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

residues of the N-terminus. The mRNAs encoding both isoforms have been detected in all mammalian species studied, suggesting that the two phosphatase C subunits may have distinct functions and interact with different regulatory subunits. Also noteworthy is the fact that although the noncoding regions of the Cα and Cβ mRNAs share no sequence homology, the 3′ untranslated regions of each isoform are highly conserved among mammals. Northern analysis revealed that the Cα mRNA is expressed at relatively high levels in various porcine, rat, bovine, and rabbit tissues, particularly in brain (Khew-Goodall & Hemmings, 1988; Kitagawa et al., 1988a; Green et al., 1987; da Cruz e Silva et al., 1987). The Cβ mRNA showed a similar pattern of expression but at levels about 8–12-fold lower than the Cα mRNA (Khew-Goodall & Hemmings, 1988). The α and β transcripts encoding the 65-kDa subunit are also expressed in a tissue-specific manner (unpublished data).

Together, these observations suggest that regulation of protein phosphatase 2A may extend from transcriptional to posttranslational events. Understanding the control mechanisms that operate to coordinate the different genes encoding the catalytic and regulatory subunits is an important step toward elucidating the physiological roles of the different holoenzymes of protein phosphatase 2A. We are examining the role of transcriptional control in regulating protein phosphatase 2A levels. We report here the isolation and structural characterization of the genes encoding the Cα and Cβ isoforms. In addition, we have shown that the differential expression of the Cα and Cβ transcripts is due, at least in part, to the different strengths of the two promoters.

Experimental Procedures

Isolation of Genomic Clones. A human leukocyte genomic library generated by partial *Mbo*I digestion (Clonetech Labs., Inc.) and a human placenta genomic library (a gift from Dr. S. Shibahara, Friedrich Miescher-Institut), both cloned in EMBL3 vector, were screened. Clones λPC6c and λPC7-2 were isolated by using the full-length porcine Cα cDNA (Stone et al., 1987). They were found to encode 3′ portions of the Cα and Cβ genes, respectively, based on their hybridization to oligonucleotide or cDNA probes specific for the α or β forms. The Cα- and Cβ-specific cDNA probes were 3′ noncoding fragments corresponding to nucleotides 940–1346 (407 bp) of the human Cα cDNA (Stone et al., 1988) and to nucleotides 996–1432 (427 bp) of the human Cβ cDNA (Hemmings et al., 1988), respectively. Subsequently, three more clones (λPC15, λPC301, and λPC11-3) were isolated by using probes encoding the 5′ end of the human Cα and Cβ cDNAs; clone λPC15 was isolated by using a 134 bp fragment corresponding to nucleotides −86 to +48 of the human Cα cDNA, clone λPC301 with a 280 bp fragment corresponding to nucleotides −20 to +260 of the human Cβ cDNA, and clone λPC11-3 with an 86 bp fragment corresponding to nucleotides −20 to +66 of the human Cβ cDNA. These clones were determined to encode Cα or Cβ by hybridization with oligonucleotide probes specific for Cα and Cβ. All cDNA probes were labeled with [α-^{32}P]dATP by using the random-primer method (Feinberg & Vogelstein, 1984), and oligonucleotide probes were labeled with T_4 polynucleotide kinase and [γ-^{32}P]ATP. Filters were hybridized under standard conditions as described in Maniatis et al. (1982).

Construction of Physical Maps of the Cα and Cβ Genes. A simple restriction map was initially constructed by using *Bam*HI, *Eco*RI, *Hin*dIII, and *Sal*I for all the genomic clones isolated, and all fragments which were found to hybridize to the respective cDNA probes were subcloned into pUC18, Bluescript (Stratagene, San Diego, CA), or M13 vectors for further mapping and sequencing. Single- or double-stranded sequencing was carried out by the dideoxynucleotide chain termination method using [α-^{35}S]dATP and Sequenase (United States Biochemical, Cleveland, OH). In the 5′ untranslated and the 5′ flanking regions, sequences were determined in both directions, and where the G+C content of the sequence was particularly high, sequencing reactions were also carried out using dITP in place of dGTP.

The sequences of the individual exons were determined by sequencing from within each exon toward the boundaries using oligonucleotide primers complementary to the known cDNA sequences. About 100–200 bp at each end of all introns, except the 3′ end of intron 1 and the 5′ end of intron 2 of the Cβ gene (which were absent from the clones obtained), was sequenced. The intron sizes were determined by restriction mapping using enzymes which cut only once per exon in combination with DNA hybridization using oligonucleotide probes.

Northern Hybridization Analysis. HeLa, MDA-MB231, MFM, MCF7, T-47D, Bowes, TE 671, Calu-1, EJ, A 1146, and A-431 cells were cultured in DMEM supplemented with 10% fetal calf serum at 37 °C, 6% CO_2. Total cellular RNA was isolated by using the guanidine isothiocyanate–acid phenol method (Chomczynski & Sacchi, 1987) followed by selective precipitation of the RNA in 2.5M LiCl. Twenty micrograms of each RNA was electrophoresed on 1% formaldehyde–agarose gels and electrotransferred onto nylon filters (Zeta-Probe, Bio-Rad). The filters were hybridized and washed as previously described (Khew-Goodall & Hemmings, 1988). The relative levels of the Cα and Cβ mRNAs were obtained by densitometric scanning of the autoradiograms.

RNase Protection and S1 Nuclease Analyses. Antisense RNA probes for RNase protection analysis were made by subcloning restriction fragments from nucleotide −1400 to +13 of the α gene and from −1295 to +68 or −740 to −160 of the β gene (Figure 4) into the pGEM4 vector (Promega, Madison, WI). The resulting plasmids were linearized at positions −315 (*Ava*I) and −650 (*Sca*I) for the α and β probes, respectively, and transcribed with SP6 polymerase in the presence of 50 μCi of [α-^{32}P]GTP. Full-length transcription products were gel-purified, and 5×10^4 cpm of each probe was hybridized to 100 μg of total RNA isolated from HeLa cells. Hybridization and digestion with RNases A and T_1 were essentially carried out as described by Melton et al. (1984) except that 20 mM NaCl and 100 mM LiCl were used in the digestion buffer.

S1 nuclease mapping was carried out essentially as described by Berk and Sharp (1977). The probes used were a 3 kbp *Eco*RI–*Xmn*I fragment and a 1.3 kbp *Hin*dIII–*Pvu*II fragment for the α and β probes, respectively (see Figure 4, bottom). The Cα and Cβ probes were labeled at the 5′ end of the *Xmn*I and *Pvu*II sites, respectively, using T_4 polynucleotide kinase and [γ-^{32}P]ATP. Poly(A+) RNA was prepared from total RNA by oligo(dT)–cellulose chromatography (Aviv & Leder, 1972); 10–20 μg of poly(A+) RNA was hybridized with 4×10^4 cpm of each probe.

