# Structural and Functional Analysis of a Growth-regulated Gene, the Human Calcyclin*

Sergio Ferrari‡, Bruno Calabretta‡, J. Kim deRiel‡, Renata Battini‡, Franco Ghezzo‡, Evelyne Lauret‡, Constance Griffin§, Beverly S. Emanuel§, Fiorella Gurrieri‡, and Renato Baserga‡

From the ‡Department of Pathology and the Fels Research Institute, Temple University Medical School, Philadelphia, Pennsylvania 19140 and the §Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104

Calcyclin was originally defined as a cDNA clone (2A9) whose cognate RNA is growth-regulated and whose sequence shows strong similarities to the sequences of the S-100 protein, a calcium-binding protein, as well as to a subunit of the major cellular substrate for tyrosine kinase. Using the full-length cDNA, we have now isolated from a human genomic library several phages containing calcyclin sequences. One of the phages, ch. 28-10, contains the entire calcyclin gene, plus extensive flanking sequences. The calcyclin gene is a unique copy gene and has 3 exons. The 5' flanking sequence has been characterized, both structurally and functionally. Besides a TATA box, it contains, in the region proximate to the cap site, GC boxes and a sequence with a strong homology to the enhancer core of the SV40 promoter. Other enhancer-like elements are found scattered in both the 5' and 3' flanking regions. The proximate 5' flanking region is very active in driving the transient expression of linked reporters in transfection experiments. Finally, the calcyclin gene has been localized to the long arm of human chromosome 1, near the ski oncogene.

Calcyclin is the name that was given to a gene originally identified as a cDNA clone (2A9) whose cognate RNA is growth-regulated (1). The steady state levels of cytoplasmic mRNA recognized by a calcyclin probe are increased when $G_o$ cells are stimulated to proliferate by either serum (1, 2), platelet-derived growth factor, or epidermal growth factor, but not by platelet-poor plasma or insulin (3). The increase in steady state levels of cytoplasmic mRNA occurs also when the cells are stimulated by serum in the presence of concentrations of cycloheximide that completely suppress protein synthesis (4), indicating that the calcyclin gene is inducible in the absence of the products of other growth factor inducible genes. The levels of calcyclin mRNA are also increased and deregulated in human acute myeloid leukemias (5, 6). The calcyclin cDNA sequence has strong homologies to the sequences of the S-100 protein and of a subunit of a large complex that serves as a cellular substrate for tyrosine kinase (3).

These and other considerations (see "Discussion") suggest that this gene is a reasonable candidate for an important role in cell cycle progression. We have isolated a genomic clone of human calcyclin and in the present paper we present its full sequence, including the introns and the 5' and 3' proximate flanking sequences. We also report the structural and functional characterization of the calcyclin promoter, and the localization of the calcyclin gene to human chromosome 1.

## MATERIALS AND METHODS

*Screening of a Genomic Library*—A human genomic library of placental DNA cloned in the vector Charon 28 was a kind gift of Dr. P. Leder (Harvard University). Fifty thousand recombinants were screened with a full-length calcyclin cDNA clone according to the procedure described by Maniatis *et al.* (7). The DNAs of 6 positive individual clones were purified after the tertiary screening. By restriction digestion analysis using the restriction enzymes EcoRI, BamHI, and HindIII and hybridization with the calcyclin cDNA ($^{32}$P labeled by the nick translation technique) it was found that they were indistinguishable and therefore represented the only species of λ clones with sequence homologies to the calcyclin cDNA clone. The BamHI restriction analysis pattern of the selected clones showed two fragments, one of 1.7 kb[1] and the other of 3.0 kb, that hybridized to the calcyclin cDNA (the 1.7-kb fragment stronger than the 3.0-kb fragment).

*Southern Blot Analysis of the Calcyclin Gene in Genomic DNA*—Chromosomal DNA was extracted from frozen human spleen as described by Gross-Bellard *et al.* (8) with minor modifications. Twenty μg of DNA were digested with the restriction enzymes EcoRI, BamHI, HindIII, PstI, and PvuII and blotted by the method of Southern (9). The $^{32}$P-labeled calcyclin cDNA was hybridized to the digested genomic DNA by standard methods.

Fig. 1 shows that the hybridizing restriction fragments generated in human genomic DNA by BamHI, 1.7 and 3.0 kb (*lane A*), are identical to those in phage ch. 28-10 (*lane B* of Fig. 1).

*Mapping of Phage λ Ch. 28-10*—One of the 6 clones selected, λ ch. 28-10, was studied in detail. We used the "Phage Lambda Mapping Quick-kit" (Collaborative Research). Briefly, fragments of λ ch. 28-10 which had been partially digested with the restriction endonucleases EcoRI, BamHI, PstI, and NcoI were hybridized to a $^{32}$P-labeled oligonucleotide 5'-d(GGGCGGCGACCT)-3' that is complementary to the phage right end. After hybridization the fragments were separated by 0.6% agarose gel electrophoresis and visualized by autoradiography. Determinations of the fragment lengths identified the position of each restriction site within the insert. We were able to construct a physical map with the relative positions of these restriction enzymes which formed the basis for our subcloning strategy.

*Subcloning of the BamHI and EcoRI Fragments of Ch. 28-10 λ DNA*—EcoRI digestion gave a single band, 6 kb in size, in ch. 28-10, that hybridized to a calcyclin probe. This EcoRI 6-kb fragment and the two BamHI fragments mentioned above (1.7 and 3.0 kb in size) were subcloned into the polylinker of the pUC9 vector by standard procedures.

The positive recombinant plasmids (pG2A9 B 3.0, pG2A9 B 1.7,

[1] The abbreviations used are: kb, kilobase pair; bp, base pair.

and pG2A9 E 6.0) were also checked by Southern blot analysis and large scale preparations of plasmid DNA were obtained by the CsCl ultracentrifugation technique.

*Endonuclease Restriction Analysis of the 3.0- and 1.7-kb BamHI Fragments and the 6.0-kb EcoRI Fragment*—The 3.0- and 1.7-kb BamHI fragments and the 6.0-kb EcoRI fragment were mapped from the recombinant plasmids mentioned above with partial digestion of several different end-labeled fragments with a panel of restriction endonucleases.

*Sequence Analysis*—The following plasmid inserts were sequenced: pG2A9 B 1.7, pG2A9 E 6.0, and pG2A9 Pr 1.5 which is a subclone of pG2A9 B 3.0 extending from the ScaI to the BamHI site (see Fig. 2). All fragments were sequenced by the Maxam and Gilbert (10) sequencing technique. Both strands were sequenced and the sequencing was repeated at least once. The cap site was determined by S1 nuclease assay (7).

*Deletion Mutant Analysis of the Promoter Region*—We used for the deletion analysis of the promoter region the recombinant plasmid pG2A9 Pr 1.5 containing the first exon and the promoter region. The insert extends from the HincII/ScaI site at the 5′ end to the BamHI site at the 3′ end as described in the legend to Fig. 3. For the different deletions the plasmid was digested with HindIII and NcoI, XhoI, or SphI, and the ends were religated. For the Sma deletions, pG2A9 Pr 1.5 was digested with SmaI and the two fragments were recovered. We called the 0.3-kb fragment the SmaI fragment, and the other fragment the Sma vector (see Fig. 3). The Sma vector was religated, while the SmaI fragment was cut with BamHI and cloned in the HincII/BamHI site of the pUC9 polylinker.

