing Generates Multiple Pref-1 Transcripts. Through the sequencing of multiple pref-1 cDNA clones and RT–PCR, we demonstrate that alternate splicing serves to generate at least six forms of the pref-1 transcript. Alternate splicing is known to occur in a variety of genes (Andreadis et al., 1987) and often alters the function of the resultant protein, for example, by the addition kinase domains (Toyoshima et al., 1993) or phosphorylation sites (Najjar et al., 1993) or by serving as a mechanism for the creation of soluble and membrane-bound forms of the same protein (Goodwin et al., 1990). Removal of a single EGF-like domain via alternate splicing occurs in the extracellular matrix protein aggrecan (Baldwin et al., 1989; Fulop et al., 1993). In addition, several *Drosophila* EGF-like genes express multiple transcripts presumably due to alternate splicing (Fleming et al., 1990; Kopczynski et al., 1988).

Although the functional differences of the alternately spliced products of the *pref-1* gene are currently under investigation, several hypotheses can be considered based on the demonstrated role of alternate splicing in other proteins. In each of the alternate forms of pref-1, all or a portion of the extracellular juxtamembrane region is deleted and results in a repositioning of the EGF-like repeats closer to the cell membrane. Forms

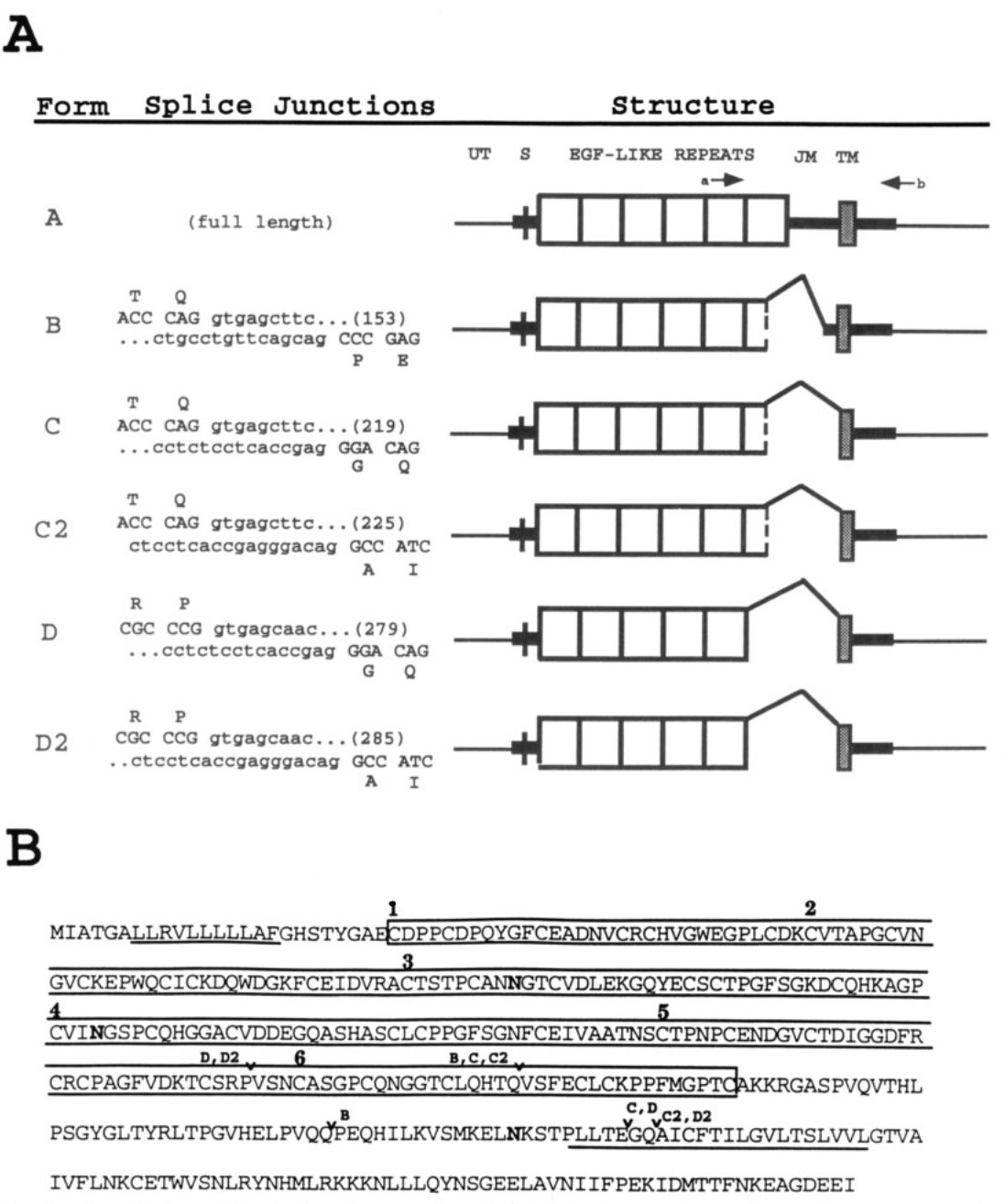

FIGURE 5: Splice junctions of alternately spliced pref-1 transcripts. (A) The six alternate forms of the pref-1 transcript as discussed in the text are listed at the left. The sequence at splice junctions is shown with deleted sequences in lower case letters and coding sequences in upper case letters. The corresponding amino acid encoded is shown above the codon. The total size of the deleted region, in base pairs, is in parentheses. To the right is the schematic illustration of the predicted products: UT, untranslated region; S, signal sequence; EGF-like repeats; JM, juxtamembrane; TM, transmembrane; C, cytoplasmic domain. The region deleted is shown by connecting lines, and the dashed line indicates a partial sixth EGF-like repeat. The arrows indicate approximate positions of the primer pair used in the RT–PCR reactions. (B) The alternate forms are indicated in the context of the complete amino acid sequence of the pref-1 