Volume 15 Number 15 1987 Nucleic Acids Research

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

Lee Ratner*, Benjamin Thielan and Tucker Collins[1]

Departments of Medicine, and Microbiology and Immunology, Washington University, St Louis, MO 63110 and [1]Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA

Received April 21, 1987; Revised and Accepted July 1, 1987 Accession no. Y00389

ABSTRACT

The c-sis gene encodes the B polypeptide chain of platelet-derived growth factor (PDGF), and is expressed in a number of normal and pathological conditions. In order to study the control of synthesis of the human c-sis product, we have initiated a study of two regions of this genetic locus which regulate transcription and translation. A clone of the 5' portion of the gene was obtained which included 1361 nucleotides upstream of the RNA initiation site. Transcriptional promoter activity of this region was demonstrated in normal and transformed cells using a plasmid with the sequences upstream of the c-sis RNA initiation site fused to an indicator gene, chloramphenicol acetyl transferase. Experiments were also performed to identify other possible regulatory regions of the c-sis gene. These data demonstrated that a portion of the c-sis first exon encoding the 5' untranslated region of the c-sis mRNA inhibited synthesis of the PDGF B product in vitro. These results define regions of the c-sis gene whose activity may be important in the regulation of transcription and translation under normal conditions and in the pathogenesis several human diseases.

INTRODUCTION

Human platelet-derived growth factor (PDGF) is a 30 kilodalton (kd) heat-stable, cationic glycoprotein initially identified in platelets (see references 1-6 for reviews). It accounts for most of the mitogenic activity in human serum for fibroblasts, and may play an important role in mesenchymal cell growth both during embryogenesis and in wound repair (7). Endothelial cells (8), macrophages (9), and smooth muscle cells (10,11) also produce PDGF-like proteins. Altered c-sis expression by any of these cell types may change the amount or composition of the secreted mitogen and play an important role in pathophysiologic processes such as inflammation, thrombosis, and atherosclerosis.

PDGF is composed of structurally related A and B polypeptide chains each of which may form a homodimer with mitogenic action

(12-15).** The PDGF B polypeptide is encoded by the c-sis proto-oncogene, the cellular homologue of the oncogene transduced by both the simian sarcoma virus and Parodi-Irgenes feline sarcoma virus (16-23). This gene has been mapped to the q11-qter region of human chromosome 22 (24,25). The PDGF A polypeptide has been mapped to region pter-q22 of human chromosome 7; it has no homology to previously identified transforming genes (26).

The c-sis gene is transcriptionally active in several normal cell types as well as in a number of transformed cell lines. Although not expressed in most normal cells (4), transcripts have been identified in developing placenta (27), cultured endothelial cells (8,30,31), activated monocytes (9), and some smooth muscle cells (10,11). This restricted tissue specificity suggests that the c-sis gene may contain structural elements that are required for efficient cell-specific transcription. C-sis mRNAs have also been identified in a number of malignant cell types of glial and neural origin (32,33), sarcomas (32,34,35), melanomas (36), and human T-lymphotropic virus type I (HTLV-I) transformed cells (37). The role of c-sis transcription in tumorigenesis, however, remains to be defined. A c-sis cDNA isolated from an HTLV-I infected lymphoid cell line was capable of transforming murine NIH 3T3 cells which gave rise to tumors in syngeneic mice (38). The predicted amino acid sequence of the PDGF B polypetide encoded by this cDNA was identical to that derived from normal cells, suggesting that quantitative rather than qualitative alterations are important for the transforming phenotype (28-30). Thus, the level of c-sis mRNA may be an important determinant of the level of growth factor expression. Regulation may occur at the levels of transcription, processing, or degradation of c-sis mRNA. Furthermore, additional control at translational levels is not excluded by these findings. To explore these possibilities, we have cloned the 5' portion of the human c-sis gene and have utilized these sequences to examine their role in the regulation of transcription and translation.

** The terminology of Westermark and colleagues as well as Ross and collaborators has been adopted for the designation of PDGF polypeptide chains (2,4,13,26,27). This is opposite to that described in ref. 3,21,28,29.

MATERIALS AND METHODS

Cell Lines

HOS, HT1080, and SKES cells were obtained from the ATCC. A172 cells were provided by A. Fisher (National Cancer Institute, Bethesda, Maryland). Bovine aortic endothelial cells (BAEC) were cultured from calf thoracic aortas as previously described (39) and generously provided by M. Gimbrone (Harvard University, Boston, Massachusetts).

Recombinant DNA Clones

High molecular weight DNA was extracted from peripheral blood mononuclear cells of a normal human donor, and partially digested with Mbo I (40). The sample was size-fractionated by sucrose gradient centrifugation and cloned in the Bam HI arms of bacteriophage J1, provided by G. Shaw (U. of Alabama, Birmingham). The recombinant phage library was screened with a probe derived from a c-sis cDNA clone corresponding to nucleotides 47-779 of exon 1 (30). A single positive plaque was identified after screening 400,000 plaques and the phage clone was designated λSIS-A. A restriction map was generated using the exon 1 probe described above, as well as probes corresponding to exons 1-7 derived from the Sac II - Sac I fragment of plasmid pSM1 (29), and exons 3-5 derived from the Bam HI fragment of plasmid L33-M (21), a bacteriophage J1 DNA probe, and a probe from total DNA from the same sample used for cloning.

The 5' 2.0 kilobases (kb) of the λSIS-A insert was subcloned into the Bam HI site of SP65 (Promega Scientific) and designated pSIS-1. All nucleotides of the pSIS-1 insert were sequenced at least once on each strand by the partial chemical cleavage method (41).

Primer Extension Mapping of the c-sis Transcriptional Initiation Site

A 121 base pair (bp) primer (nucleotides 33-153, Fig. 2) was made by digesting 30 μg of pSIS-1 with Bss HII, labeling with 10 U T4 polynucleotide kinase and 300 μCi ^{32}P- γ -ATP (7000 Ci/mmole), and subsequent digestion with Pst I. The primer was isolated by polyacrylamide gel electrophoresis. 50,000 cpm of primer was mixed with 5 μg of poly A(+) RNA from HOS cells, 5 or 50 μg of poly A(+) RNA from A172 cells, or 12,800 cpm of primer was mixed

with no RNA, in 40 mM PIPES, pH 6.4, 400 mM NaCl, 1 mM ethylene diamine tetraacetic acid (EDTA), 80% (vol/vol) formamide. The samples were heated at 90 degrees C for 10 minutes and then incubated at 64 degrees C for 3 hrs. A 185 bp double-stranded Bss HII fragment (nucleotides 170-354) was included in the primer preparation as an internal size marker for the gels. Under these hybridization conditions, nucleotides 170-354 did not hybridize to c-sis mRNA. Samples were ethanol precipitated and extended with 90 U of avian myeloblastosis reverse transcriptase (BRL) as previously described (42). Samples were denatured and electrophoresed on a 7 M urea, 5% (wt/vol) polyacrylamide gel. Partial chemical cleavage products electrophoresed in parallel were used as size markers.

Measurements of Transcriptional Promoter Activity

Plasmids used for this analysis were constructed as follows. Plasmid pSV0-CAT (43) was digested with Bam HI, blunt-ended with Klenow fragment of E. coli DNA polymerase I, and then digested with Hind III (Fig. 3). The 1.7 kb fragment with CAT and SV40 polyadenylation sequences was ligated to Bam HI and Pvu II digested SP65, yielding SP65-CAT3. The 403 bp Pst I fragment of pSIS-1 (nucleotides -366 to 37) was cloned into the Pst I site of SP65CAT3 in the same or opposite transcriptional orientation as the CAT gene in plasmids SIS-CAT1 and SIS-CAT2, respectively. Plasmids used as positive controls for the analysis included pSV2-CAT and RSV-CAT which have the CAT gene under the transcriptional control of the simian virus (SV) 40 promoter or Rous sarcoma virus (RSV) long terminal repeat sequences (LTR), respectively (43,44).