*Cloning of the CAT Gene in the Deletion Mutant Plasmids*—The CAT gene (11), previously recloned in the SmaI site of pUC9 was cut with BamHI, which cleaved in the polylinker at the 5′ end and in the 3′ end of the CAT gene after the SV40 polyadenylation signal. This BamHI fragment was cloned in the BamHI site of the polylinker of the various deletion mutants described above. However, the deletion SmaI vector plasmid was instead linearized with SmaI and then the BamHI CAT fragment was cloned in the SmaI site after filling the BamHI ends. After transformation positive colonies were selected by colony hybridization to the $^{32}$P-labeled BamHI CAT fragment. Both orientations of the CAT gene were identified in each deletion plasmid using the restriction enzyme EcoRI that cuts asymmetrically in the CAT gene.

*Cloning of the TK Gene in the Deletion Mutant Plasmids*—The intact coding sequence of the human thymidine kinase gene was obtained from pTKR2, a derivative of pTK11 (12) which contains the 1.4-kb fragment of pTK11 subcloned in pUC9. The insert in pTKR2 was released with PstI and EcoRI, purified and recut with ScaI that cuts 28 nucleotides into the 5′ noncoding region of the thymidine kinase sequence. The ScaI/EcoRI fragment was cloned in the SmaI/EcoRI restriction sites of the polylinker of the pUC 9 vector (the orientation of the fragment is SmaI at the 5′ and EcoRI at the 3′ end). After selection, the recombinant plasmid was digested with BamHI (there are two BamHI sites; one in the polylinker and the other one in the 3′ untranslated region of the TK gene) and the relevant fragment was cloned in the BamHI sites of the deletion mutant plasmids, NcoI, SphI, XhoI, and SmaI, described above. For the deletion mutant SmaI vector we used the same strategy as described for the CAT gene. Both orientations of the TK gene were selected for each deletion mutant plasmid using the restriction enzyme PvuII.

*Transfection*—tk-ts13 cells (which have a high efficiency of transfection) were transfected in suspension as described in detail by Shen *et al.* (13). In these cells, the background, by [$^3$H]thymidine incorporation, is virtually zero. Selection of TK$^+$ colonies was carried out by standard methods in hypoxanthine/aminopterin/thymidine medium.

*CAT Assay*—This was carried out as originally described by Gorman *et al.* (11).

*In Situ Hybridization*—The 2A9 (calcyclin) probe cDNA was $^3$H-labeled by nick translation to a specific activity of $3 \times 10^7$ cpm/μg of DNA. Metaphase chromosome spreads were prepared from peripheral blood lymphocytes of normal males. *In situ* hybridization was performed using a protocol described in detail elsewhere (14). After exposure for 14–21 days, slides were developed and banded using a modified Wright's Giemsa method (15).

## RESULTS

*Isolation of a Genomic Clone containing the Human Calcyclin Gene*—We have previously reported the isolation from an

Okayama-Berg library (16) of a cDNA clone containing the full-length coding sequence of the human calcyclin gene (pG2A9). Using the full-length calcyclin cDNA we have screened a human placental genomic DNA library for genomic clones carrying the calcyclin genes. Several clones were isolated and one of them, ch. 28-10, was shown to contain all the coding sequences of the calcyclin gene. The Southern blot in Fig. 1 shows the similarity of the digested DNA from the human genome and the phage ch. 28-10. By digestion with BamHI, both the genomic DNA and the insert from ch. 28-10 give two bands that hybridize to the calcyclin cDNA probe (Fig. 1), 3.0 and 1.7 kb in size, respectively. Other restriction enzymes also gave identical phages and genomic bands hybridizing to the calcyclin probe (not shown). The two BamHI fragments mentioned above were used for subcloning into pUC9 and further studies. In addition, a 6.0-kb EcoRI fragment that hybridized to the calcyclin probe was also subcloned into pUC9. By mapping the location of the 5′ end of the cDNA with a synthetic oligodeoxynucleotide and of the last exon with the 3′ end of p2A9 cDNA we were able to determine that phage ch. 28-10 contained sufficient 5′ and 3′ flanking sequences to constitute what is probably most of the functional calcyclin gene, and the insert of phage ch. 28-18 was therefore sequenced as described under "Materials and Methods."

Fig. 2 shows the physical map of the human calcyclin gene locus. The gene consists of three exons, 44, 160, and 232 base pairs, respectively, separated by two introns, respectively, of 587 and 372 base pairs. The major restriction sites are also indicated. Since there are two BamHI restriction sites in the intron between the first and second exons, two bands are expected to hybridize to the calcyclin probe in BamHI digests of genomic DNA. This, and the single 6.0-kb fragment in EcoRI digests that also hybridizes to a calcyclin probe (not shown), indicate that there is a single copy of the human calcyclin gene. Fig. 3 shows the strategy that was followed to sequence 3676 base pairs of the insert in phage ch. 28-10. Briefly, three fragments were cloned in pUC9: two BamHI fragments (1.7 and 3.0 kb) and a 6.0-kb EcoRI fragment (see
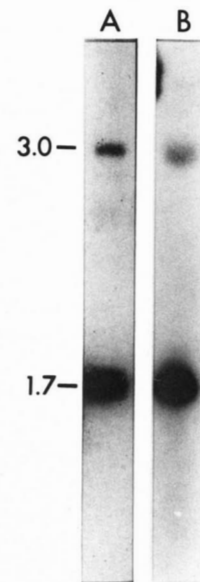


