

## The *c-sis* gene encodes a precursor of the B chain of platelet-derived growth factor

Ann Johnsson, Carl-Henrik Heldin, Åke Wasteson, Bengt Westermark<sup>1</sup>, Thomas F. Deuel<sup>2</sup>, Jung San Huang<sup>2</sup>, Peter H. Seeburg<sup>3</sup>, Alane Gray<sup>3</sup>, Axel Ullrich<sup>3</sup>, Geoffrey Scrase<sup>4</sup>, Paul Stroobant<sup>4</sup> and Michael D. Waterfield<sup>4\*</sup>

Institute of Medical and Physiological Chemistry, and <sup>1</sup>Department of Pathology, University of Uppsala, Uppsala, Sweden, <sup>2</sup>Departments of Medicine and Biological Chemistry, Washington University School of Medicine, The Jewish Hospital of St. Louis, MO 63110, USA, <sup>3</sup>Genentech Incorporated, 460 Point San Bruno Boulevard, San Francisco, CA, USA, and <sup>4</sup>Imperial Cancer Research Fund, Lincoln's Inn Fields, London, WC2A 3PX, UK

\*To whom reprint requests should be sent  
Communicated by M.Fried

**The relationship between platelet-derived growth factor (PDGF) and the proto-oncogene *c-sis* has been determined by amino acid sequence analysis of PDGF and nucleotide sequence analysis of *c-sis* genomic clones. The nucleotide sequences of five regions of the human *c-sis* gene which are homologous to sequences of the transforming region (*v-sis*) of simian sarcoma virus (SSV) were determined. By alignment of the *c-sis* and *v-sis* nucleotide sequences the predicted amino acid sequence of a polypeptide homologous to the putative transforming protein p28<sup>sis</sup> of SSV was deduced. Both predicted sequences use the same termination codon and additional coding sequences may lie 5' to the homologous regions. Amino acid sequence analysis of the PDGF B chain shows identity to the amino acid sequence predicted from the *c-sis* sequences over 109 amino acid residues. Polymorphism may exist at two amino acid residues. These results suggest that *c-sis* encodes a polypeptide precursor of the B chain. A partial amino acid sequence of the PDGF A chain is also described. This chain is 60% homologous to the B chain and cannot be encoded by that part of *c-sis* which has been sequenced but could be encoded by sequences which lie 5' to the five regions of *v-sis* homology in *c-sis*, or at a separate locus.**

**Key words:** PDGF structure/nucleotide analysis/amino acid sequence analysis/simian sarcoma virus oncogene

### Introduction

Platelet-derived growth factor (PDGF) is the major mitogen in serum for connective tissue derived cells in culture (for reviews see Westermark *et al.*, 1983; Deuel and Huang, 1983). The protein has an apparent mol. wt. of ~30 000 and perhaps contains two types of polypeptide chains linked by disulphide bonds (Deuel *et al.*, 1981a; Johnsson *et al.*, 1982). Sequence analysis of PDGF has revealed a striking homology between the sequence of one polypeptide chain of PDGF and the predicted sequence of p28<sup>sis</sup>, the putative transforming protein of simian sarcoma virus (SSV) (Waterfield *et al.*, 1983a; Doolittle *et al.*, 1983; Devare *et al.*, 1983). This structural similarity suggests that cells transformed by SSV may produce a growth factor resembling PDGF not only in structure but also in function. SSV non-producer cells synthesise a

polypeptide having antigenic sites recognised by antiserum to human PDGF (Deuel *et al.*, 1983; Robbins *et al.*, 1983). Extracts of SSV-transformed cells also contain a mitogen whose activity is neutralised by a PDGF antiserum (Deuel *et al.*, 1983). In SSV-transformed cells p28<sup>sis</sup> is proteolytically processed to generate a disulphide linked dimer (Robbins *et al.*, 1983). Growth factors having structural and functional similarities to PDGF have been detected in the media of human osteosarcoma and glioma derived cell lines (Westermark and Wasteson, 1975; Heldin *et al.*, 1980; Betsholz *et al.*, 1983; Nister *et al.*, 1984) and an SV40-transformed BHK cell line (Burk, 1976; Bourne and Rozengurt, 1976; Dicker *et al.*, 1981).