Promoter Analysis. The pSV2CAT plasmid (Gorman, 1985) was modified by replacing the *Acc*I–*Hin*dIII fragment of the SV40 promoter with a 31 bp polylinker containing multiple restriction sites to generate the p0-CAT plasmid. An *Acc*I–*Nae*I (nt −1160 to −20) fragment of the Cα gene and an *Eco*RI–*Bam*HI (nt −1286 to −23) fragment from the Cβ gene were ligated into p0-CAT to generate the plasmids pα-CAT and pβ-CAT. The promoter-CAT plasmids were transfected into HeLa cells by the calcium phosphate proce-

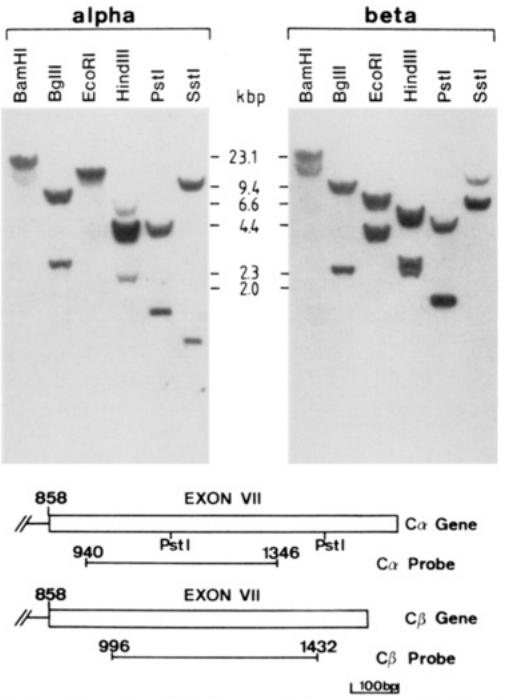


FIGURE 1: Southern hybridization analysis of human DNA. Genomic DNA (20 μg) isolated from HeLa cells was digested to completion by using the indicated restriction enzymes and separated on 0.7% agarose gels. The filters were hybridized to the Cα-specific (left panel) or Cβ-specific (right panel) cDNA probes (shown below). DNA preparation and hybridization conditions were as described by Maniatis et al. (1982). The Cα- and Cβ-specific probes were generated from the 3′ noncoding regions. The numbers correspond to nucleotide positions shown in panels A and B of Figure 3 for the Cα and Cβ genes, respectively. With the exception of the *Pst*I site in the Cα probe, none of the restriction sites used in the Southern analyses were present in the regions spanned by the Cα- and Cβ-specific probes.

dure, and the CAT activity was assayed 48 h after transfection (Gorman, 1985). The assays were carried out at three different protein concentrations, and only values in the linear range of the assay were used. Protein concentrations were determined by the method of Bradford (1976).

Results

Complexity of the Phosphatase 2A Gene Family. To determine the number of genes encoding each isoform of the protein phosphatase 2A C subunit, Southern analyses of human genomic DNA were carried out using Cα- and Cβ-specific cDNA probes derived from the 3′ noncoding regions of the human Cα (nt 940–1346; Stone et a., 1988) and Cβ (nt 996–1432; Hemmings et al., 1988) cDNAs (see Figure 1). The sequences encoded by both these probes were contained entirely within exon VII and do not contain any sites for the restriction enzymes used in the Southern analyses with the exception of a *Pst*I site located within the region used for the Cα-specific probe. Different sets of hybridizing fragments were observed with each of the two probes. When the Cα-specific probe was used, *Bam*HI, *Bgl*II, *Eco*RI, and *Sst*I digestions gave two hybridizing bands each, with one band usually hybridizing more strongly. Digestion with *Pst*I resulted in three hybridizing bands; the presence of a *Pst*I site within the region spanned by the Cα probe would account for two of the three bands observed (the small 350 bp band is barely visible in this autoradiogram presumably due to inefficient transfer). *Hin*dIII digestion also gave three hybridizing bands. These results are consistent with the α-specific probe hybridizing to two different genes, the Cα gene and a related gene or pseudogene (see below) which contains a *Hin*dIII site within the region spanned by the probe. When the Cβ-specific probe was used, digestion with five of the six enzymes gave two hybridizing bands; *Hin*dIII digestion gave three hybridizing bands. These results also suggest that there are two genes hybridizing to the Cβ probe, one of which contains a *Hin*dIII site within the region spanned by the probe (see below).

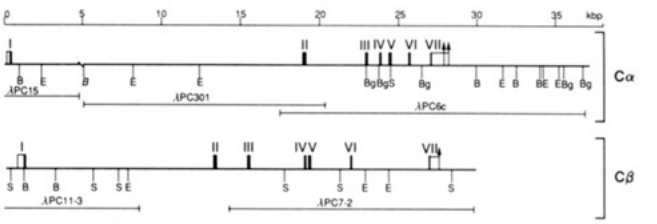
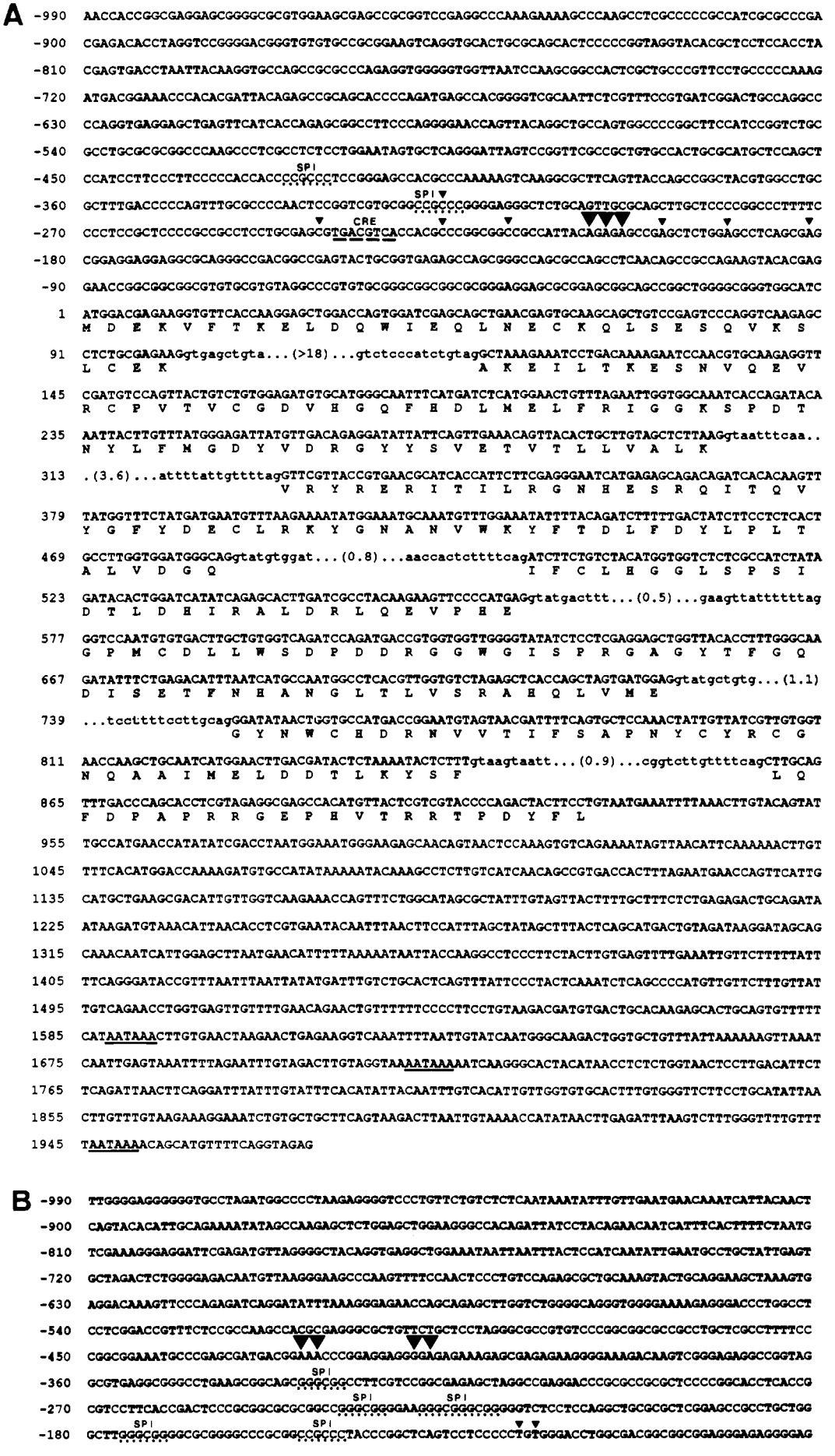


FIGURE 2: Structure of the Cα and Cβ genes. The organization of the Cα and Cβ genes is shown. Introns are indicated by solid lines and exons by boxes above the solid lines. Filled and open boxes represent translated and untranslated sequences, respectively. Arrowheads indicate the positions of the polyadenylation signals (AATAAA). A partial restriction map is shown where B = *Bam*HI, Bg = *Bgl*II, E = *Eco*RI, and S = *Sst*I. In the Cα gene, *Bgl*II and *Sst*I sites were only mapped for clone λPC6c. The locations of the λ genomic clones isolated are illustrated below the gene structures. The location of exon II of the Cβ gene was determined by polymerase chain reaction (Saiki et al., 1985) amplification of the region between exons II and III using oligonucleotide primers within these exons and genomic DNA from HeLa cells as template.

