of action of polyene macrolide antibiotics.

ACKNOWLEDGMENTS

We thank Mr. Woubeshet Ayenew and Mr. Tom Stephan for their excellent technical assistance.

Registry No. AMB, 1397-89-3; K^+, 7440-09-7; Cl^-, 16887-00-6; cholesterol, 57-88-5.

REFERENCES

Andreoli, T. E. (1973) *Kidney Int. 4*, 337-345.
Archer, D. B. (1976) *Biochim. Biophys. Acta 436*, 68-76.
Bolard, J. (1986) *Biochim. Biophys. Acta 864*, 257-304.
Brajtburg, J., Powderly, W. G., Kobayashi, G. S., & Medoff, G. (1990) *Antimicrob. Agents Chemother. 34*, 183-188.
Bramhall, J., Hofmann, J., DeGuzman, R., Montestruque, S., & Schell, R. (1987) *Biochemistry 26*, 6330-6340.
Cafiso, D. S., & Hubbell, W. L. (1978) *Biochemistry 17*, 187-195.
Cybulska, B., Mazerski, J., Borowski, E., & Gary-Bobo, C. M. (1984) *Biochem. Pharmacol. 33*, 41-46.
Cybulska, B., Herve, M., Borowski, E., & Gary-Bobo, C. M. (1986) *Mol. Pharmacol. 29*, 293-298.
Dencher, N. A., Burghaus, P. A., & Grzesiek, S. (1986) *Methods Enzymol. 127*, 746-760.
Ellena, J. F., Hutton, W. C., & Cafiso, D. S. (1985) *J. Chem. Soc. 107*, 1530-1537.
Ellena, J. F., Hill, B. D., McGarvey, G. J., & Cafiso, D. S. (1987) *Biophys. J. 51*, 529a.
El-Mashak, E. M., & Tsong, T. Y. (1985) *Biochemistry 24*, 2884-2888.
Fyles, T. M., Kaye, K. C., James, T. D., & Smiley, D. W. M. (1990) *Tetrahedron Lett. 31*, 1233-1236.
Gary-Bobo, C. M. (1989) *Biochimie 71*, 37-47.
Hartsel, S. C., Perkins, W. R., McGarvey, G. J., & Cafiso, D. S. (1988) *Biochemistry 27*, 2656-2660.
Hauser, H., Phillips, M. C., & Stubbs, M. (1972) *Nature 239*, 342-344.
Hope, M. J., Bally, M. B., Mayer, L. D., Janoff, A. S., & Cullis, P. R. (1986) *Chem. Phys. Lipids 40*, 89-107.
Hsu Chen, C.-C., & Feingold, D. S. (1973) *Biochem. Biophys. Res. Commun. 51*, 972-977.
Jain, M. K. (1972) in *The Bimolecular Membrane: A System*, pp 162-163, Van Nostrand Reinhold Co., New York.
Kano, K., & Fendler, J. H. (1978) *Biochim. Biophys. Acta 509*, 289-299.
Kleinberg, M. E., & Finkelstein, A. (1984) *J. Membr. Biol. 80*, 257-269.
Lewis, E. N., & Levin, I. W. (1989) *Biophys. J. 55*, 103a.
Marty, A., & Finkelstein, A. (1975) *J. Gen. Physiol. 65*, 515-526.
Milhaud, J., Hartmann, M.-A., & Bolard, J. (1989) *Biochimie 71*, 49-56.
Ockman, N. (1974) *Biochim. Biophys. Acta 395*, 263-282.
Ramos, H., de Murciano, A. A., Cohen, B. E., & Bolard, J. (1989) *Biochim. Biophys. Acta 982*, 303-306.
Singleton, W. S., Gray, M. S., Brown, M. L., & White, J. L. (1965) *J. Am. Oil Chem. Soc. 42*, 53-56.
Stillwell, W., Brengle, B., Hester, P., & Wassall, S. R. (1989) *Biochemistry 28*, 2798-2804.
Szoka, F., & Papahadjopoulos, D. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 4194-4198.
Terwilliger, T. C., Weissman, L., & Eisenberg, D. (1982) *Biophys. J. 37*, 353-361.
Tschopp, J. (1984) *J. Biol. Chem. 259*, 7857-7863.
Whyte, B. S., Peterson, R. P., & Hartsel, S. C. (1989) *Biochem. Biophys. Res. Commun. 164*, 609-614.

Genomic Sequence and Organization of Two Members of a Human Lectin Gene Family[†,‡]

Michael A. Gitt and Samuel H. Barondes*

Department of Psychiatry, Langley Porter Psychiatric Institute, University of California, San Francisco, 401 Parnassus Avenue, San Francisco, California 94143-0984

Received June 13, 1990; Revised Manuscript Received September 21, 1990

ABSTRACT: We have isolated and sequenced the genomic DNA encoding a human dimeric soluble lactose-binding lectin. The gene has four exons, and its upstream region contains sequences that suggest control by glucocorticoids, heat (environmental) shock, metals, and other factors. We have also isolated and sequenced three exons of the gene encoding another human putative lectin, the existence of which was first indicated by isolation of its cDNA. Comparisons suggest a general pattern of genomic organization of members of this lectin gene family.

Many animal tissues contain soluble lactose-binding lectins (S-Lac lectins)[1] (Barondes, 1984). In the rat, at least nine have been identified and partially characterized (Cerra et al., 1985; Leffler et al., 1989; Hinek et al., 1988). Although all bind lactose, there is evidence for considerable specificity among them (Leffler & Barondes, 1986; Sparrow et al., 1987). For example, three S-Lac lectins from lung (Leffler & Barondes, 1986; Sparrow et al., 1987) differ greatly in their relative binding of a series of naturally occurring mammalian glycoconjugates, suggesting biologically significant lectin-glycoconjugate interactions.

The best studied of the S-Lac lectins is a dimer with subunit molecular weight of about 14 000, here referred to as L-14. It is abundant in many cell types where it often comprises more

[†] Supported by grants from the NIH and NSF.

[‡] The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J05303.

[1] Abbreviations: S-Lac lectin, soluble lactose-binding lectin; SSC, standard saline citrate buffer; L-14, S-Lac lectin of subunit M_r ~14 000; L-30, S-Lac lectin of M_r ~30 000; bp, base pair(s); kbp, kilobase pair(s); PCR, polymerase chain reaction; IgE, immunoglobulin E; BSA, bovine serum albumin; SDS, sodium dodecyl sulfate.

