Simian Sarcoma Virus *onc* Gene, v-*sis*, Is Derived from the Gene (or Genes) Encoding a Platelet-Derived Growth Factor

Abstract. *The transforming protein of a primate sarcoma virus and a platelet-derived growth factor are derived from the same or closely related cellular genes. This conclusion is based on the demonstration of extensive sequence similarity between the transforming protein derived from the simian sarcoma virus* onc *gene, v-*sis, *and a human platelet-derived growth factor. The mechanism by which v-*sis *transforms cells could involve the constitutive expression of a protein with functions similar or identical to those of a factor active transiently during normal cell growth.*

Recently, investigations of the genetic alterations that cause normal cells to become malignant have focused on a small set of cellular genes. Acute transforming retroviruses have substituted viral genes necessary for replication with these discrete segments of host genetic information (*1*). When incorporated within the retroviral genome, these transduced cellular sequences, termed *onc* genes, acquire the ability to induce neoplastic transformation. The discovery that independent virus isolates have recombined with the same or closely related cellular proto-oncogenes (*2*) implies that only a limited number of cellular genes are capable of acquiring transforming properties under these conditions. Proto-oncogenes can also be activated to become transforming genes by mechanisms independent of transduction by retrovirus. Genetic changes as small as point mutations (*3*), as well as DNA rearrangements such as transpositions (*4*) and chromosomal translocations (*5*), have been implicated in this process.

The profound cellular alterations induced by the activated cellular transforming genes have some similarities to the growth promoting actions of hormones and growth factors. Each exerts pleiotropic effects on cellular metabolism, including the induction of sustained cell replication. However, previous efforts to demonstrate any functional or evolutionary relatedness between transforming gene products and any growth factor have been unsuccessful. Interest in understanding the actions of *onc* genes has led to concerted efforts to isolate, amplify, and sequence such genes. At the same time, investigations oriented toward understanding the function of growth factors have included basic characterization of these potent molecules through their isolation, purification, and primary structure determination. We now demonstrate that the transforming protein of a primate sarcoma virus and a platelet-derived growth factor are derived from the same or closely related cellular genes.

Initially, simian sarcoma virus (SSV) was isolated from a fibrosarcoma of a woolly monkey; SSV is the only known sarcoma virus of primate origin (*6*). The virus has been characterized in tissue culture (*7*), and its integrated DNA provirus has been cloned in infectious form (*8*). Physical and biological characterization of its genome has localized its transforming gene to its cell-derived *onc* sequence, v-*sis* (*8, 9*). The SSV genome, including the v-*sis* region, has been sequenced (*10*). Moreover, antibodies prepared against small peptides derived from the v-*sis* sequence have been used to identify the 28,000-dalton v-*sis* gene product, p28*sis*, in SSV-transformed cells (*10, 11*).

Platelet-derived growth factor (PDGF) is a heat-stable (100°C), cationic (*p*I, 9.8) protein (*12*). It circulates in blood stored in the α granules of platelets and is released into serum during blood clotting (*13*). It represents the major protein growth factor of human serum and is a potent mitogen for connective tissue and glial cells in culture (*14, 15*). Unreduced, active PDGF exhibits multiple forms ranging in size from 28,000 to 35,000 daltons (*12, 16, 17*). Reduction of PDGF produces inactive, smaller peptides ranging in size from 12,000 to 18,000 daltons (*15, 16*).

Recently, amino acid sequence analysis of the amino-terminal portions of both active human PDGF and its inactive, reduced peptides has revealed the presence of two homologous peptides (PDGF-1 and PDGF-2) in active PDGF preparations (*18*). These peptides are identical at 8 of 19 positions near their amino termini, with no sequence gaps required for the homology alignment. Whether the active PDGF preparation is composed of a single protein formed by disulfide linkage of these two peptides or is composed of two proteins, each of which consists of a disulfide-linked dimer of one of the peptides, is not yet known.