FIG. 1. **Southern blot analysis of human genomic DNA and of the ch. 28-10 phage.** Human genomic DNA (from spleen), *lane A*, and the DNA of ch. 28-10 (*lane b*) were digested with BamHI and blotted as described (9). The blot was hybridized to the nick-translated full-length cDNA of plasmid p2A9 (calcyclin insert). Size in kilobases.
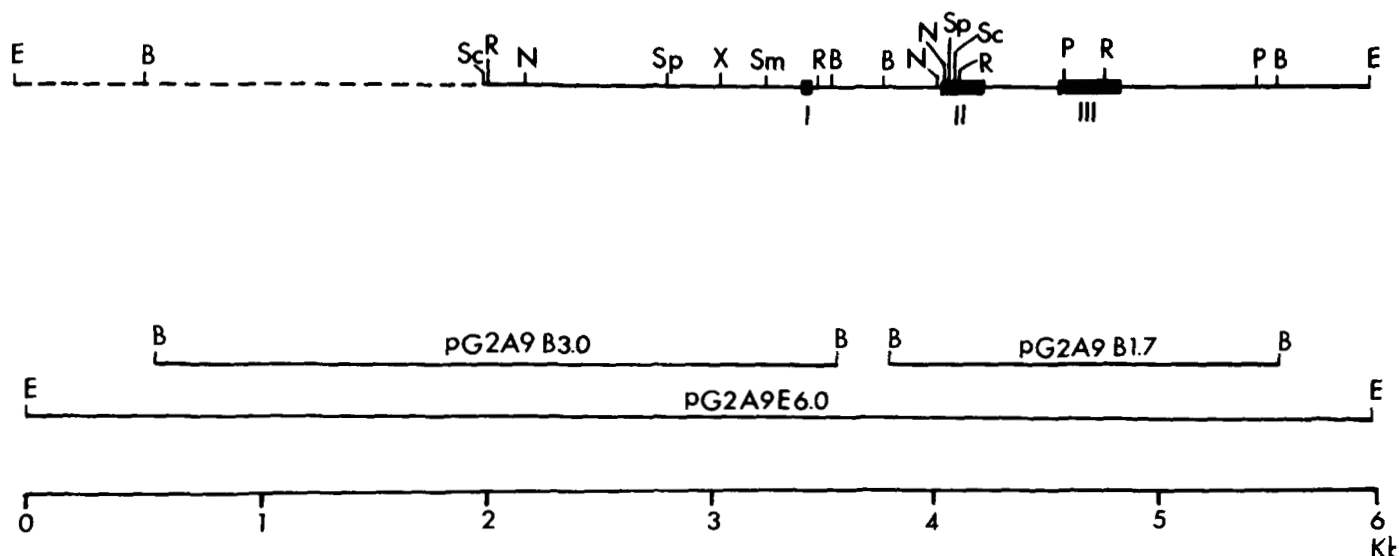
FIG. 2. **Physical map of the calcyclin gene.** The physical map is shown at the *top*. The 3 exons are given as *boxes*. Below are the 3 plasmids used for restriction analysis and sequencing (see "Materials and Methods"). The *inserts* all derive from phage ch. 28-10. *E, EcoRI; B, BamHI; Sc, ScaI; R, RsaI; Sp, SphI; X, XhoI; Sm, SmaI; P, PstI; N, NcoI.*
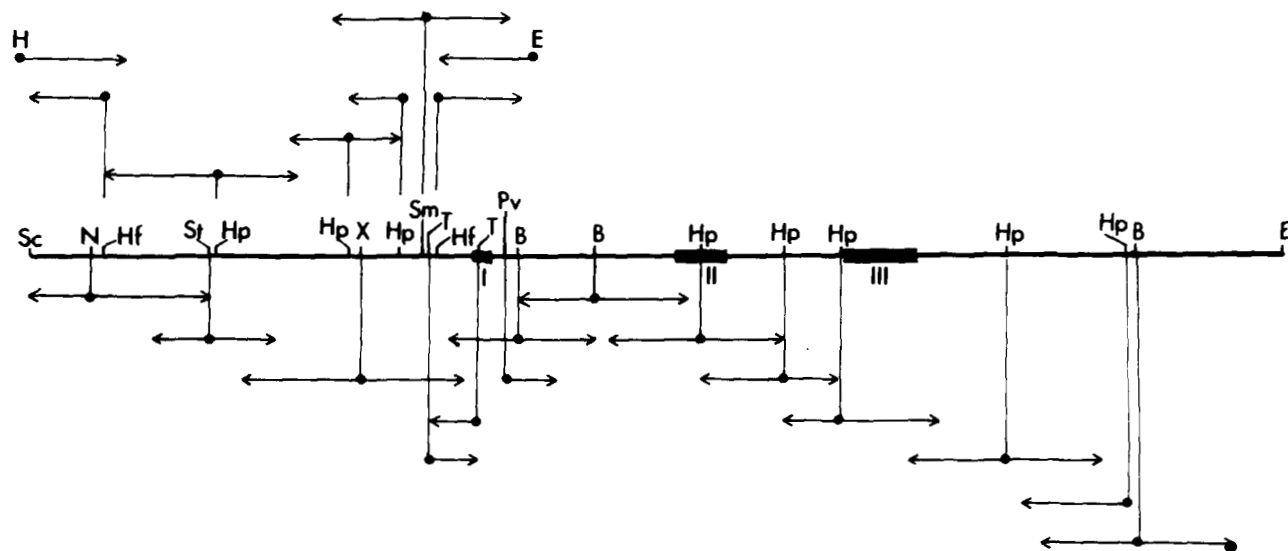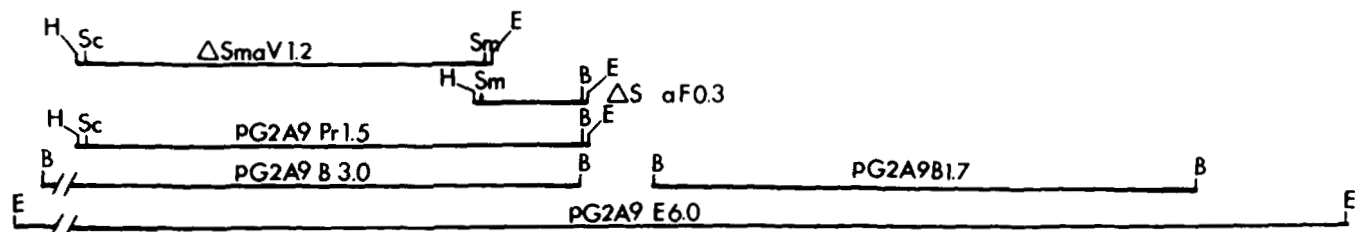


FIG. 3. **Strategy used to sequence the calcyclin gene.** The plasmids used are shown in the *upper half* of the figure. Three of the plasmids are described in the legend to Fig. 2, while pG2A9 Pr 1.5 is derived from pG2A9 B 3.0. The *2 top plasmids* are as follows: *SmaV 1.2* is the *ScaI-SmaI* fragment from pG2A9 Pr 1.5; *SmaF 0.3* is the *SmaI-BamHI* fragment from the same plasmid. *H; HindIII; Hf, HinfI; St; StuI; Hp, HpaII; Pv, PvuII; T, TaqI.* Others are as described in the legend to Fig. 2. In the 2 smaller plasmids, *HindIII* and *EcoRI* are part of the pUC9 polylinker.

"Materials and Methods"). The inserts of these subclones were then sequenced as described in the legend to Fig. 3 (the *EcoRI* and pG2A9 B 3.0 fragments were only partially sequenced, from the 3' end to the *ScaI* site in the 5' flanking region).

