
Genetic organization of the *c-sis* transcription unit

Ans M.W. Van den Ouweland, Jan J.M. Van Groningen, Jack A. Schalken¹, Han W. Van Neck, Henri P.J. Bloemers and Wim J.M. Van de Ven

Molecular Oncology Section, Department of Biochemistry, University of Nijmegen, Geert Grooteplein Noord 21, 6525 EZ Nijmegen, The Netherlands and ¹Department of Urology, University of Nijmegen, The Netherlands

Received November 11, 1986; Accepted January 8, 1987

ABSTRACT

The structure and genetic organization of the transcription unit of the *c-sis* proto-oncogene was determined. Comparative nucleotide sequence analysis of the exon sequences of feline and human *c-sis* revealed a very high degree of homology. The cap site as well as the poly(A)-addition site of the *sis* transcript of each species was identified and found in similar positions. An insert of 4 amino acids was found in the deduced translational product of feline *c-sis* and it was located at the amino-terminus of the region that constituted the platelet-derived growth factor domain. An insert of 149 bp present at the 5' end of exon 7 of human *c-sis* but absent in the simian sarcoma virus *v-sis* oncogene was also present in the feline *c-sis* proto-oncogene.

INTRODUCTION

Upon the demonstration that the function and the amino acid sequence of one chain of the human platelet-derived growth factor (PDGF) was nearly identical to the oncogene product of the simian sarcoma virus (SSV) (1,2,3), studies on the cellular *c-sis* counterpart got great impetus. Sequences of the *c-sis* proto-oncogene appeared also to be acquired by another acutely transforming retrovirus, namely the Parodi-Irgens strain of feline sarcoma virus (PI-FeSV) (4). The *c-sis*-derived sequences in PI-FeSV and SSV are expressed in different contexts of the retroviral genomes. PI-FeSV encodes a 76 kda *gag-sis* polyprotein (5) and SSV a 28 kda protein that contains a short NH₂-terminal leader sequence derived from the viral *env* gene (6,7). In both cases, constitutive expression of these PDGF-like mitogens seems capable of inducing neoplastic transformation (4,8).

In previous studies, the structure and genetic organization of the human (9,10,11) and feline (12,13) *c-sis* proto-oncogenes were described. Using nucleotide sequence analysis of genomic and cDNA clones, the distribution of the 7 human *c-sis* exons was obtained (14,15,16). In the region upstream of the first human *sis* exon, a "TATA" box and consensus sequences for a potential Sp1 binding site were found. At similar positions in the feline *c-sis*

proto-oncogene, the same sequences were present. The upstream region exhibited promoter-like activity when tested in an assay in which coding sequences for bacterial chloramphenicol-acetyl-transferase were placed under its control (13). Therefore, it possibly represents the promoter region of the *c-sis* locus and controls the synthesis of the *c-sis* mRNA (13). Further evidence for this possibility could be obtained by precisely defining the *sis* transcript. In the present report, we provide such evidence. We present data from nuclease S1 protection experiments and comparative nucleotide sequence analysis of *sis* cDNA clones and define the 5' and 3' boundaries of the feline and human *sis* transcripts.

MATERIALS AND METHODS

Biological materials: Cells were grown in Dulbecco's modification of Eagle's medium (DMEM) supplemented with 10% calf serum (GIBCO). Cell lines used in this study included HeLa (American Type Culture Collection (ATCC) CCL 2), ML-1 (17) and PC3 (ATCC CRL 1435). Feline placentas were obtained from the Central Animal Laboratory of the University of Nijmegen, Nijmegen.

Recombinant DNA clones: Isolation of human and feline *c-sis* cosmid clones and pA073, a human *c-sis* specific subclone, was described previously (12). pHN1 is a recombinant clone of λ gt11 and a 0.6 kbp *sis*-specific cDNA isolated from an oligo(dT)-primed human cDNA library. A44, a subclone of pA0121 (12), consists of a 243 bp AluI/AluI DNA restriction fragment subcloned in M13mp11. A53 is a recombinant of M13mp11 containing a 235 bp AluI/AluI DNA restriction fragment isolated from pA0144 (13). A 1.3 kbp BamHI/HindIII DNA restriction fragment, which represents a 3'-specific portion of *c-sis*, was isolated from MB70 (12) and subcloned in pUC18 resulting in pPHS1. pAct is an actin-specific clone described by Dodemont *et al.* (18). *E.coli* strain HB101 was used for the propagation of plasmids and the recombinant M13 bacteriophages were propagated in *E.coli* strain JM109.