A comparison of the PDGF amino-terminal sequences with the sequence of p28*sis* reveals a striking similarity between these proteins. This similarity was discovered by one of us (R.F.D.) during a search for sequence homology between the PDGF amino-terminal sequences and the other protein sequences in the Newat sequence data base at the University of California, San Diego (*19*). Subsequently, we have extended the sequence comparisons with additional PDGF primary

```
p28sis     1  MTLTWQGDPIPEELYKMLSGHSIRSFDDLQRLLQGDSGKEDGAELDLNMT  50

p28sis    51  RSHSGGELESLARGKRSLGSLSVAEPAMIAECKTRTEVFEISRRLIDRTN  100
PDGF-2     1                  SLGSLTIAEPAMIAECKTREEVFCICRRL?DR??  34
PDGF-1     1                      SIEEAVPAVCKTRIVIYEISRRELD???    28

p28sis   101  ANFLVWPPCVEVQRCSGCCNNRNVQCRPTQVQLRPVQVRKIEIVRKKPIF  150
PDGF-2    35  ??????PPCVEVKRCTGCCNNRNVKCRPSQVQLRP?QVRKIEIVRK[     80
PDGF-1    29  ANFL[                                               32

p28sis   151  KKATVTLEDHLACKCEIVAAARAVTRSPGTSQEQRAKTTQSRVTIRTVRV  200
PDGF-2
PDGF-1

p28sis   201  RRPPKGKHRKCKHTHDKTALKETLGA                          226
PDGF-2                                  ]
PDGF-1                                  ]
```

Fig. 1. Sequence similarity between p28*sis* and PDGF. The p28*sis* sequence is from (*10*); the PDGF sequences are from (*18, 20*). Residue identity between the p28*sis* and PDGF sequences is indicated by the solid lines between the sequences. A question mark indicates that no amino acid sequence assignment has yet been made for that position; the brackets indicate no sequence is yet available for the included segments. The box around p28*sis* positions 65 and 66 indicates a possible proteolytic processing position for generation of a fragment of p28*sis* corresponding to PDGF-2. Single letter abbreviations for the amino acid residues are as follows: A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine.

sequence obtained from peptide fragments generated by cyanogen bromide cleavage of native PDGF (*20*).

The sequence alignments and matches derived from these searches are shown in Fig. 1. The sequence comparisons are summarized as follows:

1) PDGF-1 matches p28*sis* at 18 of 29 positions identified by protein sequencing.

2) PDGF-2 matches p28*sis* at 26 of 31 positions at the PDGF-2 amino terminus and at 35 of 39 positions at the amino terminus of a 14,000-dalton PDGF-2 cyanogen bromide fragment (total match at 61 of 70 identified positions).

3) No sequence gaps due to insertions or deletions were used in these alignments.

This extensive sequence homology leaves little doubt that the v-*sis* transforming gene, as isolated from the woolly monkey, is the result of a viral recombination with the host cell gene or genes encoding PDGF or a very similar protein. The match between the 70 identified PDGF-2 residues and the corresponding segment of p28*sis* is 87.1 percent. Since the v-*sis* gene is thought to have arisen from within the genome of a woolly monkey (*6*), a member of the family Cebidae (New World monkeys), and the PDGF was isolated from human platelets, most, if not all, of the observed amino acid differences could represent species differences. This hypothesis is consistent with the known amino acid sequence similarity for myoglobin [90.8 percent identity (*21*)] and fibrinopeptides A and B [70.0 percent identity (*22*)] from humans and cebids. Moreover, seven of the nine observed differences can be derived from single base changes.

An examination of PDGF-2 and p28*sis* sequences reveals several additional facts consistent with a close relation between the two proteins. While the peptide molecular size of the 226 constituent amino acids of p28*sis* is 25,414 daltons, removal of the 66-residue fragment that precedes the section matching PDGF-2 would yield a 160-residue protein with a molecular size of 18,056 daltons, essentially the same value calculated for PDGF-2 on the basis of sodium dodecyl sulfate–polyacrylamide gel electrophoresis. Evidence for processing of p28*sis* near this point comes from studies with antiserums specific for the amino and carboxyl termini of p28*sis* (*23*). The cleavage point, based on the p28*sis* sequence, would be signaled by a double-basic (Lys-Arg) (Lys, lysine; Arg, arginine) sequence at positions 65–66. The processed 160-residue viral protein would be very basic, mainly because of an unusual number of arginines. Its calculated *p*I is about 10, a match for the observed value of 9.8 reported for active PDGF, and at neutral *p*H it would have a net charge of +20. It would contain nine cysteines; the odd number is consistent with external disulfide bonding, as in the disulfide-linked dimer observed for native PDGF. The predicted secondary structure (*24*) and hydropathy index (*25*) of the viral protein are consistent with a water-soluble globular protein, about one-third α-helical, containing no membrane-spanning segments.