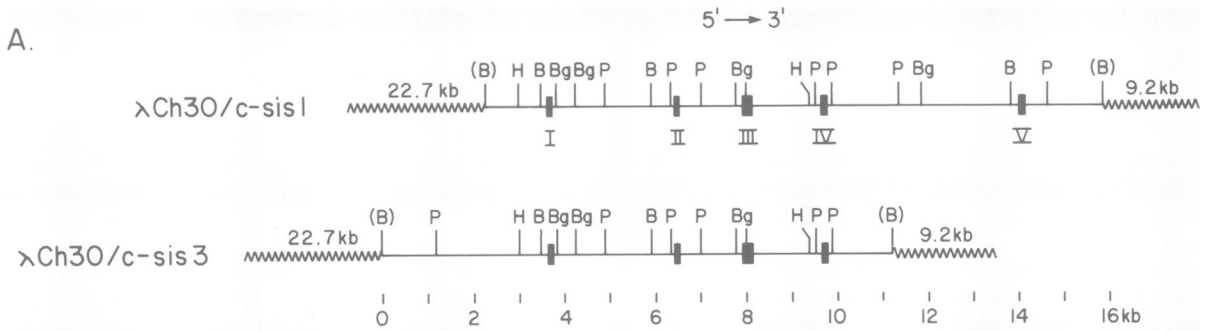
Through the use of *v-sis* probes, transcripts of human *c-sis* have been detected in a number of cell lines derived from glioblastomas and sarcomas, but not in normal fibroblasts or in melanoma and carcinoma cell lines (Eva *et al.*, 1982). This cellular specificity corresponds to the known responsiveness of mesenchymal cells and glial cell lines (but not epithelial cells) to the mitogen PDGF through expression of specific receptors (for a recent review see Westermark *et al.*, 1983; Deuel and Huang, 1983). These results suggest that activation of the gene(s) encoding PDGF, or a closely related molecule, may play a role in the control of the proliferation of certain cells.

The human proto-oncogene *c-sis* has been mapped to a locus on chromosome 22 (Dalla Favera *et al.*, 1982; Swan *et al.*, 1982) and a genomic clone which contains five regions of sequence that hybridise to *v-sis* has been characterised and partially sequenced (Wong-Staal *et al.*, 1981; Dalla Favera *et al.*, 1981; Josephs *et al.*, 1983). Here we report the isolation and partial sequence of related genomic clones which contain part of the putative coding region of *c-sis*. The predicted sequence of the *c-sis* protein product is compared with the amino acid sequence of the two polypeptide chains of human PDGF. It is shown that the PDGF B chain is identical over the 109 residues sequenced to a region of the predicted *c-sis* protein product, thus establishing that *c-sis* encodes a precursor of this chain of human PDGF. A partial sequence of the PDGF A chain is also presented. The A chain is ~60% homologous to the B chain and cannot be encoded by the region of *c-sis* which has been sequenced.

### Results and Discussion

#### *Isolation and nucleotide sequence of c-sis clones*

A 72 nucleotide long, double-stranded probe, which was synthesized on the basis of the partial *c-sis* sequence by Josephs *et al.* (1983), was used to screen a partial *Sau3A* human genomic library (provided by W.I. Wood). Two recombinant lambda clones were isolated which contained overlapping human genomic sequences. The restriction map of these clones is shown in Figure 1A. The DNA sequence of restriction fragments from regions shown to be homologous to *v-sis* nucleotide sequences by Josephs *et al.* (1983) were sequenced by sub-cloning into M13 and then sequenced by the method



of Sanger *et al.* (1977). The nucleotide sequence of five of the six regions of homology between *c-sis* and *v-sis* established by Dalla Favera *et al.* (1981) is shown in Figure 1B together with the nucleotide sequence of *v-sis* and the amino acid sequences of the predicted polypeptides.

