Requirement for a Signal Sequence in Biological Expression of the v-*sis* Oncogene

Abstract. *The protein encoded by the simian sarcoma virus oncogene (v-*sis*) contains a signal sequence, derived from the envelope gene of the parental retrovirus, which is required for transformation. Removal of the proposed signal sequence was correlated with loss of biological activity. This activity was restored to inactive deletion mutants by fusion with the coding region for a heterologous signal sequence. Biological activity of v-*sis *was also abolished by either a small deletion within the coding region of the signal sequence or by a point mutation introduced by site-directed mutagenesis.*

Simian sarcoma virus, SSV, is an acutely transforming retrovirus which carries the oncogene v-*sis* (*1–5*). Recently, the predicted amino acid sequence of the v-*sis* gene product, $p28^{sis}$, was shown to have extensive sequence homology with one polypeptide chain of platelet-derived growth factor, PDGF (*6, 7*). The protein $p28^{sis}$ is immunologically related to PDGF (*8*), and a PDGF-like mitogenic activity was found in lysates of SSV-transformed cells (*9*). The initial v-*sis* translation product is processed at the NH_2- and COOH-termini to generate lower molecular weight proteins. Although these lower molecular weight proteins are recognized by antisera to native PDGF (*8*), it is not known which form of the v-*sis* gene product is responsible for transformation.

DNA sequencing of the SSV genome shows that acquisition of the v-*sis* sequence occurred largely at the expense of the retroviral envelope gene, *env* (*2–5*). Only a small portion of the *env* gene remains upstream of v-*sis*, and shares the same reading frame. There are six ATG codons (A, adenine; T, thymine; G, guanine) in the *env-sis* open reading frame of 813 nucleotides (Fig. 1). The first three ATG codons are from the retained portion of the *env* gene, and the other three lie within v-*sis*.

In order to determine which ATG codon initiates translation in the *env-sis* open reading frame, we constructed a series of deletions in the NH_2-terminal region (Fig. 1). The effect of each deletion on the biological activity of v-*sis* was determined by focus formation assay in NIH 3T3 cells. In order to obtain expression of these deleted genes, we used an expression vector derived from Moloney murine leukemia virus (M-MuLV) (*10*). In this vector, an inserted gene is substituted for the *env* gene of M-MuLV. Transcription of the inserted gene and splicing of its messenger RNA (mRNA) should be identical to that of the M-MuLV *env* gene (*11*). When assayed in this vector, the inserted gene must provide its own ATG codon for initiation of translation.

The deleted v-*sis* genes were inserted into the M-MuLV expression vector and introduced into NIH 3T3 cells with the calcium phosphate coprecipitation technique (*12*). The results of standard focus assays are shown in Fig. 1. A deletion in which only the first ATG codon was removed (pDD147) was as efficient in transformation as the parental clone pDD120. However deletion of both the first and second ATG codons (pDD143 and pDD149) resulted in complete loss of activity. Similarly, more extensive deletions, which included the third, fourth, and sixth ATG codons, also resulted in loss of biological activity. Many of these deletions leave most of the PDGF-related region of v-*sis* intact. Therefore, the presence of the PDGF-related region is not sufficient to account for the biological activity of the v-*sis* gene product.