Despite advances in identifying cellular genes with transforming potential, little is known about proto-oncogene function or how the altered counterparts of these genes disrupt normal growth regulation. One major family of *onc* genes codes for protein kinases with specificity for phosphorylation of tyrosine residues (*2*); another has an associated GTP-binding (guanosine triphosphate) activity (*26*). Moreover, previous computer searches have revealed distant homology between the *src* gene product and cyclic adenosine monophosphate kinase (27) and possibly between the B*lym* gene product and transferrin (*28*). However, until now none of the identified *onc* genes has been shown to correspond to a cellular gene encoding a protein with known physiological function. Any such linkage might have far-reaching consequences with respect to understanding the mechanisms of action of the transforming gene itself.

Our studies reveal a high degree of relatedness between the transforming gene product of a primate sarcoma virus and a potent growth factor for human fibroblasts, smooth muscle cells, and glial cells. The mechanism by which this *onc* gene transforms cells may involve the constitutive expression of a protein with functions similar to those of a factor active transiently during normal cell growth.

This relationship raises intriguing possibilities about the mechanism of action at the molecular level of both the *onc* gene product and the growth factor. (i) Is the PDGF activity inherent in the dimeric structure of native PDGF or in the conformation and structure of a single chain? If PDGF functions as a dimer, then does the viral protein also require this structure to be active? (ii) Does the viral protein function as an intra- or extracellular signal? PDGF has a specific receptor bound to the cell membrane and presumably acts through it. Does the viral protein act through the same or a similar receptor (*29*)? (iii) Is proteolytic processing of the viral protein to give a peptide similar in size to the corresponding PDGF peptide essential to the function of the viral protein? (iv) How general is the phenomenon of an *onc* gene coding for a protein whose sequence is similar or identical to that of a normal growth factor? (v) Can antiserums specific for *onc* gene-related growth factors inhibit the growth of associated tumors (*30*)?

RUSSELL F. DOOLITTLE
Department of Chemistry, University of California, San Diego, La Jolla 92093

MICHAEL W. HUNKAPILLER
LEROY E. HOOD
Division of Biology, California Institute of Technology, Pasadena 91125

SUSHILKUMAR G. DEVARE
KEITH C. ROBBINS
STUART A. AARONSON
Laboratory of Cellular and Molecular Biology, National Cancer Institute, Bethesda, Maryland 20205

HARRY N. ANTONIADES
Center for Blood Research and *Department of Nutrition, Harvard University School of Public Health, Boston, Massachusetts 02115*

References and Notes

1. "Viral oncogenes," *Cold Spring Harbor Symp. Quant. Biol.* **44** (1980).
2. *RNA Tumor Viruses*, R. Weiss, N. Tiche, H. Varmus, J. Coffin, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1982).
3. C. J. Tabin *et al.*, *Nature (London)* **300**, 143 (1982); E. P. Reddy, R. K. Reynolds, E. Santos, M. Barbacid, *ibid.*, p. 149; E. Taparowsky, Y. Suard, O. Fasano, K. Shimizu, M. Goldfarb, M. Wigler, *ibid.*, p. 762.
4. G. Rechavi, D. Givol, E. Cannani, *ibid.*, p. 607.
5. R. Dalla-Favera *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 7824 (1982); R. Taub *et al.*, *ibid.*, p. 7837; S. Crews, R. Barth, L. Hood, J. Prehn, K. Calame, *Science* **218**, 1319 (1982); L. J. Harris, P. D'Eustachio, F. H. Ruddle, K. B. Marcu, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 6622 (1982); G. L. C. Shen-Ong, E. J. Keath, S. P. E. Piccoli, M. D. Cole, *Cell* **29**, 443 (1982).
6. G. J. Thielen, D. Gould, M. Fowler, D. L. Dungworth, *J. Natl. Cancer Inst.* **47**, 881 (1971).
7. S. A. Aaronson, *Virology* **52**, 562 (1973).
8. K. C. Robbins, S. G. Devare, S. A. Aaronson, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 2918 (1981).
9. E. P. Gelmann, F. Wong-Staal, R. A. Kramer, R. C. Gallo, *ibid.*, p. 3373; F. Wong-Staal *et al.*, *Nature (London)* **294**, 273 (1981); K. C. Robbins, R. L. Hill, S. A. Aaronson, *J. Virol.* **41**, 721 (1982).
10. S. G. Devare, E. P. Reddy, J. D. Law, K. C. Robbins, S. A. Aaronson, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 731 (1983).
11. K. C. Robbins, S. G. Devare, E. P. Reddy, S. A. Aaronson, *Science* **218**, 1131 (1982).
12. H. N. Antoniades, C. D. Scher, C. D. Stiles, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1809 (1979).
13. D. R. Kaplan, F. C. Chao, C. D. Stiles, H. N. Antoniades, C. D. Scher, *Blood* **53**, 1043 (1979).
14. R. Ross, J. Glomset, B. Kariya, L. Harker, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 1207 (1974); C. D. Scher, R. C. Shepard, H. N. Antoniades, C. D. Stiles, *Biochim. Biophys. Acta* **560**, 217 (1979).
15. C. H. Heldin, B. Westermark, A. Wasteson, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 3722 (1979).
16. ———, *Biochem. J.* **193**, 907 (1981); H. N. Antoniades, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 7314 (1981).
17. T. F. Deuel, J. S. Huang, R. T. Proffit, J. U. Baenziger, D. Chang, B. B. Kennedey, *J. Biol. Chem.* **256**, 8896 (1981); E. W. Raines and R. Ross, *ibid.*, pp. 257 and 5154.
18. H. N. Antoniades and M. W. Hunkapiller, *Science* **220**, 963 (1983).
19. R. F. Doolittle, *ibid.* **214**, 149 (1981). The data