Homology regions II, III and IV are each delimited by splice donor and acceptor sites located by comparison of the *c-sis* and *v-sis* sequences, while the 5th region of homology lacks a donor splice site at its 5' end. As previously reported by Devare *et al.* (1983) there is no candidate initiator codon, promoter and associated sequences or signal sequence upstream from the end of the first region of *c-sis* homologous to *v-sis* (I, see Figure 1), and the donor splice site at a position equivalent to the codon for amino acid -60 in Figure 1 may possibly be that used to delimit homology region I. The open reading frame of *c-sis* predicts an amino acid sequence homologous to that of the predicted *v-sis* protein product and is terminated within homology region V at a TAG codon exactly homologous to the TAA codon in the *v-sis* sequence. The amino acid sequence of the predicted *c-sis* polypeptide differs at 11 positions from the predicted *v-sis* polypeptide previously published. To establish the full extent of similarity of the predicted and observed protein sequences further analyses of PDGF protein structure were undertaken.

#### Analysis of component polypeptides of PDGF

Two different approaches have been used to separate and analyse the amino acid sequence of the component chains of PDGF – preparative SDS polyacrylamide gel electrophoresis (Antoniades and Hunkapiller, 1983; Doolittle *et al.*, 1983) and h.p.l.c. techniques (Johnsson *et al.*, 1982; Waterfield *et al.*, 1983a). In these studies sequence analysis was used to assign the polypeptides to two groups which may be derived from two putative precursor chains, A and B. Due to incomplete purification and lack of overlapping sequences the precise sequence of the putative A and B chain precursors has not been conclusively established (Antoniades and Hunkapiller, 1983; Doolittle *et al.*, 1983; Waterfield *et al.*, 1983a).

The PDGF A chain has been defined as that eluting early from reverse phase h.p.l.c. developed with a gradient of propanol, and the B chain as that eluting late from these columns (Johnsson *et al.*, 1982). The A and B chains have the amino-terminal sequences 'NH<sub>2</sub>-SIEE-' and 'NH<sub>2</sub>-SLGS-', respectively, (Waterfield *et al.*, 1983a). The B chain shows extensive identity to a region of the sequence of the putative transforming protein of simian sarcoma virus (Waterfield *et al.*, 1983a; Doolittle *et al.*, 1983). [This chain was called I by Waterfield *et al.*, and 1a (and b) by Doolittle *et al.*, 1983]. To establish whether all the component polypeptides found in PDGF result from cleavage of A and B chains we have carried out further structural analysis of PDGF.

PDGF from platelet concentrates purified by the method of Johnsson *et al.* (1982) yielded material with an apparent mol. wt. of 26 000–30 000 on SDS polyacrylamide gels. Following reduction and alkylation of cysteine residues with

[<sup>14</sup>C]-iodoacetamide, peptides were fractionated by reverse phase h.p.l.c. in 2 M guanidine HCl, acetic acid using a gradient of *n*-propanol (Johnsson *et al.*, 1982; Waterfield *et al.*, 1983a).

