Hiroshima J. Med. Sci.
Vol.42, No.1, 47~52, March, 1993
HIJM 42-6

Nucleotide Sequence of the 5' Region of the Human Platelet-derived Growth Factor A-chain Gene

Yasuo TAKIMOTO[1),+,*], Wen Yu LI[1)], Zhao Yi WANG[1)], Benton D. TONG[1)] and Thomas F. DEUEL[1,2)]

[1)]Department of Medicine, [2)]Department of Biology Chemistry, Jewish Hospital at Washington University Medical Center, 216 South Kingshighway, St. Louis, MO 63110

ABSTRACT

We have isolated and characterized genomic clones of the 5' region of the human PDGF A-chain gene. The putative "TATAA" box was identified 876 bp upstream of the translation initiation site. A significant number of potential regulatory elements were identified in DNA sequences upstream of the "TATAA" box and were also found within the introns sequenced. The DNA sequence results differ significantly from those reported for the PDGF B-chain gene, suggesting a basis for the expression of the PDGF A-chain gene and the B-chain gene under very different conditions.

***Key words**: PDGF-A chain, 5'-Region, Nucleotide sequence*

The platelet-derived growth factor (PDGF) is a heterodimeric 30 kD glycoprotein released from α granules of activated platelets. PDGF is the major mitogen of mesodermal cells and a potent chemotactic agent for neutrophils, monocytes, fibroblasts, and smooth muscle cells[7)]. The PDGF B-chain gene is encoded by the c-sis gene, the normal cellular homolog of the v-sis oncogene of the simian sarcoma virus (SSV)[8,30)], and has been mapped to human chromosome 22q11[9)]. A PDGF A-chain cDNA clone (D-1) was isolated and sequenced first from a human malignant glioma cell library[2)]. However, cDNA clones isolated from a human endothelial cell library differed from that clone. The endothelial cDNA clones lacked a 69 bp insert included in the A-chain clones isolated from the glioma cell line, resulting in the loss of an 18 amino acid, highly basic region containing the predicted COOH-terminus and the substitution of the COOH-terminus with 3 entirely different amino acid residues[28)]. Genomic sequencing subsequently confirmed alternative splicing as the basis for the differences in the clones[26)]. The PDGF A-chain gene shares significant homology to the v-sis gene but as yet the A-chain gene has no significant homology to other previously identified transforming genes. It has been mapped to human chromosome 7[2)]. A-chain homodimers with strong mitogenic activity have been isolated from a human osteosarcoma cell line[16)] and subsequently the expression of the A-chain has been demonstrated in many different cell lines, including normal human endothelial cells, placenta, and certain tumor cell lines[2)]. The regulation of expression of A- and B-chain genes, however, appears to be quite different, as shown by differences in the A- and B-chain mRNA levels in endothelial cells in response to the endothelial cell growth factor[26)] or by the selective induction of the A-chain gene after PDGF or EGF stimulation of normal human fibroblasts[22)]. The basis for the differences in regulation is not known. In order to understand the mechanism of regulation of the A-chain gene, we cloned the A-chain genomic DNA and initiated its nucleotide sequencing study. We report here the location of the putative TATAA box for the human PDGF A-chain gene, nearly 1.8 kb upstream of the coding sequences, and the complete sequence of the first two exons and the first intervening sequence of the gene.

MATERIALS AND METHODS

Materials

T4 DNA ligase, T4 polynucleotide kinase, and DNA polymerase I (Klenow fragment) were obtained from Bethesda Research Laboratories. T4 DNA polymerase was from New England Bio-Labs. Calf intestinal alkaline phosphatase was obtained from Pharmacia. Restriction enzymes were obtained from Bethesda Research Laborato-

[+]*Present address. Department of Internal Medicine, Hiroshima Asa General Hospital, 2-1-1 Kabe, Asakita-ku, Hiroshima 731-02, Japan.*
[*]*Author to whom correspondence should be directed*

ries and New England BioLabs. The EMBL-3 phage genomic DNA library constructed from MboI partially digested human leukocyte DNA from Clontech was used for isolation of genomic clones. Sequencing kits with klenow fragment and reverse transcriptase were obtained from Promega, with sequenase from United States Biochemical Corporation, and with 7-deaza dGTP from Pharmacia. α^{32}P d-ATP (specific activity 800 Ci/mmol) and γ^{32}P ATP (specific activity 3000 Ci/mmol) were from New England Nuclear.