*Sequence of the Calcyclin Gene*—The entire sequence of the calcyclin gene, including the introns and the proximate 5' and 3' flanking sequences, is shown in Fig. 4. The introns are indicated by *lower case letters*. A recognizable TATAA box is indicated at nucleotide −29 upstream of the cap site which

```
                                                                   -1371            .                    .
                                            AGTACTCGGTGTTCCTGAGGATGCTGTGCATGGCCTACAACGACTTCTTTC
                                                 ScaI
-1320                 .                    .                    .                    .                    .
TAGAGGACAACAAGTGACCAGGGCTGCCCTCCACCCTCACCCTCCACCCTTTGCTGCTGACCTCGGCTGCTCCTCTCACAGACCCTCTTTGGCCCCTGCCCTCCTCTCCCTCCCAGATGG

-1200                 .                    .                    .                    .                    .
ACCCTTCCATGGGAGGAAATAAAGTTTCCATCGCAGGTGCTGGGAGTCTGGTTTTGAAGCTGTCTTGTCTACCTTGGCCTGGGGAGAGGGGAGCACAGGAAGGGTCTCTCCTTGAGTGGG
     NcoI
-1080                 .                    .                    .                    .                    .
TTGAGACAGCTTCTGCCTCTGGGGGTTAGGGTCCTGGGCTCCCACTGCATTCCTCTCCTTCTTTGGTGTGGACGTCATTGGTTTTGTCATGGCTTAGTTTTGCCTGCCTGGAAAATGGGG

-960                  .                    .                    .                    .                    .
AAGTTAGGCCAGGCGGGAACTCTGCAAGGATGCAGAGGAAGTTAAGAGGGAAAGTTGCTTTGAGAGGAGGACACTGGGAGGGGTTGGGAGTGGCTCCTGAGGGCGGTGATAGGCAGGCAG

-840                  .                    .                    .                    .                    .
GCCTGACTTGTCCACAGCTCACCCGGAGGCCACCTTGGCAGCACCTGTAGGAAGGGCATGTCTGGCCTCCACACCAGCCCCCTCCCTCTTCACCATTTCCCCTTCAATAGCACCACTCTC

-720                  .                    .                    .                    .                    .
ATCATCTATGGGGGACAGTGCTTTCTTCTCTCCCTGCCTCCTCCATCAAAATCTTTTCTCAGGGGAGGGTCTGAAAAGGCCTTCACTCCCCCGTAAATAACGAATGGTGCTTACAGGGCT

-600                  .                    .                    .                    .                    .
GGGCTCCCACGTGCATGCACATTAACACCAAAGGTGCTGTAGTGAATGGAATTTGGGGCACTGAGGGGAAGGCGTGGAGGTGTTGGTAGGAACTTGTTGCTGGTGGGGGATGGGCGCCGT
          SphI
-480                  .                    .                    .                    .                    .
AGATATCCTTTACACCACTGGCTACTCCCCCTATCTCCTCTGGGGTGACCCTGAGTATCCTCTGTGGGACACCGGCATCCTGTGAGGCGCCCTCCTTGCCCACATTGACGCTGCGCTGGC
                                                                                                                        X
-360                  .                    .                    .                    .                    .
TCGAGGGTCACATTCACGGTCTGGCAGAGGAAGCAGGGGTGACCGCCGCAGTCCTCCTCCTGCTCCCCTTGCCGAGTCACGTGTCACGAAGAGCAAACTGAGCAAACTGAGCTGCGCAGA
hoI
-240                  .                    .                    .                    .                    .
TGAGGGGAGACTCGTCACCAGGCGTGCAGTGGGCACTGCTGGGCTCCCCCATCCCGTCCTAACCCGGAACAGCCCCGGGCAGGAGGCGTGGAAAGTCGAGGGGGTAAACCGCGAATGTGC
                                                                                          SmaI
-120                  .                    .                    .                    .                    .
GTTGTGTAAGCCACGGCGCAGGGTGGGGCGCGGGCGGGACTTGGGCGGGCGGGGTGGGCTTGGCCGAGCTGGCCTCCGGGGCACCGACCGCTATAAGCCAGTCGGACTGCGACACAGCC

1                    .                    .                    .                    .                    .              120
CATCCCCTCGACCGCTCGCGTCGCATTTGGCCGCCTCCCTACCGgtgagttctctccaggagccctgggtactttccagggccagctgccctcacgctgggggtccagccatcccctgcc

                     .                    .                    .                    .                    .              240
cagttcagccgctggatccagactggggccatctgtggcgctcccccgctggagggatagtcaggagcagcagtgctgtgccaggcaggccttgggctaagggatcgcaatggggtgtgc
     BamHI
                     .                    .                    .                    .                    .              360
tcttttggggtgcggaagggagtgccctgggtgtgtcattgccaccatgtgtggccctgtgaagctgtgtttaagctgcctttgcagcctccattcccctccctgcccagccatactcc

                     .                    .                    .                    .                    .              480
tcaacttctggatcccctgaaggacagttctcagctgtgcccaaagctactgttcctatatgcttcttagaatccttaagccacctctcttgccttggccctagtgtgctctctccttcc

                     .                    .                    .                    .                    .              600
ccttcagccctgggctgtctcctgatgccattgtgtgtggcctgagactgggtggttccaaaggaggcggggctagtgcaggcagcattattggggtgtgtgggtgagaagtccttgctc

                     .                    .                    .                    .                    .              720
ccatggcactgactaggccctctgctgccagCTCCAAGCCCAGCCCTCAGCCATGGCATGCCCCCTGGATCAGGCCATTGGCCTCCTCGTGGCCATCTTCCACAAGTACTCCGGCAGGGA

                     .                    .                    .                    .                    .              840
GGGTGACAAGCACACCCTGAGCAAGAAGGAGCTGAAGGAGCTGATCCAGAAGGAGCTCACCATTGGCTCGgtgagtggcctcctccccaggacccctttccccacccttgtcctttggaa

                     .                    .                    .                    .                    .              960
gcaaggattagggggagagagaggtgccaggtgcatctgactcacatttacccacattctgaggccctggtccacatgtagaccctgagctgtagaccccactctcccagcgggtaggggat

                     .                    .                    .                    .                    .              1080
gcttccagccggatatccatctctccaaatgaggaccagtaactgagaagtatctgaggagaagcaatgccaaagtgacatgggtccttggtgatgagggagcacagagccacttgcaga

                     .                    .                    .                    .                    .              1200
gaggattgcctaggaggggggaaggggaagaatccagggttgtcatcaccactgagtatggatttcacattctaacacattagAAGCTGCAGGATGCTGAAATTGCAAGGCTGATGGAAGA

                     .                    .                    .                    .                    .              1320
CTTGGACCGGAACAAGGACCAGGAGGTGAACTTCCAGGAGTATGTCACCTTCCTGGGGGCCTTGGCTTTGATCTACAATGAAGCCCTCAAGGGCTGAAAATAAATAGGGAAGATGGAGAC

                     .                    .                    .                    .                    .              1440
ACCCTCTGGGGGTCCTCTCTGAGTCAAATCCAGTGGTGGGTAATTGTACAATAAATTTTTTTTGGTCAAATTTACCCTTGCGTCTTGGCTTCCGAATGATTTCTGTTCCTCCTTGGCTTA

                     .                    .                    .                    .                    .              1560
GTGGGACACCAGCCATTGGAAGATTTGCTCACGGTCAACCTCTGAAAATGACTCATTGACTCGCCAGGCCAGAGGACCCACCCTGACAAGGCTGCCTCTAGCGCGTAAGGTGCCTTTATG

                     .                    .                    .                    .                    .              1680
TGAATGAGGAGAGATGCCCCTCTTGGCAACGCCATCCTAAGGAAAGGCTCAAGTGGTTTCCAGTAGAGAGAGTCCTGGGATGAGCTTGGAGATGGAAATGGTCCTTTGGGCCGGGATGTG

                     .                    .                    .                    .                    .              1800
ATGGGGTTTGGGGGCCTGGAAGTGAGGCAGAGATAGTTCCAGAGGCTCCCAGATGTGTTTTGCTCTGGGTGTGGCAAGAGGGGCCTTGGGGTGGGGCAAGTCCCTTTCTCATCACAGCGC

                     .                    .                    .                    .                    .              1920
AGGGGTTAGATAGGGCACATCTGAGATGCCTGAGGCTTGGCTCAGGGAGTTTCCTACACCAGTGAGGACGCTGTGTGACTGAGTCTACTGCGGCTGCCCAGGTCCCAGGTGGAGTGGGGG

                     .                    .                    .                    .                    .              2040
AGGCACACTCTTGGAGTGTGTCCCGTCATTCAGGGTGAGGGCTTTTTGTTGGAACGGTGGTCTGAGGAGCTGGCAGCTGCACCAACACGTGAACCACGGGGTGTTCAGTAATGGGGCGGG

                     .                    .                    .                    .                    .              2160
GTATCCCTGCAGCCTCAGCGTAATGACTCACCCGGCACTTCCACGGGATCCAGCCTGGATCTCAGCCCCCATCAGAGAAGATGACTAATTGAATCATTGTCCATCATCTGGATTAGTGTT

                     .                    .                    .                    .                    .              2280
TTAAGGCAGAAGGGAAGAGGATAAGGAGGGTAAACGCTGTTTCCGGGTGATGCCACATCATTAAGCCTCTCTAGGCCTAGTCCGAGCTGGGCAAGTTTACCTCTAGCTTCTGGGGAAGAG

ATCTTGACTTTAGATGGAGA
```