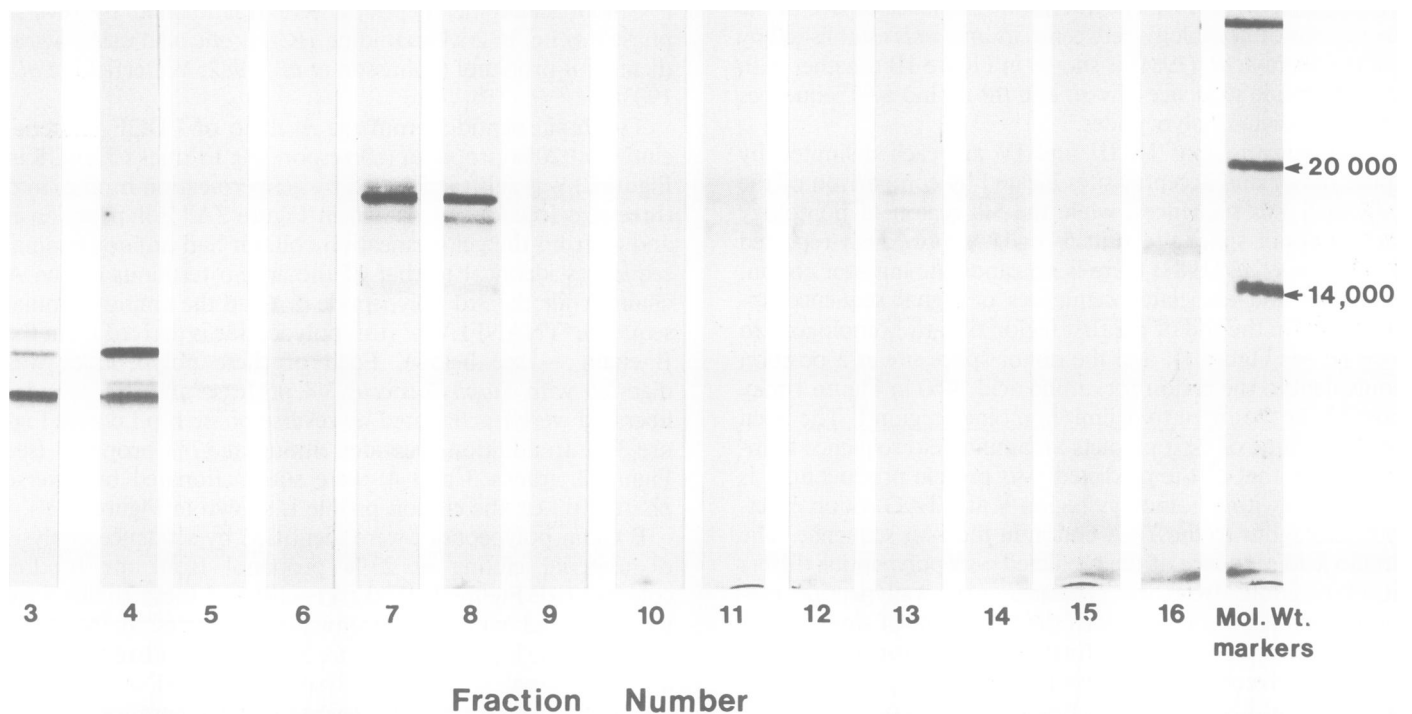
To obtain peptides from the A chain of PDGF, material eluting at 20% propanol (corresponding to tracks 7 and 8 in Figure 2) was subfractionated by gel permeation h.p.l.c. into three size fractions as shown in Figure 3A. Polypeptides a<sub>1</sub> and a<sub>2</sub> from the gel permeation column had amino-terminal sequences identical to that of the amino terminus of the A chain, while the 3rd polypeptide (a<sub>3</sub>) had the amino-terminal sequence 'TNANFLV---' (this polypeptide is derived from the B chain – see below). Each of these polypeptides was digested with *Staphylococcus* V8 protease and the peptides liberated were fractionated by reverse phase h.p.l.c. (see Figure 3B). In addition, peptides eluting at 17% propanol (see Figure 2, tracks 3 and 4) were subfractionated by reverse phase h.p.l.c.; the elution profile is shown in Figure 4.

B chain polypeptides were identified by sequence analysis of material eluting at 23% propanol from the h.p.l.c. columns (see Figure 2, tracks 11–14) and these chains were then digested with V8 enzyme and peptides analysed by reverse phase h.p.l.c. (see Figure 5). Amino acid analysis and amino-terminal sequence analysis was carried out on each peptide recovered from the reverse phase separations.