DNA sequence analysis: DNA fragments were ligated into the polylinker region of M13 mp8-11 (19). Sequencing of the DNA fragments was according to the dideoxy method of Sanger *et al.* (20). Gel readings were recorded, edited and compared using the Staden programs (21).

RNA isolation and hybridization: Total cellular RNA was isolated using the lithium-urea procedure described by Auffray and Rougeon (22). Ten μ g of oligo(dT)-cellulose purified mRNA was glyoxalated, fractionated on 1% agarose gels (23) and transferred to Hybond-N (Amersham). Isolation of DNA probes, their nick-translation and hybridization of Southern blots were car-

ried out as described before (24). Hybridization of Northern blots was performed according to the method described by Church and Gilbert (25).

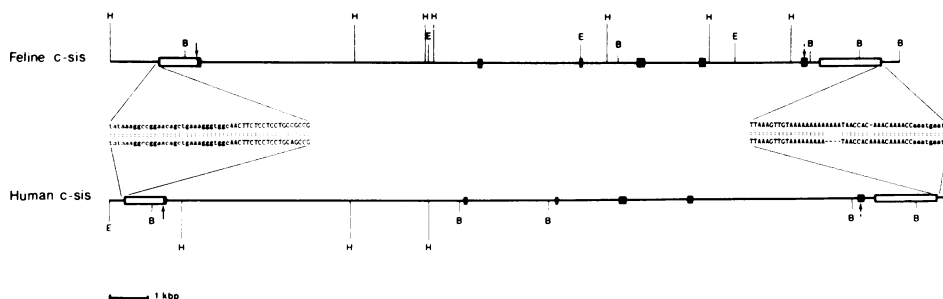
Construction and screening of cDNA libraries: Oligo(dT)-primed cDNA libraries of ML-1 or feline placenta mRNA were constructed in λ gt11 as described by Roebroek *et al.* (26). About 1.2×10^6 plaques obtained upon infection of *E.coli* Y1090 (27), were screened as described by Hanahan and Meselson (28).

Nuclease S1 analysis: Nuclease S1 mapping was carried out as described by Van Leen *et al.* (29). Oligo(dT)-selected mRNA (5 μ g) from feline placenta or human PC3 cells were hybridized with the inserts of A53 and A44, respectively. The inserts of A53 and A44 were labeled using the method of Sanger (20) and purified by electrophoresis in a 6% polyacrylamide/7 M ureum gel. Hybridizations were performed at 55°C. Nuclease S1 (Boehringer) concentrations used in the experiments were 25 units/ μ l and incubations were performed at 37°C.

RESULTS AND DISCUSSION

Genetic organization of feline c-sis.

In an approach to define the c-sis transcription unit, we decided not to limit our studies to the sis transcript of only the human species but to include also the feline sis transcript. We reasoned that by such an extension species-specific characteristics could be eliminated and a more general description of the c-sis transcription unit could be obtained. Detailed genomic information about the c-sis locus, however, was only available for human c-sis. To obtain the necessary genomic data about the feline c-sis locus, the genomic organization of the feline sis transcription unit was determined by comparative analysis of previously described cosmid clones (12,13). Overlapping inserts of these clones contained the complete feline or human c-sis proto-oncogene (12,13). Based upon the known exon distribution in human c-sis (15,16), the location of the feline c-sis exons was determined accurately by Southern blot analysis of the inserts of the feline cosmid clones using human DNA fragments containing c-sis exons as molecular probes (data not shown). In control experiments, DNA fragments identified as putative sis exons were tested for their ability to detect c-sis transcripts in Northern blot analyses (data not shown). Upon identification of DNA fragments that contained the feline c-sis exons, the nucleotide sequence of them was determined. The topological distribution of the feline and human c-sis exons is given in Fig. 1 and reveals great similarity. In Fig. 2,