FIG. 4. **Sequence of the human calcyclin gene.** The nucleotides are numbered from the cap site = +1. The GC boxes, the TATA box, and the sequences with strong homologies to the SV40-core enhancers are *underlined* (*thin line*). Introns are in *lower case letters*. Polyadenylation signal and the initiation and stop codons are *underlined*. The restriction enzymes of the promoter region relevant to subsequent experiments are indicated.

was determined by S1 nuclease assay (not shown) and is numbered +1. The termination codon TGA is located at nucleotide 1295 and the polyadenylation signal, AAUAAA, is *underlined* at nucleotide residue 1370, about 20 base pairs upstream of the termination of the cDNA. The consensus nucleotides for introns, GT at the beginning of an intron and AG at the end of the intron, are present. The putative amino acid sequence of the calcyclin gene has already been reported in a previous paper (3). The first ATG codon in the cDNA that is followed by an open reading frame is located at nucleotide 651 of the gene, so that the first exon of the calcyclin gene seems to be untranslated. As already pointed out in a previous paper (3), this ATG codon is considered to be the correct one for three reasons: 1) there are no other ATG codons prior to it, in any frame of the cDNA; 2) it is surrounded by canonical nucleotides (17); and 3) it gives a reading frame with maximum homologies to the subunits of the S-100 proteins.

The first putative $Ca^{2+}$-binding domain (nucleotides 710-751) is in the second exon, while the second binding site (nucleotides 1202-1240) is in the third exon, so that the introns separate the two domains.

*Structural Analysis of the Flanking Regions of the Human Calcyclin Gene*—The salient structural features of the flanking regions of the calcyclin gene can be summarized as follows. 1) There is a conventional TATAA box at nucleotide residue −29, as already mentioned, upstream of the initiation site. 2) There is no clear-cut CAAT box. There are at least three GC boxes around nucleotide residues extending from −89 to −67, although two of them are overlapping. An additional GGCGGG sequence occurs at residues 546 in the first intron.

## TABLE I

*Sequence similarities between the promoter region of the calcyclin gene and other promoter elements*

The numbers refer to the nucleotide residues of the calcyclin gene as indicated in Fig. 4. Numbers in parenthesis indicate the number of nucleotides identical to the consensus sequences. Asterisks indicate dissimilarities.

| | |
|---|---|
| GGCGGG | N-ras gene (6/6) |
| GGCGGG | −874 to −879 |
| | −83 to −88 |
| | −73 to −78 |
| | −69 to −74 |
| TGGAATATTC‾CGG | Heat-shock element |
| TGGAA-AGTCGAGG | −140 to −152 (10/14) |
| GGTGTGGAAAGTCC | SV40 enhancer |
| GGCGTGGAAAGTCG | −143 to −156 (12/14) |
| GGTGTGGA-CGTC | −935 to −946 (11/13) |
| GGTGTGGCAAG | 1748 to 1758 (10/11) |
| GGTGGGGCAAGTC | 1770 to 1782 (11/13) |

These GGCGGG boxes or their inverted complement CCGCCC are present in several other promoter regions notably those of SV40 (18), the Harvey ras gene (19-22), and others (see "Discussion"). The GC boxes are thought to be the binding sites for a transcriptional factor called Sp1 (23, 24). The whole region from −252 to the TATA box is GC rich. 3) There is an enhancer-like structure at nucleotide residue −144. This enhancer has a strong homology to the C2 component of the SV40 enhancer (25), and to the consensus sequence for other viral and cellular enhancers (26, 27). In addition to this enhancer-like structure which is strongly homologous to the core sequence of other enhancers reported in the literature, there are other enhancer-like structures in the 5′ and 3′ flanking regions of the calcyclin gene. Sequence similarities between the SV40 enhancer sequences and some of the enhancer-like sequences in the flanking regions of the calcyclin genes are summarized in Table I. 4) Upstream of the GC boxes there is also a pyrimidine-purine alternation TGTGCG, which has also been described in the SV40 enhancer (25). 5) There is also a purine-rich region which is located roughly around nucleotide −328. Purine-rich regions have been proposed to play a role in the response of a gene to serum regulation (28). However, the purine-rich sequence in the 5′ flanking region of the calcyclin gene is imperfect and is interrupted by a C. 6) Repeat sequences have also been investigated. There are several, both inverted and direct, but, at the present moment, we cannot attribute to them any significance. No Alu sequences have been found in the entire sequenced fragment. 7) Downstream from the polyadenylation signal there is also a TCAAA sequence that approximates the TTCAAA sequence that is often found downstream of the 3′ end of eukaryotic genes between 30 and 60 base pairs from the poly(A) signal (29).

*Functional Analysis of the Calcyclin Promoter*—To delineate which of the sequences described above may be important in the regulation of the calcyclin gene, we proceeded to a functional analysis of the 5′ flanking sequence, using hybrid genes in which various portions of the 5′ flanking region of the calcyclin gene were fused to the coding sequence of the chloramphenicol acetyltransferase (CAT) gene (11). For this purpose we cloned upstream of the CAT coding sequence the entire calcyclin sequence extending from the *Hinc*II/*Sca*I at residue −1371 to nucleotide residue 134 (*Bam*HI), which is in the first intron of the calcyclin gene. Progressive deletions were then carried out using unique restriction sites of the 5′ flanking region. These deletions are summarized in Fig. 5 which gives the remaining 5′ flanking sequences upstream of the CAT gene. The chimeric genes were transfected into exponentially growing tk-ts13 cells and the results of these experiments are summarized also in Fig. 5. The actual CAT assays are presented in Fig. 6. There are modest differences in the efficiency of CAT gene transcription, whether the full 5′ flanking sequence of the calcyclin gene is used or the
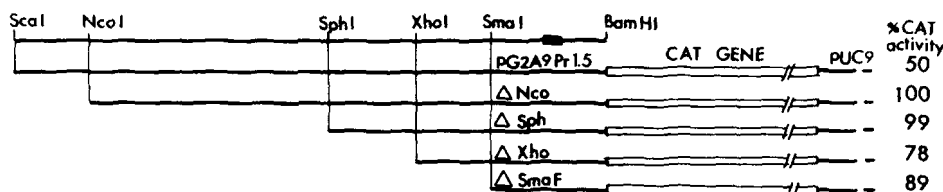