0006-2960/91/0430-82$02.50/0

than 0.1% of the soluble protein, but is not expressed in others (Barondes, 1984). cDNA encoding this lectin has been cloned and sequenced from a variety of sources (Gitt & Barondes, 1986; Clerch et al., 1988; Raz et al., 1988; Abbott & Feizi, 1989; Abbott et al., 1989; Couraud et al., 1989: Hirabayashi et al., 1989; Merkle et al., 1989; Wilson et al., 1989; Cooper & Barondes, 1990; Hynes et al., 1990). The cDNA encoding a related lectin from chicken skin has also been cloned and sequenced (Ohyama et al., 1986). Another related lectin that has been studied extensively has an apparent subunit molecular weight of 29000–35000 on polyacrylamide gel electrophoresis and a C-terminal half with considerable homology to L-14 (Jia & Wang, 1988; Liu et al., 1985). Its sequence has been determined from clones from mouse lung (Jia & Wang, 1989) and rat leucocytes (Liu et al., 1985; Albrandt et al., 1987). Since estimates of its molecular weight range around 30000, it will be referred to as L-30. The partial amino acid sequence of several other S-Lac lectins from rat intestine indicates that they too are related proteins (Leffler et al., 1989). In addition, we have previously isolated the cDNA encoding a human putative lectin that is yet another member of this family (Gitt & Barondes, 1986).

One of the features of S-Lac lectins is that several show marked variations in expression with development (Beyer & Barondes, 1982). For example, L-14 is actively synthesized at a specific stage in differentiation of cultured mouse muscle cells (Cooper & Barondes, 1990), suggesting a role in muscle maturation. In addition, glucocorticoid regulation of L-14 (Clerch et al., 1987) and chicken skin lectin (Oda et al., 1989) has been reported, and evidence relating L-14 to neoplasia and metastasis has been presented (Raz et al., 1987).

To learn more about the structure, organization, and regulation of expression of S-Lac lectins, we undertook the isolation and sequencing of the human genes for L-14 and for the related putative lectin that we previously identified (Gitt & Barondes, 1986). We here report the nucleotide sequence of the human L-14 gene including all exons and introns and over 250 base pairs of the upstream region. One significant finding is that upstream sequences suggest control by multiple regulatory elements including those that respond to glucocorticoids, heat (environmental) shock, and metals. We also describe the isolation and partial structural studies of the gene encoding the related putative lectin (Gitt & Barondes, 1986). Our results indicate that it is organized like L-14, providing further evidence for its relationship with the S-Lac lectin family.

Experimental Procedures

Materials. Unless otherwise noted, all DNA-modifying and restriction enzymes were purchased from Bethesda Research Laboratories (Gaithersburg, MD), all radiolabeled nucleotides were obtained from Amersham (Arlington Heights, IL), and all other reagents were from Sigma (St. Louis, MO).

Screening of Human Genomic Library. A human leukocyte-derived genomic library in an engineered derivative of the pJB8 cosmid vector, graciously provided by Drs. C. Lau and Y. W. Kan (Lau & Kan, 1983), was screened separately in duplicate with clone 1 (pHL14-1) and clone 2 (pHL14-2) probes (Gitt & Barondes, 1986), using the method of Hanahan and Meselson (1980). The clone 1 probe (P1) was prepared by *Hin*cII/*Eco*RI excision of 360 bp of the cDNA insert from pHL14-1 (Gitt & Barondes, 1986). The clone 2 probe (P-L-14) was prepared by excision of the 200 bp cDNA insert of pHL14-2 (Gitt & Barondes, 1986) by *Eco*RI. Probes were radiolabeled with [α-^{32}P]dCTP, using the BRL nick translation kit for the P1 probe, and by the random primer polymerization technique of Feinberg and Vogelstein (1984) for the PL-14 probe, using Amersham DNA polymerase 1 and random primers from Pharmacia (Piscataway, NJ).

Blots were prehybridized at 41 °C in 50.8% formamide, 50 mM sodium phosphate, pH 7.0, 1 × Denhardt's (0.02% polyvinyl pyrrolidone-360, 0.02% Ficoll 400, and 0.02% BSA), 1.5 mM EDTA, 3 × SSC (Maniatis et al., 1982), 20 μg/mL denatured salmon sperm DNA, and 150 μg/mL yeast tRNA and then hybridized in the same solution plus 10% dextran sulfate, also at 41 °C. Filters were washed in 2 × SSC/1 × Denhardt's for 45 min at room temperature, then twice in 0.1 × SSC/0.1% SDS at 50 °C for 45 min, air-dried, and autoradiographed using Kodak X-OMAT AR film and American Bionetics (Emeryville, CA) lanthanum oxybromide (blue III) intensifying screens. Areas on the filter replicas corresponding to positive signals were excised, and cell suspensions were then plated at varying densities and rescreened. This procedure was repeated until 100% of the colonies on a plate gave positive signals in hybridization.

Purification of Cloned DNA. DNA was purified from overnight cultures of the purified positive clones by the alkaline–SDS lysis method (Close & Rodriguez, 1982). Restriction digests, electrophoresis, ligation, transformation of competent *Escherichia coli* XL-1 Blue (Stratagene, La Jolla, CA), and other manipulations followed descriptions by Maniatis et al. (1982).

Southern Blotting. After electrophoresis, gels were photographed, then treated sequentially at room temperature with 2.0 M HCl for 15 min, 0.5 M NaOH/1.5 M NaCl for 30 min, 0.5 M Tris-HCl/1.5 M NaCl, pH 7.5, for 30 min, and 10 × SSC for 10 min, and transferred overnight in 10 × SSC to 0.45-μm nylon membranes (Micron Separations Inc., Westborough, MA). The membranes were then irradiated with ultraviolet light in a Stratagene (La Jolla, CA) Stratalinker for a total energy of 0.12 J. The blots were then prehybridized at 65 °C in 15% deionized formamide, 0.2 M Na_2HPO_4, pH 7.2, 1 mM EDTA, 1% BSA, and 7% SDS and incubated overnight at 58 °C in the same solution containing at least 10^6 cpm of probe. Blots were then washed at 58 °C in 2 × SSC/1% SDS for 45 min and then twice in 0.2 × SSC/0.1% SDS again at 58 °C and autoradiographed after air-drying.

Subcloning. Cosmid 182, containing the gene for L-14, was digested with *Pst*I, and the 1.5-kb PL-14-hybridizing fragment was subcloned into pBluescript (Stratagene, La Jolla, CA). The cosmid was also doubly digested with *Eco*RI and *Bam*HI, yielding a 1.0-kb hybridizing fragment, which was cloned into M13mp19 (Messing & Vieira, 1982). Additionally, a 1.0-kb *Pst*I/*Eco*RI fragment was cloned into M13mp19 and M13mp18. These four clones contained all the genomic DNA that hybridized with PL-14.

Since the PL-14 cDNA probe was not full-length, we used cosmid DNA as template and oligonucleotide primers described below in the polymerase chain reaction to isolate genomic DNA containing the remainder of the gene. PCR-generated fragments were sequenced directly by using double-strand sequence techniques, and also subcloned into pBluescript and sequenced by the Sanger technique.

The two P1-hybridizing *Pst*I fragments of cosmid 42, containing the three exons of the clone 1 gene, were cloned into pBluescript. The resulting plasmids were sequenced with vector primers.

Deletions. Nested deletions of the *Pst*I/*Eco*RI 1.0-kb insert in M13 were generated by the method described by Dale et al. (1985). For nested deletions of clones in pBluescript vectors, we followed the manufacturer's instructions, using

*Kpn*I and *Xho*I to cut the vector prior to *Exo*III digestion.