Cloning

A-chain genomic clones were obtained from the EMBL-3 phage genomic DNA library using a synthetic oligonucleotide (A-2), corresponding to nucleotides 389-424 of the D-1 cDNA clone[2], and labeled with the T4 polynucleotide kinase and γ^{32}P ATP. Four genomic clones were identified and designated as λ21, λ24, λ31, and λ40. These clones overlapped with each other;clones λ24, λ31, and λ40 contained the same fragment.

The 3.5 kb BamH1 fragment, which hybridized

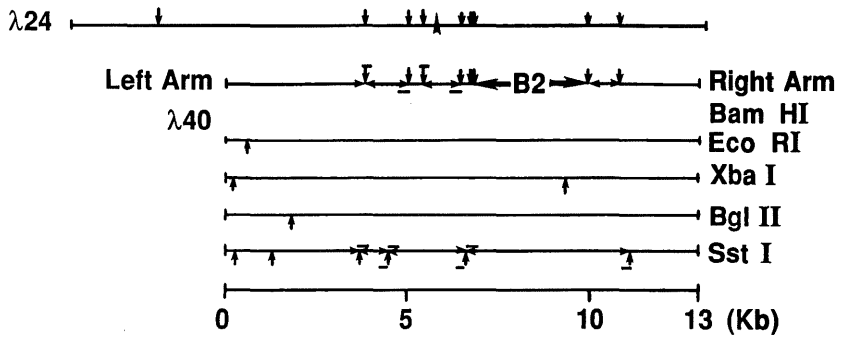

Fig. 1. Restriction map of λ40 and λ24. λ24, λ31, and λ40 contained the same fragment. The B2 fragment is indicated by the thick arrow. Fragments subcloned are indicated by arrows (←→) and those sequenced are indicated by small bars (-). There was no Sal 1 and Hind III cut site.

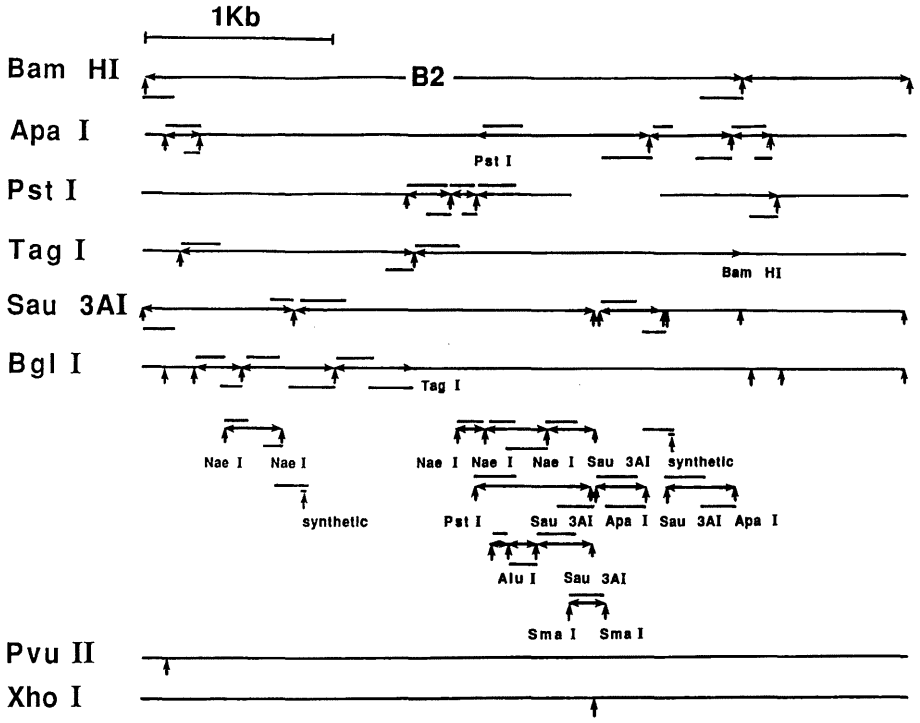

Fig. 2. Restriction map and sequencing strategy for B2. The nucleotide sequence was determined by a double-strand DNA sequencing method using the T3, T7, mp, mpR, and two synthetic primers. Fragments subcloned are indicated by arrows (←→) and those sequenced are indicated by small bars (-). All fragments sequenced overlapped each other.