Structure and Organization of the Cα and Cβ Genes. Two human genomic libraries (prepared from leukocyte and placenta DNA) were screened, and several clones encoding different portions of the Cα and Cβ genes were isolated. Physical maps of the two genes were constructured from these clones, and the regions of the genes spanned by these clones are shown in Figure 2. Clones λPC15, λPC301, and λPC6c span 28 kbp of the Cα gene that includes all the exons and at least 1 kbp of 5′ flanking sequence. However, a portion of intron 1 of unknown size was not encoded by any of the clones isolated.

Clones λPC11-3 and λPC7-2 span 21 kbp of the Cβ gene. After extensively screening the two libraries, we were unable to isolate clones that bridged the regions spanned by λPC7-2 and λPC11-3; this region includes exon II. The full length of the Cβ gene was, therefore, established by Southern analysis of human genomic DNA using DNA probes derived from the 3′ end of λPC11-3 and the 5′ end of λPC7-2. Both these probes hybridized to common *Sst*I (10.5 kbp) and *Eco*RI (15 kbp) fragments. The full length of the Cβ gene was detemined to be 27 kbp by using this analysis.

In the course of screening these libraries, we also identified several clones which encoded two different pseudogenes. One of these shared significant sequence homology (>80%) with the Cα gene both in the coding and in the 3′ noncoding regions. However, the exon–intron structure of this gene was completely different from that of the Cα and Cβ genes; sequence analysis (data not shown) revealed four exons spread over approximately 5 kbp. The putative coding regions contained three in-frame termination codons (in exons I, II, and III) and is therefore unlikely to encode a functional protein. Physical mapping (data not shown) revealed that this gene accounts for the second set of hybridizing bands detected by the Cα probe in Figure 1. The presence of a *Hin*dIII site within the region hybridizing to the probe resulted in an addition hybridizing band observed upon digestion of genomic DNA with *Hin*dIII. The second pseudogene (data not shown) appears to be the result of retroinsertion of a partially processed Cβ transcript. It contained a 7 bp deletion at the boundary of exons II and III (which were fused in this gene) and would result in a frame-shift. This Cβ pseudogene probably accounts for the second set of hybridizing bands detected by the Cβ-specific probe (Figure 1).