Polymerase Chain Reaction. The following oligonucleotides were synthesized based on the cDNA sequence of full-length cDNA for human L-14 (Couraud et al., 1989; Abbott & Feizi, 1989; Hirabayashi et al., 1989): (A) CAAGCCATGATT-GAGTCC, upstream sequence (−); (B) GGGCTGA-AAGCTTGTGCGCCTGCCCGGG, upstream sequence (+); (C) CCTGAATCTCAAACCTGG, exon 2 (+); (D) GCGTCAGGAGCCACCTCGCCTCGCACTCG, exon 2 (−); (E) GCAGGATCCGGTTGTTGCTGTCTTTGCC, exon 3 (−); where (+) and (−) denote sense and antisense oligonucleotides, respectively.

We also synthesized adaptor and $(dT)_{17}$-adaptor oligonucleotides for use in isolating the upstream region of the L-14 gene, using a protocol of Frohman et al. (1988).

The PCR reactions were done following the method of Frohman et al. (1988) using a Perkin-Elmer Cetus (Norwalk, CT) thermal cycler, with the following modifications. For Taq polymerase buffer, we used 61 mM Tris-HCl, pH 8.8, at room temperature, 6.1 mM $MgCl_2$, 155 μg/mL BSA, and 15 mM $(NH_4)_2SO_4$. We used 1 μg of human genomic DNA (Clontech, Palo Alto, CA) or cosmid DNA in the amplifications. Just prior to enzyme addition, the solution was boiled for 10 min and then immediately placed on ice. The samples were then placed at 72 °C for addition of 2.5 units of Taq polymerase (Perkin-Elmer Cetus) and overlaid with 30 μL of mineral oil. Annealing was done for 40 s at 64 °C for oligonucleotides B and D, and at 52 °C for oligonucleotides A, C, and E, adaptor, and $(dT)_{17}$-adaptor oligonucleotides. The annealing was followed by a 4-min extension period at 72 °C. The DNA was denatured for 40 s at 96 °C for the first five cycles and at 94 °C for the following 35 cycles. Products were separated on a 1% agarose gel, excised, and electroeluted.

PCR of the 5′ Nontranslated Region. Cosmid 182 DNA was digested with *Eco*RI, boiled for 10 min, immediately placed on ice, and then tailed with deoxyadenosine triphosphate using terminal deoxynucleotidyl transferase, following the manufacturer's instructions. Tailing proceeded for 10 min at 37 °C. The DNA was then boiled for 10 min, placed on ice immediately, and used for PCR, using adaptor, $(dT)_{17}$-adaptor, and oligonucleotide A.

Sequencing. Clones in pBluescript (Stratagene) or M13 (Messing & Vieira, 1982) were sequenced using Sequenase (United States Biochemicals, Cleveland, OH) by the modification of the Sanger method (Sanger et al., 1977) described by the enzyme manufacturer. Single-strand DNA was prepared as described in the pBluescript manual, using R408 helper phage. Direct sequencing without cloning into M13 or pBluescript was by the method described by Kraft et al. (1988). The Bionet SEQ, GEL, and GENED programs (Intelligenetics, Mountain View, CA) were used to analyze the sequences obtained with respect to homologies and secondary structural features.

Results

Using the cDNA probes PL-14 and P1 to screen 6×10^5 colonies of a human genomic library in a cosmid vector, we found 8 clones that hybridized with PL-14 and 12 that hybridized with P1. After the colonies were purified in successive screenings at lower density, DNA was isolated from several of them and digested with *Eco*RI. Southern blots were then examined with the cDNA probes used to screen the library. The genomic clones that hybridized with PL-14 gave rise to a band of 6.4 kb and another of varying size, ranging from 2.9 to 3.5 kb. Of the clones that hybridized with P1, all gave the same hybridization pattern, consisting of a major band of 6.3 kb and two minor bands of 4.4 and 2.4 kb. The genomic L-14 clone (cosmid 182) that gave rise to the largest variable band and one of the genomic clones that hybridized with P1 (cosmid 42) were selected for further study.

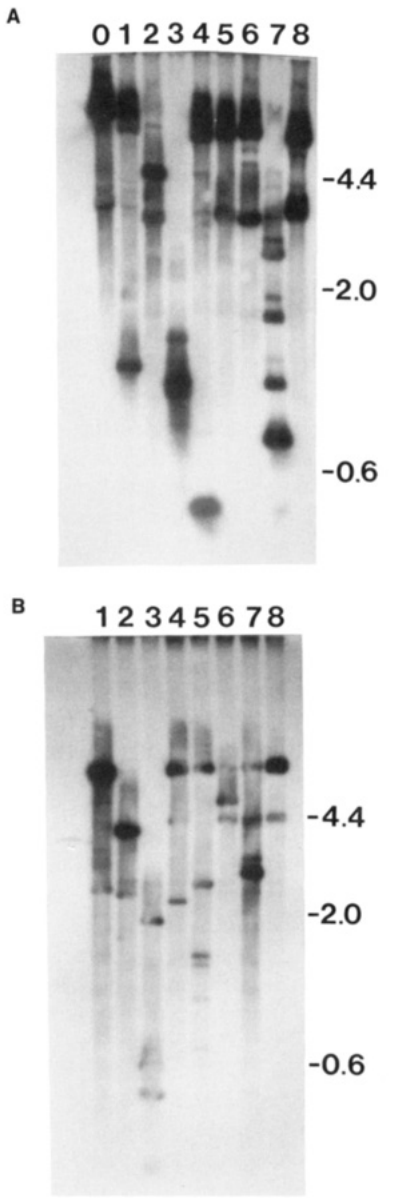

FIGURE 1: Southern blots of clones 182 (A) and 42 (B). DNA purified from these genomic clones was digested with restriction enzymes and electrophoresed. Blots were hybridized to the complementary radiolabeled cDNA probe. The lanes for both blots contain DNA restricted with the following enzymes: lane 0, *Bam*HI; lane 1, *Bam*HI/*Eco*RI; lane 2, *Hin*dIII/*Eco*RI; lane 3, *Pst*I/*Eco*RI; lane 4, *Kpn*I/*Eco*RI; lane 5, *Sst*I/*Eco*RI; lane 6, *Sal*I/*Eco*RI; lane 7, *Sma*I/*Eco*RI; lane 8, *Eco*RI. Size markers are in kilobases.

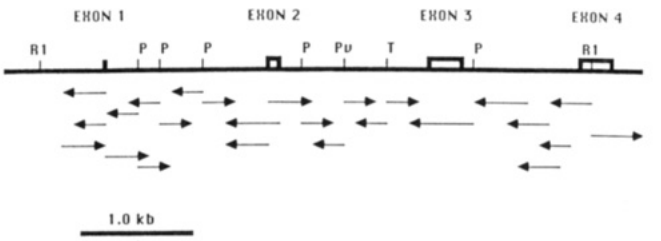

FIGURE 2: Strategy for sequencing clone 182. The strategy is depicted by arrows under a partial restriction map of the genomic DNA. Exonic regions are denoted by rectangles. Some of the restriction sites used in cloning are also indicated (R1 = *Eco*RI, P = *Pst*I, Pv = *Pvu*II, T = *Taq*I, and B = *Bam*HI).

