a molecular tracer (*21*). Specific questions concerning the seasonal input of DOM that might result from either phytoplankton productivity or litterfall and increased leaching could be evaluated by this technique. Moreover, the degree of oxidation of the lignin signature might be used to distinguish inputs of soil organic matter from fresh leachate. For river systems draining different vegetative regions, lignin analysis could reveal the geographic sources of DOM by identification of the regionally characteristic vascular plant sources. Also, differences in the lignin signature of humic and fulvic acids could be exploited to determine their relative conservation during estuarine mixing (*22*). Finally, the production of lignin-derived phenols from the CuO oxidation of marine humic substances would be an unambiguous indicator that terrestrially derived carbon is present in the marine dissolved organic pool (*23*).

JOHN R. ERTEL
JOHN I. HEDGES
School of Oceanography, University of Washington, Seattle 98195

EDWARD M. PERDUE
School of Geophysical Sciences, Georgia Institute of Technology, Atlanta 30332

References and Notes

1. M. Maybeck, in "Flux of Organic Carbon by Rivers to the Ocean," report of a workshop, Woods Hole, Mass., 21 to 25 September 1980 (Department of Energy, Washington D.C., 1981), p. 219.
2. J. I. Hedges, *ibid.*, p. 109.
3. E. M. Thurman and R. L. Malcolm, *Environ. Sci. Technol.* **15**, 463 (1981); J. H. Reuter and E. M. Perdue, *Geochim. Cosmochim. Acta* **41**, 325 (1977).
4. C. R. Lytle and E. M. Perdue, *Environ. Sci. Technol.* **15**, 224 (1981).
5. M. S. Sweet and E. M. Perdue, *ibid.* **16**, 692 (1982).
6. R. F. Christman and M. Ghassemi, *J. Am. Water Works Assoc.* **58**, 723 (1966); R. F. Christman, in *Organic Matter in Natural Waters*, D. W. Hood, Ed. (University of Alaska, College, 1974), p. 181.
7. J. Pempkowiak and R. Pockington, in *Aquatic and Terrestrial Humic Materials*, R. F. Christman and E. T. Gjessing, Eds. (Ann Arbor Science, Ann Arbor, Mich., 1983), p. 371.
8. J. I. Hedges and P. L. Parker, *Geochim. Cosmochim. Acta* **40**, 1019 (1976); J. I. Hedges and D. C. Mann, *ibid.* **43**, 1803 (1979); *ibid.*, p. 1809.
9. J. I. Hedges, J. R. Ertel, E. B. Leopold, *ibid.* **46**, 1869 (1982).
10. J. I. Hedges and J. R. Ertel, *Anal. Chem.* **154**, 178 (1982).
11. K. V. Sarkanen and C. H. Ludwig, *Lignins* (Wiley, New York, 1971).
12. E. B. Leopold, R. Nickmann, J. I. Hedges, J. R. Ertel, *Science* **218**, 1305 (1982).
13. We characterize vanillyl phenols by 3-methoxyl substitution, syringyl phenols by 3,5-dimethoxyl substitution, and cinnamyl phenols by *trans*-propenoic acid substitution.
14. E. M. Perdue, in *Chemical Modeling in Aqueous Systems*, E. A. Jenne, Ed. (Symposium Series 93, American Chemical Society, Washington, D.C., 1979), p. 94.
15. J. R. Ertel and J. I. Hedges, in *Aquatic and Terrestrial Humic Materials*, R. F. Christman and E. T. Gjessing, Eds. (Ann Arbor Science, Ann Arbor, Mich., 1983), p. 143.
16. _____, in preparation.
17. H. Blaschke, *Soil Biol. Biochem.* **11**, 581 (1979).
18. Two-thirds of the vanillyl phenols in lignin are contained in structures not released by CuO oxidation, whereas syringyl structural units are released with 90 percent efficiency (*8*). Thus all vanillyl phenol yields were multiplied by 3.
19. R. L. Crawford, *Lignin Biodegradation and Transformation* (Wiley, New York, 1981).
20. Aromatic rings containing vicinal hydroxyl groups are cleaved by CuO oxidation. This apparently is the reason why soil organic matter, although predominantly derived from vascular plants, produces few lignin-derived phenols by this technique (*16*).
21. R. F. Mantoura and E. M. S. Woodward, *Geochim. Cosmochim. Acta* **47**, 1293 (1983).
22. E. R. Sholkovitz, E. A. Boyle, N. B. Price, *Earth Planet. Sci. Lett.* **40**, 130 (1978).
23. Recent measurements (*21*) and lignin analyses of dissolved humic substances in the Baltic Sea (*7*) indicate that terrestrial carbon could be a significant fraction of marine dissolved organic carbon.
24. Supported by NSF grants OCE8023970 and OCE8219294. This is contribution 1358 from the School of Oceanography, University of Washington.