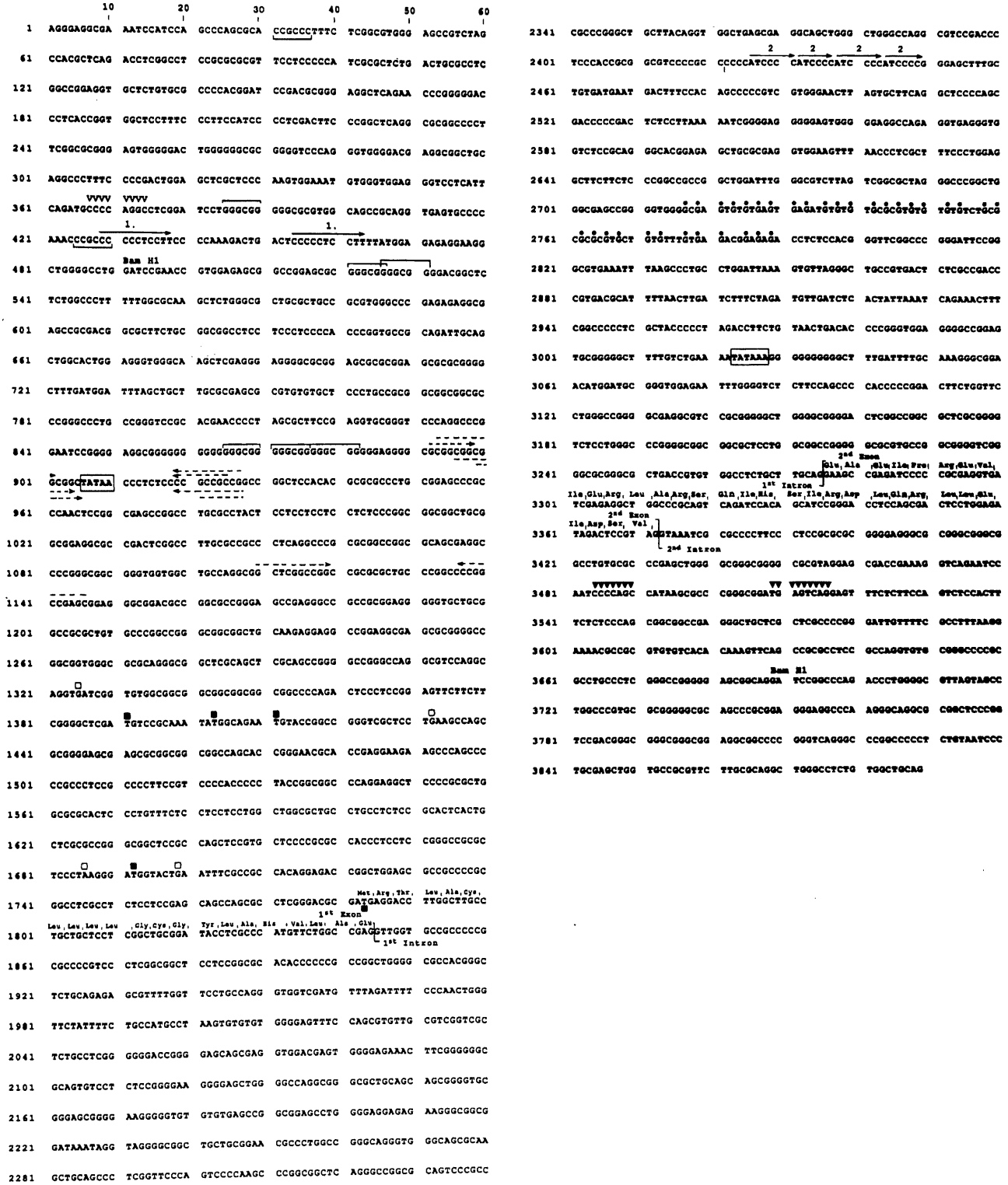

Fig. 3. Nucleotide sequence of the PDGF A-chain genomic DNA. The putative "TATAA" sequences are boxed-in. Closed squares indicate the start of codons and open squares the end of codons. Tandem repeat sequences are indicated with arrows (→) and inverted repeat sequences with dotted arrows (⇢). Splicing junctions are indicated by (⌉). The consensus sequences for the potential SP1 binding site (GGGCGG) are indicated by (⌐) and the "CCGCCC" sequeces similar to the one found in the SV40 promoter is indicated by (⌴). The consensus sequences for the potential AP1 and AP2 binding site are indicated by (▽). Closed circles mark the poly dGdN stretch.

to the A-2 probe, was subcloned into pUC 19 plasmid and designated as B2 (Fig. 1). A precise restriction enzyme map of the B2 fragment was established (Fig. 2). Small fragments from B2 were then obtained by digestion with specific restriction enzymes, and were subcloned into Blue-Script plasmid (Stratagene) for sequencing.

DNA Sequence Analysis

The nucleotide sequence was determined by the dideoxy nucleotide method using double-strand DNA[6]. 7-deaza dGTP and dITP were used for fragments in highly G-C-rich regions with G-C compression. In order to sequence areas with high degrees of secondary structure, the sequencing reactions were done with reverse transcriptase at 60°C or with sequenase. Orientation of the individual DNA fragments was established on the basis of overlapping sequences.

RESULTS