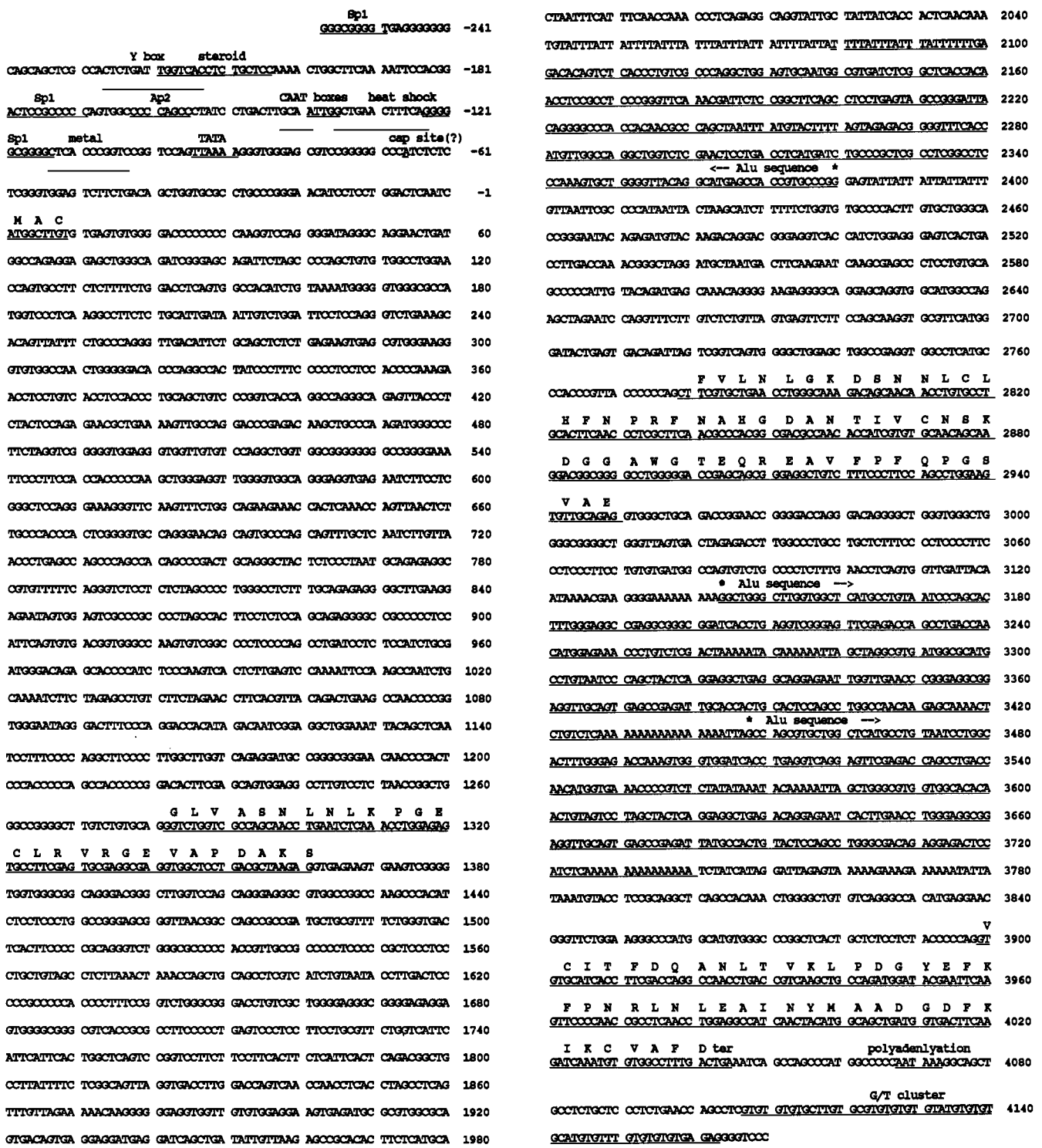

FIGURE 3: Sequence of the L-14 gene. Numbering starts at the translational start codon ATG, and the upstream sequence is denoted by negative numbers. Exons are identified by the translated amino acids noted above. Alu sequences and their orientation are represented by arrows over the corresponding sequences. Other features are noted above the sequence with lines below denoting extent of the specific sequence.

Cosmids 182 and 42 were doubly restricted with a panel of enzymes along with *Eco*RI, electrophoresed, and examined by Southern blotting. The results (Figure 1) guided our initial subcloning of the genomic fragments. PL-14 did not hybridize with cosmid 42, and P1 did not hybridize with cosmid 182, supporting the conclusion that each was encoded by a different gene.

The complete sequence of the human L-14 gene of cosmid 182 was determined by using the strategy summarized in Figure 2 and is presented in Figure 3. The gene contains four exons and three introns, the latter with several Alu sequences. The intron–exon junction sequences are all consistent with consensus sequences (YYYYYYYYYYNCAG/G and AG/GTAAGT for intron/exon and exon/intron, respectively, where Y = C or T, and N = A, C, G, or T; Padgett et al., 1986). The last two introns have three full copies of the Alu-type repetitive sequence (Deininger et al., 1981; Kariya et al., 1987). The intron 3 Alu sequence is oriented antiparallel, and the intron 4 Alu sequences are oriented head-to-tail and parallel to the direction of lectin gene transcription.

The predicted amino acid sequence of the exons is identical with that observed by direct sequencing of amino acids of L-14 isolated from human placenta (Hirabayashi & Kasai, 1988), except that the N-terminal methionine coded for by the clone