12 September 1983; accepted 8 November 1983

Human Proto-Oncogene Nucleotide Sequences Corresponding to the Transforming Region of Simian Sarcoma Virus

Abstract. *The nucleotide sequences of the six regions within the normal human cellular locus (c-*sis*) that correspond to the entire transforming region of the simian sarcoma virus (SSV) genome (v-*sis*) were determined. The regions are bounded by acceptor and donor splice sites and, except for region 6, resemble exons. Region 6 lacks a 3′ donor splice site and terminates −5 base pairs from the 3′ v-*sis*-helper-viral junction. This is consistent with a model proposing that SSV was generated by recombination between proviral DNA of simian sarcoma associated virus and proto-*sis *and that introns were spliced out subsequently from a fused viral-*sis *messenger RNA. This also suggests that the 3′ recombination occurred within an exon of the woolly monkey (*Lagothrix*) genome. The open reading frames predicting the v-*sis *and c-*sis *gene products coincide with the stop codon of c-*sis *located 123 nucleotides into the fifth region of homology. The overall nucleotide homology was 91 percent with substitutions mainly in the third codon positions within the open reading frame and with greatest divergence within the untranslated 3′ portion of the sequences. The predicted protein products for v-*sis *and c-*sis *are 93 percent homologous. The predicted c-*sis *gene product is identical in 31 of 31 amino acids to one of the published sequences of platelet-derived growth factor. Thus, c-*sis *encodes one chain of human platelet-derived growth factor.*

The simian sarcoma virus (SSV) is the only acutely transforming retrovirus (causing disease within a short latency period) that has been obtained from primates. After isolation of this virus from a fibrosarcoma of a pet woolly monkey (*Lagothrix*) (*1*), the molecularly cloned genome was shown to contain 1.0 kilobase (kb) of woolly monkey–derived cellular sequences that are responsible for the transforming potential of the virus (*2–4*). Since SSV is related to the nonacutely transforming gibbon ape leukemia viruses, infection of the woolly monkey with a virus from this group probably preceded the recombination event giving rise to SSV (*5*).

The complete nucleotide sequence of the viral transforming gene as well as immunological data have shown that the transforming product of SSV is a 28,000 dalton protein, $p28^{sis}$, that is probably encoded mostly by sequences derived from woolly monkey cellular DNA (*6*). Several amino acid residues at the amino terminus of p28 are thought to be coded by the envelope (*env*) gene of the helper virus derived sequences.

The human cellular homolog of the SSV genome, c-*sis*, contains 1.0 kb of coding regions, distributed over 12.5 kb, that are interspersed by at least four introns (*7*). *Sis*-related messenger RNA's (mRNA's) of 4.2 kb can be detected in sarcomas, glioblastomas, and some human T-cell leukemia virus (HTLV)–infected cell lines (*8*). These results and the nucleotide sequence data of the 5′ cellular (c-*sis*) gene indicate that additional exons must be present 5′ and may also be present 3′ to the v-*sis* region of homology (*9*).

Recent sequence comparisons indicated that the predicted v-*sis* transforming product is highly homologous to human platelet-derived growth factor (PDGF) (*10, 11*). It is likely that most of the sequence differences are species specific; however, it is not known whether additional changes could have occurred which give v-*sis* its transforming potential. Although v-*sis* is known to transform mouse fibroblasts in vitro (*8*), addition of PDGF to the cells does not cause morphological transformation (*12*). Thus, it is not known whether transformation is due to intracellular expression of a functionally normal or altered *sis* gene product. Transformed cells of mesenchymal origin that express the *sis* message correlate well with tissues known to be mitogenically stimulated by PDGF, that is, fibroblasts and glial cells (*8, 13*). The restricted expression of *sis* in a large

Table 1. Flanking sequences of v-*sis* homologous regions in cloned λL33. The left-hand column shows the acceptor splice sequences CAG] while the right-hand column lists the donor splice sequences [GT directly adjacent to the regions indicated. Region 6 lacks a 3′ donor sequence.

Acceptor	c-*sis* region	Donor
GTGTCTGCCCGGCAG]	1	[GTAAATGGAATCCCG
TGACTTCTCCTGCAG]	2	[GTAAGACTGAGACAC
TTCATCTCCTCCCAG]	3	[GTGCGTAGGCTCCGG
CCTCTCCAGCTCCAG]	4	[GTAACCACCTTTCCA
CTTTTTCTTGTGCAG]	5	[GTGAGGGCCAGGCGG
CCACCCTTCCATCAG]	6	ACTCCATGGCTAGGA