Measurements of transcriptional promoter activity were performed by transfecting 10 μg of plasmid DNA onto 90% confluent 100 mm plates of HOS cells or 25 μg of of plasmid DNA onto 80% confluent cultures of bovine aortic endothelial cells, using calcium phosphate precipitation (44). The HOS cells were glycerol shocked 24 hrs after adding the DNA-calcium phosphate precipitate. Cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% (vol/vol) fetal calf serum (Gibco), 110 mg/L pyruvate, 100 u/ml penicillin, and 100 μg/ml streptomycin. The HOS cells were split into 150 mm plates 24 hrs after glycerol

shock, and were harvested 24 hrs after splitting. The endothelial cells were simply washed 12 hrs after transfection, and harvested 36 hrs later. 20 μl of cell lysate was incubated a 180 μl reaction in 140 mM Tris-Cl, pH 7.8, 2.4 mM acetyl CoA, 0.05 μCi 14C-chloramphenicol (53 mCi/mmole, New England Nuclear) for 45 min at 37 degrees C. Samples were extracted with ethyl acetate, dried , and spotted on a Silica gel 60 thin layer chromatography (TLC) plate. The solvent for chromatography was chloroform:methanol (95:5). TLC plates were treated with Enhance spray (New England Nuclear) and exposed to XRP-5 xray film (Kodak) with an intensifier screen at -70 degrees C for 16 hrs. In the HOS experiment, individual spots were cut and counted, and the percent 14C-chloramphenicol acetylated determined. In replicate transfections of the same DNA sample, the standard error was less than 6%.

Translation of c-sis Sequences

Several deletion clones of 5'untranslated c-sis sequences were constructed for this analysis (Fig. 4). The endothelial c-sis cDNA clone, B2-1 comprising the B chain open reading frame and 980 bp of 5'untranslated region (30) was isolated after Eco RI digestion and ligated to Eco RI digested SP64 (Promega Scientific, ref. 40). This clone lacks the first 46 bp of the c-sis gene following the RNA initiation site, and is designated Δ1-46. Restriction enzyme sites in the multiple cloning region of SP64, located just before the c-sis sequences were utilized for creation of the deletions. Δ1-46 was digested with either Xho I and Sal I, Bam HI, Pst I, or Bss HII, and then religated to produce deleted clones Δ1-480, Δ1-773, Δ 1-890, and Δ148-974, respectively (deletion end-points shown in Fig. 3 and Fig. 4). The names of the clones indicate the c-sis sequences deleted in each case.

In vitro transcription reactions (25 μl) contained 2 μg of Mst II linearized DNA template, 40 mM Tris-Cl, pH 7.5, 6 mM magnesium chloride, 2 mM spermidine, 10 mM dithiotreitol, 100 μg/ml bovine serum albumin, 25 U RNasin (Promega Scientific), and 0.5 mM CTP, UTP, and ATP. m^7GpppG was present in the reaction mixtures at a concentration of 2.0 mM and the GTP concentration was 0.1 mM. The transcription reaction was begun by the addition of 10 U of SP6 polymerase. Following incubation for 2 hrs at 40 degrees C, 5 U RNase-free DNase I (Promega Scientific) was added

and the reaction continued for 30 min. at 37 degrees C. The mixture was phenol extracted and ethanol precipitated twice with 1.0 M ammonium acetate, once with 0.15 M potassium acetate, and washed with 70% ethanol. Aliquots of the RNAs were denatured with 2.0 M formaldehyde and 50% (vol/vol) formamide containing 0.04 M morpholinopropanesulfonic acid, 20 mM sodium acetate, and 1 mM EDTA (1 X MOPS) by heating at 65 degrees C for 10 min. prior to electrophoresis on a 1.0% agarose minigel containing 1.0 M formaldehyde and 1 X MOPS. Following electrophoresis, RNA transcripts were visualized by eluting the formaldehyde with distilled water prior to staining with ethidium bromide.

In vitro translation was performed with 0.5 μg of each capped transcript in a 60 min. reaction at 25 degrees C in 30 μl of a micrococcal nuclease-treated wheat germ cell-free translation system (Bethesda Research Laboratories) containing 100 mM potassium, 2.0 mM magnesium, 30 μCi ^{35}S-methionine (800 Ci/mmole), and 8 U of RNasin (Promega Scientific) according to the manufacturer's recommendations. Peptidyl-tRNAs were removed following the incubation period by treating the translation products with RNase A (200 μg/ml) at 25 degrees C, for 10 min. The incubation mixtures were precipitated with 10% (wt/vol) trichloracetic acid (TCA), washed twice with 5% (wt/vol) TCA, solubilized in Laemmli sample buffer (0.125 M Tris-Cl, pH 6.8, 2% (wt/vol) SDS, 0.7 M 2-mercaptoethanol, 10% (vol/vol) glycerol, and 0.04% (wt/vol) bromphenol blue), and analyzed on a 10-15% (wt/vol) gradient SDS polyacrylamide gel, prior to processing by autoradiography.

RESULTS

The c-sis RNA initiation site

A recombinant phage clone, λSIS-A, was obtained which includes the 5' portion of the human *c-sis* gene and 1361 base pairs (bp) upstream of the RNA initiation site. This clone also includes exons 1-4 and overlaps with a previously described normal human genomic clone, λ33 with *c-sis* exons 2-7 (21). The latter clone includes all portions of the human genome which are homologous to *v-sis*. These two clones completely define the *c-sis* gene.

The 5' portion of λSIS-A was subcloned and designated pSIS-1.