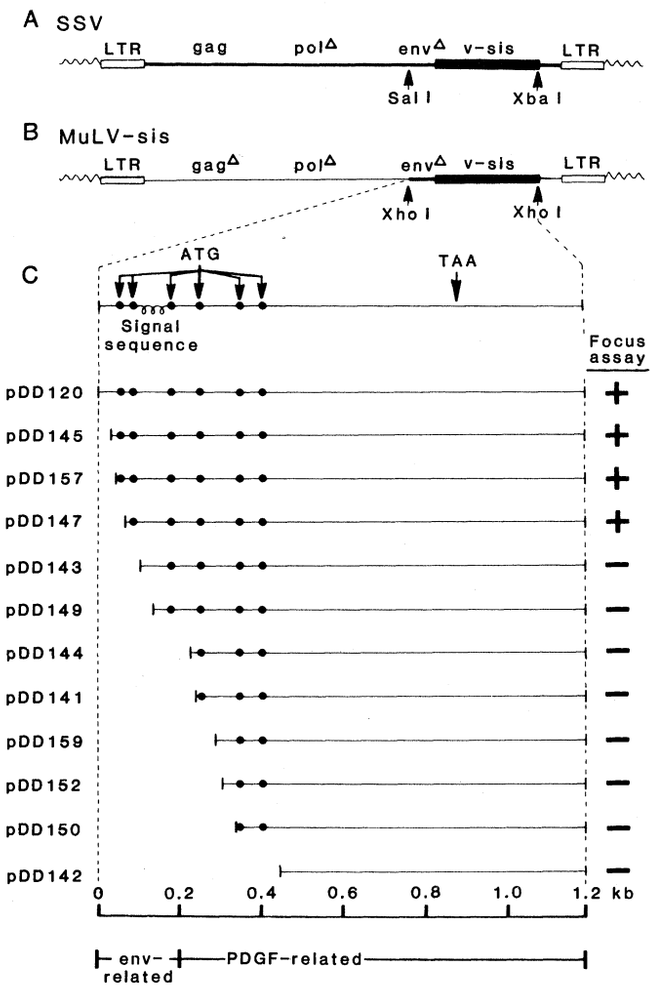


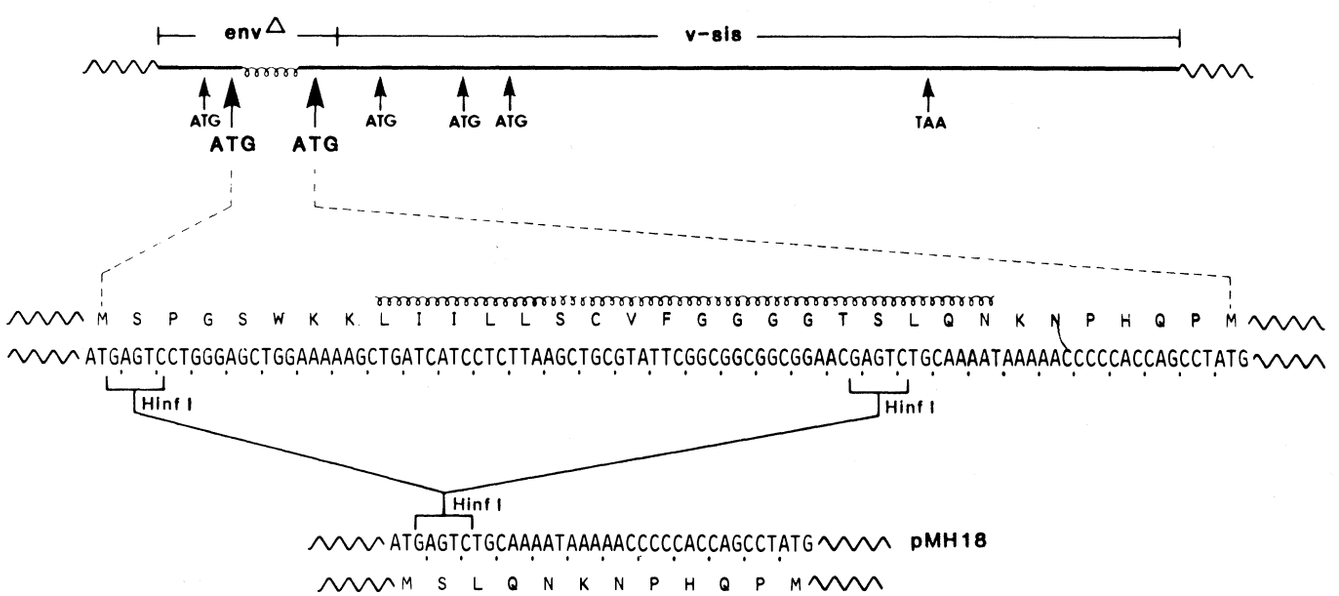
Fig. 1. NH_2-terminal deletion analysis of the *env-sis* coding region. (A) The SSV genome structure. Open triangles indicate the presence of a deletion within the gene. (B) Structure of the M-MuLV–derived expression vector. This vector contains deletions in both the *gag* and *pol* genes of M-MuLV, and the M-MuLV *env* gene has been completely removed. The restriction sites flanking the *env-sis* coding region of SSV were converted to Xho I sites by the use of synthetic oligonucleotides and inserted into the unique Xho I site of the vector. The correct orientation of the *env-sis* coding region for expression of the v-*sis* gene product was confirmed by restriction enzyme mapping. The v-*sis* gene used in our studies was obtained from the C60 clone of SSV which is biologically inactive due to an inversion in the SSV promoter (*2*). This inversion, however, left the v-*sis* region and adjacent SSV sequences intact. (C) Structure of the *env-sis* coding region. The open reading frame of 813 nucleotides contains a total of six ATG codons as shown by the closed circles (*18*). The recombination point in SSV between SSAV-derived sequences and c-*sis*–derived sequences is shown at the bottom of the figure as the junction between "*env*-related" and "PDGF-related" sequences, respectively. The proposed coding region for the signal sequence is indicated by a curlicue. The nucleotide sequence of each deletion endpoint was determined by Maxam and Gilbert sequencing (*19*). For focus assays, the calcium phosphate coprecipitates were added to subconfluent monolayers of NIH 3T3 cells and scored for focus formation after 1 week (*12*). The (+) notation indicates a minimum value of 1.1×10^4 focus-forming units (ffu) per picomole of cloned DNA. A (−) indicates a maximum value of 7.0×10^1 ffu/pmole. Conditioned media from each focus assay was collected and titered onto fresh monolayers of NIH 3T3 cells. In all cases where a positive result was obtained in the DNA transfection focus assay, the conditioned media exhibited a virus titer of at least 9×10^4 ffu/ml. When a negative result was obtained in the DNA transfection assay, the conditioned media exhibited a virus titer of less than 5×10^2 ffu/ml.