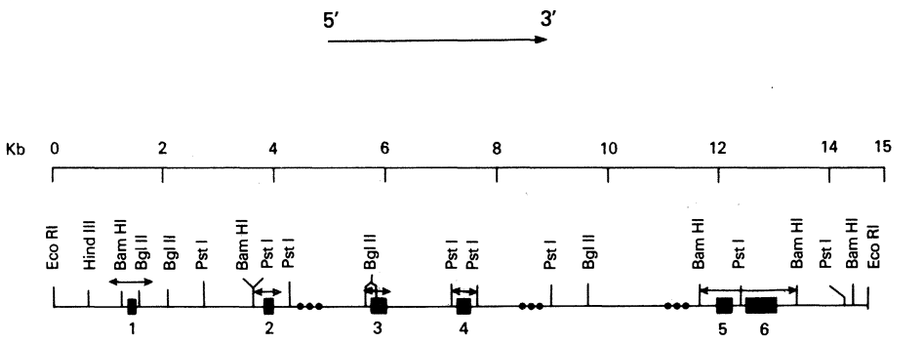

Fig. 1. The organization of clone λL33 showing the regions (black boxes numbered 1 through 6) homologous to v-*sis*. Arrows indicate the regions sequenced. Dots show the approximate locations of the Alu family of repeated sequences (*7*).

number of cell types suggests that the 4.2 kb c-*sis* product may play a role in transformation. This is further suggested by the finding that *sis* is located on a region of human chromosome 22 (*14*) that is translocated in 90 percent of chronic myelogenous leukemias (CML's) (*15*) and in some Ewing sarcomas (*16*).

To further define the relations among c-*sis*, v-*sis*, and PDGF, we studied the nucleotide sequence of the region of c-*sis* homologous to our v-*sis* sequences, the amino acid sequences of the predicted products, and the homology of the predicted c-*sis* product to the published sequences of PDGF.

Our human c-*sis* clone λL33 (Fig. 1) is 14.5 kb in length and contains 1.0 kb of v-*sis* homologous sequences (c-*sis* regions). These are represented as black boxes numbered 1 through 6 which are separated by five nonhomologous stretches of DNA. We previously determined the nucleotide sequence of region 1 (*9*) and have now determined the nucleotide sequences of regions 2 through 6, these regions including their flanking sequences shown below the arrows in Fig. 1. Each region is bordered by acceptor CAG] and donor [GT splice sites and, except for region 6, these regions resemble complete exons (*17*) (Table 1). Region 6 lacks a donor splice site at the 3′ end and resembles only a 5′ portion of an exon. No polyadenylation signal (AATAAA) (*18*) could be found in the 250 bases downstream from this region. These data support a model in which RNA processing occurred as a step during the recombination events that generated SSV (*9*) and suggest that the 3′ recombination event occurred within the woolly monkey exon sequences homologous to the human c-*sis* region 6.

Figure 2 shows the alignment of our v-*sis* nucleotide sequence (top line) and the predicted translation product (second line) to the c-*sis* nucleotide sequence and predicted translation product (third and fourth lines, respectively). Here the homologies in the c-*sis* nucleotide sequence and translation product are indicated by dashes and asterisks, respectively, while the differences from v-*sis* are indicated by substituting the changes. Of the restriction enzyme sites indicated by solid triangles, Bgl II, Pvu II, Sma I, and Bst EII are conserved between v-*sis* and c-*sis*. The v-*sis* sequences at positions 1 through 58 are derived from the helper-viral envelope region (*6*). The c-*sis* sequences are arranged in a spliced fashion and begin −6 base pairs from the 5′ v-*sis* helper viral junction at position 59. Splice points are indicated by the large open triangles. We were able to place the splice point for regions −1 and 1 at position 58 by comparison to the sequence of a human c-*sis* complementary DNA (cDNA) clone (*19*). The splice points for regions 1 and 2 at position 155, regions 2 and 3 at position 245, regions 3 and 4 at position 451, regions 4 and 5 at position 596, and regions 5 and 6 at position 749 were determined by comparison to the v-*sis* sequences. The number of nucleotides in regions 1 to 6 are 97, 90, 206, 145, 153, and 300 base pairs, respectively. We are reluctant to call these regions exons since comparisons have been made to woolly monkey–derived sequences and alternative splice sites may be utilized in the human mRNA.

Our entire v-*sis* sequence derived from clone C60 (*2*) confirms the previously predicted open reading frame reported by Devare *et al.* (*6*). However, a guanine to adenine change at position 118 in our open reading frame results in a codon change from aspartate to asparagine. Several other sequence differences were found in the 3′ noncoding region (not shown).

The spliced c-*sis* sequences are 991 base pairs long and contain an open reading frame with the same phase and terminator codon position as v-*sis*. The terminator codons at position 718 for v-*sis* and c-*sis* are TAA and TAG, respectively. The open reading frame of c-*sis* covers 660 nucleotides and terminates 123 base pairs within c-*sis* region 5. The viral open reading frame begins with an initiator ATG at position 40 located 5′ to the v-*sis* helper viral junction (*6*). The human c-*sis* open reading frame also extends further 5′ to the v-*sis* region of homology based on the larger size of the human message, the presence of splice points at the 5′ border region 1, and lack of a promoter signal (TATAAA) in the 250 base pairs upstream from this region (*9*). This has been confirmed by determinations of the nucleotide sequences of a c-*sis* cDNA clone (*19*).