A
-990 AACCACCGGCGAGGAGCGGGGCGCGTGGAAGCGAGCCGCGGTCCGAGGCCCAAAGAAAAGCCCAAGCCTCGCCCCCGCCATCGCGCCCGA
-900 CGAGACACCTAGGTCCGGGGACGGGTGTGTGCCGCGGAAGTCAGGTGCACTGCGCAGCACTCCCCCGGTAGGTACACGCTCCTCCACCTA
-810 CGAGTGACCTAATTACAAGGTGCCAGCCGCGCCCAGAGGTGGGGGTGGTTAATCCAAGCGGCCACTCGCTGCCCGTTCCTGCCCCCAAAG
-720 ATGACGGAAACCCACACGATTACAGAGCCGCAGCACCCCAGATGAGCCACGGGGTCGCAATTCTCGTTTCCGTGATCGGACTGCCAGGCC
-630 CCAGGTGAGGAGCTGAGTTCATCACCAGAGCGGCCTTCCCAGGGGAACCAGTTACAGGCTGCCAGTGGCCCCGGCTTCCATCCGGTCTGC
-540 GCCTGCGCGCGGCCCAAGCCCTCGCCTCTCCTGGAATAGTGCTCAGGGATTAGTCCGGTTCGCCGCTGTGCCACTGCGCATGCTCCAGCT
-450 CCATCCTTCCCTTCCCCCACCACCCCGCCCTCCGGGAGCCACGCCCAAAAAGTCAAGGCGCTTCAGTTACCAGCCGGCTACGTGGCCTGC
-360 GCTTTGACCCCCAGTTTGCGCCCCAACTCCGGTCGTGCGGCCGCCCGGGGAGGGCTCTGCAGTTGCGCAGCTTGCTCCCCGGCCCTTTTC
-270 CCCTCCGCTCCCCGCCGCCTCCTGCGAGCGTGACGTCACCACGCCCGGCGGCCGCCATTACAGAGAGCCGAGCTCTGGAGCCTCAGCGAG
-180 CGGAGGAGGAGGCGCAGGGCCGACGGCCGAGTACTGCGGTGAGAGCCAGCGGGCCAGCGCCAGCCTCAACAGCCGCCAGAAGTACACGAG
-90 GAACCGGCGGCGGCGTGTGCGTGTAGGCCCGTGTGCGGGCGGCGGCGCGGGAGGAGCGCGGAGCGGCAGCCGGCTGGGGCGGGTGGCATC
1 ATGGACGAGAAGGTGTTCACCAAGGAGCTGGACCAGTGGATCGAGCAGCTGAACGAGTGCAAGCAGCTGTCCGAGTCCCAGGTCAAGAGC
M D E K V F T K E L D Q W I E Q L N E C K Q L S E S Q V K S
91 CTCTGCGAGAAGgtgagctgta...(>18)...gtctcccatctgtagGCTAAAGAAATCCTGACAAAAGAATCCAACGTGCAAGAGGTT
L C E K A K E I L T K E S N V Q E V
145 CGATGTCCAGTTACTGTCTGTGGAGATGTGCATGGGCAATTTCATGATCTCATGGAACTGTTTAGAATTGGTGGCAAATCACCAGATACA
R C P V T V C G D V H G Q F H D L M E L F R I G G K S P D T
235 AATTACTTGTTTATGGGAGATTATGTTGACAGAGGATATTATTCAGTTGAAACAGTTACACTGCTTGTAGCTCTTAAGgtaatttcaa..
N Y L F M G D Y V D R G Y Y S V E T V T L L V A L K
313 .(3.6)...attttattgttttagGTTCGTTACCGTGAACGCATCACCATTCTTCGAGGGAATCATGAGAGCAGACAGATCACACAAGTT
V R Y R E R I T I L R G N H E S R Q I T Q V
379 TATGGTTTCTATGATGAATGTTTAAGAAAATATGGAAATGCAAATGTTTGGAAATATTTTACAGATCTTTTTGACTATCTTCCTCTCACT
Y G F Y D E C L R K Y G N A N V W K Y F T D L F D Y L P L T
469 GCCTTGGTGGATGGGCAGgtatgtggat...(0.8)...aaccactcttttcagATCTTCTGTCTACATGGTGGTCTCTCGCCATCTATA
A L V D G Q I F C L H G G L S P S I
523 GATACACTGGATCATATCAGAGCACTTGATCGCCTACAAGAAGTTCCCCATGAGgtatgacttt...(0.5)...gaagttatttttttag
D T L D H I R A L D R L Q E V P H E
577 GGTCCAATGTGTGACTTGCTGTGGTCAGATCCAGATGACCGTGGTGGTTGGGGTATATCTCCTCGAGGAGCTGGTTACACCTTTGGGCAA
G P M C D L L W S D P D D R G G W G I S P R G A G Y T F G Q
667 GATATTTCTGAGACATTTAATCATGCCAATGGCCTCACGTTGGTGTCTAGAGCTCACCAGCTAGTGATGGAGgtatgctgtg...(1.1)
D I S E T F N H A N G L T L V S R A H Q L V M E
739 ...tcctttttccttgcagGGATATAACTGGTGCCATGACCGGAATGTAGTAACGATTTTCAGTGCTCCAAACTATTGTTATCGTTGTGGT
G Y N W C H D R N V V T I F S A P N Y C Y R C G
811 AACCAAGCTGCAATCATGGAACTTGACGATACTCTAAAATACTCTTTgtaagtaatt...(0.9)...cggtcttgttttcagCTTGCAG
N Q A A I M E L D D T L K Y S F L Q
865 TTTGACCCAGCACCTCGTAGAGGCGAGCCACATGTTACTCGTCGTACCCCAGACTACTTCCTGTAATGAAATTTTAAACTTGTACAGTAT
F D P A P R R G E P H V T R R T P D Y F L
955 TGCCATGAACCATATATCGACCTAATGGAAATGGGAAGAGCAACAGTAACTCCAAAGTGTCAGAAAATAGTTAACATTCAAAAAACTTGT
1045 TTTCACATGGACCAAAAGATGTGCCATATAAAAATACAAAGCCTCTTGTCATCAACAGCCGTGACCACTTTAGAATGAACCAGTTCATTG
1135 CATGCTGAAGCGACATTGTTGGTCAAGAAACCAGTTTCTGGCATAGCGCTATTTGTAGTTACTTTTGCTTTCTCTGAGAGACTGCAGATA
1225 ATAAGATGTAAACATTAACACCTCGTGAATACAATTTAACTTCCATTTAGCTATAGCTTTACTCAGCATGACTGTAGATAAGGATAGCAG
1315 CAAACAATCATTGGAGCTTAATGAACATTTTTAAAAATAATTACCAAGGCCTCCCTTCTACTTGTGAGTTTTGAAATTGTTCTTTTTATT
1405 TTCAGGGATACCGTTTAATTTAATTATATGATTTGTCTGCACTCAGTTTATTCCCTACTCAAATCTCAGCCCCATGTTGTTCTTTGTTAT
1495 TGTCAGAACCTGGTGAGTTGTTTTGAACAGAACTGTTTTTTCCCCTTCCTGTAAGACGATGTGACTGCACAAGAGCACTGCAGTGTTTTT
1585 CATAATAAACTTGTGAACTAAGAACTGAGAAGGTCAAATTTTAATTGTATCAATGGGCAAGACTGGTGCTGTTTATTAAAAAAGTTAAAT
1675 CAATTGAGTAAATTTTAGAATTTGTAGACTTGTAGGTAAAATAAAAATCAAGGGCACTACATAACCTCTCTGGTAACTCCTTGACATTCT
1765 TCAGATTAACTTCAGGATTTATTTGTATTTCACATATTACAATTTGTCACATTGTTGGTGTGCACTTTGTGGGTTCTTCCTGCATATTAA
1855 CTTGTTTGTAAGAAAGGAAATCTGTGCTGCTTCAGTAAGACTTAATTGTAAAACCATATAACTTGAGATTTAAGTCTTTGGGTTTTGTTT
1945 TAATAAAACAGCATGTTTTCAGGTAGAG
B
-990 TTGGGGAGGGGGGTGCCTAGATGGCCCCTAAGAGGGGTCCCTGTTCTGTCTCTCAATAAATATTTGTTGAATGAACAAATCATTACAACT
-900 CAGTACACATTGCAGAAAATATAGCCAAGAGCTCTGGAGCTGGAAGGGCCACAGATTATCCTACAGAACAATCATTTCACTTTTCTAATG
-810 TCGAAAGGGAGGATTCGAGATGTTAGGGGCTACAGGTGAGGCTGGAAATAATTAATTTACTCCATCAATATTGAATGCCTGCTATTGAGT
-720 GCTAGACTCTGGGGAGACAATGTTAAGGGAAGCCCAAGTTTTCCAACTCCCTGTCCAGAGCGCTGCAAAGTACTGCAGGAAGCTAAAGTG
-630 AGGACAAAGTTCCCAGAGATCAGGATATTTAAAGGGAGAACCAGCAGAGCTTGGTCTGGGGCAGGGTGGGGAAAAGAGGGACCCTGGCCT
-540 CCTCGGACCGTTTCTCCGCCAAGCCACGCGAGGGCGCTGTTCTGCTCCTAGGGCGCCGTGTCCCGGCGGCGCCGCCTGCTCGCCTTTTCC
-450 CGGCGGAAATGCCCGAGCGATGACGGAAACCCGGAGGAGGGGAGAGAAAGAGCGAGAGAAGGGGAAAGACAAGTCGGGAGAGGCCGGTAG
-360 GCGTGAGGCGGGCCTGAAGCGGCAGCGGGCGGCCTTCGTCCGGCGAGAGCTAGGCCGAGGACCCGCGCCGCGCTCCCCGGCACCTCACCG
-270 CGTCCTTCACCGACTCCCGCGGCGCGCGGCCGGGCGGGGAAGGCGGGCGGGGGTCTCCTCCAGGCTGCGCGCTCGGAGCCGCCTGCTGG
-180 GCTTGGGGCGGGGCGCGGGGCCCGCGGCCGCCCTACCCGGCTCAGTCCTCCCCCTGTGGGACCTGGCGACGGCGGCGGAGGGAGAGGGGAG