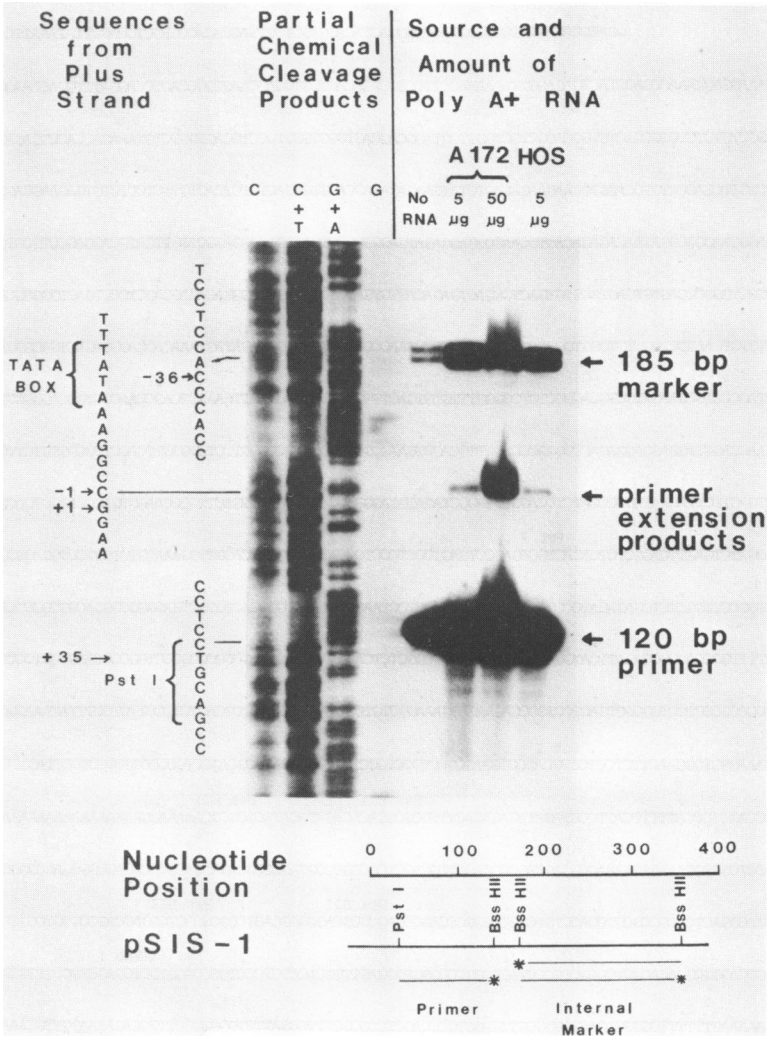
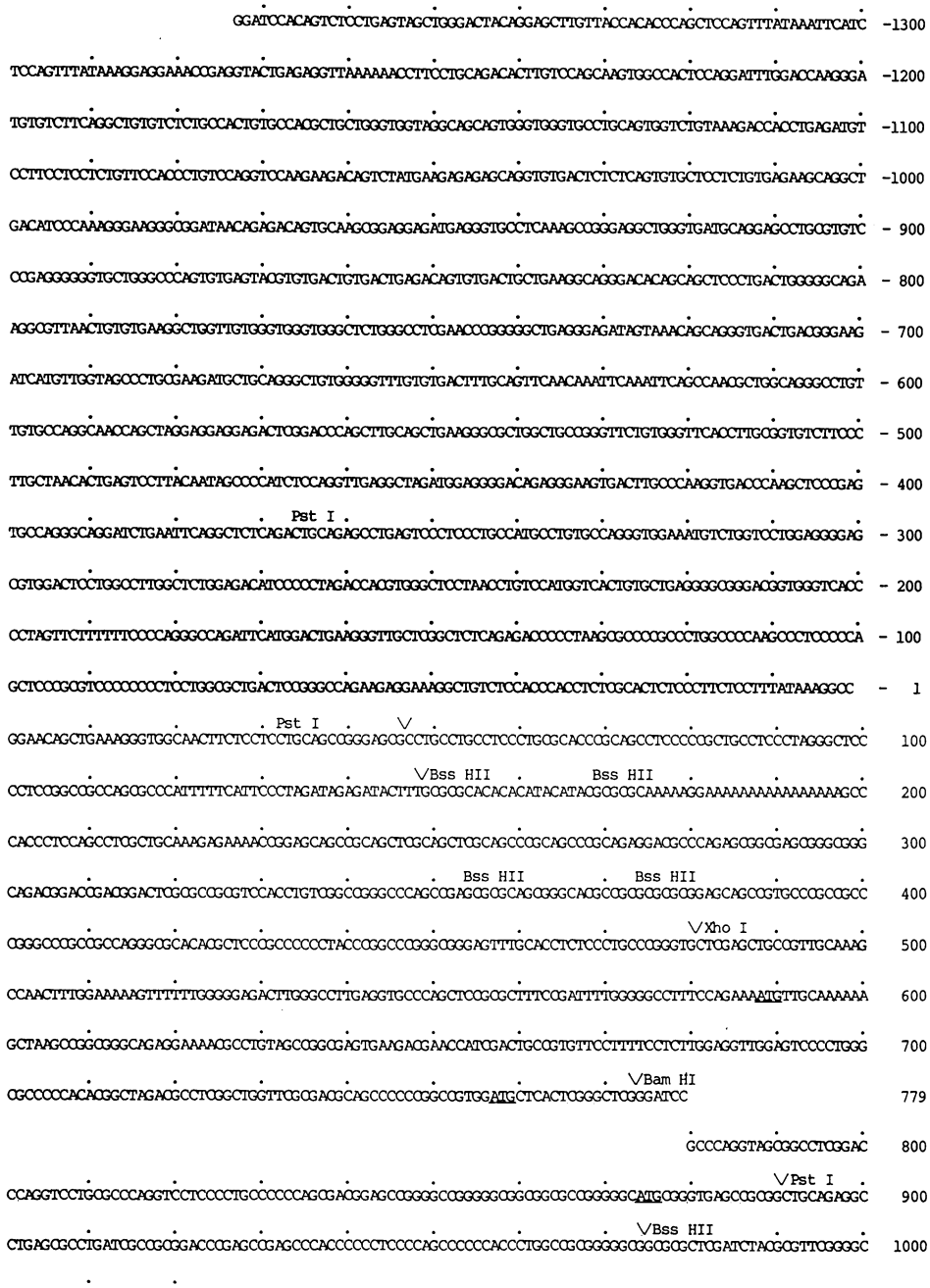


Fig. 1. Primer extension mapping of the c-sis transcriptional initiation site. A 121 nucleotide primer (nucleotides 33-153) was annealed with poly A(+) RNA from the indicated sources or with no RNA, as described in the Materials and Methods section. Samples were extended with reverse transcriptase and electrophoresed on a denaturing polyacrylamide gel. Partial chemical cleavage products electrophoresed in parallel were used as size markers. Two 185 nucleotide fragments were included in the samples as size markers. At the bottom of the figure is shown the nucleotide positions, relevant restriction enzyme sites in pSIS-1, and the DNA fragments labeled at the end indicated by an asterisk which were used as primer and internal size marker.