We found 43 base pair differences between v-*sis* and c-*sis* within the open reading frames, giving a nucleotide sequence homology of 93 percent. The differences occurred mostly at third codon positions and resulted in 18 amino acid changes out of 220 encoded. The predicted products are completely homologous in the center of the open reading frames between the alanine residue at codon position 259 and glutamic acid 535, which suggests a functional conservation within this region. In addition to being 93 percent homologous, the predicted products are hydrophilic and contain a relatively large number of basic residues, findings that are consistent for the *p*I (9.7 to 10.2) of human PDGF (*10, 11, 20*). Slightly greater sequence differences were found in the 3′ untranslated region with 51 mismatches and gaps in 331 nucleotides, accounting for a nucleotide sequence homology of 85 percent. A gap of two nucleotides in v-*sis* was found at position 742 and gaps of 2 and 12

nucleotides were found in c-*sis* at positions 772 and 963, respectively. The sequence within the gap at 963 was bounded by the inverted repeat sequence CA/TG. It is possible that this symmetry was involved in a deletion or insertion event at this position (*21*). The homology between c-*sis* and v-*sis* ends −5 base pairs from the v-*sis* helper viral junction. No short repeat sequence could be found at this position related to the 5′ helper viral junction. Furthermore, no homology to any 3′ helper virus sequence was found. Divergence from the woolly monkey sequences may thus have occurred at this location prior to the recombination event. This may have obscured any sequences that may have participated in strand selection—the proposed mechanism for regenerating the 3′ viral sequences in oncogene-containing retroviruses (*22*). Alternatively, the proposed mechanism could be more random than once thought. Another explanation, although less likely, is that post-recombination divergence could have occurred. The sequence analysis of a woolly monkey *sis* clone may answer these questions.

The homology of the predicted amino acid sequences of our clones to sequences of human PDGF is shown in Fig. 3. The homology to human PDGF begins with serine at position 67 (position 238 in the sequence of Fig. 2) (see Fig. 3). This follows the c-*sis* posttranslational-processing signal arginine-arginine which is a conservative change from lysine-arginine seen in the predicted viral product, suggesting that the upstream c-*sis* sequences encode the amino terminal portion of a PDGF precurser molecule that is processed (*6, 10, 11, 23*). Two groups of investigators have reported PDGF sequences under different terminologies (*10, 11, 24*). We have listed the sequences of Antoniades and Hunkapiller under the original designations 1a, 1b, 2a, and 2b, and we refer to the five peptides partially sequenced by Waterfield *et al.* as peptides I to V. The PDGF peptide I sequence completely agreed with the c-*sis* predicted product in 31 of 31 amino acid residues. Therefore, c-*sis* encodes this portion of the PDGF molecule. The threonine and isoleucine at 72 to 73 of human PDGF, which were found to differ from the v-*sis* predicted product, are in fact coded by the c-*sis* sequence.

Peptides 2a and 2b, although containing the threonine and isoleucine at positions 72 to 73 were reported to contain