These results suggest that translation of the v-*sis* gene product is initiated at either the first or second ATG codon in the *env-sis* coding region.

There is an uninterrupted stretch of 18 codons (hydrophobic residues 22 to 39 counting from the first ATG codon) in the *env-sis* coding region located between the second and the third ATG codons (Fig. 2). These 18 codons are derived from the NH_2-terminal coding region of the *env* gene of simian sarcoma associated virus (SSAV), which is the parental retrovirus of SSV. Little is known about the processing of the *env* gene product of SSAV. In M-MuLV, the envelope glycoprotein gp70 is synthesized as a precursor protein, $Pr80^{env}$ (*13, 14*). This precursor contains an NH_2-terminal signal sequence of 33 amino acids which is cleaved during the translocation of the nascent polypeptide chain across the rough endoplasmic reticulum. The M-MuLV *env* signal sequence includes a region of 11 hydrophobic amino acids. In order to determine whether the region of hydrophobic amino acids in the *env-sis* gene product of SSV plays an analogous role in directing membrane translocation, we constructed a small deletion mutant (pMH18) in the *env-sis* coding region between the first and second ATG codons (Fig. 2). Using two Hinf I sites that bracket this region, we deleted 63 nucleotides (21 codons), thereby maintaining the correct reading frame across the new junction. All the ATG codons of the parental gene are retained in this mutant and remain in frame with the PDGF-related region. This deletion mutant was incapable of inducing focus formation in NIH 3T3 cells, thus demonstrating that the Hinf I–deleted region, containing a putative signal sequence, is required for synthesis of a transforming gene product from the v-*sis* coding region.

We have also used oligonucleotide site-directed mutagenesis to construct a point mutant in the proposed signal sequence (*15*). In this mutant, the valine codon (GTA) for the eighth amino acid in the hydrophobic region (Fig. 2) was changed to a glutamic acid codon (GAA). The mutant thus differs from the wild type by only a single T to A base substitution in the *env*-derived portion of the *env-sis* coding region. When this mutant was assayed for biological transformation using our M-MuLV–derived expression vector, it was unable to transform NIH 3T3 cells.

As a further test of our hypothesis that a functional signal sequence is required for the biological activity of v-*sis*, we restored biological activity to inactive deletion mutants by the addition of a heterologous signal sequence (Fig. 3). We constructed pDD151, an M-MuLV–derived expression vector (in a manner analogous to that outlined in the legend to Fig. 1) in which an inserted coding region would be fused to the *env* gene. The location of the Xho I site was moved so that it replaced the BstE II site at nucleotide 5923 in the *env* gene of M-MuLV (*14*). This resulted in the retention of 152 nucleotides of the M-MuLV *env* gene, including the ATG initiation codon and the coding region for the signal sequence of M-MuLV $Pr80^{env}$. Three dif-