GGATCCACAGTCTCCTGAGTAGCTGGGACTACAGGAGCTTGTTACCACACCCAGCTCCAGTTTATAAATTCATC -1300
TCCAGTTTATAAAGGAGGAAACCGAGGTACTGAGAGGTTAAAAAACCTTCCTGCAGACACTTGTCCAGCAAGTGGCCACTCCAGGATTTGGACCAAGGGA -1200
TGTGTCTTCAGGCTGTGTCTCTGCCACTGTGCCACGCTGCTGGGTGGTAGGCAGCAGTGGGTGGGTGCCTGCAGTGGTCTGTAAAGACCACCTGAGATGT -1100
CCTTCCTCCTCTGTTCCACCCTGTCCAGGTCCAAGAAGACAGTCTATGAAGAGAGAGCAGGTGTGACTCTCTCAGTGTGCTCCTCTGTGAGAAGCAGGCT -1000
GACATCCCAAAGGGAAGGCGGATAACAGAGACAGTGCAAGCGGAGGAGATGAGGGTGCCTCAAAGCCGGGAGGCTGGGTGATGCAGGAGCCTGCGTGTC - 900
CCGAGGGGGGTGCTGGGCCCAGTGTGAGTACGTGTGACTGTGACTGAGACAGTGTGACTGCTGAAGGCAGGGACACAGCAGCTCCCTGACTGGGGGCAGA - 800
AGGCGTTAACTGTGTGAAGGCTGGTTGTGGGTGGGTGGGCTCTGGGCCTCGAACCCGGGGGCTGAGGGAGATAGTAAACAGCAGGGTGACTGACGGGAAG - 700
ATCATGTTGGTAGCCCTGCGAAGATGCTGCAGGGCTGTGGGGGTTTGTGTGACTTTGCAGTTCAACAAATTCAAATTCAGCCAACGCTGGCAGGGCCTGT - 600
TGTGCCAGGCAACCAGCTAGGAGGAGGAGACTCGGACCCAGCTTGCAGCTGAAGGCGCTGGCTGCCGGGTTCTGTGGGTTCACCTTGCGGTGTCTTCCC - 500
TTGCTAACACTGAGTCCTTACAATAGCCCCATCTCCAGGTTGAGGCTAGATGGAGGGACAGAGGAAGTGACTTGCCCAAGGTGACCCAAGCTCCCGAG - 400
Pst I
TGCCAGGGCAGGATCTGAATTCAGGCTCTCAGACTGCAGAGCCTGAGTCCCTCCCTGCCATGCCTGTGCCAGGGTGGAAATGTCTGGTCCTGGAGGGGAG - 300
CGTGGACTCCTGGCCTTGGCTCTGGAGACATCCCCCTAGACCACGTGGGCTCCTAACCTGTCCATGGTCACTGTGCTGAGGGGCGGGACGGTGGGTCACC - 200
CCTAGTTCTTTTTTCCCCAGGGCCAGATTCATGGACTGAAGGGTTGCTCGGCTCTCAGAGACCCCCTAAGCGCCCCGCCCTGGCCCCAAGCCCTCCCCCA - 100
GCTCCCGCGTCCCCCCCCCTCCTGGCGCTGACTCCGGGCCAGAAGAGGAAAGGCTGTCTCCACCCACCTCTCGCACTCTCCCTTCTCCTTTATAAAGGCC - 1
Pst I
GGAACAGCTGAAAGGGTGGCAACTTCTCCTCCTGCAGCCGGGAGGGCCTGCCTGCCTCCCTGCGCACCCGCAGCCTCCCCCGCTGCCTCCCTAGGGCTCC 100
Bss HII
Bss HII
CCTCCGGCCGCCAGCGCCCATTTTTCATTCCCTAGATAGAGATACTTTGCGCGCACACACATACATACGCGCGCAAAAAGGAAAAAAAAAAAAAAAAGCC 200
CACCCTCCAGCCTCGCTGCAAAGAGAAAACCGGAGCAGCCGCAGCTCGCAGCTCGCAGCCCGCAGCCCGCAGAGGACGCCCAGAGCGGCGAGCGGGCGGG 300
Bss HII
Bss HII
CAGACGGACCGACGGACTCGGCGCCGCGTCCACCTGTCGGCCGGGCCCAGCCGAGCGCGCAGCGGGCACGCCGCGCGCGCGGAGCAGCCGTGCCCGCCGCC 400
Xho I
CGGGCCCGCCGCCAGGGCGCACACGCTCCCGCCCCCCTACCCGGCCCGGGCGGGAGTTTGCACCTCTCCCTGCCCGGGTGCTCGAGCTGCCGTTGCAAAG 500
CCAACTTTGGAAAAAGTTTTTTGGGGGAGACTTGGGCCTTGAGGTGCCCAGCTCCGCGCTTTCCGATTTTGGGGGCCTTTCCAGAAAATGTTGCAAAAAA 600
GCTAAGCCGGCGGGCAGAGGAAAACGCCTGTAGCCGGCGAGTGAAGACGAACCATCGACTGCCGTGTTCCTTTTCCTCTTGGAGGTTGGAGTCCCCTGGG 700
Bam HI
CGCCCCCACACGGCTAGACGCCTCGGCTGGTTCGCGACGCAGCCCCCCGGCCGTGGATGCTCACTCGGGCTCGGGATCC 779
GCCCAGGTAGCGGCCTCGGAC 800
Pst I
CCAGGTCCTGCGCCCAGGTCCTCCCCTGCCCCCCAGCGACGGAGCCGGGGCCGGGGGCGGCGGCGCCGGGGGCATGCGGGTGAGCCGCGCTGCAGAGGC 900
Bss HII
CTGAGCGCCTGATGCCGCGGACCCGAGCCGAGCCCACCCCCCTCCCCAGCCCCCCACCCTGGCCGCGGGGGCGGCGCGCTCGATCTACGCGTTCGGGGC 1000
CCCGCGGGGCCGGGCCCCGAGTCGGC 1026

This subclone was used to define the c-sis RNA initiation site (Fig. 1). For this experiment, an end-labeled (-) strand 121 bp DNA fragment was generated by digestion of pSIS-1 with Bss HII and Pst I. This fragment was annealed with poly A(+) RNA from the human osteogenic sarcoma cell line, HOS, or the glioblastoma cell line, A172. The annealed DNA fragment was extended with reverse transcriptase and the size of the extended products determined by denaturing polyacrylamide gel electrophoresis in comparison to partial chemical cleavage products from a DNA fragment labeled at the same Bss HII site. Two labeled 185 nucleotide fragments were included as internal size markers for this analysis. A single extended product of 153 nucleotides was identified in each reaction. Evidence that this extended product denotes the RNA initiation site is: a) the identical size of the extended products obtained using either HOS or A172 poly A(+) RNA, b) the increased intensity of the band for the 153 nucleotide fragment after annealing with 50 μg of A172 poly A(+) RNA compared to 5.0 μg poly A(+) RNA, and c) the lack of a band for a DNA fragment of this size when performing the experiment in the absence of RNA. The CAP site is only 46 bp upstream of the 5'terminus of a previously characterized c-sis cDNA (30).

Sequences located 5' of the c-sis RNA initiation site

The structure of the insert in pSIS-1 was determined by nucleotide sequencing and is presented in Fig. 2. The RNA initiation site is preceded by a TATA box at position -10 to -7. No sequence consistent with the CCAAT consensus sequence (45) was identified upstream of the c-sis RNA initiation site, similar to that found in many other eukaryotic promoters. Located at -220 to -211 is a decanucleotide sequence which is identical to the Sp1

Fig. 2. Sequences surrounding the c-sis transcriptional initiation site. The sequence of the pSIS-1 insert (nucleotides - 1373 to 779) and downstream sequences of exon 1 as derived from the B2-1 cDNA clone (nucleotides 779 to 1026, ref. 30) are shown. Nucleotide positions are shown to the right of each line, and position +1 corresponds to the c-sis CAP site. Above each line of sequences are indicated a dot for every tenth nucleotide and restriction enzyme sites relevant to the analyses shown in Fig. 1, 3, and 4. The deletion end-points of the clones used in Fig. 4 are also shown above the sequence lines. The three ATG codons upstream of the authentic ATG initiating the B chain open reading frame are underlined.

binding consensus sequence (G/T)(G/A)GGCG(G/T)(G/A)(G/A)(C/T) (46); a similar sequence is located on the minus strand of the B chain promoter located at position -128 to -119. In this same region of the plus strand is the sequence CCGCCC. This sequence is repeated six times in the 21 bp repeats in the SV40 early promoter. Located at position -435 is a purine-rich region containing the sequence GGAAGTGA. This sequence is identical to the consensus sequence in the adenovirus E1a enhancer (47) and is also found in the 5' flanking sequences of the adenovirus major late protein gene, polyoma virus early gene, avian sarcoma virus, Friend spleen focus-forming virus, and mouse mammary tumor virus LTRs, and human α_2 and β -interferon genes (47,48). The longest direct repeat sequences preceding the RNA initiation site are found at positions -1317 to -1304 and -1298 to -1285. No long regions of purine-pyrimidine asymmetry which may form a Z-DNA structure are found.