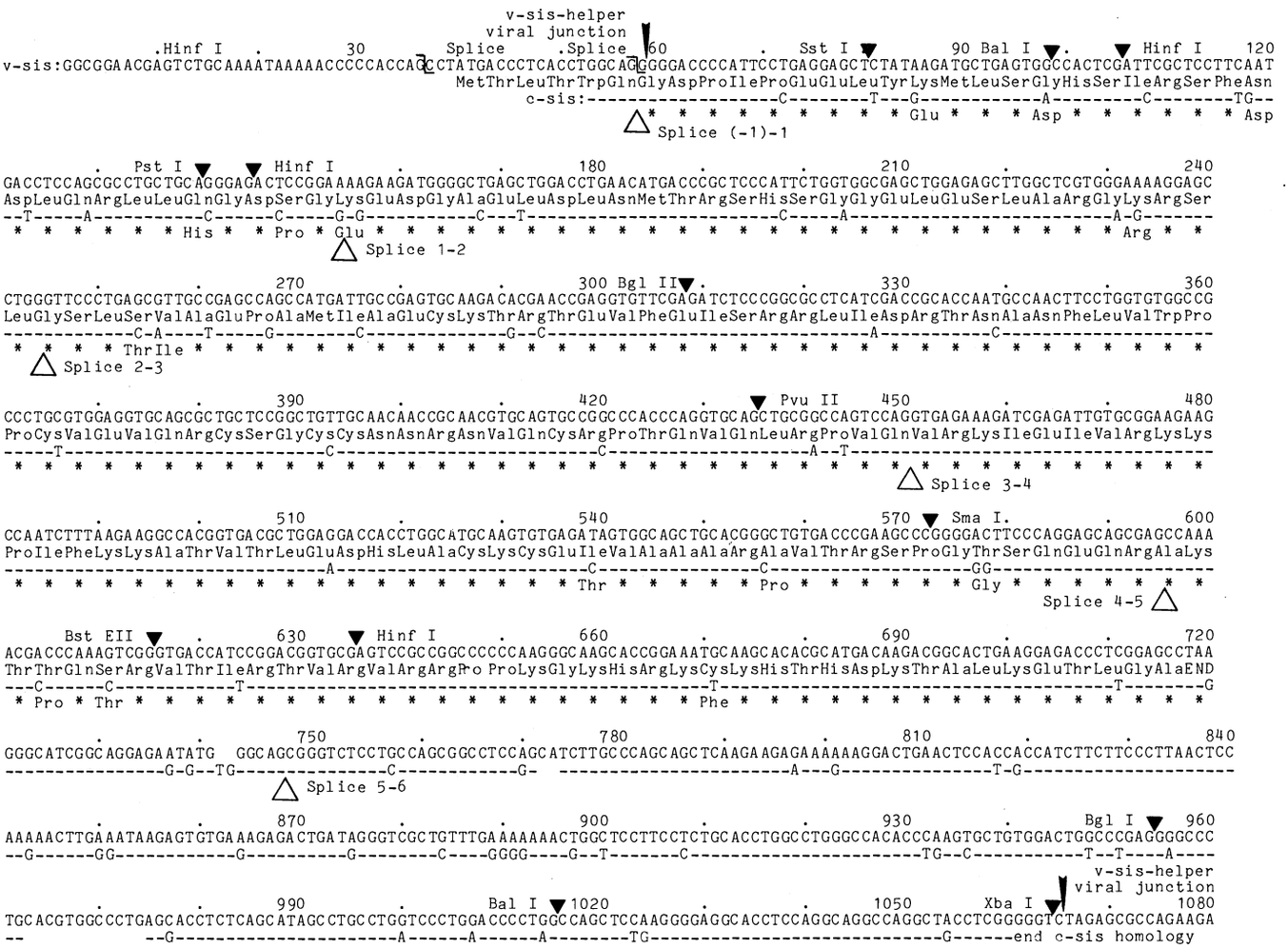

Fig. 2. Nucleotide sequence and translation product comparison of v-*sis* and spliced c-*sis*. The v-*sis* and c-*sis* nucleotide sequences are on lines 1 and 3 while the respective translation products are on lines 2 and 4. Dashes and asterisks show homology of the nucleotide sequences and translation products of c-*sis* to the respective v-*sis* sequences while differences are indicated by substitution. Restriction enzyme sites are indicated by solid triangles. Splice points are indicated by the large open triangles. Sequencing was as previously described by Maxam and Gilbert (*27*) except for the small Pst I fragment containing region 4 which was subcloned into phage M13 and sequenced by the dideoxy chain termination technique as described by Sanger *et al.* (*28*). Abbreviations for the nucleic acid residues are: A, adenine; C, cytosine; G, guanine; and T, thymine. Abbreviations for amino acid residues are: Ala, alanine; Arg, arginine; Asn, asparagine; Asp, aspartic acid; Cys, cysteine; Glu, glutamic acid; Gln, glutamine; Gly, glycine; His, histidine; Ile, isoleucine; Leu, leucine; Lys, lysine; Met, methionine; Phe, phenylalanine; Pro, proline; Ser, serine; Thr, threonine; Trp, tryptophan; Tyr, tyrosine; and Val, valine.

glutamic acid at position 86 instead of threonine and cysteines at 90 and 92 instead of glutamic acid and serine. If real, these differences could reflect a polymorphism in the pooled population from which PDGF was isolated or could indicate that a genetic region other than c-*sis* encodes these peptides. The extent of the differences makes the first explanation unlikely, since one change from glutamic acid to cysteine requires changes in all three base pairs of the codon $GA^{A}_{G} \rightarrow TG^{T}_{C}$ while the others require two or one base pair change. The second explanation is unlikely since the peptides have such high homology to the predicted v-*sis* and c-*sis* products and peptide I. A more likely explanation is that the protein sequence is inaccurate and that peptide I is the correct sequence. We therefore hold that peptides I, 2a, and 2b are encoded by c-*sis*.