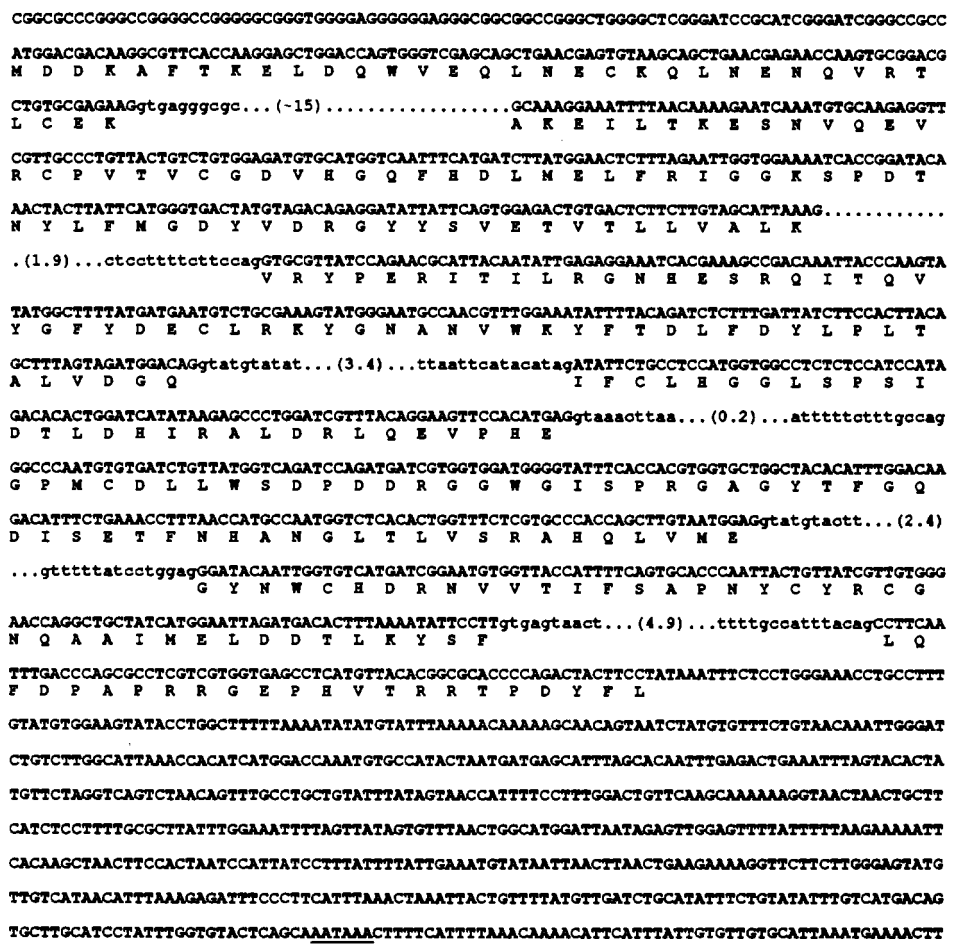


FIGURE 3: Nucleotide and deduced amino acid sequences of the Cα (A) and Cβ (B) genes. The top line shows the nucleotide sequence with the upper case letters denoting exon and 5′ flanking sequences and lower case letters denoting intron sequences. The amino acid sequence is shown in single-letter code below. The sizes of the introns in kilobase pairs are shown in parentheses. The major transcription start sites are indicated by large inverted triangles and the minor start sites by small inverted triangles. The A of the translation initiation codon, ATG, is designated as +1, and numbering does not include intron sequences. Consensus sequences for Sp1 binding (GGGCGG, dotted underline), CRE (Cα only) (TGACGTCA, dashed underline), and polyadenylation signal (AATAAA, solid underline) are indicated. The following differences were observed between the gene and published cDNA sequences: (1) Nucleotides 1–14 of the Cα cDNA of Arino et al. (1988) did not correspond to those in the equivalent positions on the Cα gene. (2) nt −54 to −61 of the Cα gene were absent in the corresponding positions of the Cα cDNA of Arino et al. (1988). This deletion was not present in the cDNA sequence of Stone et al., (1988) which is in complete agreement with the gene sequence. (3) nt −20 to −8 of the Cβ cDNA (Hemmings et al., 1988) did not correspond to those in the equivalent positions of the Cβ gene. (4) nt 1140–1143 (3′ noncoding) of the Cβ gene were absent in the corresponding positions of the Cβ cDNA of Hemmings et al. (1988) but were present in the Cβ cDNA of Arino et al. (1988). (5) The A at position 873 of the Cβ gene is a G in the Cβ cDNA (Hemmings et al., 1988).

The sequences of the Cα and Cβ genes are presented in Figure 3A,B. Both genes are organized into seven exons (see Figure 2) with the location of the introns being identical in the two genes (Figure 3A,B). Introns 1–5 are located between codons while intron 6 splits a codon between the second and third nucleotides. All the intron–exon splice junctions conform to splice-site consensus sequences (Padgett et al., 1986). In both genes, exons II–VII are clustered whereas exon I is separated from the rest of the gene by a large intron (>15 kbp). Exon I encodes all of the 5′ untranslated sequence plus the first 34 amino acids of the N-terminus of the protein. Exon VII begins 70 bp 5′ of the translation termination signal and contains all of the 3′ untranslated region. In the Cα gene, exon VII also contains two additional polyadenylation signals (AATAAA: Proudfoot & Brownlee, 1976; Fitzgerald & Shenk, 1981) located 120 and 365 bp downstream of the first.

The transcription start sites of the Cα and Cβ genes were mapped by RNase protection and S1 nuclease analyses. When the Cα probe was used, a major fragment of 218 bp and several minor fragments of 332, 255, 240, 232, 213, 205, and 195 bases were observed (Figure 4). This suggests the presence of a major start site at position −205 and several distinct minor start sites located both upstream and downstream of the major start site (Figure 3A). Alternatively, the Cα probe may also hybridize to one or more related genes giving rise to the minor bands. In contrast, the Cβ probe gave a broad band of between 490 and 520 bases implicating the use of several clustered start sites between positions −422 and −452 with similar frequencies. Two minor, but distinct, start sites were observed at positions −125 and −127. To obtain a more precise location of the major cap sites, a shorter Cβ probe was used. This analysis revealed two major bands corresponding to positions −409 and −423 and several minor start sites in the same region (data not shown). It should be noted that the autoradiogram shown in Figure 4 has been overexposed to reveal the minor bands and thus the Cα transcript does not appear to be greater than 10-fold more abundant than the Cβ transcript, as observed by Northern analysis (see Figure 5). Shorter exposures showed that RNase protection and S1 nuclease analyses (data not shown) agree with the data obtained by Northern analysis. S1 nuclease analysis using Cα and Cβ probes that extend further upstream (see Experimental Procedures) gave protected fragments identical with those obtained by RNase protection experiments (data not shown), thereby confirming

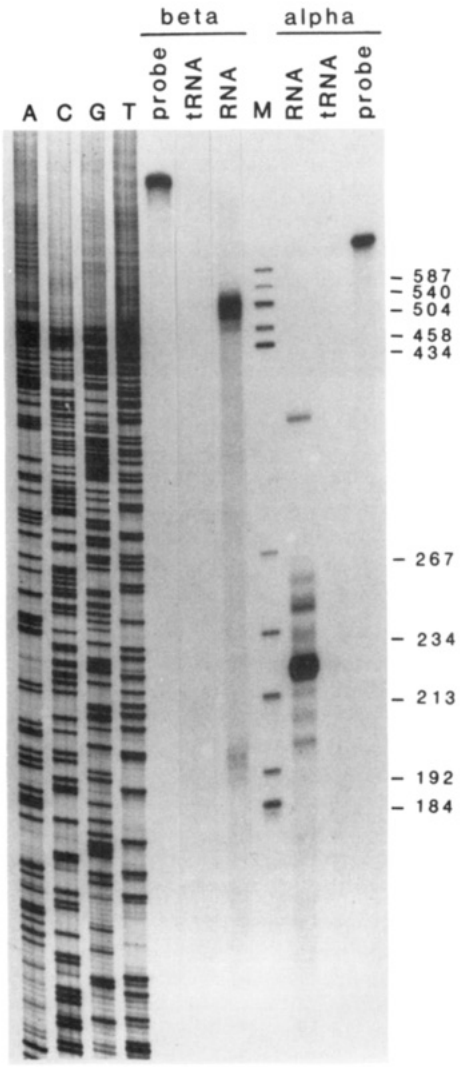


FIGURE 4: Start site mapping by RNase protection analysis. cRNA probes generated to the 5′ end of the Cα or Cβ genes (bottom) were hybridized to total RNA (100 μg) or tRNA (100 μg) and digested with RNases A and T_1. The protected fragments were resolved by electrophoresis on 8% polyacrylamide gels containing 8 M urea together with an aliquot of each of the undigested probes. Lanes marked A, C, G, and T are a set of sequencing reactions obtained by using a known DNA template used as a molecular weight marker. Lane M contains pBR322 DNA cleaved with *Hae*III and labeled with polynucleotide kinase and [γ-^{32}P]ATP. The autoradiogram has been overexposed to reveal the minor bands and therefore does not reflect the quantitative difference between the Cα and Cβ major transcripts. The restriction sites used for subcloning and linearization to generate the Cα and Cβ cRNA probes are shown below. The numbers correspond to nucleotide positions as shown in Figure 3.