base searched included 145,581 amino acid residues comprising 684 individual sequences in the Newat list and 121,098 residues from 1081 sequences in the 1978 Dayhoff collection [*Protein Segment Dictionary 78*, M. O. Dayhoff, L. T. Hunt, W. C. Barker, R. M. Schwartz, B. C. Orcutt, Eds. (National Biomedical Research Foundation, Silver Spring, Md., 1978)].
20. PDGF-A (18,000 daltons) was prepared by reduction, alkylation, and gel electrophoresis of active PDGF (*18*). It consists of roughly equal amounts of PDGF-1 and -2. Treatment of this preparation with cyanogen bromide followed by gel electrophoresis yielded two protein bands (H. N. Antoniades and M. W. Hunkapiller, unpublished data). The larger of these, 18,000 daltons, contained only PDGF-1 sequence, indicating that PDGF-1 was not cleaved by the cyanogen bromide treatment, but that PDGF-2 was. The smaller band, 14,000 daltons, contained a single sequence that must be derived from PDGF-2 by cleavage at two sites near the amino terminus of PDGF-2, one of which is presumably at Met^{12} (methionine) of the PDGF-2 sequence. Amino acid sequence analysis was performed as described (*18*) with the Caltech gas phase protein sequenator.
21. H. Dene, J. Sazy, A. E. Romero-Herrera, *Biochim. Biophys. Acta* **625**, 133 (1980).
22. G. L. Wooding and R. F. Doolittle, *J. Hum. Evol.* **1**, 553 (1972).
23. S. A. Aaronson and K. C. Robbins, unpublished data.
24. P. Y. Chou and G. D. Fasman, *Annu. Rev. Biochem.* **47**, 251 (1978); J. Garnier, D. J. Osgothorpe, B. Robson, *J. Mol. Biol.* **120**, 97 (1978).
25. J. Kyte and R. F. Doolittle, *J. Mol. Biol.* **157**, 105 (1982).
26. T. Y. Shih, M. O. Weeks, H. A. Young, E. M. Scolnick, *Virology* **96**, 64 (1979).
27. W. C. Barker and M. O. Dayhoff, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2836 (1982).
28. G. Goubin, D. F. Goldman, J. Luce, P. E. Neiman, G. M. Cooper, *Nature (London)* **302**, 114 (1983).
29. M. D. Sporn and G. J. Todaro, *N. Engl. J. Med.* **303**, 878 (1980); D. T. Graves, A. J. Owen, H. N. Antoniades, *Cancer Res.* **43**, 83 (1983).
30. A preliminary announcement of the results reported here was made at a symposium on "The Role of Oncogenes in Carcinogenesis" at the 74th annual meeting of the American Society of Biological Chemists, San Francisco, Calif., 5 to 9 June 1983.
31. We thank T. Hunkapiller for preparation of Fig. 1, and the American Red Cross Blood Services, Northeast Section, for the supply of clinically outdated human platelets. Supported in part by NIH research grants RR00757 (R.F.D.) and CA30101 (H.N.A.) and by grants from the Council for Tobacco Research, USA, Inc. (H.N.A.) and the Weingart Foundation (L.E.H.).