FIG. 5. **Deletion analysis of the promoter region of the calcyclin gene.** The various fragments generated by the indicated restriction enzymes were cloned in the pUC9 polylinker as described under "Materials and Methods." The CAT gene was then cloned in the *Bam*HI site of the polylinker. tk-ts13 cells were transfected with the chimeric genes and the CAT activity was determined by standard methods (11). The results, normalized to the *Nco*I plasmid, are shown next to each plasmid.
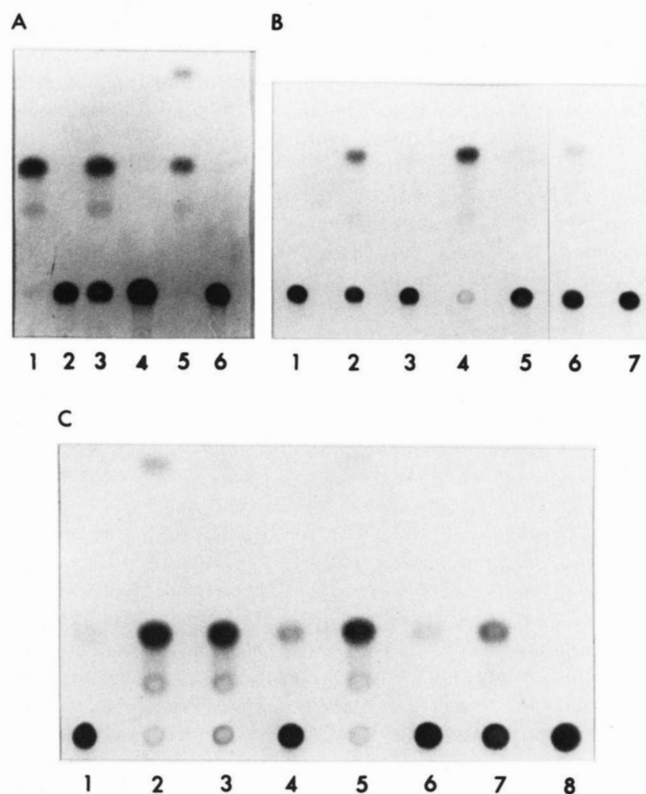
FIG. 6. **Promoter activity of various chimeric plasmids carrying deletions in the 5′ flanking region of the calcyclin gene.** CAT activity was determined by standard methods (11). The deletions are explained in the legend to Fig. 5. *Panel A: lane 1*, positive control; *2*, negative control; *3*, pG2A9 Pr 1.5 with CAT gene in correct orientation; *4*, same as *lane 3*, but with the CAT gene in the opposite orientation; *5*, pSVU2CAT; *6*, mock. *Panel B: lane 1*, mock; *lane 2*, Δ*Xho*I with the CAT gene in the correct orientation: *lane 3*, same as *lane 2*, with the CAT gene in the opposite orientation: *lanes 4* and *5*, Δ*Sph*I with the CAT gene in correct and opposite orientations, respectively; *lane 6*, positive control; *lane 7*, negative control. *Panel C: lane 1*, pUC-CAT; *lane 2*, pSV2CAT; *lanes 3* and *4*, Δ*Sma*F with CAT gene in correct and opposite orientations, respectively; *lanes 5* and *6*, *Nco*I, with the CAT gene in the correct and opposite orientations, respectively; *lane 7*, positive control; *lane 8*, negative control.

smaller fragments down to the small fragment which includes only the 167 base pairs just above the cap site. This small fragment, though, includes the TATAA box, the GC boxes, and the enhancer-like structure. Constructs in which the direction of the CAT gene was reversed were inactive (Fig. 6). The entire 5′ flanking sequence used in these experiments consistently gave lower values than the various deletions. This was confirmed in another experiment in which the calcyclin promoter and its deletion mutants were used to drive another linked reporter, the human thymidine kinase cDNA (12). After transfection of exponentially growing tk-ts13 cells (13) with the chimeric plasmids, colonies were selected in hypoxanthine/aminopterin/thymidine medium, the plates were fixed after 14 days, and colonies were counted. The results are summarized in Table II. Again, the promoter with a deletion upstream of the *Nco*I site (nucleotide −1194), was more active than the entire 5′ flanking region used in these experiments or than some of the smaller fragments. No colonies were obtained when the TK cDNA was cloned in the opposite direction to the promoter or its partial deletions (summarized in one row in Table II) or when it had no eukaryotic promoter at all.

The *Sma*V vector was also slightly active in the CAT assay. The *Sma*V vector includes the whole 5′ flanking region except

TABLE II

*Activity of deletion mutants of the promoter region of the calcyclin gene*

The numbers in the first column refer to the nucleotide residues in Fig. 4. The enzyme in the second column refers to the deletions described in Figs. 5 and 6. The last column gives the number of colonies 14 days after tk-ts13 cells were transfected with the appropriate constructs (all carrying the human thymidine kinase cDNA) and selected in hypoxanthine/aminopterin/thymidine medium.

| Nucleotide residues | Restriction site used for deletion | Number of colonies in HAT[a] |
|---|---|---|
| −1370 to +134 | None | 220 |
| −1194 to +134 | *Nco*I | 360 |
| −588 to +134 | *Sph*I | 334 |
| −361 to +134 | *Xho*I | 264 |
| −167 to +134 | *Sma*I | 266 |
| Various constructs | (Reversed cDNA) | 0 |
| pTK11 (TK cDNA driven by SV40 promoter) | | 1444 |
| No promoter | | 0 |

[a] HAT, hypoxanthine/aminopterin/thymidine medium.
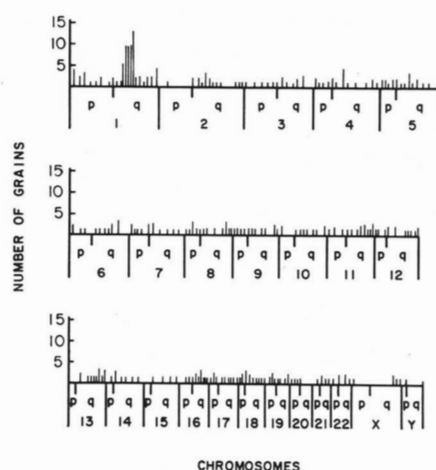


FIG. 7. **Grain distribution from two *in situ* hybridization experiments.** The histogram shows the grain distribution in the combined 185 metaphase spreads using the 2A9 cDNA clone as probe. *Abscissa* represents the chromosomes in their relative size proportions, and the *ordinate* shows the distribution and number of silver grains.

for the last 169 base pairs proximate to the cap site (called by us the *Sma* fragment, see above). Yet this 5′ flanking region, devoid of TATAA box, GC boxes, and enhancer-like structure still gave positive results in the CAT test. On the other hand, the activity of this promoter devoid of the *Sma* fragment, was less when tested in the TK assay in which the coding region of the human TK gene is used as a linked reporter for the various deletions of the calcyclin promoter (not shown).