#### The amino acid sequence of the B chain in PDGF

The B chain peptide fragments analysed (see Figures 2 and 5) could be aligned to form a sequence of 109 amino acids which was identical to that predicted by the nucleotide sequence analysis of *c-sis* (see Figure 6). This sequence data differs at two positions from that previously published (Waterfield *et al.*, 1983a). In the present study residues 34 and 101 were found to be asparagine and threonine respectively whereas we had previously assigned serine or threonine at 34 and isoleucine at 101. Inspection of original sequence data confirms the previous assignments and we conclude that polymorphic variant genes exist in the population. The carboxyl terminus of the B chain studied here was assigned as threonine at residue 109 on the basis of the composition of peptide d2 which could not have been derived from cleavage of the B chain by V8 protease. Presumably cleavage by a trypsin like enzyme had occurred C-terminal to residue 110 and the C-terminal arginine was removed by a carboxypeptidase. The amino terminus is presumably formed by cleavage at the dibasic sequence located immediately amino-terminal to the processed B chain amino terminus. No peptides corresponding to sequences from the polypeptide predicted from the *c-sis* nucleotide sequence, located N- or C-terminal to the 109 residue amino acid sequence have been found; this suggests that the entire PDGF B chain has been sequenced. It corresponds to a mol. wt. of ~12 000, in apparent contrast to the mol. wt. estimates from SDS-gel electrophoresis (14 500–16 000) (Johnsson *et al.*, 1982; Waterfield *et al.*, 1983a). This discrepancy may arise from the anomalous behaviour during SDS-gel electrophoresis.

**Fig. 1A.** Restriction maps of two *c-sis* λCh 30 clones. Sites of cleavage by restriction endonucleases (B = *Bam*HI; BG = *Bgl*II; H = *Hind*III; P = *Pst*I; (B) indicates positions where partial *Sau*3A digested DNA fragments were inserted into the *Bam*HI site of lambda CH30) are shown together with the location of 5 regions (shown I–V) of homology between *c-sis* and *v-sis* (Josephs *et al.*, 1983) nucleotide sequences. **1B.** Comparison of the nucleotide sequences of five regions of homology between *c-sis* and *v-sis* nucleotide sequences and the deduced amino acid sequences of the putative polypeptides encoded. Differences between *v-sis* (Josephs *et al.*, 1983) and *c-sis* nucleotide sequences are indicated by the substituted base shown below the *c-sis* sequence. Differences in the predicted polypeptide sequences are shown as stippled areas with the substituted amino acid in p28<sup>SIS</sup> (the putative *v-sis* transforming protein) shown below the predicted *c-sis* polypeptide sequence. Amino acid numbers are given using 1 as the amino terminus of the B chain polypeptide of PDGF. The carboxyl terminus of the B chain is residue 109 (see text and Figure 6).



**Fig. 2.** Purification of polypeptides from PDGF. SDS polyacrylamide gel analysis of fully reduced and alkylated polypeptides separated by reverse phase h.p.l.c. Aliquots of different fractions from the reverse phase h.p.l.c. (Johnsson *et al.*, 1982; Waterfield *et al.*, 1983a) were subjected to SDS-gel electrophoresis (Blobel and Dobberstein, 1975) using gradient gels consisting of 13–18% polyacrylamide. Protein was visualised by silver staining (Morrissey, 1981).

The sequence of the B chain derived here differs from that of Antoniadis and Hunkapiller (1983) and Doolittle *et al.* (1983) at three positions. We consider our sequence authentic since it was obtained from several overlapping fragments and is identical to the predicted amino acid sequence of the *c-sis* locus.

Analysis of other peptides showed that cleavage of the B chain occurs at residues 32 and 79 which are homologous to positions 98 and 145 of p28<sup>sis</sup> (Waterfield *et al.*, 1983a) creating a B chain with three peptides (of 32, 47 and 30 residues) linked by disulphide bonds. At present it is not possible to evaluate whether these cleavages have any biological significance; it is most likely that they have occurred during the purification of PDGF.