The structure of sequences preceding the RNA initiation site in the c-sis gene in cells expressing c-sis transcripts (HTLV-I infected lymphoid, HOS and HT1080 sarcoma, and A172 glioblastoma cell lines) was compared to that of cells in which the gene is transcriptionally inactive (SKES sarcoma cell line and non-activated peripheral blood mononuclear cells). This was done by Southern blot hybridization using a probe complementary to nucleotides 1019 - 1068. Rearrangements or amplification of the c-sis gene were not detected in these samples which were examined after digestion with Hind III, Bst EII, or several other restriction enzymes (data not shown).

Transcriptional promoter activity of the c-sis gene

The region upstream of the RNA start site was also tested for transcriptional promoter activity. The 403 bp Pst I fragment (nucleotides -366 to +37) was cloned into the Pst I site of the plasmid SP65-CAT3, 5' to the bacterial chloramphenicol acetyltransferase (CAT) gene in the correct and incorrect transcriptional orientations in plasmids SIS-CAT1 and SIS-CAT2, respectively (Fig. 3a).

Transfection of SIS-CAT1 into HOS cells resulted in CAT activity in the cell extracts which gave rise to 14.5% conversion of 14C-chloramphenicol to its acetylated metabolites (Fig. 3b).

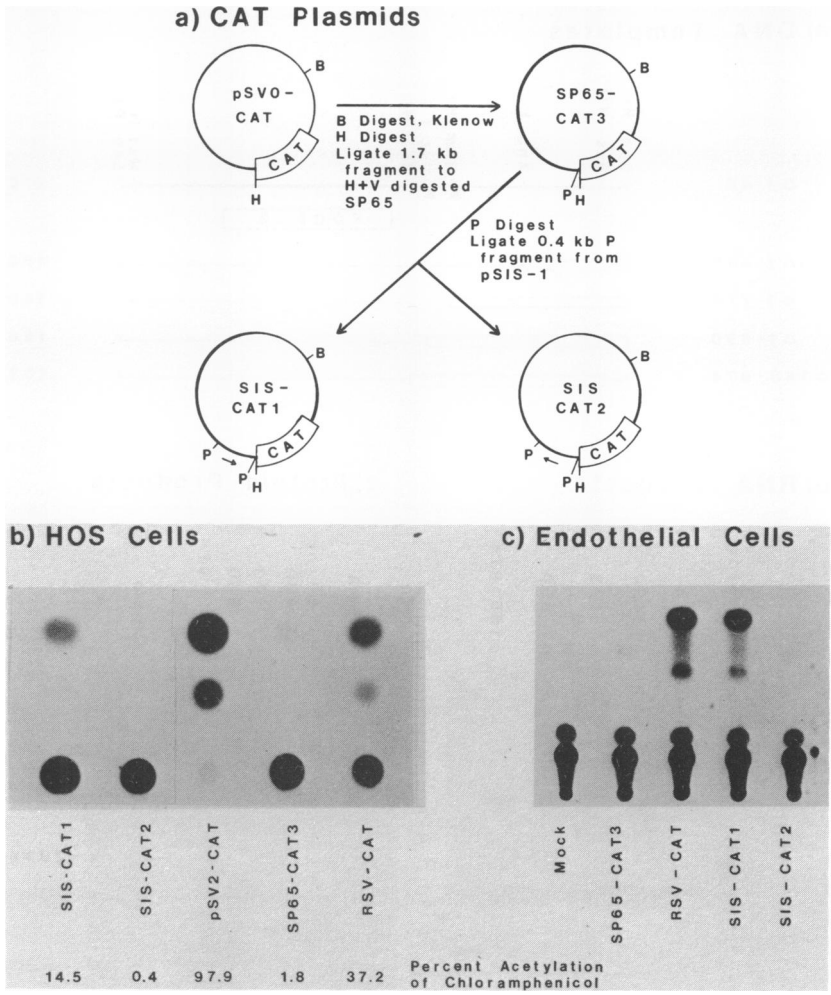


Fig. 3. Transcriptional promoter activity of the c-sis gene. a) The method of construction of plasmids SIS-CAT1 and SIS-CAT2 with the c-sis sequences in the same or opposite transcriptional orientation relative to the CAT gene, respectively, is shown. Details are provided in the Materials and Methods section. The positions of the restriction enzyme sites in the drawing are approximate. Abbreviations include Bam HI (B), Hind III (H), Pst I (P), and Pvu II (V). b) Measurements of transcriptional activity of c-sis plasmids in HOS cells. The plasmid DNA used in each case is indicated, and the experimental details are provided in the Materials and Methods section. The position of unacetylated chloramphenicol is at the bottom, and 1- and 3-acetyl chloramphenicol forms above. The percent acetylation is shown below the corresponding picture of the autoradiograph. c) Measurements of transcriptional promoter activity in endothelial cells.

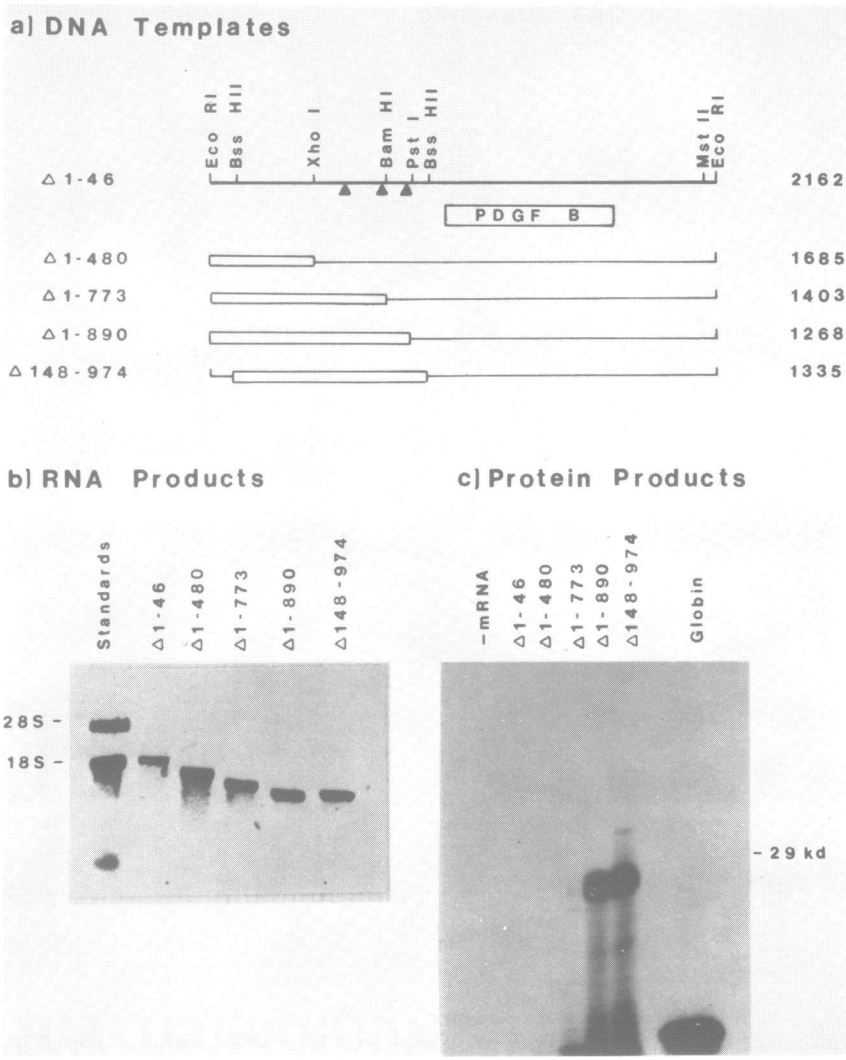


Fig. 4. Regulation of expression of the PDGF B polypeptide by 5'untranslated mRNA sequences. a) Construction of deletions of the c-sis 5'untranslated region. Details are provided in the Materials and Methods section. The name of the clone designates exon 1 sequences which have been deleted. The numbers on the right side of the figure indicate the expected size in nucleotides of the full-length transcripts produced by SP6 polymerase during in vitro transcription from the polymerase recognition site. The positions of relevant restriction enzyme sites in clone Δ1-46 are shown at the top of the figure, as well as the position of the PDGF B open reading frame initiated from the fourth ATG codon in the sequence. The positions of the first three ATG codons are