*Localization of the Calcyclin Gene to Human Chromosome 1*—Two independent *in situ* hybridization experiments were performed using normal metaphase chromosomes and the calcyclin (2A9) probe. Three-hundred forty-eight chromosomally located grains on 185 metaphases were analyzed. The predominant site of hybridization in each experiment was the long arm of chromosome 1, with a total of 46 grains (13%) at bands 1q21 → q25. The number of grains in this region was at least four times the number found on any other chromosomal segment of similar length (Fig. 7). These results indicate regional mapping of the calcyclin (2A9) gene to the q21 → q25 region of chromosome 1.

### DISCUSSION

We present here the full sequence of the human calcyclin gene (defined in Ref. 3), including its proximate 5′ and 3′

flanking sequences. We have also identified a region in the 5′ flanking sequence that has the structural elements and displays the function of a promoter, and have localized the gene to human chromosome 1.

There are two major reasons why we believe the calcyclin gene merits a careful study: it is growth-regulated and its coding sequence suggests a calcium-binding, calcium-modulated protein.

The calcyclin gene was originally identified as the 2A9 cDNA clone, whose cognate RNA level was increased by serum (1). Since then, its growth regulation has been clearly established (2–4), as already mentioned in the Introduction. Especially important are the facts that the expression of calcyclin RNA is deregulated in human acute myeloid leukemias (6), and the observation that its putative amino acid sequence has sequence similarities with the p10 (30) or p11 (31) subunit which is part of a large complex, constituting the major cellular substrate for tyrosine kinase (32).

Its role as a calcium-binding protein has yet to be demonstrated. However, in a previous paper (3) we have reported the sequence of a full-length cDNA clone, corresponding to this gene, to which we gave the name of calcyclin. There is a 55% sequence similarity between the coding sequence of the calcyclin cDNA and the coding sequences of the $\alpha$ and $\beta$ subunits of the bovine S-100 protein (33–35). This is a calcium-binding protein (for a review, see Ref. 36), that has often been used as a marker of human malignancy, especially malignant melanoma and tumors of neuroendocrine origin (37, 38), and also has considerable homology with a p10 (30) or p11 (31) subunit as mentioned above. The sequence similarities of the putative amino acid sequence of calcyclin are particularly striking in those regions that in the S-100 protein have been identified as the calcium-binding sites. Here the homology between the S-100 $\alpha$ or $\beta$ subunit and calcyclin reaches a level of 91% (3).

Because of the role that calcium and tyrosine kinase are thought to play in the regulation of cell proliferation (39–41), the fact that calcyclin seems to be coding for a calcium-binding protein with sequence similarities to the S-100 protein and to a component of a tyrosine kinase cellular substrate suggests a role of this gene in the control of cell proliferation. This suggestion is supported by the finding that its RNA level is growth regulated and is deregulated in human acute myeloid leukemias (3, 5, 6). Finally, the calcyclin gene is not induced by Adenovirus infection, even under conditions when cellular DNA synthesis is stimulated (42) and its mRNA is not detectable in human T lymphocytes (43).

We have now isolated a genomic clone that includes the entire calcyclin gene and some of the flanking sequences. The calcyclin gene is a compact gene covering only 1392 nucleotides that comprise three exons and two introns. The first exon is short and does not contain an initiation AUG codon and is presumably untranslated. The putative $Ca^{2+}$-binding domains are separated by introns, in accordance with Gilbert's rule (44) that exons represent functional domains. Our sequence includes 1370 bp upstream of the cap site and 1200 bp of 3′ flanking sequence. In this paper, we have concentrated our attention on the 5′ flanking sequence. Upstream of the cap site are 3 GC boxes extending from −71 to −90. These sequences have been shown to be binding sites for the transcription factor Sp1 (23, 24). The GC boxes (GGCGGG or CCGCCC) are found in several other promoters. In some of these, the GC boxes are accompanied by the absence of a TATA box, notably in the Ha-ras-1 gene (21, 22), the N-ras gene (45), the ribosomal proteins' promoters (46), and the human epidermal growth factor receptor gene (22) which also

lack a CAAT box. While the promoter region of the calcyclin gene lacks a CAAT box, it has a classic TATAA box, in the right position upstream of the cap site, a feature it shares with other GC-rich promoters like those of the SV40 (18), human c-sis (47), the human thymidine kinase gene (48), the human dihydrofolate reductase gene (49), and c-myc (22). The size of exon 1, 44 bp, is very similar to the size of exon 1 in the human Ha-ras-1 gene, 40 bp (21), and both of them are untranslated. Among the various $Ca^{2+}$-binding, $Ca^{2+}$-modulated proteins (36), only one genomic sequence has been published, that of the chicken calmodulin gene (50). According to the revised sequence,[2] the 5′ flanking region of the chicken calmodulin gene has a TATAA box (TTTATGA) about 25 bp upstream of the cap site, and, 5′ to it, a region with a very high GC content. The chicken calmodulin gene, however, has six exons. A pseudogene of calmodulin has been described in the rat genome (51). We have no detectable evidence of pseudogenes of calcyclin in the human genome.

There are at least four sequences in the flanking regions of the calcyclin gene that bear resemblance to the consensus core sequence of enhancers, especially viral enhancers (25–27, 52, 53). The sequence similarity is especially striking if one takes the C2 component of the SV40 enhancer (25) GGTGTGGAAAGTC, where 12 out of 14 nucleotides are identical to a sequence just upstream of the GC-rich region of the calcyclin promoter. Other similar sequences can be found, two of them in the 3′ flanking region. Their function and significance have not yet been determined. Overlapping with the enhancer at −140 to −152 is a sequence that has similarity (10 out of 14 nucleotides) to a consensus sequence for the heat-shock element (54). Another heat-shock like element is located about 200 bp upstream of the first one, where 9 out of 12 nucleotides are identical to the dyad sequence CTGGAAT/ATTCCCG described in the human HSP70 promoter (28). The 5′ flanking region of the calcyclin gene has promoter function, *i.e.* it can drive linked reporters in a transient expression assay after transfection of the appropriate chimeric plasmids. The 178 bp upstream of the NcoI restriction site in the promoter region seem to decrease the activity of the calcyclin promoter. Although the decrease is modest (50%), it has been repeatedly and consistently observed, both with the CAT gene and the TK cDNA as the linked reporters. We cannot say at this point that this fragment includes negative regulatory elements, but this possibility has to be investigated, especially since the expression of calcyclin mRNA depends on the cell type (43). At any rate, the analysis of deletion mutations given in this paper can serve as a basis to identify the specific sequences that make this gene growth-regulated (28). In the present experiments, promoter activity was determined in exponentially growing cells. Future experiments will have to address the question of which sequences specifically activate this gene in growing cells or repress it in quiescent cells. Especially important, in this respect, is the strong activity displayed in growing cells by the *Sma* fragment which includes only 169 bp upstream of the cap site.

The calcyclin gene has been localized to the long arm of human chromosome 1, near the ski oncogene.