No differences are seen in the overlapping sequences of peptides II, 1a, and 1b (*10, 11, 24*). The composite sequence of these three peptides differs from the predicted c-*sis* product in 15 amino acids within the 34 reported amino terminal sequences and are clearly encoded by another genetic region. Peptides II and III are reported to contain serine residues at position 100 while the predicted c-*sis* and v-*sis* products have an asparagine at this position. Peptide III could be derived from peptide II because of the conserved serine residue at position 100. Peptide IV contains threonine instead of asparagine or serine at position 100. This could be due to an error in amino acid sequencing or to polymorphism, or peptide IV could be encoded by a genetic region other than peptide I, 2a, and 2b. After residue 100, the reported sequences of peptides II to IV agree completely in 42 of 42 overlapping residues with the c-*sis* predicted product. A threonine is seen in the predicted c-*sis* product at position 167 rather than isoleucine reported for peptide V. Threonine is also predicted at the equivalent position in our human *sis* cDNA clone (*19*). Therefore, the isoleucine of peptide V could be due to an error in amino acid sequence or to polymorphism, or peptide V could be encoded by some genetic region other than that for peptide I. In all, peptide V agrees with the c-*sis* predicted product in 22 of 23 reported amino acids.

The final processed PDGF has two forms, PDGF I (molecular weight 32,000) and PDGF II (molecular weight 28,000), which have essentially the same amino acid composition but differ in extent of glycosylation (*25*). When either PDGF I or PDGF II is reduced two or more different peptides can be obtained which can be classified into two groups. We define group A to include peptide I and peptides 2a and 2b of Fig. 3 on the basis of the high or complete homology of their 31 amino terminal amino acids to the v-*sis* or c-*sis* predicted products. Group B, which includes peptides II, 1a, 1b, and probably III is approximately 50 percent homologous to this region. Peptides IV and V could belong to either group A or B but may represent products encoded by other members of a family of *sis*-related genes. However, we think this is unlikely because of the extremely high homology of these peptides to the c-*sis* predicted product. The size differences among peptides within each group are thought to be due to variable processing or the effects of degradation (*10, 24*). In any case, at least two different genetic regions are responsible for encoding the group A and group B molecules. Because of the pattern after reduction, PDGF is thought to consist of a cysteine cross-linked complex comprised of at least one peptide from each group (*24*). Group A and B molecules are therefore probably derived from polypeptide chains that we suggest be called chains A and B, respectively. The c-*sis* sequence we report here encodes the sequences of group (chain) A. It is not known whether both groups are contained in a single precurser molecule, are encoded by separate mRNA's, or occur as a result of reinitiation during the translation of a common mRNA. It is interesting that the divergence among the peptides is within or upstream of coding region 3. Thus, it is conceivable that the two groups may occur as a result of alternative splicing of a more divergent upstream coding region to an acceptor splice site at or upstream from region 4.

Studies show that PDGF-like activity can be demonstrated in lysates of SSV-transformed cells and that the SSV-transforming protein product is structurally and immunologically similar to PDGF (*26*). Thus, the biological activity of the v-*sis* product stems from expression of the coding region analogous to the c-*sis* sequence reported here.

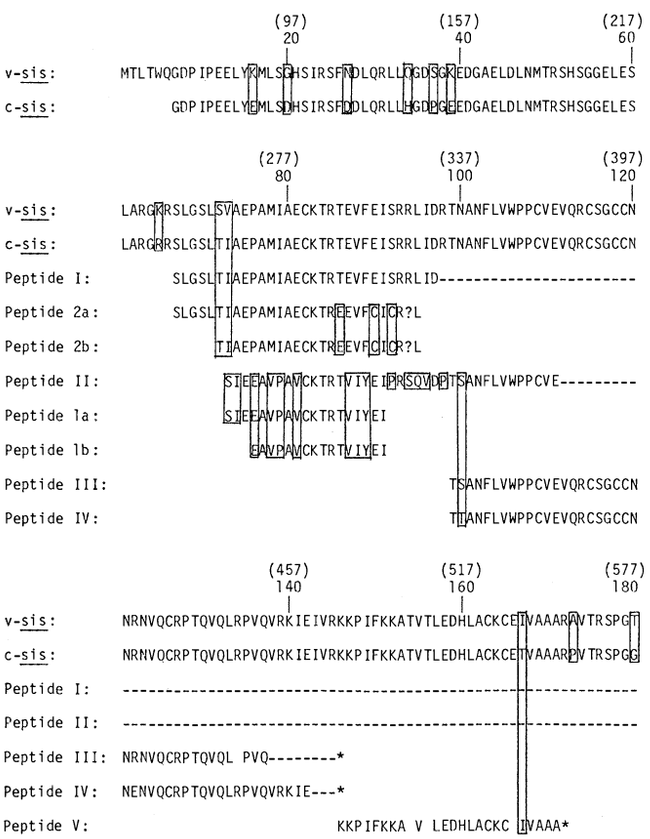

Fig. 3. Comparison of the v-*sis* and c-*sis* predicted translation products showing the region overlapping the reported sequences of PDGF. Numbering starts at the possible viral initiator ATG. The positions in the nucleotide sequences of Fig. 2 are given in parentheses. PDGF peptides I to IV and 1a, 1b, 2a, and 2b were reported previously (*10, 11, 23*). The boxes highlight the differences of the peptides from the predicted v-*sis* and c-*sis* protein products. Dashes indicate that the lengths of the individual peptides extend beyond the regions sequenced as previously described by Waterfield *et al.* (*11*).