indicated by solid triangles. b) Size analysis of SP6/c-sis transcripts. In vitro transcription was performed as described in the Materials and Methods section. An aliquot of each transcript was electrophoresed on a denaturing agarose gel and stained with ethidium bromide. The first lane shows 28S and 18S rRNA species, and the remaining lanes show the products using the indicated templates. c) In vitro translation of SP/c-sis transcripts. Each transcript was translated in a wheat germ extract as described in the Materials and Methods section. The labeled products were reduced and analyzed on a polyacrylamide gel, prior to processing by autoradiography. The first lane shows the reaction products in the absence of exogenous mRNA. The last lane shows the products using globin mRNA. The other lanes show the protein products using capped transcripts from each of the indicated templates. The position of migration of carbonic anhydrase (29 kd) is indicated.

This was inferior to the transcriptional activity of pSV2-CAT which utilizes a simian virus 40 promoter, and RSV-CAT which utilizes the Rous sarcoma virus long terminal repeat sequence (LTR) as a transcriptional promoter. However, the activity was significantly greater than that seen with SP65-CAT3 which lacks a eukaryotic transcriptional promoter, or SIS-CAT2 in which the orientation of the c-sis sequences is opposite to that of the CAT gene. SIS-CAT1 has also been found to have transcriptional promoter activity in normal cultured endothelial cells (Fig. 3c), but not in human dermal fibroblasts (our unpublished observations).

Sequences located 3' of the c-sis RNA initiation site: The 5' untranslated region

The position of the RNA initiation site delineates the size of exon 1 as 1088 bp. The fourth ATG codon is the likely initiation codon for PDGF polypeptide chain B, and precedes the long open reading frame predicted by the c-sis sequences (30). There are three upstream ATG codons in exon 1 at nucleotide positions 588, 757, and 874 (Fig. 2). These ATG sequences initiate open reading frames of 5, 10, and 12 codons, respectively. The sequences surrounding all four of the ATG codons match the consensus signal for translational initiation poorly (49).

The 5' untranslated region contains 4 copies of a 7 bp imperfect tandem repeat sequence (nucleotides 241-268), and a polypurine tract of 23 residues (nucleotides 175-198). The 5' untranslated region has a high overall G + C content (70%). In

the area immediately upstream of the PDGF coding region, the G + C content increases to 85%. The high G + C content suggests that the PDGF B chain 5' untranslated region may have considerable secondary structure *in vivo*.

To determine the effect of the *c-sis* 5' untranslated region on translation efficiency *in vitro*, the endothelial *c-sis* cDNA clone B2-1 (30) was transferred to the vector SP64. This plasmid contains the bacteriophage SP6 promoter sequence directly upstream of a multiple restriction enzyme site cloning region, and has been utilized to generate functional mRNAs for *in vitro* translation. *C-sis* RNA transcripts were synthesized with SP6 polymerase from plasmids linearized with the restriction enzyme Mst II (Fig. 4a). We chose Mst II because a) it uniquely cleaves all of the deletion constructions at the same location 390 bp downstream from the B chain open reading frame termination codon, and b) it generates 5' protruding ends which give the least amount of spurious transcription. The CAP analogue 7-methylguanosine triphospoguanosine, m7GpppG, was included in the transcription mixtures. The expected sizes of the full length transcripts generated from the constructs are indicated in Fig. 4a. Size analysis of the *in vitro* synthesized transcripts on denaturing agarose gels show that for each construction only one species was produced and that the transcripts synthesized were full length (Fig. 4b).

The relative translational efficiencies of capped transcripts synthesized *in vitro* were determined by translation in a wheat germ cell-free system, and analysis of the reduced products was by sodium dodecyl polyacrylamide gel electrophoresis (Fig. 4c). Translation of the PDGF B chain open reading frame should give rise to a 28 kd polypeptide (28). The first lane shows the translation products in the absence of added mRNA, and the next five lanes show the translation products of the transcripts from the deletion constructions, Δ 1-46, Δ 1-480, Δ 1-773, Δ 1-890, and Δ148-974. Transcripts from constructions Δ1-46 and Δ1-480 were essentially not translated. The transcript from Δ1-773 generates about ten-fold less 28 kd polypeptide than that from constructs containing larger deletions, Δ 1-890 and Δ148-974. Thus, the *c-sis* 5'untranslated region inhibits translation of the PDGF B chain open reading frame *in vitro*, and inhibition is decreased by deletion in the 5' untranslated region.

Discussion

The c-sis gene encodes the B polypeptide chain of PDGF. There is considerable evidence suggesting a role for this protein in normal development of mesencymal tissues, wound repair, chemotaxis, vasoconstriction, atherosclerosis, and tumorigenesis (reviewed in ref. 4). Most data suggests that transcriptional control of the c-sis gene is the major point of regulation of the expression of PDGF chain B, though one can not fully exclude a possible role for post-transcriptional events. Thus, to elucidate the molecular details involved in these physiological processes, a study of transcriptional and translational regulatory regions of the c-sis gene was undertaken.

The work described here together with previous studies of c-sis genomic and cDNA sequences (16,21,28,30,50) provides a comprehensive understanding of the structure of this gene. The c-sis gene spans 23 kb and includes seven exons. Exon 1 includes a long 5' untranslated region of 1026 bp, larger than that of most other eukaryotic genes (51). This region has a number of other unusual structural features. First, there are four copies of a tandemly repeated sequence of 7 bp. Second, the 5' untranslated region is G + C rich and may assume considerable secondary structure. Third, the long open reading frame which encodes PDGF polypeptide B is initiated by the fourth ATG codon in the sequence (30). Though utilization of an ATG codon for translational initiation other than the first has been shown for several viral mRNAs (52-55) and cellular mRNAs (56-59), this aspect of c-sis mRNA structure remains unusual. Sequences surrounding the fourth ATG codon in c-sis mRNA do not appear to be a closer match for the consensus signal for translational initiation than are those flanking the first and third ATG codons. Sequences surrounding the second ATG codon are least favorable for initiation in light of the pyrimidine at position -3 which is predicted to have a dominant inhibitory effect. Each of the first three ATG codons is followed by only short open reading frames. It is also possible that the secondary structure of this region of the transcript may play a role in selection of an initiator AUG codon recognized by ribosomes (60-62). In any event, this unusual mRNA structure may provide for post-transcriptional regulation of the c-sis gene similar to that demonstrated in other cases

(63-65). Thus, there appears to be a complex regulatory network controlling c-sis gene expression. This may be important in normal physiologic regulation as well as for the prevention of inappropriate expression of a proto-oncogene product which may have pathological consequences.

The data presented here maps the RNA initiation site to a unique site. The same site was identified in each of two c-sis expressing cell lines, HOS and A172. This is unlike the findings of several other eukaryotic genes in which multiple RNA initiation sites have been found (66-68).