protein. The underlined regions represent the signal peptide and transmembrane domain. The six EGF-like repeats are boxed and indicated by numbering (1–6) above the first cysteine of each repeat, and the three potential N-linked glycosylation sites are in boldface. The borders of the area deleted in each of the alternately spliced forms are shown by downward arrows above the amino acid sequence. B, C, C2, D, and D2, written to the left of the arrows, indicate the location of the 5′ donor site for that particular form and those to the right of the arrows the 3′ acceptor site for that form.

C, C2, D, and D2 all affect the sixth EGF-like domain. In forms B, C, and C2, a portion of the domain is missing and results in alteration of the triple disulfide loop structure, and in forms D and D2, all of the sixth repeat is deleted. It is possible that this sixth repeat has a distinct function in pref-1. Deletion analysis of the *Drosophila* notch protein demonstrates that not all EGF-like repeats are equal; rather, only 2 of the 36 EGF-like repeats of notch govern its interaction with the EGF-like proteins delta and serrate (Rebay et al., 1991). Additionally, in forms C, C2, D, and D2, the area deleted contains a potential site for N-linked glycosylation. It is possible that this site has a role in the function of the pref-1 protein. The deleted region is bounded by consensus splice signals. The fact that these signals are all within exon 5, and that the deletions maintain the reading frame, suggests that the splicing mechanism is likely a combination of retained intron and the use of alternate donor and acceptor splice sites. Detection of only a single 1.7 kb species of pref-1 mRNA by Northern analysis may indicate that full-length undeleted pref-1 is the most abundant form. This is supported by our analysis of multiple cDNAs for pref-1, where 10 of 15 cDNA clones analyzed encoded the undeleted form A, and suggests

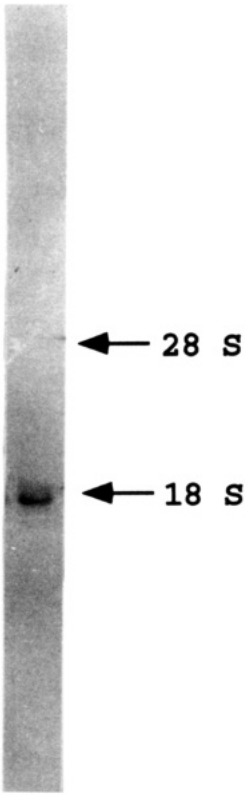

FIGURE 6: Northern analysis of the pref-1 transcript. Five micrograms of total RNA from 3T3-L1 cells was size-fractionated on a 1% agarose gel and hybridized to radiolabeled pref-1 probe. The positions of 28S and 18S ribosomal RNA bands are indicated by arrows.

that the relative intensity of the various RT–PCR products is not quantitative.