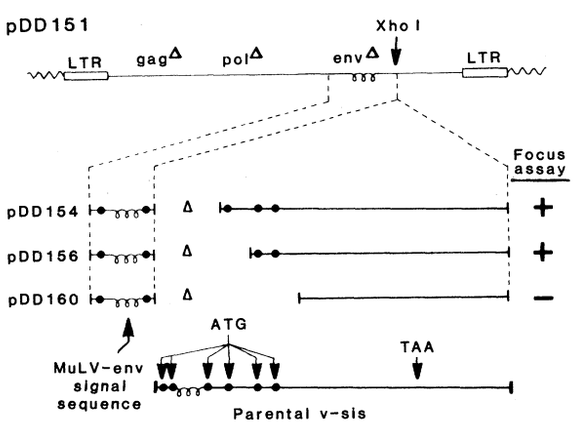


Fig. 2 (above). The structure of deletion mutation pMH18 within the proposed signal sequence. The nucleotide and predicted amino acid sequence of the *env-sis* coding region between the second and third ATG codons and the region across the restored Hinf I site in pMH18 are shown. The proposed signal sequence lies within a stretch of 18 hydrophobic residues as indicated by curlicues. Open triangles indicate the presence of a deletion within the gene. C, cysteine; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; S, serine; T, threonine; V, valine; W, tryptophan. Fig. 3 (left). Reactivation of *sis* deletion mutants by fusion with a heterologous signal sequence. Open triangles indicate the presence of a deletion within the gene. Closed circles indicate ATG codons in the *env-sis* open reading frame. Plasmids pDD154, pDD156, and pDD160 contain the NH_2-terminal deletions (see Fig. 1) from pDD141, pDD150, and pDD142, respectively. The nucleotide sequence at each *env*(MuLV)-*sis*(SSV) junction was determined (*19*), and the correct reading frame was maintained. Focus assays were carried out as described in the legend to Fig. 1.

ferent *sis* deletion mutants were inserted into this signal peptide vector so that the correct reading frame was maintained across the junction. These constructs were expected to encode an *env*(MuLV)-*sis*(SSV) fusion protein. Two of the deletion mutants, pDD154 and pDD156, regained full biological activity. The third, pDD160, was still unable to induce focus formation (Fig. 3), a result that was not surprising, as pDD160 contains a significant deletion of the PDGF-related region. The observation of full biological activity in pDD154 and pDD156, in which the PDGF-related region remained intact, confirms our hypothesis that a functional signal sequence is required for transformation by the v-*sis* gene product.

Our results were obtained by molecular manipulations of coding regions and subsequent correlations with biological activity. A more direct test of our model would be to examine the size and structure of the *sis*-related proteins synthesized in cells infected with some of the nontransforming mutants described above. To date, such experiments have yielded inconclusive results because of the lack of a highly reactive antibody to the *sis* gene product. We believe that our mutations are altering processing of the *sis* gene product, rather than transcription or translation for the following reasons: (i) A single base change in the coding region of the proposed signal sequence abolished transformation. This mutation would not be expected to affect transcription nor the initiation of translation. (ii) The deletion mutant pMH18 leaves all transcriptional signals and all six ATG codons of the *env-sis* coding region intact, yet is biologically nontransforming. (iii) RNA synthesized in vitro (*16*) from the wild-type *sis* gene and also from the nontransforming mutant pDD143 was successfully translated by means of a reticulocyte lysate system (*17*) to yield polypeptides of the anticipated molecular weights (*15*). Thus, the deleted *sis* gene in pDD143 can be transcribed to yield translatable RNA.

According to one model, transformation results from the interaction between the PDGF-related protein and the cell surface receptor for PDGF. The v-*sis* gene product is predicted to be either an integral membrane protein or a secreted protein, which would facilitate interaction with the PDGF receptor. Our experiments are consistent with the hypothesis that a signal sequence is required for transformation by v-*sis*. Further, these data indicate that the primary translation product of the *env-sis* coding region, in the absence of any proteolytic processing, would be either 33 or 30 kilodaltons depending on whether the first or second ATG codon is used. This is in contrast to other proposals that the primary translation product initiates at the third ATG codon and is about 28 kilodaltons (*4–8*).

Our results demonstrate a novel mechanism whereby viral encoded sequences may activate a cellular protooncogene. The activation of cellular sequences encoding a PDGF-related protein would occur as a result of viral signals which control transport of the viral *env* gene product. The point of integration of cellular sequences into the retroviral genome may be an important factor in the activation of those sequences into a transforming oncogene. Thus our data suggest that alteration of the cytological location of a growth factor can result in cellular transformation.

MARK HANNINK
DANIEL J. DONOGHUE*
Department of Chemistry, B-017,
University of California,
San Diego, La Jolla 92093