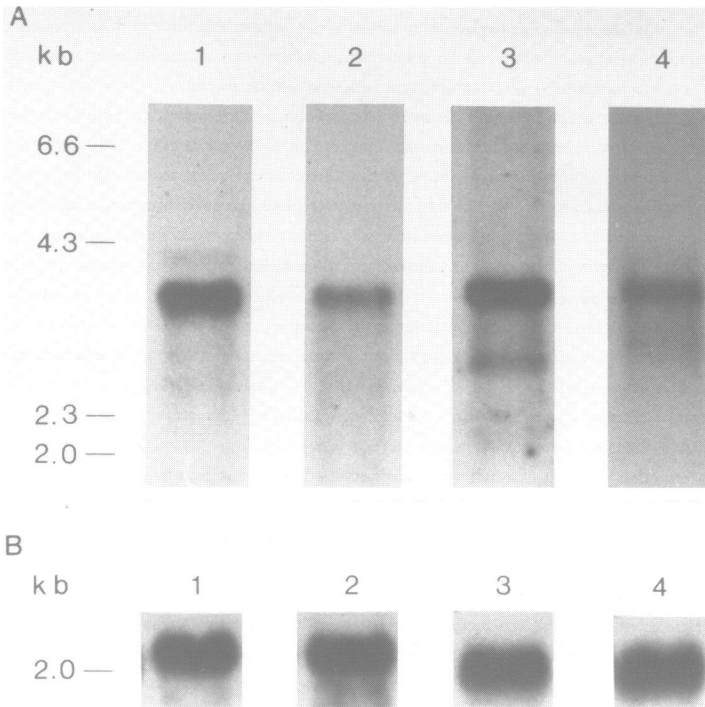


1. Topology of the feline and human *c-sis* exons. The bars in the upper and lower part of the figure represent schematic restriction endonuclease maps of the feline and human *c-sis* proto-oncogene, respectively. Open boxes represent the 5' and 3' noncoding exon sequences, while coding sequences are depicted as solid boxes. The positions of the used ATG (↓) and TGA (↑) codons are also shown. Between the restriction endonuclease maps, the nucleotide sequence of the 5' and 3' ends of the feline and human *sis* transcription units as well as their flanking sequences are given. Exon sequences are depicted as capitals.

nucleotide sequence data of the feline *sis* exons and a small portion of their flanking sequences are shown. It should be noted that results obtained from nucleotide sequence analysis of feline *sis*-specific cDNA clones that are described later in this report are already included in the figure. The position of the "TATA" consensus promoter sequence is also indicated in the figure.

Expression of the feline *c-sis* proto-oncogene.

To precisely define the exons in the *sis* transcription unit, analysis of cDNA clones prepared with *sis* mRNA was required. To select appropriate sources for *sis* mRNA, we analyzed poly(A)-selected RNA from a human prostate carcinoma (PC3) and a myelomonoblastoma (ML-1) cell line and from specimens of feline placenta (Fig. 3A). Selection of feline placenta was based upon studies by Goustin *et al.* (30) who demonstrated the presence of relatively high levels of *c-sis* mRNA in placenta tissue. As a *c-sis*-specific molecular probe, pPHS1 was used. In an attempt to compare similar amounts of poly(A)-selected RNA, RNA concentrations were estimated based upon O.D.₂₆₀ readings and ethidium bromide staining. As an additional control experiment, levels of actin transcripts were estimated by Northern blot analysis (Fig. 3B). Northern blot analysis of mRNA isolated from feline term placentas (Fig. 3A, lane 1) or from placentas containing embryos that were about 3 weeks old (Fig. 3A, lane 2) revealed the presence of a 3.5 kb *sis*-specific transcript. However, a difference in the level of expression of *sis* seemed to exist in



3. Northern blot analysis of RNA transcripts of feline placentas and human cell lines. (A) Poly(A)-selected RNA was isolated from feline term placentas (lane 1), from placentas containing 3 weeks old embryos (lane 2), human PC3 (lane 3) or human ML-1 (lane 4) cells and screened with the inserts of pPHS1 (lanes 1 and 2) and pA073 (lanes 3 and 4) as molecular probes. (B) Screening of the same Northern blot as in (A) with the insert of pAct as molecular probe. Molecular weight markers include λ DNA digested with the restriction endonuclease HindIII.