#### *Partial amino acid sequence of the A chain of PDGF*

Fragments of the A chain of PDGF yielded peptide sequences which could be aligned with the B chain as shown in Figure 6A. The following peptides were studied. Peptides b4 and b7 from the amino terminus of the A chain were formed by cleavage following glutamic acid residues 4 and 39. Peptide b8 was a fragment of peptides b4 or b7 formed by cleavage at glutamic acid residues 4 and 18 of the A chain. The amino acid sequence of this region has previously been determined (Peptide II, Waterfield *et al.*, 1983a) and compositional data obtained here (data not shown) together with additional sequence data confirm the original sequence assignments. Peptides b2, b3 and b6 were only partially sequenced but could be located C-terminal to residue 39 of the A chain by homology with the predicted p28<sup>sis</sup> amino acid sequence. Peptide b1 was fully sequenced and located by homology to p28<sup>sis</sup> at residues 71–80. This assignment assumes that there are no deletions or insertions in the unsequenced region of the A chain between residues 58 and 70. Further sequence was obtained by

analysis of peptide c3 (see Figure 4) purified from the early eluting fraction of the original h.p.l.c. fractionation in 2 M guanidine, 1 M acetic acid (fractions 3 and 4, see Figure 2) of the fully reduced and alkylated PDGF chains. This partial sequence overlapped that of peptide b1 and extended the A chain sequence to residue 91. The amino acid composition of peptide c3 suggests that the carboxyl terminus of the A chain extends ~13 residues from the region sequenced giving a total length of 104 residues.

Among the 75 amino acids that have been determined in the A chain sequence, 60% are homologous to the B chain. A possible glycosylation site Asn-Thr-Ser exists at residue 48 of the A chain. Since carbohydrate analysis has demonstrated that PDGF contains carbohydrate (Deuel *et al.*, 1981b), this site may indeed be glycosylated. The data presented here predicts that the A chain has 104 amino acids. A peptide of this length would have a mol. wt. of 11 000–12 000, whereas the mol. wt. of the A chain has been estimated to be 18 000 by SDS-gel electrophoresis (Johnsson *et al.*, 1982). This discrepancy may be due to glycosylation. The present A chain sequence differs from that of Doolittle *et al.* (1983) at a number of positions. The cyanogen bromide cleaved polypeptide used in the latter study may possibly have been derived from a mixture of A and B chains. It is also possible that polymorphism may contribute to the differences, since an analogous situation prevails in the B chain.

#### *Analysis of two forms of PDGF – PDGF I and PDGF II*

Previous studies have shown that human PDGF purified from outdated platelets can be separated into two components PDGF I and PDGF II, by gel permeation chromatography (Deuel *et al.*, 1981b; Antoniadis, 1981). These two components have apparent mol. wts. of 30 000–32 000 and 28 000–29 000 on SDS polyacrylamide gels. Sequence