STEVEN F. JOSEPHS
CHAN GUO
LEE RATNER
FLOSSIE WONG-STAAL
Laboratory of Tumor Cell Biology, National Cancer Institute, Bethesda, Maryland 20205

References and Notes

1. G. H. Theilen, D. Gould, M. Fowler, D. L. Dungworth, *J. Natl. Cancer Inst.* **47**, 881 (1971).
2. K. C. Robbins, S. G. Devare, S. A. Aaronson, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 2918 (1981); E. P. Gelmann *et al.*, *ibid.*, p. 3373.
3. F. Wong-Staal *et al.*, *Nature (London)* **294**, 273 (1981).
4. K. C. Robbins, R. L. Hill, S. A. Aaronson, *J. Virol.* **41**, 721 (1982).
5. R. C. Gallo and F. Wong-Staal, in *Viral Oncology*, G. Klein, Ed. (Raven, New York, 1980), p. 399.

6. S. G. Devare. E. P. Reddy, J. Doriacan, K. C. Robbins, S. A. Aaronson, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 731 (1983).
7. R. Dalla-Favera, E. P. Gelmann, R. C. Gallo, F. Wong-Staal, *Nature (London)* **292**, 31 (1981).
8. A. Eva *et al.*, *ibid.* **295**, 116 (1982); E. H. Westin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2490 (1982).
9. S. F. Josephs, R. Dalla-Favera, E. P. Gelmann, R. C. Gallo, F. Wong-Staal, *Science* **219**, 503 (1983).
10. R. F. Doolittle *et al.*, *ibid.* **221**, 275 (1983).
11. M. Waterfield *et al.*, *Nature (London)* **304**, 35 (1983).
12. C. D. Stiles, *Cell* **33**, 653 (1983).
13. R. Ross, J. Glomset, B. Kariya, L. Harker, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 1207 (1974); N. Kohler and A. Lipton, *Exp. Cell. Res.* **87**, 297 (1974); B. Westermark and A. Wasteson, *Adv. Metab. Disord.* **8**, 85 (1975).
14. R. Dalla-Favera, R. C. Gallo, A. Giallongo, C. M. Croce, *Science* **218**, 686 (1982); D. C. Swant *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 4691 (1982).
15. J. Groffen *et al.*, *J. Exp. Med.* **158**, 9 (1983).
16. A. Aurias, C. Rimbaut, D. Buffe, J. Dubousset, A. Mazabraud, *N. Engl. J. Med.* **309**, 496 (1983); C. Turc-Carel, I. Philip, M.-P. Berger, T. Philip, G. M. Lenoir, *ibid.*, p. 497.
17. I. Seif, G. Khoury, R. Dhar, *Nucleic Acids Res.* **6**, 3387 (1979); A. Efstratiadis *et al.*, *Cell* **21**, 653 (1980).
18. N. J. Proudfoot and G. G. Brownlee, *Nature (London)* **263**, 211 (1976).
19. M. F. Clarke *et al.*, in preparation; S. F. Josephs *et al.*, in preparation.
20. H. N. Antoniades, C. D. Scher, C. D. Stiles, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1809 (1979).
21. M. P. Calos and J. H. Miller, *Cell* **20**, 579 (1980).
22. J. M. Coffin, *J. Gen. Virol.* **42**, 1 (1979); M. P. Goldfarb and R. A. Weinberg, *J. Virol.* **38**, 136 (1981).
23. C. Lazure, N. G. Seidah, D. Pelaprat, M. Chretien, *Can. J. Biochem. Cell. Biol.* **61**, 501 (1983).
24. H. N. Antoniades and M. W. Hunkapiller, *Science* **220**, 963 (1983).
25. T. F. Deuel *et al.*, *J. Biol. Chem.* **256**, 8896 (1981).
26. T. F. Deuel, J. S. Huang, S. S. Huang, P. Stroobant, M. D. Waterfield, *Science* **221**, 1348 (1983); K. C. Robbins, H. N. Antoniades, S. G. Devare, M. W. Hunkapiller, S. A. Aaronson, *Nature (London)* **305**, 605 (1983).
27. A. M. Maxam and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560 (1977).
28. F. Sanger, S. Nicklen, S. A. Coulson, *ibid.*, p. 5463.
29. We thank R. C. Gallo for continued support, E. Orr of Genetics Institute, Boston, for performing some of the sequencing experiments, and L. S. Lloyd for technical assistance. We also thank A. Mazzuca for secretarial assistance.