parison of the size of the human and feline *c-sis* mRNAs revealed no major differences. A weak hybridization signal at about 2.6 kb was detected in RNA preparations from PC3 and ML-1 cells. The nature and origin of this transcript is not known at present but it was not seen in the feline placenta specimens. Slamon and Cline (31) have studied *sis* expression during embryonic development of the mouse. Higher levels of *sis* mRNA were found at early stages of development than at later stages. However, it should be noted that tissue specimens taken during the early stages of embryonic development included embryonic as well as extra-embryonic tissue. This in contrast to specimens from later stages in development in which only embryos were studied. The observed differences in the levels of *sis* transcription is of

interest, especially in light of the report that mRNA levels of the PDGF-receptor are higher in term placenta than in earlier stages of pregnancy (32). However, for a proper evaluation of the observations, identification of the cell types that express the PDGF receptor and the sis product is required.

Isolation of sis cDNA clones.

Using poly(A)-selected mRNA from feline term placenta or human ML-1 cells, a feline and human cDNA library was constructed in λ gt11. Initial screening of the feline cDNA library (about 1×10^6 plaques) was performed with a combination of two different probes. These included a 3'-sis-specific probe (insert of pPHS1) and a 5'-sis-specific probe (a 1.5 kbp BamHI/XhoI insert of pAO144, which contained the 3' end of the first sis exon). Screening resulted in the identification of 16 positive cDNA clones. Of these, 15 hybridized to the 3'-sis-specific probe. Since the internal EcoRI restriction endonuclease recognition sites within the cDNAs were not methylated during the construction of the cDNA libraries, the inserts of all 15 clones appeared to be located downstream of the internal EcoRI site in exon 3 of the feline c-sis proto-oncogene. Only 1 positive clone (FA0184) was isolated whose insert was located upstream of the internal EcoRI site. The nucleotide sequence of the cDNA insert of FA0184 and of 6 cDNAs that were located downstream of the internal EcoRI recognition site was determined. The insert of FA0184 starts at position 835 and stops at the EcoRI recognition site. The inserts of the 6 other cDNAs varied in length between 0.9 and 2.2 kbp. One cDNA contained a poly(A) tail and, based upon this observation, the poly(A)-addition site could be mapped at position 3355 (Fig. 2). The 3' ends of the other five cDNAs stopped at various positions (3301; 3301; 3327; 3332; 3343) (data not shown).