We determined the nucleotide sequence of the 5' region of the human PDGF A-chain gene (Fig. 3). The λ40 fragment contained the first and second exons. The putative splice site between the first and second exons was identified at positions 450–451 and between the second and third exons at positions 547–548, using the numbering system adopted for the D-1 clone[2]. The splice sites agreed with the cDNA sequence of the D-1 clone[2] and with consensus sequences of splicing[27].

Promoter Region

A "TATAA" sequence[4] was identified 487 bp upstream of the 5' terminus of the D-1 clone[2]. The region surrounding the putative "TATAA" box is highly G-C-rich. In the 60 bp upstream fragment, the G-C content was 97%, and in the 60 bp downstream fragment it was 83%, suggesting that these regions may have considerable secondary structure. On both sides of the putative "TATAA" box, there are several inverted repeats which may make hairpin loops (Fig. 3). The G-C-rich area also suggests that this region is protected from methylation[3].

5' Upstream region of the "TATAA"

Six potential SP-1 binding sites containing the consensus sequence "GGGCGG"[5] were identified within the 550 bp upstream of the "TATAA" sequence. Interestingly, three of the putative binding sites were repeated tandemly. Two "CCGCCC" sequences were found, identical to those within the SV40 early promoter[10] and in close proximity to the "TATAA" box of the c-myc gene[31] and the EGF receptor gene[17]. A 10 bp tandem repeat, "CCCCCTCCTT", was found about 450 bp upstream of the "TATAA" sequence. No "CAAT" box[4] was found within the 900 bp DNA region upstream of the "TATAA" sequence, but a "GAAT" sequence was found 65 bp upstream of the "TATAA" sequence.