29 October 1983; accepted 23 November 1983

Selenium Deficiency in Cattle Associated with Heinz Bodies and Anemia

Abstract. *Cattle grazing St. Augustine grass growing on peaty muck soils in the Florida Everglades developed anemia associated with the presence of Heinz bodies and suboptimal concentrations of selenium in blood. Selenium supplementation corrected the anemia, prevented Heinz body formation, increased the body weight of cows and calves, and elevated blood selenium. This may be the first recorded example of widespread anemia in a population due to selenium deficiency.*

Glutathione peroxidase (E.C.1.11.1.9) protects red blood cells from oxidative damage due to hydrogen peroxide and other peroxides (*1*). The enzyme catalyzes the reduction of hydrogen peroxide to water and a large range of lipid peroxides to hydroxy acids (*2*). Reduced glutathione, the unique hydrogen donor in this process, is regenerated by reduced nicotinamide adenine dinucleotide phosphate and glutathione reductase. Glutathione peroxidase is a selenium-containing enzyme (*3*); in bovines it contains 4 g-atoms of selenium per mole (*4*).

Rotruck *et al.* (*5*) found that dietary selenium in the presence of glucose reduced ascorbic acid–induced hemolysis of rat erythrocytes, oxidation of hemoglobin, and the proportion of cells with Heinz bodies. Dietary α-tocopherol had no effect on hemoglobin oxidation or Heinz body formation.

Some cases of hemolytic anemia in humans appear to be associated with reduced levels of glutathione peroxidase in erythrocytes. Investigators have attributed the reduced enzyme activity to genetic causes (*5a*); however, it is possible that dietary selenium deficiency has a significant role in some instances. We report here an investigation of selenium deficiency–associated anemia in grazing beef cattle.

We studied a herd of Santa Gertrudis cattle grazing St. Augustine grass growing on peaty muck soils in the Everglades region of Florida. When subjected to standard ranch practice [96 cows per 40-ha paddock with access to one water trough and one mineral box containing a mixture of dicalcium phosphate, sodium chloride, vitamin A, and trace elements (*6*)], most cattle thrived, but a few developed a wasting condition and anemia. A multifactorial experimental design revealed that, in the absence of mineral supplementation, the cattle developed severe cobalt and copper deficiencies, anorexia, wasting, and anemia. These conditions were, in the main, reversed by the addition of cobalt and copper to the drinking water. Clearly, provision of a trace element mixture in a single box for a herd of cattle is an inadequate method of supplementation for a significant proportion of animals. While provision of cobalt in the drinking water substantially reduced the incidence of anemic animals to about one quarter of those previously shown to be anemic, a residual number with low hemoglobin persisted, particularly through the summer.

In investigating this residual anemia, we established that the grass had a very low selenium content (< 0.02 to 0.055 ppm dry matter, $N = 20$; detection limit, 0.02 ppm). Furthermore, examination of blood smears revealed a high incidence of Heinz bodies in anemic animals. Because of the previous rat studies (*5*) showing that oxidative damage of red cells could be prevented by selenium, the following field trial was undertaken.

Santa Gertrudis heifers ($N = 144$) pregnant with their first calves were selected for uniformity and randomly divided into four groups (A to D) of 36 animals each. Each group was randomly assigned to one of four 22.7-ha pastures, each with a separate water tank of approximately 10,000 liters, the sole source of free water available. Beginning on 11 July 1980 a concentrated solution of sodium selenate was added to the water of groups B, C, and D once weekly such

Table 1. Effect of dietary selenium on hematological parameters and concentration of selenium in whole blood of Santa Gertrudis heifers. Values are means ± standard errors. For each sampling, means in the same column with different superscripts are significantly different ($P < 0.01$, analysis of variance).

Selenium intake (mg/day)	Selenium in blood (μg/dl)	Packed cell volume (percent)	Hemoglobin (grams per 100 ml)	Number of cows with Heinz bodies			
				0	+1	+2	+3
		92 days after treatment					
0.0*	0.38 ± 0.02	30.7 ± 0.56	10.5 ± 0.23	5	13	6	11
0.5	0.75 ± 0.03[a]	40.5 ± 0.43[a]	14.0 ± 0.90[a]	36	0	0	0
1.0	1.1 ± 0.03[b]	40.9 ± 0.38[a]	14.2 ± 0.10[a]	36	0	0	0
2.0	2.3 ± 0.10[c]	40.5 ± 0.44[a]	14.2 ± 0.12[a]	36	0	0	0
		1 year after treatment					
0.0	0.8 ± 0.1	32.9 ± 0.81	12.0 ± 0.26[b]	0	7	7	22
1.0	5.4 ± 0.2[a]	38.7 ± 0.43[a]	12.2 ± 0.20[b]	31	4	0	0
2.0	6.2 ± 0.3[a]	39.6 ± 0.53[a]	14.2 ± 0.17[c]	35	1	0	0
0.5 to > 4.0	5.0 ± 0.3[a]	39.8 ± 0.50[a]	13.5 ± 0.23[c]	35	1	0	0

*$N = 35$. One cow in this group died of undetermined causes before the 92 days after treatment sampling. A replacement cow of similar age was substituted to restore the number to 36.