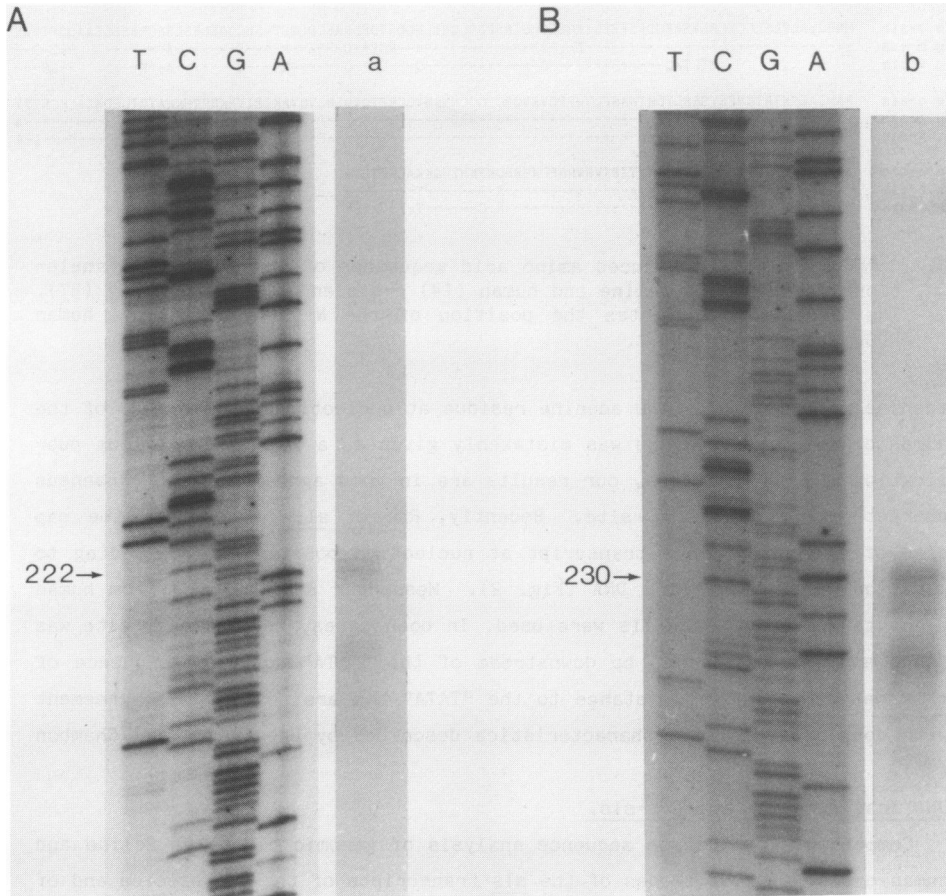
Screening of the human ML-1 cDNA library with the insert of pAO73 as a molecular probe resulted in the isolation of pHN1. To map the poly(A)-addition site in the human sis transcript, the nucleotide sequence of the 0.6 kbp insert of pHN1 was determined (data not shown). This sequence appeared to be identical to nucleotide sequences of cDNA clone pSM-1 described by Ratner et al. (15). The sequence of pSD1, a human sis-specific cDNA clone isolated by Rao et al. (16) lacked sequences at the 3' end. Comparison of the nucleotide sequence data of our human and feline sis cDNA clones indicated that the DNA region in which the poly(A)-addition site was located was well conserved. Only two small deletions in poly(A)-stretches (4 adenine residues in human and 1 in feline; see Fig. 1) were observed. In both

species, the same nucleotide constituted the poly(A)-addition site and the same site was also found by Ratner *et al.* (15). The poly(A)-addition site proposed by Rao *et al.* (16) can be found 30 nucleotides upstream of the site proposed in this paper (Fig. 1 and 2). No authentic polyadenylation signal (AATAAA) was observed in our isolates of the feline and human *c-sis* proto-oncogene. Rao *et al.* (16) have described an AATAAA sequence 18 bp downstream of their poly(A)-addition site. In our feline and human cDNA clones, a "C" instead of a "T" was present in the sequence at that position which does not represent the consensus sequence of the polyadenylation signal (Fig. 1; Fig. 2: nucleotide position 3347 to 3352).

Recently, it was suggested that the pentanucleotide sequence AUUUA in the 3' region of mRNAs is involved in mRNA degradation (33). This sequence was mainly found in mRNAs that encode growth factors such as lymphokines and cytokines (33). The sequence can be present in multiple copies. For instance, human interleukin-3 contains 6 copies (L. Dorssers, pers. comm.). The 3' region of the feline *c-sis* transcript appeared to contain 3 copies of this sequence (Fig. 2: nucleotide positions 2997, 3301 and 3314). The direct involvement of these sequences in *sis* mRNA degradation remains to be established.

Identification of the cap site.

To define precisely the 5' end of the human and feline *sis* transcripts, nuclease S1 protection experiments were performed using mRNA from specimens of feline term placenta or human PC3 mRNA. A 235 bp AluI/AluI DNA fragment (A53; Fig. 2: nucleotide position -13 to 222 of feline *c-sis*) and a 243 bp AluI/AluI DNA fragment (A44; nucleotide position 391 to 633 of human *c-sis* (13)) were used in these studies. Fig. 4 shows the results of the nuclease S1 digestion of hybrids between feline placenta RNA and A53 DNA (purified insert) and between PC3 mRNA and A44 DNA (purified insert). In analysis of the feline *sis* transcript, a hybrid of 222 nucleotides long appeared to represent a major portion of the nuclease S1 resistant molecules (Fig. 4, lane a). Some other bands were visible on the autoradiographs and they represented hybrids of different length. Similar results were obtained in nuclease S1 experiments with the human *sis* transcript (Fig. 4, lane b). The results suggest that the feline and human *sis* transcript have their cap sites in the same position. The presence of some bands of minor intensity on the autoradiographs could be explained as follows. It is possible, that more transcription start sites are present. This is more often observed, as for instance with transcripts of the genes coding for chicken $\alpha 2$ type I col-