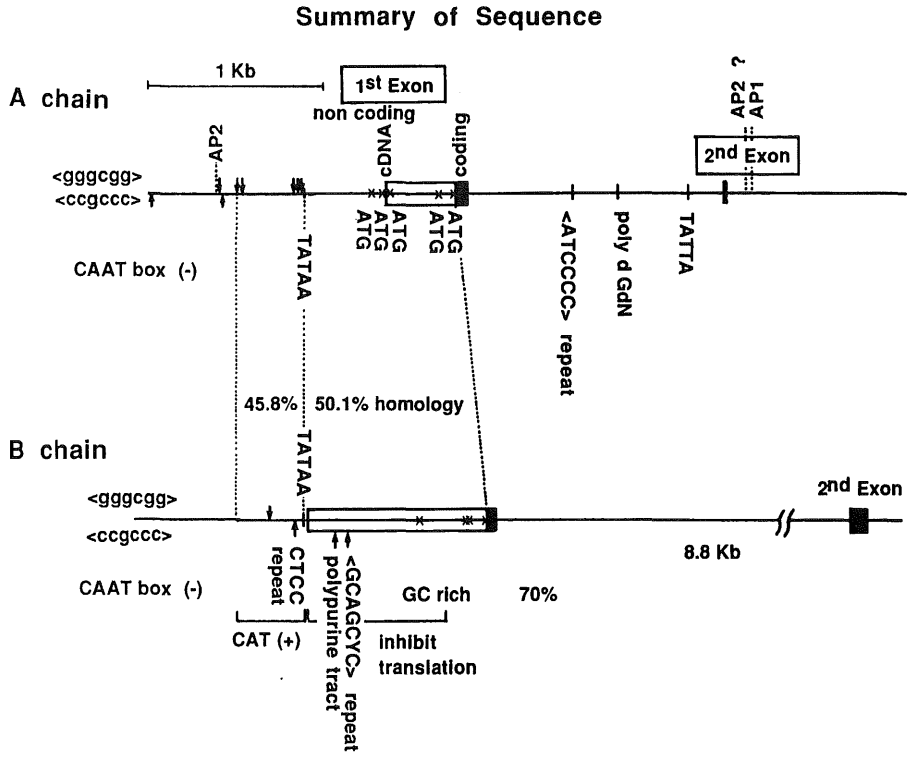

Fig. 4. Summary of PDGF A-chain sequence and comparison with that of PDGF B-chain gene. The start site of cDNA clone (D-1) was indicated by "cDNA".

From the "TATAA" Sequence to the First Intron

The translation initiation site was found 876 bp downstream of the "TATAA" sequence, implying that a long transcribed but untraslated region is present in the A-chain gene, similar to that in the PGDF B-chain gene[24,25,29]. This region was highly G-C-rich; a 10 bp inverted repeat was found 210 bp downstream of the "TATAA" sequence. This region may be important in regulating translation of the A-chain gene, a regulation potentially similar to that observed in the PDGF B-chain gene[25]. There were five "ATG" start codons found, but each of the first four were followed by stop codons. Furthermore, the sequences surrounding the first four "ATG" sites did not conform to the consensus sequences for translation initiation[19] and coded only for short polypeptide chains. The overall gene structure is very similar to that of the B-chain gene, but there were no clusters of the tandem repeat sequences of 7 bp (GCAGCT/CC, Fig. 4) which may control mRNA translation, as were found previously in the B-chain gene. The predicted first 21 amino acids of the first exon had the properties characteristic of a signal peptide sequence[20].

From the First Intron to the Second Exon

In the first intron, an "ATCCCC" sequence was repeated four times; this tandem repeat sequence may function similarly to the enhancer activity of the short repeat sequences found in the interferon-β gene[11]. About 260 bp downstream of this short repeat sequence was a poly dGdN stretch, a sequence known to form Z-DNA and also associated with enhancer activity in other genes[13,14]. A second "TATAA" sequence was found about 250 bp upstream of the second exon.

Significantly, consensus sequences of binding sites for transcription factors induced by TPA also were found[1,21]. In the second intron, a potential AP1 binding site "TGAGTCAG" was found and a potential AP2 binding site, "CCCCAGGC", was found about 535 bp upstream of the putative "TATAA" box. A closely related sequence, "CCCAGC", was identified in the second intron.

Comparisons with B-chain gene DNA sequences were made from -400 bp to the "TATAA" sequence (Fig. 4), there was about 45% homology. From the "TATAA" sequence to the start codon, the homology was about 50%. There was no Alu-sequences[18] or the consensus sequences of inducible elements of fos gene[12,15,23].

DISCUSSION

We have isolated and characterized genomic clones of the 5' region of the huma PDGF A-chain gene.