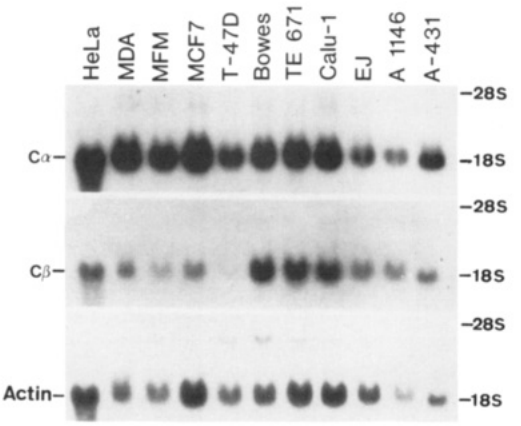


FIGURE 5: Expression of the Cα and Cβ genes in human cell lines. Total RNA (20 μg) from each of the indicated cell lines was hybridized to ^{32}P-labeled Cα- or Cβ-specific cDNA probes [3′ noncoding fragments corresponding to nt 940–1346 of the Cα cDNA (Stone et al., 1988) and nt 966–1432 of the Cβ cDNA (Hemmings et al., 1988)]. The specific activities of the probes labeled by the random-primer method (Feinberg & Vogelstein, 1984) were 2.4×10^9 (for Cα) and 2.6×10^9 (for Cβ) cpm/μg. After autoradiography, the filters were stripped and reprobed with a ^{32}P-labeled actin cDNA probe. The positions of the 18S and 28S rRNA markers are shown.

the use of multiple start sites in both genes. These data indicate the presence of an uninterrupted 5′ untranslated sequence of 205 bases for the Cα transcript and between 422 and 454 bases for the Cβ transcript. There are no upstream AUGs located in the 5′ untranslated region of the Cα mRNA. Interestingly, however, the 5′ untranslated region of the Cβ mRNA contains an open reading frame initiating with an AUG at nucleotide −431 (7 bp upstream of the 5′-most major cap site). This open reading frame extends into the coding sequence and terminates at nucleotide +60. Comparison of the translated sequence of the open reading frame with the protein database reveals no significant homology with any known protein. The significance of this open reading frame, if any, is currently being investigated.

Thus, the predicted transcript sizes are approximately 1.6, 1.7, or 1.9 kb for the Cα gene and 1.9 kb for the Cβ gene. Northern analysis using RNA isolated from several different human cell lines (Figure 5) revealed only one mRNA species of about 2 kb with the Cα- and Cβ-specific probes. These data suggest that the third polyadenylation signal in the Cα gene is preferentially used in these cell lines (but see Discussion).

Properties of the Transcripts and Features of the 5′ Flanking Sequences of the Cα and Cβ Genes. Northern analysis of total RNA isolated from 11 human cell lines (HeLa, MDA-MB231, MFM, MCF7, T-47D, Bowes, TE 671, Calu-1, EJ, A 1146, and A-431) showed large differences in the amount of each transcript as well as in the Cα to Cβ ratio (Figure 5). In general, the Cα mRNA was more abundant than the Cβ mRNA. The ratio of the Cα to Cβ transcripts, determined by densitometric scanning of the autoradiograms, ranged from 27:1 in MCF7 cells to 1:1 in A 1146 cells.

We compared the 5′ flanking sequences of the Cα and Cβ genes to identify sequence motifs responsible for the differential regulation of transcription of the two genes. This analysis revealed that the sequences of the promoter regions, although sharing some characteristics, were different. The features common to the two promoters and also the 5′ noncoding sequences were a high G+C content (extending from −1200 and −600 in the α and β genes into the first intron) and an increase in the ratio of CpG to GpC from 0.2–0.25 in the bulk of the genome (Russel et al., 1976) to 1.0 in the Cβ promoter and 0.8 in the Cα promoter. Both promoters lacked TATA and CCAAT boxes in appropriate positions for use as transcription initiation signals. The Cα gene contained an 8 bp palindrome (TGACGTCA) 26 bp upstream of the start site of transcription that is identical with the consensus sequence of the cAMP-regulatory elements (CRE) identified in many cAMP-responsive genes (Roesler et al., 1988) (Figure 3A). No such motif was observerd at an equivalent position in the

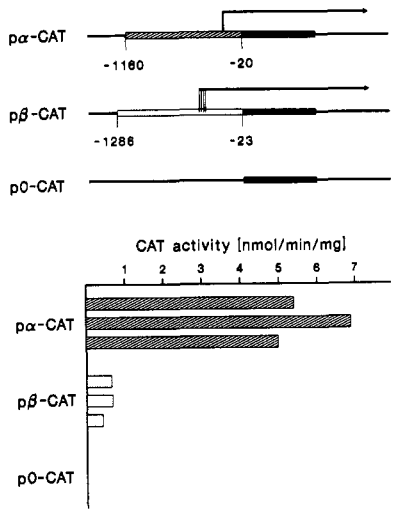


FIGURE 6: Analysis of the promoter strengths of the Cα and Cβ genes. Plasmids pα-CAT and pβ-CAT were constructed by fusing portions (indicated by the nucleotide number beneath where +1 denotes the A in the translation initiation codon, ATG) of the Cα and Cβ genes to the CAT gene as shown. p0-CAT is a promoterless plasmid. Hatched and dotted boxes represent the 5′ flanking regions of the Cα and Cβ genes, respectively, and filled boxes represent the CAT gene. The plasmids were transfected into HeLa cells, and the resulting CAT activity was determined from three independent transfections.

Cβ gene, but a truncated CRE was present within the region of the clustered start sites at postion −430 (Figure 3B). Two inverted Sp1 consensus sequences (CCGCCC; Kadanoga et al., 1986) were also observed in the region upstream of the start site of transcription in the Cα gene (Figure 3), while in the Cβ gene several Sp1 consensus sequences (GGGCGG) were observed in the 5′ untranslated region between the clustered start sites and the two weak start sites (Figure 3).

The differences observed between the Cα and Cβ promoters implied that they may have different transcriptional activities. To investigate this possibility, we fused ~1200 bp of the promoter and 5′ noncoding regions of the Cα and Cβ genes to the bacterial chloramphenicol acetyltransferase (CAT) reporter gene and used the resulting pα-CAT and pβ-CAT plasmids for transient expression analyses in HeLa cells (Figure 6). Results from three typical experiments are shown; the expression of the pα-CAT plasmid was about 8.5-fold higher than that of pβ-CAT. A similar ratio of Cα to Cβ mRNA levels was detected by Northern blot analysis in HeLa cells (Figure 5). The pα-CAT and pβ-CAT plasmids were also transfected into several other human cell lines, and in all cases, the CAT expression resulting from the Cα promoter was higher than that resulting from the Cβ promoter (data not shown). This observation is consistent with the relative levels of Cα and Cβ mRNAs detected by Northern analysis in most human cell lines (Figure 5). The expression of the pα-CAT plasmid was about 2–4-fold higher than that of pSV2CAT transfected into HeLa cells (data not shown). These promoter activities are consistent with the high levels of C subunits mRNAs generally observed. Preliminary deletion experiments showed that in both genes, about 100–150 bp of 5′ flanking sequence was sufficient to confer full promoter activity (data not shown).

Discussion

We have cloned and established the structure of the genes encoding the α and β isoforms of the human phosphatase 2A C subunit. The genes show conserved exon structures, suggesting that they have probably evolved from a common ancestral gene. Seven of the eight amino acid differences observed between these two isoforms are located within exon I; the eighth is in exon III. Furthermore, comparison of the amino acid sequence of the human phosphatase 2A Cα with those of the C subunit of several different protein phosphatases (Table I) shows that the regions with the highest homologies fall within those encoded by exons II–VI. Taken together, these data suggest that the sequences encoded by exons II–VI (the "constant" exons) may be important for substrate binding and catalysis; this high sequence homology could in part explain why protein phosphatases 1 and 2A dephosphorylate the same substrates. Exons I and VII (the "variable" exons) may encode sequences important for regulation. It will be of some interest to examine the genes of protein phosphatases 1 and 2B, which are related to the type 2A, to ascertain whether they show the same conservation of exon structure.