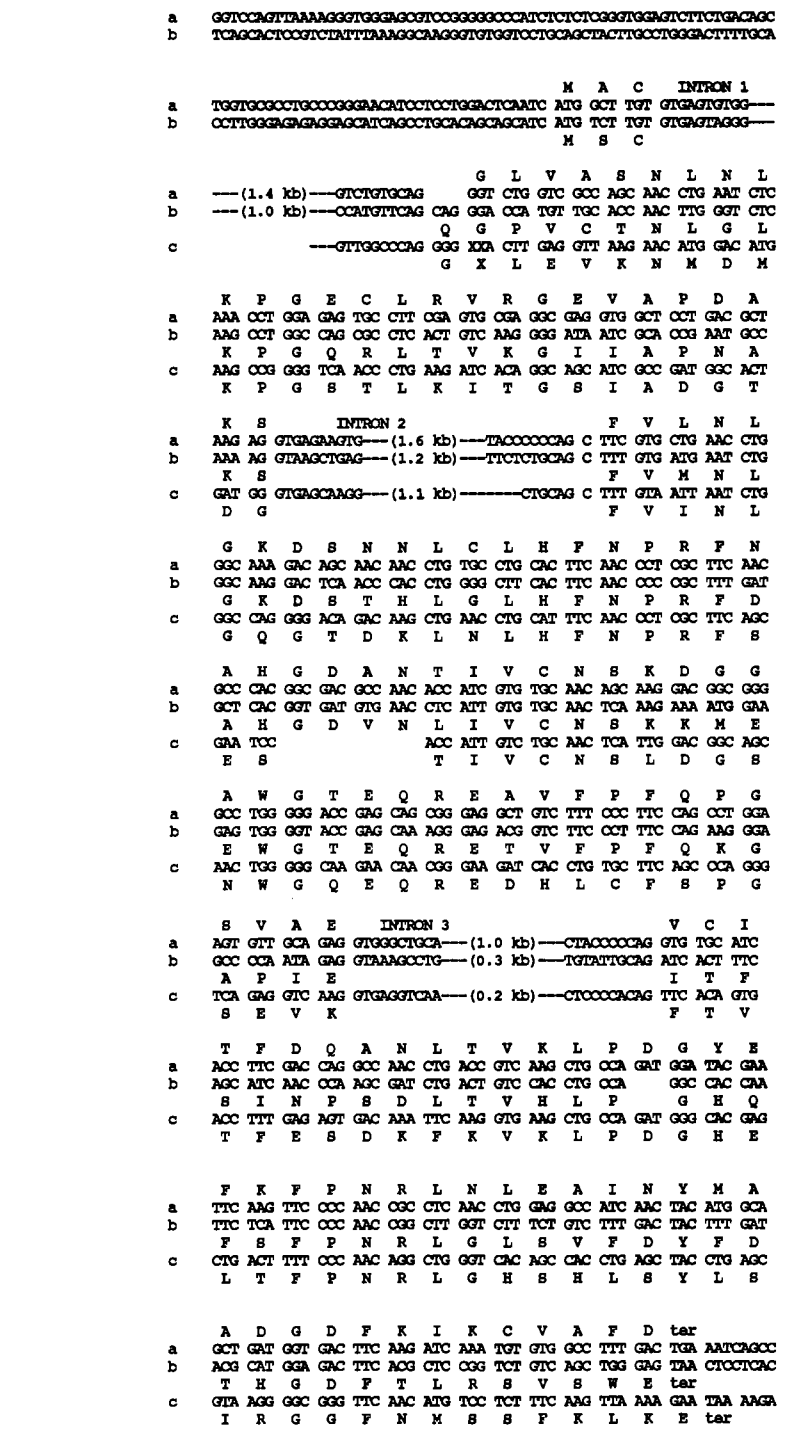

FIGURE 4: Partial sequence of clone 42 (c) and comparison with relevant portions of the sequence of clone 182 (a) and published (Ohyama & Kasai, 1988) partial genomic sequence of a chicken skin lectin (b). The features shown are exon/intron structure, sequence, and the deduced amino acid sequence.

is not present in the mature protein. The nucleotide sequence of the L-14 gene is also consistent with that found in three different studies of cDNAs encoding human L-14 (Couraud et al., 1989; Abbott & Feizi, 1989; Hirabayashi et al., 1989), but there are a number of differences from pHL14-2. The latter is, however, almost identical with L-14 cDNAs isolated from mouse tissues by Wilson et al. (1989) and in this laboratory (Cooper & Barondes, 1990). Since the cDNA library from which pHL14-2 was derived was made from a human tumor grown in a mouse, it probably arose from mouse tissue, as suggested by Wilson et al. (1989). However, the L-14 gene of cosmid 182 is clearly of human origin, having been derived

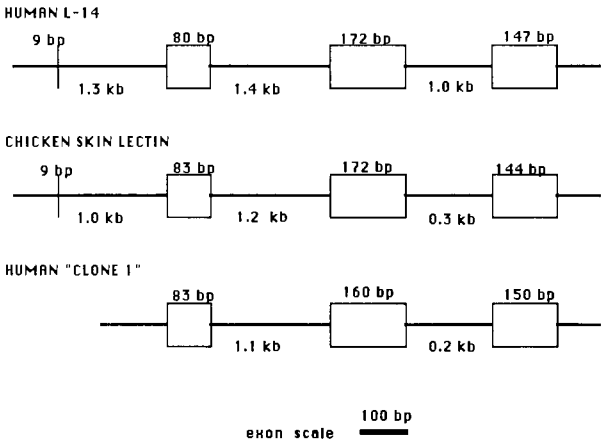

FIGURE 5: Schematic depiction of genomic organization of human L-14, the chicken skin lectin examined by Ohyama and Kasai (1988), and clone 42 (human "clone 1"). Intron dimensions are not to scale.

from a human library, with no such possibility of contamination.

Several other features of the L-14 gene are noted in Figure 3. The 3′ noncoding region of the gene has a consensus polyadenylation signal AATAAA (Fitzgerald & Shenk, 1981) 22 bp past the termination codon TGA. Another 33 bp downstream of the polyadenylation signal is an extensive copolymer sequence of alternating G's and T's, over 53 bases, with only a few substitutions. Smaller G/T clusters have been noted in the majority of eukaryotic 3′ sequences (McLauchlan et al., 1985; Platt, 1986) and may be involved in efficient processing of the mRNA (McLauchlan et al., 1985), formation of Z-DNA structure (Taboury & Taillandier, 1985), or binding to a cellular factor (Leith et al., 1988).

In the upstream region, the sequence surrounding the initiator methionine codon conforms with the consensus sequence for other genes, GGRNNATGG (Kozak, 1986), where R = A or G, and N = any nucleotide. Further upstream is a TATA box-like sequence located the correct distance from a consensus transcriptional initiation site (Breathnach & Chambon, 1981). Other putative upstream regulatory elements include three Sp1 binding site consensus sequences (Dynan & Tjian, 1985), an Ap2 binding site (Roesler et al., 1988), three CCAAT boxes (Efstratiadis et al., 1980; Graves et al., 1986), a heat shock element (Dynan & Tijian, 1985), a site similar to the consensus glucocorticoid response element (Yamamoto, 1985), a possible metal regulatory element (Karin et al., 1984), and a sequence that resembles the Y box of histocompatibility genes (Dorn et al., 1987; Kelly & Trowsdale, 1985).

We also determined some of the sequence of clone 42 from subclones of hybridizing *Pst*I fragments. Three exons were identified, with only two nucleotides difference from pHL14-1. The exon sequences, deduced amino acid sequences, and intron–exon boundaries are shown in Figure 4, sequence c.

Although we have not determined the complete structure of the clone 42 gene, these results were aligned with the related parts of clone 182 (Figure 4, sequence a) and a lectin from chicken skin (Figure 4, sequence b) whose genomic organization was reported by Ohyama and Kasai (1988). Comparison of the available information on the overall structure of these genes is shown schematically in Figure 5, and the relationship of the deduced amino acid sequence of the putative lectin encoded by clone 42 to the sequences of L-14 and L-30 is summarized in Figure 6.