These results have established a number of potentially important sites on the PDGF A-chain gene which may be involved in regulation of the expression. The promoter region analyzed revealed significant differences between the PDGF A-chain gene and the PDGF B-chain gene. These may mediate the observed differences in the regulation of the two genes.

(Received April 24, 1992)
(Accepted January 31, 1993)

REFERENCES

1. **Angel, P., Imagawa, M., Chiu, R., Stein, B., Imbra, R.J., Rahmsdorf, H.J., Jonat, C., Herrlich, P. and Karin, M.** 1987. Phorbol ester-inducible genes contain a common cis element recognized by a TPA-modulated trans-acting factor. Cell **49:** 729-739.
2. **Betsholtz, C.A., Johnsson, A., Heldin, C.-H., Westermark, B., Lind, P., Urdea, M.S., Eddy, R., Shows, T.B., Philpott, K., Mellor, A.L., Knott, T.J. and Scott, J.** 1986. cDNA sequence and chromosomal localization of human platelet-derived growth factor A-chain and its expression in tumor cell lines. Nature **320:** 695-699.
3. **Bird, A., Taggart, M., Frommer, M., Miller, O.J. and Macleod, D.** 1985. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-Rich DNA. Cell **40:** 91-99.
4. **Breathnach, R. and Chambon, P.** 1981. Organization and Expression of Eucaryotic split genes coding for proteins. Ann. Rev. Biochem. **50:** 349-384.
5. **Briggs, M.R., Kadonaga, J.T., Bell, S.P. and Tjian, R.** 1986. Purification and biochemical characterization of the promoter-specific transcription factor, Sp1. Science **234:** 47-52.
6. **Chen, E.Y. and Seeburg, P.H.** 1985. Supercoil Sequencing : A fast and simple method for sequencing plasmid DNA. DNA **4:** 165-170.
7. **Deuel, T.F.** 1987. Polypeptide growth factors : Roles in normal and abnormal cell growth. Ann. Rev. Cell Biol. **3:** 443-492.
8. **Doolittle, R.F., Hunkapiller, M.W., Hood, L.E., Davare, S.G., Robbins, K.C., Aaronson, S. and Antoniades, H.N.** 1983. Simian sarcoma virus onc gene, v-sis, is derived from the gene (or genes) encoding a platelet-derived growth factor. Science **221:** 275-277.
9. **Favera, R.D., Gallo, R.C., Giallongo, A. and Croce, C.M.** 1982. Chromosomal localization of the human homology (c-sis) of the simian sarcoma virus onc gene. Science **218:** 686-688.
10. **Fromm, M. and Berg, P.** 1982. Deletion mapping of DNA regions required for SV40 early region promoter function in vivo. J. Mol. Appl. Genet. **1:** 457-481.
11. **Fujita, T., Shibuya, H., Hotta, H., Yamanishi, K. and Taniguchi, T.** 1987. Interferon-β gene regulation: Tandemly repeated sequences of a synthetic 6 bp oligomer function as a virus-inducible enhancer. Cell **49:** 357-367.
12. **Greenberg, M.E., Siegfried, Z. and Ziff, E.B.**

1987. Mutation of the c-fos gene dyad symmetry element inhibits serum inducibility of transcription in vivo and the nuclear regulatory factor binding in vitro. Mol. Cell. Biol. **7:** 1217-1225.