Table I: Amino Acid Identity and Similarity between Human Cα (2Aα) and Other Protein Phosphatases[a]

type	overall	exon I	exons II–VI	exon VII
2Aα/2Aβ	97 (98)	79 (88)	99.6 (100)	100 (100)
2Aα/1[b]	46 (69)	22 (59)	51 (73)	14 (32)
2Aα/2B[c]	41 (63)	12 (32)	54 (73)	23 (59)
2Aα/SIT4[d]	58 (78)	33 (60)	63 (80)	22 (43)
2Aα/X[e]	46 (81)		69 (84)	30 (57)

[a] Percent identities and similarities (inside parentheses) between different protein phosphatases were calculated by using the algorithm of Needleman and Wunsch (1970). The exons correspond to those determined for the protein phosphatase 2A cα and Cβ genes. [b] Rabbit phosphatase 1 (Berndt et al., 1987). [c] Rat phosphatase 2B (Ito et al., 1989). [d] Yeast homologue of phosphatase 2A (Arndt et al., 1989). [e] Rabbit phosphatase X (da Cruz e Silva et al., 1988).

The 3′ untranslated region of each isoform, which is highly conserved between species but divergent between Cα and Cβ mRNAs (Stone et al., 1987, 1988; da Cruz e Silva & Cohen, 1987; Arino et al., 1988; Hemmings et al., 1988; Kitagawa et al., 1988a), is located entirely within exon VII. Analysis of the 3′ untranslated region of the Cα gene revealed three polyadenylation signals at positions 1588, 1708, and 1953. Northern blot analysis of several different human cell lines with the Cα-specific probe, however, revealed only a ~2-kb mRNA, indicating that in the cell lines polyadenylation occurs principally at the third signal. No sequence motifs matching the consensus for the downstream elements CAYTG (Berget, 1984) or YGTGTTYY (McLauchlan et al., 1985) were found 3′ of the first and second polyadenylation signals. However, not all genes require these downstream consensus sequences for polyadenylation to occur (Birnstiel et al., 1985). In the course of cloning the human cDNAs, we isolated 2 Cα clones, 1 of which terminates at the first polyadenylation signal and is followed by a string of 17 A's which was not found in the second clone. The second clone terminates 13 bp after the third polyadenylation signal, suggesting that both the first and third polyadenylation signals are functional (unpublished data). Only one polyadenylation signal was found in the Cβ gene, and a sequence similar to the consensus sequence, YGTGTTYY, was found 31 bp downstream of the polyadenylation signal.

The 5′ flanking sequences of both genes lack TATA and CCAAT boxes in the appropriate positions relative to the transcription start sites. This probably accounts for the multiple start sites observed by S1 and RNase protection analyses. Both promoters and 5′ untranslated regions are G+C rich and show an increase in the ratio of CpG to GpC compared to the rest of the genome. These features are common to several "housekeeping" genes which are constitutively expressed in all tissues (Dynan, 1986; Bird, 1986). That these 5′ flanking sequences function as promoters is demonstrated

by the transient expression experiments. Preliminary deletion experiments showed that between 100 and 150 bp of 5′ flanking sequence was sufficient to mediate full transcriptional activity in both genes (data not shown). Furthermore, the relative activities of these two promoters transfected into HeLa cells reflect the relative steady-state Cα and Cβ mRNA levels in these cells. Thus, the differential expression of the Cα and Cβ mRNAs appears to be transcriptionally regulated. Although the promoters of the Cα and Cβ genes have in common the above-mentioned features, they differ in their nucleotide sequence as well as the number and locations of putative Sp1 binding sites. The Cα gene also contains a putative CRE that is absent in the Cβ gene. It remains to be seen whether the CRE confers cAMP-responsiveness to the Cα gene.

The Cα and Cβ genes evidently arose by gene duplication, but the fact that their promoters are different suggests that their expression may be regulated through different transcriptional rates. Furthermore, the 3′ untranslated sequences which are different between isoforms but highly conserved between species may also play a role in the regulation of the steady-state mRNA levels. Both of these mechanisms would modulate the levels of the α and β isoforms in the cell. However, the activity and substrate specificity of protein phosphatase 2A have also been shown to be influenced by the interaction of the C subunit with different regulatory subunits [reviewed in Cohen (1989)]. Therefore, a very complex system for regulating protein phosphatase 2A exists involving (a) apparent tissue-specific transcription of different C subunit and regulatory subunit (65 kDa) isoforms and (b) assembly of the C subunits with different regultory subunits to generate various holoenzyme forms (Cohen, 1989).

Acknowledgments

We thank Dr. J. Hofsteenge for his support, Dr. S. Shibahara for his gift of human placenta DNA library, W. Zürcher for an endless supply of oligonucleotides, and Dr. F. Stutz for the modified pSV2CAT plasmid. We also thank Drs. L. Ballou, G. Goodall, J. Jiricny, and W. Filipowicz for comments on the manuscript.

Registry No. DNA (human protein phosphatase 2A catalytic subunit α isoform gene coding region), 118216-48-1; protein phosphatase 2A (human catalytic subunit α isoform reduced), 110717-81-2; DNA (human protein phosphatase 2A catalytic subunit β isoform gene coding region), 118216-47-0; protein phosphatase 2A (human catalytic subunit β isoform reduced), 118232-88-5; protein phosphatase, 9025-75-6.

References

Arino, J., Woon, C. W., Brautigan, D. L., Miller, T. B., Jr., & Johnson, G. L. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 4252–4256.

Arndt, K. T., Styles, C. A., & Fink, G. R. (1989) *Cell 56*, 527–537.

Aviv, H., & Leder, P. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 1408–1412.

Ballou, L. M., & Fischer, E. H. (1986) *Enzymes (3rd Ed.) 17*, 311–361.

Berget, S. M. (1984) *Nature 309*, 179–181.

Berk, A. J., & Sharp, P. A. (1977) *Proc. Natl. Acad. Sci. U.S.A. 12*, 721–732.

Berndt, N., Campbell, D. G., Caudwell, F. B., Cohen, P., da Cruz e Silva, E. F., da Cruz e Silva, O. B., & Cohen, P. T. W. (1987) *FEBS Lett. 223*, 340–346.

Bialojan, C., & Takai, A. (1988) *Biochem. J. 256*, 283–290.

Bird, A. P. (1986) *Nature 321*, 209–213.

Birnstiel, M. L., Busslinger, M., & Strub, K. (1985) *Cell 41*, 349–359.

Booher, R., & Beach, D. *Cell 57*, 1009–1016.

Bradford, M. M. (1976) *Anal. Biochem. 72*, 248–254.

Chomczynski, P., & Sacchi, N. (1987) *Anal. Biochem. 152*, 156–159.

Cohen, P. (1989) *Annu. Rev. Biochem. 58*, 453–508.

Crouch, D., & Safer, B. (1980) *J. Biol. Chem. 255*, 7918–7924.

da Cruz e Silva, O. B., & Cohen, P. T. W. (1987) *FEBS Lett. 226*, 176–178.

da Cruz e Silva, O. B., Alemany, S., Campbell, D. G., & Cohen, P. T. W. (1987) *FEBS Lett. 221*, 415–422.

da Cruz e Silva, O. B., da Cruz e Silva, E. F., & Cohen, P. T. W. (1988) *FEBS Lett. 242*, 106–110.

Doonan, J. H., & Morris, N. R. (1989) *Cell 57*, 987–996.

Dynan, W. S. (1986) *Trends Genet. 2*, 196–197.

Feinberg, A. P., & Vogelstein, B. (1984) *Anal. Biochem. 137*, 266–267.

Fitzgerald, M., & Shenk, T. (1981) *Cell 24*, 251–160.

Gorman, C. (1985) in *DNA Cloning—A Practical Approach* (Glover, D. M., Ed.) Vol. I, pp 143–190, IRL Press, Oxford.