In summary, we have presented a structural and functional analysis of a human gene, which we refer to as the calcyclin gene, that, because of its inducibility by growth factors, its overexpression in leukemias and its sequence, can be considered as a promising candidate among those genes that may play a role in cell cycle progression.

---

# REFERENCES

1. Hirschhorn, R. R., Aller, P., Yuan, Z., Gibson, C. W., and Baserga, R. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81,** 6004–6008
2. Rittling, S. R., Brooks, K. M., Cristofalo, V. J., and Baserga, R. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83,** 3316–3320
3. Calabretta, B., Battini, R., Kaczmarek, L., de Riel, J. K., and Baserga, R. (1986) *J. Biol. Chem.* **261,** 12628–12632
4. Rittling, S. R., Gibson, C. W., Ferrari, S., and Baserga, R. (1985) *Biochem. Biophys. Res. Commun.* **132,** 327–335
5. Calabretta, B., Kaczmarek, L., Mars, W., Ochoa, D., Gibson, C. W., Hirschhorn, R. R., and Baserga, R. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 4463–4467
6. Calabretta, B., Venturelli, D., Kaczmarek, L., Narni, F., Talpaz, M., Anderson, B., Beran, M., and Baserga, R. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83,** 1495–1498
7. Maniatis, T., Fritch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
8. Gross-Bellard, M., Oudet, P., and Chambon, P. (1973) *Eur. J. Biochem.* **36,** 32–38
9. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517
10. Maxam, A. M., and Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560
11. Gorman, C. M., Moffat, L. F., and Howard, B. H. (1982) *Mol. Cell. Biol.* **2,** 1044–1051
12. Bradshaw, H. D., Jr., and Deininger, P. L. (1984) *Mol. Cell. Biol.* **4,** 2316–2320
13. Shen, Y.-M., Hirschhorn, R. R., Mercer, W. E., Surmacz, E., Tsutsui, Y., Soprano, K. J., and Baserga, R. (1982) *Mol. Cell. Biol.* **2,** 1145–1154
14. Emanuel, B. S., Sellinger, B. T., Gudast, J., and Myes, J. C. (1986) *Am. J. Hum. Genet.* **38,** 38–44
15. Cannizzaro, L. C., and Emanuel, B. S. (1984) *Cytogenet. Cell Genet.* **38,** 308–309
16. Okayama, H., and Berg, P. (1983) *Mol. Cell. Biol.* **3,** 280–289
17. Kozak, M. (1986) *Cell* **44,** 283–292
18. Fromm, M., and Berg, P. J. (1982) *J. Mol. Appl. Genet.* **1,** 457–481
19. Capon, D. J., Chen, E. Y., Levinson, A. D., Seeburg, P. H., and Goeddel, D. V. (1983) *Nature* **302,** 33–37
20. Reddy, E. P. (1983) *Science* **220,** 1061–1063
21. Ishii, S., Merlino, G. T., and Pastan, I. (1985) *Science* **230,** 1378–1381
22. Ishii, S., Xu, Y. H., Stratton, R. R., Roe, B. A., Merlino, G. T., and Pastan, I. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 4920–4924
23. Dynan, W. S., Saffer, J. D., Lee, W. S., and Tjian, R. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 4915–4919
24. McKnight, S., and Tjian, R. (1986) *Cell* **46,** 795–805
25. Firak, T. A., and Subramanian, K. N. (1986) *Mol. Cell. Biol.* **6,** 3667–3676
26. Lang, J. S., and Spandidos, D. A. (1986) *Anti Cancer Res.* **6,** 437–450
27. Dorsch-Hasler, K., Keil, G. M., Weber, F., Jasin, M., Schaffner, W., and Koszinowski, U. H. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 8325–8329
28. Wu, B., Kingston, R. E., and Morimoto, R. I. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83,** 629–633
29. Urano, Y., Watanabe, K., Sakai, M., and Tamaoki, T. (1986) *J. Biol. Chem.* **261,** 3244–3251
30. Glenney, J. R., Jr., and Tack, B. F. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 7884–7888
31. Gerke, V., and Weber, K. (1985) *EMBO J.* **4,** 2917–2920
32. De, B. K., Misono, K. S., Lukas, T. J., Mroczkowski, B., and Cohen, S. (1986) *J. Biol. Chem.* **261,** 13784–13792
33. Isobe, T., and Okuyama, T. (1981) *Eur. J. Biochem.* **116,** 79–86
34. Isobe, T., Ishioka, N., and Okuyama, T. (1981) *Eur. J. Biochem.* **115,** 469–474
35. Kuwano, R., Usui, H., Maeda, T., Fukui, T., Yamanari, N., Ohtsuka, E., Ikehara, M., and Takahashi, Y. (1984) *Nucleic Acids Res.* **12,** 7455–7465
36. Van Eldik, L. J., Zendegui, J. G., Marshak, D. R., and Watterson, D. M. (1982) *Int. Rev. Cytol.* **77,** 1–61
37. Stefansson, K., Wollmann, R., and Jerkovic, M. (1982) *Am. J. Pathol.* **106,** 261–268
38. Kahn, H. J., Marks, A., Thom, H., and Baumal, R. (1983) *Am. J. Clin. Pathol.* **79,** 341–347
39. Hazelton, B., Mitchel, B., and Tupper, J. (1979) *J. Cell Biol.* **83,** 487–498
40. Carpenter, G., Stoschek, C. M., and Soderquist, A. M. (1982) *Ann. N. Y. Acad. Sci.* **397,** 11–17
41. Cooper, J. A., Bowen-Pope, D. F., Raines, E., Ross, R., and Hunter, T. (1982) *Cell* **31,** 263–273
42. Liu, H. T., Baserga, R., and Mercer, W. E. (1985) *Mol. Cell. Biol.* **5,** 2936–2942
43. Kaczmarek, L., Calabretta, B., and Baserga, R. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 5375–5379
44. Gilbert, W. (1985) *Science* **228,** 823–824
45. Hall, A., and Brown, R. (1985) *Nucleic Acids Res.* **14,** 5255–5268
46. Dudov, K. P., and Perry, R. P. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83,** 8545–8549
47. Van den Ouweland, A. M. W., Roebroek, A. J. M., Schalken, J. A., Claesen, C. A. A., Bloemers, H., and Van de Ven, W. J. M. (1986) *Nucleic Acids Res.* **14,** 765–778
48. Kreidberg, J. A., and Kelly, T. J. (1986) *Mol. Cell. Biol.* **6,** 2903–2909
49. Chen, M.-J., Shimada, T., Moulton, A. D., Cline, A., Humphries, R. K., Maizel, J., and Nienhuis, A. W. (1984) *J. Biol. Chem.* **259,** 3933–3943
50. Simmen, R. C. M., Tanaka, T., Ts'ui, K. F., Putkey, J. A., Scott, M. J., Lai, E. C., and Means, A. R. (1985) *J. Biol. Chem.* **260,** 907–912
51. Nojima, H., and Sokabe, H. (1986) *J. Mol. Biol.* **190,** 391–400
52. Weiher, H., Konig, M., and Gruss, P. (1983) *Science* **219,** 626–631
53. Banerji, J., Rusconi, S., and Schaffner, W. (1981) *Cell* **27,** 299–308
54. Pelham, H. R. B. (1982) *Cell* **30,** 517–528