References and Notes

1. G. H. Theilen, D. Gould, M. Fowler, D. L. Dungworth, *J. Natl. Cancer Inst.* **47**, 881 (1971).
2. E. P. Gelman, F. Wong-Staal, R. A. Kramer, R. C. Gallo, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 3373 (1981).
3. K. C. Robbins, S. G. Devare, S. A. Aaronson, *ibid.*, p. 2918.
4. S. G. Devare, E. P. Reddy, J. D. Law, K. C. Robbins, S. A. Aaronson, *ibid.* **80**, 731 (1983).
5. S. F. Josephs, C. Guo, L. Ratner, F. Wong-Staal, *Science* **223**, 487 (1984).
6. R. F. Doolittle *et al.*, *ibid.* **221**, 275 (1983).
7. M. D. Waterfield *et al.*, *Nature (London)* **304**, 35 (1983).
8. K. C. Robbins *et al.*, *ibid.* **305**, 605 (1983).
9. T. F. Deuel, J. S. Huang, S. S. Huang, P. Stroobant, M. D. Waterfield, *Science* **221**, 1348 (1983).
10. D. J. Donoghue, in preparation.
11. When inserted in the MuLV-derived expression vector, the *sis* gene should be expressed from a spliced messenger RNA very similar in structure to the *sis*-specific messenger RNA produced by SSV. In cells transformed by SSV, a spliced messenger RNA of 2.7 kilobases has been identified [A. Eva *et al.*, *Nature (London)* **295**, 116 (1982)] which presumably arises by splicing to the splice acceptor site of the SSAV *env* gene retained in the SSV genome. Although the 3′-splice acceptor site in SSV is not known, examination of the SSV nucleotide sequence reveals several potential splice acceptor sites upstream of the *env-sis* region.
12. F. L. Graham and A. J. van der Eb, *Virology* **52**, 456 (1973); P. Andersson, M. P. Goldfarb, R. A. Weinberg, *Cell* **16**, 63 (1979).
13. L. E. Henderson, T. D. Copeland, G. W. Smythers, H. Marquardt, S. Oroszlan, *Virology* **85**, 319 (1978).
14. T. M. Shinnick, R. A. Lerner, J. G. Sutcliffe, *Nature (London)* **293**, 543 (1981).
15. M. Hannink and D. J. Donoghue, data not shown.
16. M. R. Green, T. Maniatis, D. A. Melton, *Cell* **32**, 681 (1983).
17. H. R. B. Pelham and R. J. Jackson, *Eur. J. Biochem* **67**, 247 (1976).
18. The initial DNA sequence data suggested that v-*sis* was actually a *pol-env-sis* fusion gene in which the first two ATG codons shown in Fig. 1C were not present [S. G. Devare *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 3179 (1982)]. Subsequent corrections to the SSV sequence (*4*) resulted in the structure of the *env-sis* coding region as shown in Fig. 1C.
19. A. Maxam and W. Gilbert, *ibid.* **74**, 560 (1977).
20. We thank Russell Doolittle, Tony Hunter, Immo Scheffler, Bart Sefton, and Suresh Subramani for advice and criticism and Richard Bold for excellent technical assistance. This work was supported by PHS grant CA34456 awarded by the National Cancer Institute, DHSS; the Chicago Community Trust/Searle Scholars Program (D.J.D.); and a Cell and Molecular Biology Training grant, GM07313 (M.H.).
* To whom reprint requests should be addressed.

14 May 1984; accepted 24 September 1984

Constitutive Fragile Sites and Cancer

Abstract. *Breaks were observed at 51 sites in homologous chromosomes in lymphocytes from ten humans and two great apes when cells were deprived of thymidine. The incidence of breaks was enhanced by caffeine, a substance that inhibits DNA repair in replicating cells. The locations of 20 sites were correlated with breakpoints that have been related to human malignancy.*

A few specific chromosome sites tend to be expressed as gaps or breaks during routine human metaphase chromosome preparations (320 bands per haploid set). With the exception of sites 3p14.2, 6q25.3, and 16q23.2, these sites have appeared infrequently and at low levels of expression, and therefore have been difficult to characterize (*1*). Deprivation of folic acid and thymidine results in the expression of 13 of the 16 known heritable fragile sites (h-fra), and in the appearance of "spontaneous" chromosome breaks (*2*), here termed constitutive fragile sites (c-fra). By examining elongated chromosomes [670 to 850 bands per haploid set (*3*)], we have found a large family of c-fra located at precise points in homologous chromosomes of human and primate genomes. The expression of these genomic weak points is enhanced when cells are exposed to 2.2 m*M* caffeine during the last 6 hours of culture.

Blood from five males and five females ranging in age from 10 to 65 years was studied. Eight of these individuals were normal; one was mentally retarded and had h-fra Xq28.1 (human 2 in Table 1); one had acute leukemia inv(16)(p13.1q22.1), and c-fra 16q22.1 (human 3 in Table 1). We also donated blood after having taken 5 mg of folinic acid a day orally for 3 days (humans 1 and 4 in Table 1). In addition, blood was also studied from one 24-year-old female chimpanzee (*Pan troglodytes*) and one 34-year-old female gorilla (*Gorilla gorilla*). We used blood lymphocytes which