Green, D. D., Yang, S.-I., & Mumby, M. C. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 4880–4884.

Haystead, T. A. J., Sim, A. T. R., Carling, D., Honnor, R. C., Tsukitani, Y., Cohen, P., & Hardie, D. G. (1989) *Nature 337*, 78–81.

Hemmings, B. A., Wernet, W., Mayer, R., Maurer, F., Hofsteenge, J., & Stone, S. R. (1988) *Nucleic Acids Res. 16*, 11366.

Hemmings, B. A., Adams-Pearson, C., Maurer, F., Müller, P., Goris, J., Merlevede, W., Hofsteenge, J., & Stone, S. R. (1990) *Biochemistry 29*, 3166–3167.

Ito, A., Hashimoto, T., Hirai, M., Takeda, T., Shuntoh, H., Kuno, T., & Tanaka, C. (1989) *Biochem. Biophys. Res. Commun. 163*, 1492–1497.

Kadanoga, J. J., Jones, K. A., & Tjian, R. (1986) *Trends Biochem. Sci. 11*, 20–23.

Khew-Goodall, Y., & Hemmings, B. A. (1988) *FEBS Lett. 238*, 265–268.

Kitagawa,T., Tahira, T., Ikeda, I., Kikuchi, K., Tsuiki, S., Sugimura, T., & Nagao, M. (1988a) *Biochim. Biophys. Acta 951*, 123–129.

Kitagawa, T., Sakai, R., Tahira, T., Tsuda, H., Ito, N., Sugimura, T., & Nagao, M. (1988b) *Biochem. Biophys. Res. Commun. 157*, 821–827.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, 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.

Melton, D. A., Krieg, P. A., Rebagliati, M. R., Maniatis, T., Zinn, K., & Green, M. R. (1984) *Nucleic Acids Res. 12*, 7035–7056.

Needleman, S. B., & Wunsch, D. C. (1970) *J. Mol. Biol. 48*, 443–453.

Ohkura, H., Kinoshito, N., Miyatami, S., Toda, T., & Yanagida, M. (1989) *Cell 57*, 997–1007.

Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S., & Sharp, P. A. (1986) *Annu. Rev. Biochem. 55*, 1119–1159.

Pato, M. D., & Adelstein, R. S. (1983) *J. Biol. Chem. 258*, 7047–7054.

Proundfoot, N. J., & Brownlee, G. G. (1976) *Nature 263*, 211–214.

Roesler, W. J., Vandenbark, G. R., & Hanson, R. W. (1988) *J. Biol. Chem. 263*, 9063-9066.
Russel, G. J., Walker, P. M. B., Elton, R. A., & Subak-Sharpe, J. H. (1976) *J. Mol. Biol. 108*, 1-23.
Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., & Arnheim, N. (1985) *Science 230*, 1350-1354.
Stone, S. R., Hofsteenge, J., & Hemmings, B. A. (1987) *Biochemistry 26*, 7215-7220.
Stone, S. R., Mayer, R., Wernet, W., Maurer, F., Hofsteenge, J., & Hemmings, B. A. (1988) *Nucleic Acids Res. 16*, 11365.
Tung, H. Y. L., Alemany, S., & Cohen, P. (1985) *Eur. J. Biochem. 148*, 253-263.
Usui, H., Imazu, M., Maeta, K., Tsukamoto, H., Azuma, K., & Takeda, M. (1988) *J. Biol. Chem. 263*, 3752-3761.
Waelkens, E., Goris, J., & Merlevede, W. (1987) *J. Biol. Chem. 262*, 1049-1059.

Grafting of a Calcium-Binding Loop of Thermolysin to *Bacillus subtilis* Neutral Protease

Salvatore Toma,[‡] Susanna Campagnoli,[‡] Immacolata Margarit,[‡] Roberto Gianna,[‡] Guido Grandi,[*,‡] Martino Bolognesi,[§] Vincenzo De Filippis,[∥] and Angelo Fontana[∥]

Laboratories of Genetic Engineering and Microbiology, Eniricerche, 20097 San Donato Milanese, Milan, Italy, and Department of Organic Chemistry, Biopolymer Research Centre of CNR, University of Padua, 35131 Padua, Italy

Received April 23, 1990; Revised Manuscript Received August 14, 1990

ABSTRACT: The surface loop which in the *Bacillus subtilis* neutral protease (NP) extends from amino acid residue 188 to residue 194 was replaced, by site-directed mutagenesis, with the 10-residue segment which in the homologous polypeptide chain of thermolysin (TLN) binds calcium-4 [Matthews, B. W., Weaver, L. H., & Kester, W. R. (1974) *J. Biol. Chem. 249*, 8030-8044]. The mutant NP was isolated to homogeneity, and its structural, functional, calcium-binding, and stability properties were investigated. Proteolytic fragmentation with *Staphylococcus aureus* V8 protease of mutant NP was used to isolate and analyze the protein fragment encompassing the site of mutation, unambiguously establishing the effective insertion of the new 10-residue segment. Atomic absorption measurements allowed us to demonstrate that mutant NP binds three calcium ions instead of the two ions bound to wild-type NP, showing that indeed the chain segment grafted from TLN to NP maintains its calcium-binding properties. The mutant NP showed kinetic parameters essentially similar to those of the wild-type NP with Z-Phe-Leu-Ala-OH as substrate. The enzyme inactivation of mutant vs wild-type NP was studied as a function of free [Ca^{2+}]. It was found that mutant NP was much less stable than the wild-type NP when enzyme solutions were dialyzed at neutral pH in the presence of [Ca^{2+}] below 10^{-3} M. On the other hand, the kinetic thermal stability to irreversible inactivation of mutant NP, when measured in the presence of 0.1 M $CaCl_2$, was found to be increased about 2-fold over that of the wild-type NP. Thus, modulation of enzyme stability by free [Ca^{2+}] in mutant NP correlates with similar findings previously reported for thermolysin. Overall, the results obtained indicate that protein engineering experiments can be used to prepare hybrid proteins on the basis of sequence and function analysis of homologous protein molecules and show the feasibility of engineering metal ion binding sites into proteins.

An exciting potential application of protein engineering is the creation of proteins with novel functions (Oxender & Fox, 1987). Although a full extension of this technology to the de novo protein design will be possible only when a deep understanding of protein folding and structure is reached (Richards, 1986), the construction of hybrid proteins represents a first promising attempt. Hybrid proteins are those obtained when two proteins, or parts of them, are joined together to give new polypeptide structures with functions eventually in common with those of the parent molecules. The first hybrid proteins reported in the literature were obtained in a more or less empirical approach, namely, by randomly fusing the DNA coding sequences of two proteins, some of which were, for example, β-galactosidase (Weinstock et al., 1983), phosphatase (Hoffman & Wright, 1985), Gal4 repressor (Fields & Song, 1989), and others. Recently, more rational approaches to hybrid protein construction have been undertaken. For example, hybrid human-mouse antibodies were created by substituting the complementarity-determining regions of a mouse antibody with the corresponding one of a human myeloma protein (Jones et al., 1986). The feasibility of substituting one β-turn for another without substantially altering the overall protein structure was demonstrated by Hynes et al. (1989). These authors prepared a hybrid staphylococcal nuclease with a β-turn of four amino acid residues replaced with a five-residue turn from concanavalin A. A crystallographic study of the mutant nuclease revealed that the con-

* Correspondence should be addressed to this author at the Laboratories of Genetic Engineering and Microbiology, Eniricerche, Via Maritano 26, I-20097 San Donato Milanese, Milan, Italy.
‡ Eniricerche.
§ Department of Genetics, Section of Crystallography, University of Pavia, 27100 Pavia, Italy.
∥ University of Padua.

0006-2960/91/0430-97$02.50/0