13. **Hamada, H., Petrino, M.G., Kakunaga, T., Seidman, M. and Stollar, B.D.** 1984. Characterization of genomic poly (dT-dG) poly(dC-dA) sequences : structure, organization, and conformation. Mol. Cell. Biol. **4:** 2160-2621.
14. **Hamada, H., Seidman, M., Howard, B.H. and Gorman, C.M.** 1984. Enhanced gene expression by the poly(dT-dG) poly (dC-dA) sequence. Mol. Cell. Biol. **4:** 2622-2630.
15. **Hayes, T.E., Kitchen, A.M. and Cochran, B.H.** 1987. Inducible binding of a factor to the c-fos regulatory region. Proc. Natl. Acad. Sci. USA **84:** 1272-1276.
16. **Heldin, C.-H., Johnsson, A., Wennergren, S., Wernstedt, C., Betsholtz, C. and Westermark, B.** 1986. A human osteosarcoma cell line secretes a growth factor structurally related to a homodimer of PDGF A-chain. Nature **319:** 511-514.
17. **Ishii, S., Xu, Y.H., Stratton, R.H., Roe, B.A., Merlino, G.T. and Pastan, I.** 1985. Characterization and sequence of the promoter region of the human epidermal growth factor receptor gene. Proc. Natl. Acad. Sci. USA **82:** 4920-4924.
18. **Jelinek, W. R., Toomey, T. P., Leinwand, L., Duncan, C.H., Biro, P.A., Choudary, P.V., Weissman, S.M., Rubin, C.M., Houck, C.M., Deininger, P.L. and Schmid, C.W.** 1980. Ubiquitous, interspersed repeated sequences in mammalian genomes. Protein AP-2 is inhibited by SV40 T antigen. Proc. Natl. Acad. Sci. USA **77:** 1398-1402.
19. **Kozak, M.** 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucl. Acids Res. **12:** 857-872.
20. **Kreil, G.** 1981. Transfer of proteins across membranes. Ann. Rev. Biochem. **50:** 317-348.
21. **Mitchell, P.J., Wang, C. and Tjian, R.** 1987. Positive and negative regulation of transcription in vitro : enhancer-binding. Cell **50:** 847-861.
22. **Paulsson, Y., Hammacher, A., Heldin, C.-H. and Westermark, B.** 1987. Possible positive autocrine feedback in the prereplicative phase of human fibroblasts. Nature **328:** 715-717.
23. **Prywes, R. and Roeder, R.G.** 1986. Inducible binding of a factor to the c-fos enhancer. Cell **47:** 777-784.
24. **Rao, C.D., Igarashi, H., Chiu, I.-M., Robbins, K.C. and Aaronson, S.A.** 1986. Structure and sequence of the human c-sis/platelet-derived growth factor 2(SIS/PDGF 2) transcriptional unit. Proc. Natl. Acad. Sci. USA **83:** 2392-2396.
25. **Ratner, L., Thielan, B. and Collins, T.** 1987. Sequences of the 5' portion of the human c-sis gene : characterization of the transcriptional promoter and regulation of expression of the protein product by 5' untranslated mRNA sequences. Nucl. Acids Res. **15:** 6017-6036.
26. **Sclafani, R.A., Patterson, M., Rosamond, J. and Fangman, W.L.** 1988. Differential regulation of the yeast CDC 7 gene during mitosis and meiosis. Mol. Cell, Biol. **8:** 293-300.
27. **Sharp, P.A.** 1981. Speculations on RNA splicing. Cell **23:** 643-646.
28. **Tong, B.D., Auer, D.E., Jaye, M., Kaplow, J.M., Ricca, G., McConathy, E., Drohan, W. and Deuel, T.F.** 1987. cDNA clones reveal differences between human glial and endothelial cell platelet-derived growth factor A-chains. Nature **328:** 619-621.
29. **Van den Ouweland, A.M.W., Roebroek, A.J.W., Schalken, J.A., Claesen, C.A.A., Bloemers, H.P.J. and Van de Ven, W.J.M.** 1986. Structure and nucleotide sequence of the 5' region of the human and feline c-sis proto-oncogenes. Nucl. Acids Res. **14:** 765-778.
30. **Waterfield, M.D., Scrace, G.T., Whittle, N., Stroobant, P., Johnsson, A., Wasteson, A., Westermark, B., Heldin, C.-H., Huang, J.S. and Deuel, T.F.** 1983. Platelet-derived growth factor is structurally related to the putative transforming protein $p28^{sis}$ of simian sarcoma virus. Nature **304:** 35-39.
31. **Watt, R., Nishikura, K., Sorrentino, J., ar-Rushdi, A., Croce, C.M. and Rovera, G.** 1983. The structure and nucleotide sequence of the 5' end of the human c-myc oncogene. Proc. Natl. Acad. Sci. USA **80:** 6307-6311.