In summary, we have cloned the murine gene encoding pref-1 and determined the exon–intron structure of the gene. We have defined the transcription start site, and demonstrated the promoter activity of its 5′ flanking sequence. Furthermore, we have identified and analyzed six forms of the pref-1 transcript that arise by alternate splicing within exon 5. The analysis of the pref-1 gene we present will allow further studies on the evolutionary relationship between genes of the EGF-like family of proteins and determination of functional differences between various forms of the pref-1 protein expressed in preadipocytes, and can serve as a valuable model system to address the mechanism of gene down-regulation during adipocyte differentiation.

ACKNOWLEDGMENT

We thank N. Moustaid and J. Chow for help with the transfection of the pPREF-LUC construct. We also thank N. Gekakis for advice on primer extension.

REFERENCES

Andreadis, A., Gallego, M. E., & Nadal-Ginard, B. (1987) *Annu. Rev. Cell Biol. 3*, 207–242.

Anson, D. S., Choo, K. H., Rees, D. J. G., Giannelli, F., Gould, K., Huddleston, J. A., & Brownlee, G. G. (1984) *EMBO J. 3*, 1053–1060.

Aratani, Y., & Kitagawa, Y. (1988) *J. Biol. Chem. 263*, 16163–16169.

Baldwin, C. T., Reginato, A. M., & Prockop, D. J. (1989) *J. Biol. Chem. 264*, 15747–15750.

Bell, G. I., Fong, N. M., Stempien, M. M., Wormsted, M. A., Caput, D., Ku, L., Urdea, M. S., Rall, L. B., & Sanchez-Pescador, R. (1986) *Nucleic Acids Res. 14*, 8427–8446.

Blasband, A. J., Rogers, K. T., Chen, X., Azizkhan, J. C., & Lee, D. C. (1990) *Mol. Cell. Biol. 10*, 2111–2121.

Braiser, A. R., Tate, J. E., & Habener, J. F. (1989) *BioTechniques 7*, 1116–1121.

Bucher, P., & Trifonov, E. N. (1986) *Nucleic Acids Res. 14*, 10009–10026.

Calvo, J. C., Rodbard, D., Katki, A., Chernick, S., & Yanagishita, M. (1991) *J. Biol. Chem. 266*, 11237–11244.

Carpenter, G., & Cohen, S. (1990) *J. Biol. Chem. 265*, 7709–7712.

Chirgwin, J., Przybla, A., MacDonald, R., & Rutter, W. (1979) *Biochemistry 18*, 5294–5299.

Coffman, C., Harris, W., & Kinter, C. (1990) *Science 249*, 1438–1441.

de Herreros, A. G., & Birnbaum, M. J. (1989) *J. Biol. Chem. 264*, 19994–19999.

Delgadillo-Reynoso, M. G., Rollo, D. R., Hursh, D. A., & Raff, R. A. (1989) *J. Mol. Evol. 29*, 314–327.

Dowbenko, D. J., Diep, A., Taylor, B. A., Lusis, A. J., & Lasky, L. A. (1991) *Genomics 9*, 270–277.

Ellisen, L. W., Bird, J., West, D. C., Soreng, A. L., Reynolds, T. C., Smith, S. D., & Sklar, J. (1991) *Cell 66*, 649–661.

Fehon, R. G., Kooh, P. J., Rebay, I., Regan, C. L., Xu, T., Muskavitch, M. A. T., & Artavanis-Tsakonas, S. (1990) *Cell 61*, 523–534.

Fen, Z., Dhadly, M. S., Yoshizumi, M., Hilkert, R. J., Quertermous, T., Eddy, R. L., Shows, T. B., & Lee, M. (1993) *Biochemistry 32*, 7932–7938.

Fleming, R. J., Scottgale, T. N., Diederich, R. J., & Artavanis-Tsakonas, S. (1990) *Genes Dev. 4*, 2188–2201.

Foster, D. C., Yoshitake, S., & Davie, E. W. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 4673–4677.

Fulop, C., Walcz, E., Valyon, M., & Glant, T. T. (1993) *J. Biol. Chem. 268*, 17377–17383.

Gidoni, D., Dynan, W. S., & Tjian, R. (1984) *Nature 312*, 409–413.

Goodwin, R. G., Friend, D., Ziegler, S. F., Jersy, R., Falk, B. A., Ginpel, S., Cosman, D., Dower, S. K., March, C. J., Namen, A. E., & Park, L. S. (1990) *Cell 60*, 941–951.