The CAP site is preceded by a TATA box at positions -10 - -7, which is a somewhat shorter separation than is found in most eukaryotic genes where the TATA box is generally 20-30 bp from the RNA initiation site (45). The start site mapped here disagrees with two previous reported positions for the B chain transcriptional start site at positions 15 and 21, respectively (69, 70). Premature termination of reverse transcription or RNA folding could account for some of the smaller primer extension and S1 nuclease resistant products, since in all reports the start site is unique. Two GGGCGG boxes are found in opposite orientations at positions -128 - -119 and -220 - -211 relative to the RNA start site. Though similar GGGCGG boxes are found in many other eukaryotic genes (71-77), these sequences are generally closer to the CAP site. We must await further analysis of transcriptional activity of the c-sis gene in order to determine whether these sequences mediate binding of an SP1-like protein.

Only short direct repeat sequences are found 5' to the c-sis RNA initiation site. The longest repeat sequences identified are 14 bp in length located at positions -1317 to -1304 and -1298 to -1285. In addition, there is a sequence at -435 similar to an enhancer found in the adenovirus E1a gene, and several other genes. The role, if any, of these sequences in transcriptional activity from the c-sis gene will be determined by mutational analysis.

Sequences within the 366 bp region upstream of the RNA initiation site demonstrate transcriptional promoter activity. This provides further proof for the assignment of the RNA initiation site. In addition it provides reagents for an analysis of

the functional role of the individual sequence elements described above in transcriptional activity of the c-sis gene. Though activity was demonstrated for the SIS-CAT plasmid in HOS and endothelial cells, no attempt was made to quantitate relative transcriptional activity in different tissue types. Tissue specificity of c-sis transcription, and definition of sequences responding to trans regulatory factors are currently under investigation.

Lastly, we have found that sequences in the 5' untranslated mRNA sequences are capable of inhibiting the expression of the c-sis protein product in vitro. Deletion of a portion of these sequences results in a considerably higher level of translation. The relative role of secondary structure and upstream ATG codons in this process remains to be elucidated. Similar data supporting a role for translational regulation in vivo has also been obtained (our unpublished data). These findings provide a framework for a comprehensive analysis of the regulation of expression of the c-sis gene product. Such studies will likely be important in understanding the pathogenesis of atherosclerosis and certain aspects of neoplastic development.

NOTE

Two publications have appeared while this work was in progress which describe more limited sequence data upstream of the c-sis RNA initiation site in agreement with those reported here (69,78).

ACKNOWLEDGEMENTS

We thank S. Josephs and B. Hahn for providing high molecular weight samples, P. Reddy for the oligonucleotide probe, J. Boss for assistance with the in vitro transcription reactions, D. Tanen and P. Riendeau for technical assistance, and J. Pober and S. Orkin for enthusiastic support. This work was supported by a contractural agreement between Washington University and the Monsanto Co., by an institutional American Cancer Society grant to Washington University, and by NIH grant HL35716 to T.C. L.R. is a Hartford Foundation fellow.

*To whom reprint requests should be addressed at: Box 8125, 660 S.Euclid, Washington University, St Louis, MO 63110, USA

REFERENCES

1. Deuel, T.F. and Huang, J.S. (1984). Blood 64:951-958.
2. Heldin, C.-H. and Westermark, B. (1984). Cell 37:9-20.
3. Ratner, L., Josephs, S.F., and Wong-Staal, F. (1985). Ann. Rev. Microbiol. 39:419-449.
4. Ross, R., Raines, E.W., and Bowen-Pope, D.F. (1986). Cell 46:155-169.
5. Scher, C.D., Shepard, R.C., Antoniades, H.N., and Stiles, C.D. (1979). Biochem. Biophys. Acta 560:217-241.
6. Stiles, C.D. (1983). Cell 33:653-655.
7. Kohler, N. and Lipton, A. (1974) Exp. Cell Res. 87:297-301.
8. Barrett, T.B., Gajdusek, C.M., Schwartz, S.M., McDougall, J.K., and Benditt, E.P. (1984). Proc. Natl. Acad. Sci. U.S.A. 81:6772-6774.
9. Martinet, Y., Bitterman, P.B., Morenx, J.-F., Grotendorst, G.R., Martin, G.R., and Crystal, R.G. (1986). Nature 319:158-160.
10. Seifert, R.A., Schwartz, S.M., and Bowen-Pope, D.F. (1984). Nature 311:669-671.
11. Walker. L.N., Bowen-Pope, D.F., Ross, R., and Reidy, M.A. (1986). Proc. Natl. Acad. Sci. U.S.A. 83:7311-7315.
12. Deuel, T.F., Huang, J.S., Proffitt, R.T., Baenziger, J.U., Chang, D., and Kennedy, B.B. (1981). J. Biol. Chem. 256:8896-8899.
13. Heldin, C.-H., Johnsson, A., Wennergren, S., Wernstedt, C., Betsholtz, C., and Westermark, B. (1986). Nature 319:511-514.
14. Heldin, C.-H., Westermark, B., and Wasteson, A. (1981). Biochem. J. 193:907-913.
15. Stroobant, P. and Waterfield, M.D. (1984). EMBO J. 3:2963-2967.
16. Chiu, I.-M., Reddy, E.P., Givol, D., Robbins, K.C., Tronick, S.R., and Aaronson, S.A. (1984). Cell 37:123-129.
17. Devare, S.G., Reddy, E.P., Law, D.J., Robbins, K.C., and Aaronson, S.A. (1983). Proc. Natl. Acad. Sci. U.S.A. 80:731-735.
18. Doolittle, R.F., Hunkapiller, M.W., Hood, L.E., Devare, S.G., Robbins, K.C., Aaronson, S.A., and Antoniades, H.N. (1983). Science 221:275-277.
19. Gelmann, E.P., Wong-Staal, F., Kramer, R.A., and Gallo, R.C. (1981). Proc. Natl. Acad. Sci. U.S.A. 78:3373-3377.
20. Johnsson, A., Heldin, C.-H., Wasteson, A., Westermark, B., Deuel, T.F., Huang, J.S., Seeburg, P.H., Gray, A., Ullrich, A., Scrace, G., Stroobant, P., and Waterfield, M.D. (1984). EMBO J. 3:921-928.
21. Josephs, S.F., Guo, C., Ratner, L., and Wong-Staal, F. (1984). Science 223:487-491.
22. 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). Nature 304:35-39.
23. Besmer, P., Snyder, H.W., Murphy, J.E., Hardy, W.D., and Parodi, A. (1983). J. Virol. 46:606-613.
24. Dalla Favera, R., Gallo, R.C., Giallongo, A., and Croce, C.M. (1982). Science 218:686-688.
25. Harper, M.E., Franchini, G., Love, J., Simon, M.I., and Wong-Staal, F. (1983). Nature 304:169-171.
26. Betsholtz, C., 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). Nature 320:695-699.
27. Goustin, A.S., Betsholtz, C., Pfeifer-Ohlsson, S., Persson, H., Rydnert, J., Bywater, M., Holmgren, G., Heldin, C.-H., Westermark, B., and Ohlsson, R. (1985). Cell 41:301-312.
28. Josephs, S.F., Ratner, L., Clarke, M.F., Westin, E.H., Reitz, M.S., and Wong-Staal, F. (1984). Science 225:636-639.
29. Ratner, L., Josephs, S.F., Jarrett, R., Reitz, M.S., and Wong-Staal, F. (1985). Nucl. Acids Res. 13:5007-5018.
30. Collins, T.D., Ginsburg, D., Boss, J.M., Orkin, S.H., and Pober, J.S. (1985). Nature 316:748-750.
31. Jaye, M., McConathy, E., Drohan, W., Tong, B., Deuel, T., and Macaig, T. (1985). Science 228:882-885.
32. Eva, A., Robbins, K.C., Andersen, P.R., Srinivasan, A., Tronick, S.R., Reddy, E.P., Ellmore, N.W., Galen, A.T., Lautenberger, J.A., Papas, T.S., Westin, E.H., Wong-Staal, F., Gallo, R.C., and Aaronson, S.A. (1982). Nature 295:116-119.
33. Van Zoelen, E.J.J., van de Ven, W.J.M., Franssen, H.J., van Oostewaard, T.M.J., van der Saag, P.T., Heldin, C.-H., and de Laat, S.W. Mol. Cell. Biol. 5:2289-2297.
34. Betsholtz, C., Westermark, B., Ek, B., and Heldin, C.-H. (1984). Cell 39:447-457.
35. Graves, D.T., Owen, A.J., Barth, R.K., Tempst, P., Winoto, A., Fors, L., Hood, L.E., and Antoniades, H.N. (1984). Science 226:972-974.
36. Westermark, B., Johnsson, A., Paulsson, Y., Betsholtz, C., Heldin, C.-H., Herlyn, M., Rodeck, U., and Koprowski, H. (1986). Proc. Natl. Acad. Sci. U.S.A. 83:7197-7200.
37. Westin, E.H., Wong-Staal, F., Gelmann, E.P., Dalla Favera, R., Papas, T.S., Lautenberger, J.A., Eva, A., Reddy, E.P., Tronick, S.R., Aaronson, S.A., and Gallo, R.C. (1982). Proc. Natl. Acad. Sci. U.S.A. 79:2490-2494.
38. Clarke, M.F., Westin, E., Schmidt, D., Josephs, S.F., Ratner, L., Wong-Staal, F., Gallo, R.C., and Reitz, M.S. (1984). Nature 308:464-467.
39. Gimbrone, M.A., Jr. (1976). T.H. Spaet (ed.), Progress in Hemostasis and Thrombosis, vol. 3, Grune and Stratton, New York, pp. 1-18.
40. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, New York.
41. Maxam, A. and Gilbert, W. (1980). Meth. Enzymol. 65:499-560.
42. Broome, S. and Gilbert, W. (1985). Cell 40:537-546.
43. Gorman, C.M., Moffat, L.F., and Howard, B.H. (1982). Mol. Cell. Biol. 2:1044-1051.
44. Gorman, C.M., Merlino, G.T., Willingham, M.C., Pastan, I., and Howard, B.H. (1982). Proc. Natl. Acad. Sci. U.S.A. 79:6777-6781.
45. Breatnach, R. and Chambon, P. (1981). Ann. Rev. Biochem. 50: 349-383.
46. Briggs, M.R., Kadonaga, J.T., Bell, S.P., and Tijan, R. (1986). Science 234:47-52.
47. Hearing, P. and Shenk, T. (1983). Cell 33:695-703.
48. Goodbourn, S., Zinn, K., and Manitatis, T. (1985). Cell 41:509-521.
49. Kozak, M. (1986). Cell 44:282-292.
50. Gazit, A., Hisanaga, I., Chiu, I.-M., Srinivasan, A., Yaniv,