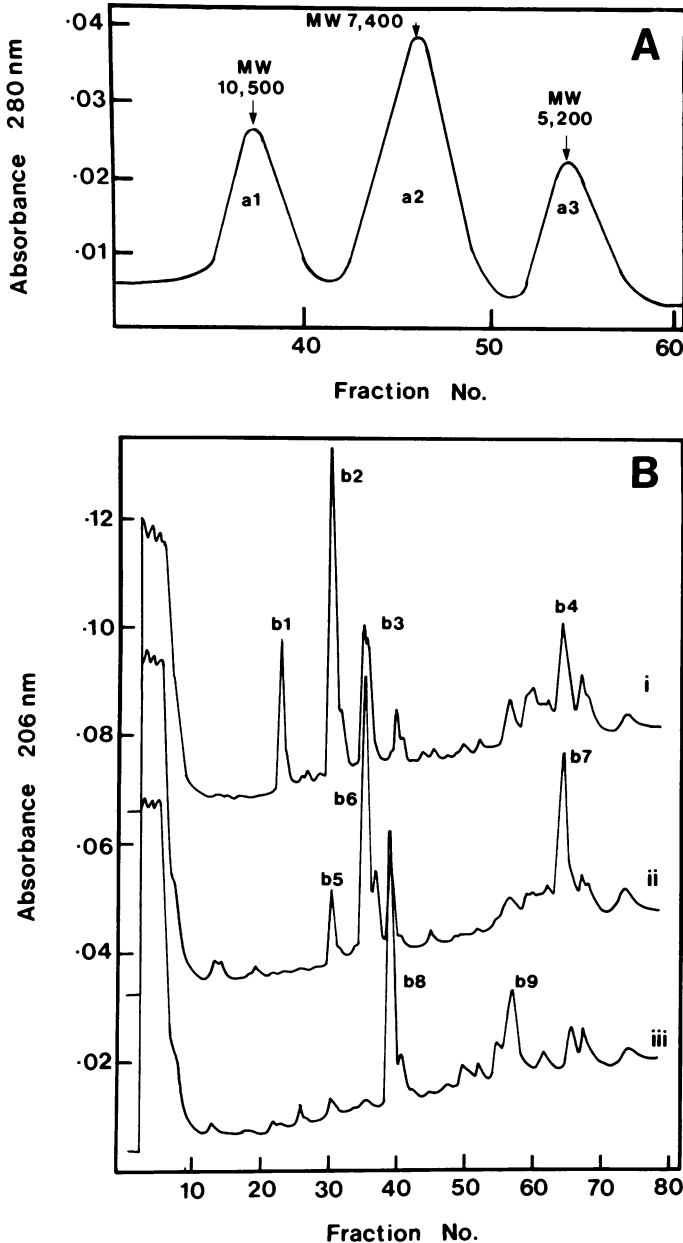


Fig. 3. (A) Gel permeation h.p.l.c. separation in 6 M guanidine of polypeptides derived from materials eluted in fraction 7 and 8 of the separation shown in Figure 2. The absorbance of the eluate at 280 nm is shown. The apparent mol. wts., for the eluted peptides  $a_1$ ,  $a_2$  and  $a_3$  calculated using standard polypeptides of known mol. wts. (Ui, 1979) are indicated by arrows. Polypeptides  $a_1$ ,  $a_2$  and  $a_3$  were subjected to sequence analysis and then further digested with V8 protease. Peptides liberated were separated by reverse phase h.p.l.c. using 0.1% trifluoroacetic acid and acetonitrile gradients. (B) The absorbance profiles at 206 nm of digests of polypeptides  $a_1$ ,  $a_2$  and  $a_3$  (from A above) are shown as i, ii and iii, respectively. Peptide numbers ( $b_1$ – $b_9$ ) refer to peptides located in PDGF by analysis of their amino acid sequence or of their amino acid composition (see text and Figure 6).

analysis of PDGF I and II was carried out to find out if these two PDGF fractions contained unique polypeptides, and this showed that both fractions contained the same component polypeptides having at least 8 distinctive amino terminal sequences (see Table I). The analysis of the relative amounts of these sequences was made using recoveries of amino acids present at unique positions within the first five residues. The results showed that PDGF I, as well as PDGF II, contains

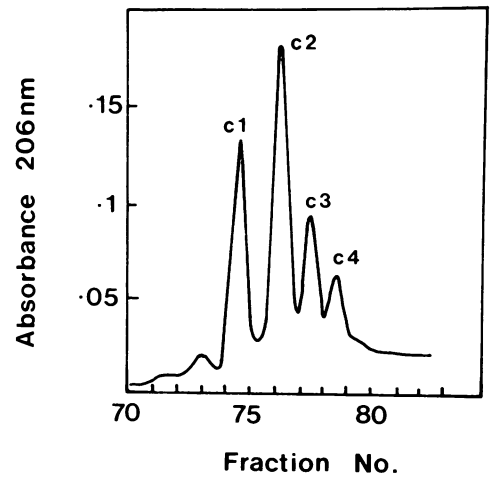


Fig. 4. Reverse phase h.p.l.c. separation of peptides from fractions 3 and 4 of material purified as shown in Figure 2 was made using 0.1% trifluoroacetic acid and acetonitrile gradients. The absorbance of the eluate at 206 nm is shown.