Green, H., & Kehinde, O. (1974) *Cell 1*, 113–116.

Green, H., & Kehinde, O. (1975) *Cell 5*, 19–27.

Helman, L. J., Sack, N., Plon, S. E., & Israel, M. A. (1990) *Nucleic Acids Res. 18*, 685.

Higashiyama, S., Abraham, J. A., Miller, J., Fiddes, J. C., & Klagsbrun, M. (1991) *Science 251*, 936–939.

Ignotz, R. A., & Massague, J. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 8530–8534.

Imagawa, M., Chui, R., & Karin, M. (1987) *Cell 51*, 251–260.

Kidd, S., Kelley, M. R., & Young, M. W. (1986) *Mol. Cell. Biol. 6*, 3094–3108.

Kiss, I., Deak, F., Holloway, R. G., Jr., Delius, H., Mebuss, K. A., Frimberger, E., Argraves, W. S., Tsonis, P. A., Winterbottom, N., & Goetinck, P. F. (1989) *J. Biol. Chem. 264*, 8126–8134.

Kopczynski, C. C., Alton, A. K., Fechtel, K., Kooh, P. J., & Muskavitch, M. A. T. (1988) *Genes Dev. 2*, 1723–1735.

Krusius, T., Gehlsen, K. R., & Ruoslahti, E. (1987) *J. Biol. Chem. 262*, 13120–13125.

Laborda, J., Sausville, E. A., Hoffman, T., & Notario, V. (1993) *J. Biol. Chem. 268*, 3817–3820.

Massague, J. (1990) *J. Biol. Chem. 265*, 21393–21396.

Mitchell, P. J., Wang, C., & Tjian, R. (1987) *Cell 50*, 847–861.

Najjar, S. M., Accili, D., Philippe, N., Jernberg, J., Margolis, R., & Taylor, S. I. (1993) *J. Biol. Chem. 268*, 1201–1206.

Ntambi, J. M., Buhrow, S. A., Kaestner, K. H., Christy, R. J., Sibley, E., Kelly, T. J., & Lane, M. D. (1988) *J. Biol. Chem. 263*, 17291–17300.

Ny, T., Elgh, F., & Lund, B. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 5355–5359.

Patthy, L. (1987) *FEBS Lett. 214*, 1–7.

Patthy, L. (1988) *J. Mol. Biol. 202*, 689–696.

Paulauskis, J. D., & Sul, H. S. (1988) *J. Biol. Chem. 263*, 7049–7054.

Ploos van Amstel, H. K., Reitsma, P. H., Van der Logt, C. P. E., & Bertina, R. M. (1990) *Biochemistry 29*, 7853–7861.

Plowman, G. D., Green, J. M., McDonald, V. L., Neubauer, M. G., Disteche, C. M., Todaro, G. J., & Shoyab, M. (1990) *Mol. Cell. Biol. 10*, 1969–1981.

Rebay, I., Fleming, R. J., Fehon, R. G., Cherbas, L., Cherbas, P., & Artavanis-Tsakonas, S. (1991) *Cell 67*, 687–699.

Rubin, C. S., Hirsch, A., Fung, C., & Rosen, O. M. (1978) *J. Biol. Chem. 253*, 7570–7578.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Shapiro, M. B., & Senapathy, P. (1987) *Nucleic Acids Res. 15*, 7153–7174.

Siegelman, M. H., Cheng, I. C., Weissman, I. L., & Wakeland, E. K. (1990) *Cell 61*, 611–622.

Smas, C. M., & Sul, H. S. (1993) *Cell 73*, 725–734.

Smith, P. J., Wise, L. S., Berkowitz, R., Wan, C., & Rubin, C. S. (1988) *J. Biol. Chem. 263*, 9402–9408.

Sudhof, T. C., Goldstein, J. L., Brown, M. S., & Russell, D. W. (1985) *Science 228*, 815–822.

Toyoshima, H., Kozutsumi, H., Maru, Y., Hagiwara, K., Furuya, A., Mioh, H., Hanai, N., Takaku, F., Yazaki, Y., & Hirai, H. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 5404–5408.

Wharton, K. A., Johansen, K. M., Xu, T., & Artavanis-Tsakonas, S. (1985) *Cell 43*, 567–581.

Yochem, J., & Greenwald, I. (1989) *Cell 58*, 553–563.

Yochem, J., Weston, K., & Greenwald, I. (1988) *Nature 335*, 547–550.