3 June 1983; revised 12 June 1983

Rapid Changes in Tree Leaf Chemistry Induced by Damage: Evidence for Communication Between Plants

Abstract. *Potted poplar ramets showed increased concentrations and rates of synthesis of phenolic compounds within 52 hours of having 7 percent of their leaf area removed by tearing, as did undamaged plants sharing the same enclosure. Damaged sugar maple seedlings responded in a manner similar to that of the damaged poplars. Nearby undamaged maples had increased levels of phenolics and hydrolyzable and condensed tannin within 36 hours, but exhibited no change in rates of synthesis. An airborne cue originating in damaged tissues may stimulate biochemical changes in neighboring plants that could influence the feeding and growth of phytophagous insects.*

Attacks by herbivores on plants may reduce the quality of the plant tissues for subsequent feeding (*1, 2*). It has not been known whether such changes occur quickly enough to reduce ongoing defoliation. We report here that poplar (*Populus* × *euroamericana*) ramets and sugar maple (*Acer saccharum* Marsh) seedlings exhibit elevated concentrations of phenols and increased protein binding of phenolic compounds in leaf extracts within 75 hours of being mechanically damaged. Synthesis of these compounds apparently increases in response to damage. Recently, Rhoades (*3*) presented bioassay results suggesting that a factor released by damaged trees stimulates changes in the leaf quality of undamaged neighbors. We also report chemical evidence of such communication between trees: leaves of undamaged individuals in the same enclosure as damaged plants exhibit similar chemical changes.

In two separate experiments 45 1.5-month-old poplar ramets and 45 4-month-old sugar maple seedlings grown under controlled conditions (*4*) were placed in two gas-tight Plexiglas enclosures (*5*) inside a growth chamber. One enclosure housed 15 true control plants, while the other housed 15 experimental and 15 "communication control" plants. Two leaves on each experimental plant were damaged (*6*), while the leaves of both control groups were untouched. At the time of damage, $^{14}CO_2$ was introduced into both enclosures and then scrubbed from the air (*7*) before the first of three sampling intervals (*8*). At each interval five plants from each group were removed from the chamber for analyses of four chemical leaf traits known to influence insect feeding and to be affected by herbivore damage in other tree species (*1, 2*). Four leaves from each seedling (*9*) were extracted and analyzed for total phenolics, hydrolyzable and condensed tannins, and tanning activity (*10*). Carbon-14 was assayed in leaf extracts and in phenolics precipitated by aqueous lead acetate (*11*).

At the first sampling interval all seedlings had elevated chemical measures, probably because of metabolic stimulation by elevated CO_2 (*12*). Hence the effect of damage must be considered against the dynamic background of the true controls. Damaged plants and communication controls exhibited chemical changes that were significantly different from patterns seen in the true controls.

Total phenolics in leaf extracts from poplar ramets suffering damage to approximately 7 percent of their leaf area increased 123.3 percent within 52 hours ($P = .001$) and decreased to control levels by 100 hours (Fig. 1A). Communication controls produced leaf extracts with phenolic contents elevated 57.6 percent ($P = .001$) in less than 52 hours (Fig. 1A). Damaged poplars incorporated more photosynthate into phenolics than did controls; ^{14}C in phenolic precipitate increased 150 percent by 52 hours (Fig. 1B). Undamaged communication controls also showed increased phenolic synthesis; ^{14}C in phenolics increased 128 percent ($P = .01$) after neighboring trees were damaged (Fig. 1B).

Damaged sugar maple seedlings had a significantly higher total phenolic content at 75 hours than did true controls (Fig. 1C). The protein binding capacity of leaf extracts from damaged seedlings was 33 percent higher (equivalent to 10 percent tannic acid by dry weight) than that of leaves from true controls at 75 hours (Fig. 1D). The concentration of hydrolyzable tannin increased in all plants by 36 hours (Fig. 1E), perhaps because of elevated photosynthetic rates under conditions of enriched CO_2 (*12*). At 75 hours hydrolyzable tannin remained elevated in leaves from damaged trees, while it was reduced in leaves from controls. No changes in condensed tannins were evident in leaves from damaged trees or true controls.

Phenolic and hydrolyzable tannin contents in leaf extracts from undamaged communication controls kept in the same air space as damaged trees were significantly higher than those in true controls by 36 hours and remained significantly higher at 75 hours (Fig. 1, C and D). These extracts also had significantly higher concentrations of condensed tannin than extracts from true controls, increasing 21.4 percent after 36 hours to the dry weight equivalent of 0.04 percent red oak tannin (*10*). The level of condensed tannin declined to that of the true controls after 75 hours.

As in the damaged poplar seedlings, ^{14}C was actively incorporated into phenolic compounds by damaged maple seedlings: ^{14}C counts were 2.5 times greater in precipitated phenolics than in true controls at 75 hours ($P = .03$). However, the number of ^{14}C counts in phenolics from communication control leaves did not differ significantly from