Discussion

In this paper, we report the first complete genomic sequence

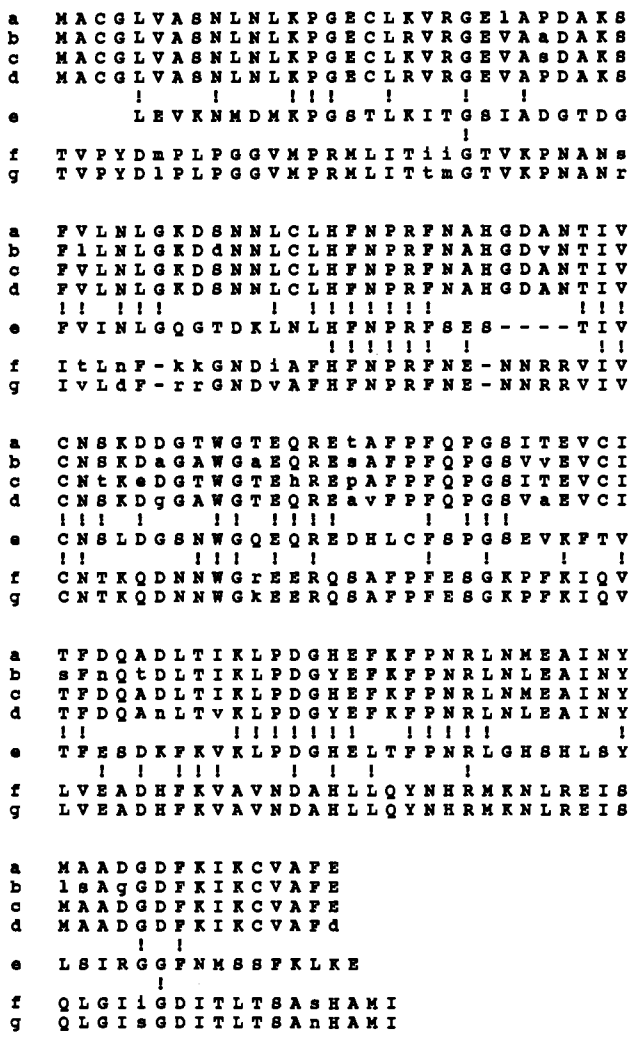

FIGURE 6: Comparison of protein sequences of members of the mammalian S-Lac lectin family. Two extensively studied lectins, L-14 (a–d) and L-30 (f, g), are compared with the partial sequence deduced from the cDNA designated "clone 1" (e) whose genomic counterpart is clone 42. The examples of L-14 (a–d) are from (a) rat (Clerch et al., 1988), (b) cow (Abbott et al., 1989), (c) mouse (Wilson et al., 1989; Cooper & Barondes, 1990), and (d) human (Couraud et al., 1989; Abbott & Feizi, 1989; Hirabayashi et al., 1989; this paper). Identical residues in all four are in capitals. The examples of L-30 are from (f) rat (Liu et al., 1985) and (g) mouse (Jia & Wang, 1988). The residues that are the same in both are in capitals. The available deduced sequence of the human "clone 1" putative lectin (e) that was derived from studies of pHL14-1 (Gitt & Barondes, 1986) and clone 42 is compared with the L-14 and L-30 sequences, with exclamation points indicating identities.

of an S-Lac lectin. When coupled with previous work with a chicken skin lectin and with the distinct putative lectin encoded by cDNA clone 42, some general features of the genomic organization of this family of structurally related lectins are suggested.

The gene for L-14 consists of four exons, splitting the regions of the coding sequence that are homologous between different lectins (Gitt & Barondes, 1986; Clerch et al., 1988; Paroutaud et al., 1987) into several possible domains. As in previous studies of cDNA, the genomic sequence shows no evidence of a secretory signal peptide, supporting evidence that secretion of L-14 is regulated by a novel mechanism (Cooper & Barondes, 1990). Exon 3 contains the most extensive region of homology between the soluble lectins including the apparently invariant HFNPRF and the residues WGTEQRE thought to be involved in galactoside binding (Paroutaud et al., 1987; Levi & Teichberg, 1981). The presence of Alu sequences in the introns surrounding exon 3 (Figure 3) and their implicated role in in vivo recombination (Lehrman et al., 1986; Jagadeeswaran et al., 1982; Henthorn et al., 1986; Nicholls et al., 1987) suggest the possible use of the exon 3 domain as a functional unit for glycoconjugate binding. Further evidence for this hypothesis is seen in the sequence of L-30, which has been shown to be an IgE binding protein (Liu et al., 1985; Albrandt et al., 1987). In L-30, the sequence homologous to L-14 exon 3 is surrounded by sequences homologous to the IgG Fc receptor gene (Albrandt et al., 1987).

The genomic structure of human L-14 is very similar to that of the chicken skin lectin published by Ohyama and Kasai (1988). Except for the added triplet specifying glutamine in exon 2 of chicken, the intron/exon boundaries occur at exactly the same positions in these two lectins. This is especially notable because the chicken skin lectin (Kasai et al., 1989) is distinct from chicken lactose lectin I (Barondes, 1984) which is probably the chicken form of L-14. In support of this relationship, both L-14 and chicken lactose lectin I are dimeric, are prominent in muscle, and exhibit similar developmental regulation (Barondes, 1984; Cooper & Barondes, 1990). Our partial results with the putative human lectin encoded by clone 42 support the generality of this organization, although the relationship of the clone 42 gene to the chicken skin lectin is not yet clear.

The upstream region of the L-14 gene suggests that it may be regulated in many ways. A possible heat shock element is found at position −130, diverging from the consensus (CNNGAANNTTCNNG; Dynan & Tjian, 1985), by only one nucleotide, a property shared with several other known heat-shock controlled genes (Hickey et al., 1986; Zuker et al., 1984). However, it should be noted that the presence of a heat-shock consensus sequence may not be sufficient for heat-shock regulated expression. Many heat-shock regulated genes also display developmental expression independent of heat-shock stimuli. Although no data exist for developmental expression in the human, L-14 is developmentally regulated in chicken and rat (Barondes, 1984; Beyer & Barondes, 1982; Oda & Kasai, 1983; Clerch et al., 1987).

A putative steroid-binding site is located at −210. The consensus sequence (T/CGGTNA/TCAA/TNTGTT/CCT; Yamamoto, 1985) differs from the lectin gene sequence at three nucleotides. There is evidence for glucocorticoid regulation of synthesis of rat L-14 (Clerch et al., 1987) and of chicken skin lectin (Oda et al., 1989). Yet another putative regulatory element is the metal regulatory element, of consensus sequence T/CG/TCGCCCGGCT/CC (Karin et al., 1984). The lectin sequence diverges at only two nucleotides, and is located at −110. We also found a sequence related to the Y box of histocompatibility genes (Dorn et al., 1987; Kelly et al., 1985), overlapping the putative steroid-binding site. It has the consensus sequence TTCTGATTGGTTAC of which the L-14 gene shares 12 of 14 nucleotides. A possible relationship of L-14 to the histocompatibility system is of interest since lectins have been implicated in a primitive defense system by binding surface glycoconjugates of invading organisms, thus facilitating phagocytosis (Ofek & Sharon, 1988).