A., Tronick, S.R., Robbins, K.C., and Aaronson, S.A. (1984). Cell 39:89-97.
51. Kozak, M. (1984). Nucl. Acids Res. 12:857-872.
52. Briand, J.-P, Gerard, K., and Guilley, H. (1978). Proc. Natl.-Acad. Sci. U.S.A. 75:3168-3172.
53. Clerx-van Haaster, C., Akashi, H., Auperin, D., and Bishop, D. (1982). J. Virol. 41:119-128.
54. Ghosh, P.K., Reddy, V.B., Swinscoe, J., Choudhary, P., Lebowitz, P., and Weissman, S.M. (1978) J. Biol. Chem. 253: 3643-3647.
55. Hefti, E., Bishop, D.H.L., Dubin, D.T., and Stollar, V. (1978). J. Virol. 17:149-151.
56. Gough, N., Metcalf, D., Gough, J., Grail, D., and Dunn, A.R. (1985). EMBO J. 4:645-653.
57. Hendy, G., Kronenberg, H., Potts, J., Jr., and Rich, A. (1981). Proc. Natl. Acad. Sci. U.S.A. 78:7365-7369.
58. Royer-Pokora, B., Kunkel, L.M., Monaco, A.P., Goff, S.C., Newburger, P.E., Baehner, R.L., Cole, P.S., Curnutte, J.T., and Orkin, S.H. (1986). Nature 322:32-38.
59. Stanton, L.W., Schwab, M., and Bishop, J.M. (1986). Proc. Natl. Acad. Sci. U.S.A. 83:1772-1776.
60. Kozak, M. (1980). Cell 19:79-90.
61. Pavlakis, G.N., Lockard, N.E., Vamvakopoulis, N., Rieser, L., RajBhandary, V.L., and Vournakis, J.N. (1980). Cell 19:91-102.
62. Pelletier, J. and Sonenberg, N. (1985). Cell 40:515-526.
63. Hinnebusch, A.G. (1984). Proc. Natl. Acad. Sci. U.S.A. 81: 6442-6446.
64. Kahana, C. and Nathans, D. Proc. Natl. Acad. Sci. U.S.A. 82: 1673-1677.
65. Thieros, G., Penn, M.D., and Greer, H. (1984). Proc. Natl. Acad. Sci. U.S.A. 81:5096-5100.
66. Ishii, S., Kadonaga, J.T., Tijan, R., Brady, J.N., Merlino, G.T., and Pastan, I. Scinece 232:1410-1414.
67. Ishii, S., Xu, Y.-H., Stratton, R.H., Roe, B.A., Merlino, G.T., and Pastan, I. (1985). Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924.
68. Leonard, W.J., Depper, J.M., Minoru, K., Kronke, M., Pfeffer, N.J., Svetlik, P.B., Sullivan, M., and Greene, W.C. (1985). Science 230:633-639.
69. Rao, C.D., Hisanaga, I., Chiu, I.-M., Robbins, K.C., and Aaronson, S.A. (1986). Proc. Natl. Acad. Sci. U.S.A. 83: 2392-2396.
70. Van den Ouweland, A.M.W., Groningen, J.J.M., Schalken, J.A., Van Neck, H.W., Bloemers, H.P.J., and Van de Ven, W.J.M. (1987) Nucl. Acids Res. 15:959-970.
71. Dynan, W.S., Sazer, S., Tjian, R., and Schimke, R.T. (1986). Nature 319:246-248.
72. Dynan, W.S. and Tjian, R. (1983). Cell 35:79-87.
73. Gidoni, D., Kadonaga, J.T., Barrera-Saldana, H., Takahashi, K., Chambon, P., and Tjian, R. (1985). Science 230:511-517.
74. Jones, K.A., Kadonaga, J.T., Luciw, P.A., and Tjian, R. (1986). Science 232:755-759.
75. Jones, K.A. and Tjian, R. (1985). Nature 317:179-182.
76. Kelley, D., Colecough, C., and Perry, R.P. (1982). Cell 29:681-689.
77. McKnight, S.L. (1982). Cell 31:355-365.
78. Van den Ouweland, A.M.W., Roebroek, A.J.M., Schalken, J.A., Claesen, C.A.A., Bloemers, H.P.J., and Van de Ven, W.J.M. (1986). Nucl. Acids Res. 14:765-778.