4. Nuclease S1 analysis of the feline and human sis transcripts. (A) Uniformly 32 P-labeled insert of A53 was hybridized with feline placenta mRNA. Upon treatment of the hybrid with nuclease S1, the product was analyzed under denaturing conditions (lane a). The four lanes at left show a portion of the nucleotide sequence of A53 to estimate molecular weights. (B) Nuclease S1 resistant hybrid products of human PC3 mRNA and A44. A portion of nucleotide sequence of A44 is shown at the left.

lagen, epidermal growth factor receptor and γ -crystallins (29,34,35). It is also possible that minor fractions of the double stranded hybrids are not properly digested by nuclease S1.

Based upon our data, the cap site of the sis transcript could be placed at the adenine residue at the nucleotide position 31 bp downstream of the "TATA" box. It was reported that most known eukaryotic mRNAs start with an

It should be noted that the 5' end of the 12 bp insertion resembles the consensus sequence of a 3' splice junction. From analysis of our feline sis-specific cDNAs it can be concluded that this splice site is not used during splicing of feline sis precursor RNA. Theoretically, however, it is possible that the cDNAs described in this paper represent only a not fully spliced sis precursor RNA. The fact that no cDNAs without the insertion were isolated argues against the latter possibility.

When compared to the v-sis sequences of SSV, an insertion was found at the 5' end of exon 7 of the feline c-sis proto-oncogene (Fig. 2: nucleotide position 1753 to 1895). This insertion was also present in the human c-sis proto-oncogene. It has been suggested (15) that because of differences in splicing of the human and woolly monkey sis precursor RNA, the sequences described above are deleted from the woolly monkey sis mRNA. Based upon the presence of this region in feline sis-specific cDNA clones, it is more likely that a deletion in the SSV genome occurred during or subsequently to its generation.

From the structure of the human and feline sis transcripts described in this paper, a consensus structure for the sis transcription unit can be deduced. The data strongly support our assignment of the promoter region that controls sis expression.

ACKNOWLEDGEMENTS

We thank Adrie van Bokhoven, Inge van Renesse, Peter Hendriksen and Saskya Ebeling for their excellent technical assistance. This work was supported by the Netherlands Cancer Foundation, Contract No. NUKC Bioch. 85-9.

REFERENCES

1. 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.
2. 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.
3. Robbins, K.C., Antoniades, H.N., Devare, S.G., Hunkapiller, M.W. and Aaronson, S.A. (1983) *Nature* 305, 605-608.
4. Irgens, K., Wyers, M., Moraillon, A., Parodi, A. and Fortuny, V. (1973) *C. R. Acad. Sci. Paris* 26, 1783-1786.
5. Besmer, P., Snyder, H.W., Jr., Murphy, J.E., Hardy, W.D., Jr. and Parodi, A. (1983) *J. Virol.* 46, 606-613.
6. Hannink, M and Donoghue, D.J. (1984) *Science* 226, 1197-1199.
7. King, C.R., Giese, N.A., Robbins, K.C. and Aaronson, S.A. (1985) *Proc. Natl. Acad. Sci. USA* 82, 5295-5299.
8. Kelly, J.D., Raines, E.W., Ross, R. and Murray, M.J. (1985) *EMBO J.* 4, 3399-3405.
9. Chiu, I-M., Reddy, E.P., Givol, D., Robbins, K.C., Tronick, S.R. and Aaronson, S.A. (1984) *Cell* 37, 123-129.
10. Josephs, S.F., Guo, C., Ratner, L. and Wong-Staal, F. (1984) *Science* 223, 487-491.