In addition to the regulatory elements found immediately upstream of the lectin gene, the presence of intronic Alu sequences and the possibility of downstream Z-DNA suggested by a G/T cluster downstream of the polyadenylation signal may also add other levels of regulation. Alu sequences have been shown to bind protein factors (Perelygina et al., 1987), and at least some appear to modulate expression (Saffer & Thurston, 1989). Z-DNA has been shown to bind several

specific proteins (Leith et al., 1988) and modify nucleosome structure (Casanovas & Azorin, 1987), both of which may have an effect on gene expression.

Notable features in intron 1 include some small direct repeats, which may be targets for binding cellular factors. One, TCCTCTCCAG at position 873, is also found with one base change at positions 798 and 947. The other, AGAGGGGC, is at positions 827 and 884. The significance of these sequences is not known, but the first sequence resembles a consensus intron/exon boundary which raises the possibility of an alternative splicing site. Intron 1 also contains several Sp1 binding site consensus sequences. Such sequences can play a role in regulation even downstream of the transcriptional initiation site.

The significance of these multiple potential mechanisms for regulation of the L-14 gene remains to be determined. They are consistent with evidence that the role of this lectin may differ in developing mature tissues and in the many cell types in which it is expressed.

Acknowledgments

We thank Michael Frohman and Michael Dush for their assistance in the polymerase chain reaction and Mark Abbott for information relating to the 5′ end of the L-14 gene.

Registry No. DNA (human lactose-binding lectin L-14 gene), 130726-87-3; lectin L-14 (human lactose-binding protein moiety reduced), 128394-29-6; lectin (human clone β-galactoside-binding precursor protein moiety reduced), 122463-53-0.

References

Abbott, W. M., & Feizi, T. (1989) *Biochem. J. 259*, 291–294.
Abbott, W. M., Mellor, A., Edwards, Y., & Feizi, T. (1989) *Biochem. J. 259*, 283–290.
Albrandt, K., Orida, N. K., & Liu, F.-T. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 6859–6863.
Barondes, S. H. (1984) *Science 223*, 1259–1264.
Beyer, E. C., & Barondes, S. H. (1982) *J. Cell Biol. 92*, 23–27.
Breathnach, R., & Chambon, P. (1981) *Annu. Rev. Biochem. 50*, 349–383.
Casanovas, J. M., & Azorin, F. (1987) *Nucleic Acids Res. 15*, 8899–8918.
Cerra, R. F., Gitt, M. A., & Barondes, S. H. (1985) *J. Biol. Chem. 260*, 10474–10477.
Clerch, L. B., Whitney, P. L., & Massaro, D. (1987) *Biochem. J. 245*, 683–690.
Clerch, L. B., Whitney, P., Hass, M., Brew, K., Miller, T., Werner, R., & Massaro, D. (1988) *Biochemistry 27*, 692–699.
Close, T. J., & Rodriguez, R. L. (1982) *Gene 20*, 305–316.
Cooper, D. N. W., & Barondes, S. H. (1990) *J. Cell Biol.* (in press).
Couraud, P. O., Casentini-Borocz, D., Bringman, T. S., Griffith, J., McGrogan, M., & Nedwin, G. E. (1989) *J. Biol. Chem. 264*, 1310–1316.
Dale, R. M., McClure, B. A., & Houchins, J. P. (1985) *Plasmid 13*, 31–41.
Deininger, P. L., Jolly, D. J., Rubin, C. M., Friedmann, T., & Schmid, C. W. (1981) *J. Mol. Biol. 151*, 17–33.
Dorn, A., Bollekens, J., Staub, A., Benoist, C., & Mathis, D. (1987) *Cell 50*, 863–872.
Dynan, W. S., & Tjian, R. (1985) *Nature 316*, 774–778.
Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. Z., DeRiel, J. K., Forget, B. G., Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C., & Proudfoot, N. J. (1980) *Cell 21*, 653–668.
Feinberg, A. P., & Vogelstein, B. (1984) *Anal. Biochem. 137*, 266–267.
Fitzgerald, M., & Shenk, T. (1981) *Cell 24*, 251–260.
Frohman, M. A., Dush, M. K., & Martin, G. R. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 8998–9002.
Gitt, M. A., & Barondes, S. H. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 7603–7607.
Graves, B. J., Johnson, P. F., & McKnight, S. L. (1986) *Cell 44*, 565–576.
Hanahan, D., & Meselson, M. (1980) *Gene 10*, 63–67.
Henthorn, P. S., Mager, D. L., Huisman, T. H. J., & Smithies, O. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 5194–5198.
Hickey, E., Brandon, S. E., Potter, R., Stein, G., Stein, J., & Weber, L. A. (1986) *Nucleic Acids Res. 14*, 4127–4145.
Hinek, A., Wrenn, D. S., Mecham, R. P., & Barondes, S. H. (1988) *Science 239*, 1539–1541.
Hirabayashi, J., & Kasai, K.-I. (1988) *J. Biochem. 104*, 1–4.
Hirabayashi, J., Ayaki, H., Soma, G., & Kasai, K. (1989) *Biochim. Biophys. Acta 1008*, 85–91.
Hynes, M. A., Gitt, M. A., Barondes, S. H., Jessell, T. M., & Buck, L. B. (1990) *J. Neurosci. 10*, 1004–1013.
Jagadeeswaran, P., Tuan, D., Forget, B. G., & Weissman, S. M. (1982) *Nature 296*, 469–470.
Jia, S., & Wang, J. L. (1988) *J. Biol. Chem. 263*, 6009–6011.
Karin, M., Haslinger, A., Holtgreve, H., Richards, R. I., Krauter, P., Westphal, H. M., & Beato, M. (1984) *Nature 308*, 513–519.
Kariya, Y., Kato, K., Hayashizaki, Y., Himeno, S., Tarui, S., & Matsubara, K. (1987) *Gene 53*, 1–10.
Kasai, K., Hirabayashi, J., Ohyama, Y., & Sakakura, Y. (1989) Abstract for International Symposium on Lectins, Nof Ginosar, Israel.
Kelly, A., & Trowsdale, J. (1985) *Nucleic Acids Res. 13*, 1607–1621.
Kozak, M. (1986) *Cell 44*, 283–292.
Kraft, R., Tardiff, J., Krauter, K. S., & Leinwand, L. A. (1988) *BioTechniques 6*, 544–549.
Lau, Y.-F., & Kan, Y. W. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 5225–5229.
Leffler, H., & Barondes, S. H. (1986) *J. Biol. Chem. 261*, 10119–10126.
Leffler, H., Masiarz, F. R., & Barondes, S. H. (1989) *Biochemistry 28*, 9222–9229.
Lehrman, M. A., Russell, D. W., Goldstein, J. L., & Brown, M. S. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3679–3683.
Leith, I. R., Hay, R. T., & Russell, W. C. (1988) *Nucleic Acids Res. 16*, 8277–8289.
Levi, G., & Teichberg, V. I. (1981) *J. Biol. Chem. 256*, 5735–5740.
Liu, F. T., Albrandt, K., Mendel, E., Kulczycki, A., & Orida, N. K. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 4100–4104.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
McLauchlan, J., Gaffney, D., Whitton, J. L., & Clements, J. B. (1985) *Nucleic Acids Res. 13*, 1347–1368.
Merkle, R. K., Zhou, Q., Schultz, T. K., Harper, W. B., & Cummings, R. D. (1989) *Arch. Biochem. Biophys. 274*, 404–416.
Messing, J., & Vieira, J. (1982) *Gene 19*, 269–276.
Nicholls, R. D., Fischel-Ghodsian, N., & Higgs, D. R. (1987) *Cell 49*, 369–378.
Oda, Y., & Kasai, K. (1983) *Biochim. Biophys. Acta. 761*, 237–245.

Oda, Y., Ohyama, Y., Obinata, A., Endo, H., & Kasai, K.-I. (1989) *Exp. Cell Res. 182*, 33–43.
Ofek, I., & Sharon, N. (1988) *Infect. Immun. 56*, 539–547.
Ohyama, Y., & Kasai, K.-I. (1988) *J. Biochem. 104*, 173–177.
Ohyama, Y., Hirabayashi, J., Oda, Y., Ohno, S., Kawasaki, H., Suzuki, K., & Kasai, K. (1986) *Biochem. Biophys. Res. Commun. 134*, 51–56.
Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S., & Sharp, P. A. (1986) *Annu. Rev. Biochem. 55*, 1119–1150.
Paroutaud, P., Levi, G., Teichberg, V. I., & Strosberg, A. D. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 6345–6348.
Perelygina, L. M., Tomilin, N. V., & Podgornaya, O. I. (1987) *Mol. Biol. Rep. 12*, 111–116.
Platt, T. (1986) *Annu. Rev. Biochem. 55*, 339–372.
Raz, A., Meromsky, L., Zvibel, I., & Lotan, R. (1987) *Int. J. Cancer 39*, 353–360.
Raz, A., Carmi, P., & Pazerini, G. (1988) *Cancer Res. 48*, 645–649.
Roesler, W. J., Vandenbark, G. R., & Hanson, R. W. (1988) *J. Biol. Chem. 263*, 9063–9066.
Saffer, J. D., & Thurston, S. J. (1989) *Mol. Cell. Biol. 9*, 355–364.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Sparrow, C. P., Leffler, H., & Barondes, S. H. (1987) *J. Biol. Chem. 262*, 7383–7390.
Taboury, J. A., & Taillandier, E. (1985) *Nucleic Acids Res. 13*, 4469–4483.
Wilson, T. J. G., Firth, M. N., Powell, J. T., & Harrison, F. L. (1989) *Biochem. J. 261*, 847–852.
Yamamoto, K. R. (1985) *Annu. Rev. Genet. 19*, 209–252.
Zuker, C., Cappello, J., Lodish, H. F., George, P., & Chung, S. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 2660–2664.

Structure and Transcriptional Regulation of Protein Phosphatase 2A Catalytic Subunit Genes†

Yeesim Khew-Goodall, Regina E. Mayer, Francisca Maurer, Stuart R. Stone, and Brian A. Hemmings*
Friedrich Miescher-Institut, Postfach 2543, Basel CH-4002, Switzerland
Received May 15, 1990; Revised Manuscript Received August 30, 1990

ABSTRACT: The α and β isoforms of the human protein phosphatase 2A catalytic subunit are encoded by distinct genes whose expression appears to be differentially regulated. To obtain a better understanding of the mechanism(s) that regulate(s) the expression of these two transcripts, we have cloned the genes encoding both isoforms. Both genes (each ~30 kbp) are composed of seven exons and six introns which intervene at identical locations, suggesting that they were derived from a common ancestral gene. However, the 5′ upstream regions as well as the regions encoding the 5′ and 3′ untranslated sequences of each mRNA are different. The promoters of both genes are very G+C rich and lack the TATA and CCAAT sequences typical of many housekeeping genes. The Cα gene contains several potential Sp1 binding sites and a potential cAMP-responsive element. Northern analysis using RNAs isolated from several different human cell lines showed that the steady-state Cα mRNA was, in general, more abundant than the Cβ mRNA. To determine whether the promoters regulate the differential Cα and Cβ RNA expression, they were fused to the reporter gene chloramphenicol acetyltransferase and transiently expressed in HeLa cells. Expression from the Cα promoter was 7–10 times stronger than that from the Cβ promoter, which paralleled the endogenous Cα and Cβ mRNA levels in HeLa cells. These data suggest that the steady-state levels of the Cα and Cβ mRNAs, are due, at least in part, to different promoter activities.

Protein phosphatases play an essential role in the regulation of many processes ranging from cellular metabolism to involvement in cell cycle events [reviewed in Cohen (1989)]. Mutations resulting in cell cycle arrest in the fungus *Aspergillus nidulans* (Doonan & Morris, 1989) and in yeast (Ohkura et al., 1989; Booher et al., 1989) have been shown to be located within a protein phosphatase 1 gene, demonstrating that this phosphatase is required for completion of the cell cycle. Furthermore, the discovery that protein phosphatases 1 and 2A are both strongly inhibited by okadaic acid, a non-phorbol ester tumor promoter, has led to the proposal that protein phosphatases may play a role in tumor suppression (Bialojan & Takai, 1988; Haystead et al., 1989). However, the mechanisms responsible for the activation–inactivation of protein phosphatase activity in these processes are poorly understood.

Several types of protein phosphatases have been identified at the protein level [reviewed by Ballou and Fischer (1986) and Cohen (1989)]. Three different holoenzyme forms of protein phosphatase 2A have been purified from a number of tissues and in all cases consist of a catalytic (C) subunit of 36 kDa in association with one or more regulatory subunits of 55, 65, and 72 kDa (Crouch & Safer, 1980; Pato & Adelstein, 1983; Tung et al., 1985; Waelkens et al., 1987; Usui et al., 1988). The C subunit has been cloned from various species, and at least two isoforms (termed Cα and Cβ) have been identified (Green et al., 1987; Stone et al., 1987, 1988; da Cruz e Silva et al., 1987; da Cruz e Silva & Cohen, 1987; Arino et al., 1988; Hemmings et al., 1988; Kitagawa et al., 1988a,b). Furthermore, two isoforms of the 65-kDa subunit have also been cloned (Hemmings et al., 1990). Thus, the structure of protein phosphatase 2A is likely to be even more complex than initially perceived.

The amino acid sequences of the Cα and Cβ proteins, deduced from the cDNA sequences, are 97% identical, with 7 of the 8 amino acid differences being clustered in the first 30

† The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J05297.
* Author to whom correspondence should be addressed.

0006-2960/91/0430-89$02.50/0