11. 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.
12. Van den Ouweland, A.M.W., Breuer, M.L., Steenbergh, P.H., Schalken, J.A., Bloemers, H.P.J. and Van de Ven, W.J.M. (1985) *Biochim. Biophys. Acta* 825, 140-147.
13. 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.
14. Collins, T., Ginsburg, D., Boss, J.M., Orkin, S.H. and Pober, J.S. (1985) *Nature* 316, 748-750.
15. Ratner, L., Josephs, S.F., Jarrett, R., Reitz, M.S., Jr. and Wong-Staal, F. (1985) *Nucl. Acids Res.* 13, 5007-5018.
16. Rao, C.D., Igarashi, H., Chiu, I-M., Robbins, K.C. and Aaronson, S.A. (1986) *Proc. Natl. Acad. Sci. USA* 83, 2392-2396.
17. Takeda, K., Minowada, J. and Bloch, A. (1981) *Cancer Res.* 42, 5152-5157.
18. Dodemont, H.J., Soriano, P., Quax, W.J., Ramaekers, F., Lenstra, J.A., Groenen, M.A.M., Bernardi, G. and Bloemendal, H. (1982) *EMBO J.* 1, 167-171.
19. Messing, J. and Vieira, J. (1982) *Gene* 19, 269-276.
20. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
21. Staden, R. (1982) *Nucl. Acids Res.* 10, 2951-2961.
22. Auffray, C. and Rougeon, F. (1980) *Eur. J. Biochem.* 107, 303-314.
23. McMaster, G.K. and Carmichael, G.G. (1977) *Proc. Natl. Acad. Sci. USA* 74, 4835-4838.
24. Schalken, J.A., Van den Ouweland, A.M.W., Bloemers, H.P.J. and Van de Ven, W.J.M. (1985) *Biochim. Biophys. Acta* 824, 104-112.
25. Church, G.M. and Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* 81, 1991-1995.
26. Roebroek, A.J.M., Schalken, J.A., Bussemakers, M.J.G., Van Heerikhuizen, H., Onnekink, C., Debruyne, F.M.J., Bloemers, H.P.J. and Van de Ven, W.J.M. (1986) *Molec. Biol. Rep.* 11, 117-125.
27. Huynh, T.V., Young, R.A. and Davis, R.W. (1985) in *DNA cloning techniques, a practical approach* (Glover, D. ed), IRL Press, Oxford.
28. Hanahan, D. and Meselson, M. (1980) *Gene* 10, 63-67.
29. Van Leen, R.W., Kastrop, P.M.M., Van Roozendaal, K.E.P. and Schoenmakers, J.G.G. (1986) *Eur. J. Biochem.* 157, 203-208.
30. 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.
31. Slamon, D.J. and Cline, M.J. (1984) *Proc. Natl. Acad. Sci. USA* 81, 7141-7145.
32. Yarden, Y., Escobedo, J.A., Kuang, W-J., Yang-Feng, T.L., Daniel, T.O., Tremble, P.M., Chen, E.Y., Ando, M.E., Harkins, R.N., Francke, U., Fried, V.A., Ullrich, A. and Williams, L.T. (1986) *Nature* 323, 226-232.
33. Shaw, G. and Kamen, R. (1986) *Cell* 46, 659-667.
34. Vogeli, G., Ohkubo, H., Sobel, M.E., Yamada, Y., Pastan, I. and De Crombrughe, B. (1981) *Proc. Natl. Acad. Sci. USA* 78, 5334-5338.
35. Ishii, S., Xu, Y-H., Stratton, R.H., Roe, B.A., Merlino, G.T. and Pastan, I. (1985) *Proc. Natl. Acad. Sci. USA* 82, 4920-4924.
36. Breathnach, R and Chambon, P (1981) *Ann. Rev. Biochem.* 50, 349-383.
37. Devare, S.G., Reddy, E.P., Law, J.D., Robbins, K.C. and Aaronson, S.A. (1983) *Proc. Natl. Acad. Sci. USA* 80, 731-735.
38. Betsholtz, C., Johnsson, A., Heldin, C-H., Westermark, B., Lind, P., Urdeda, M.S., Eddy, R., Shows, T.B., Philpott, K., Mellor, A.L., Knott, T.J. and Scott, J. (1986) *Nature* 320, 695-699.