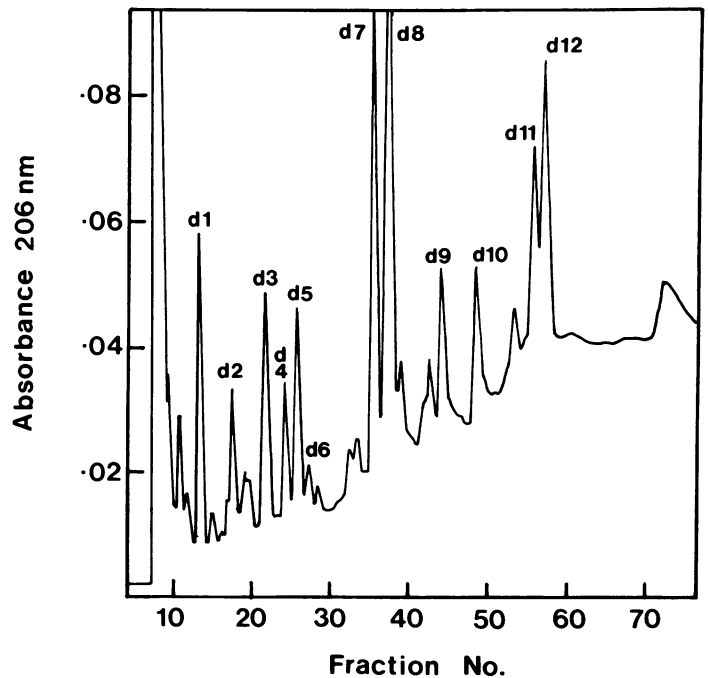
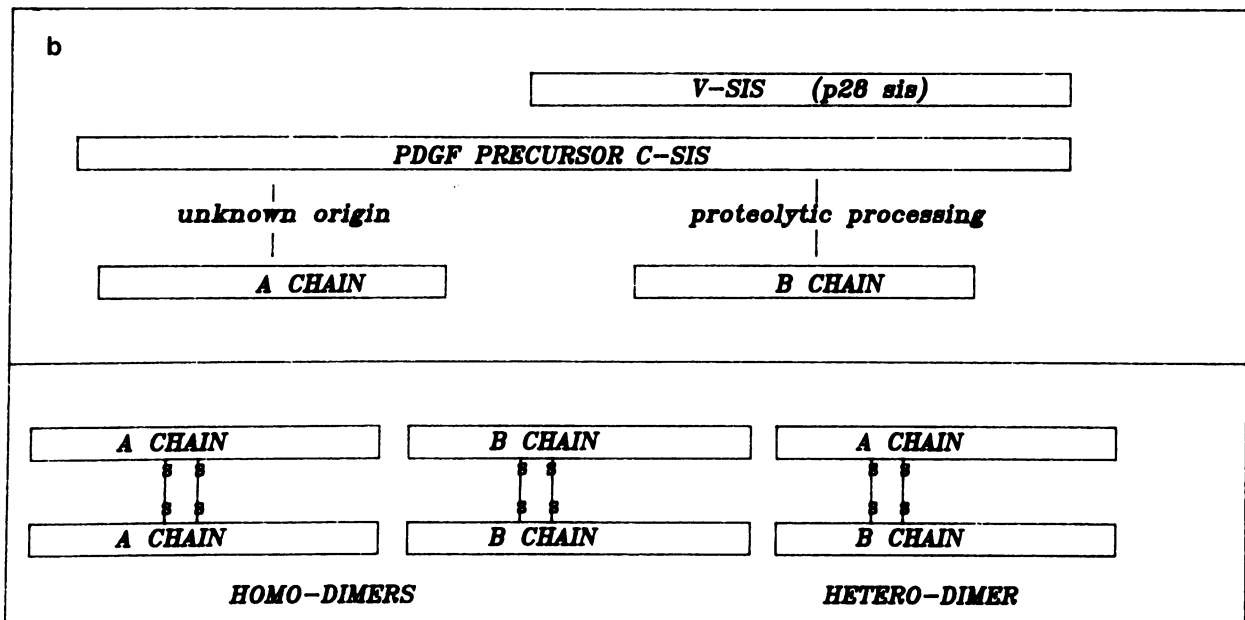
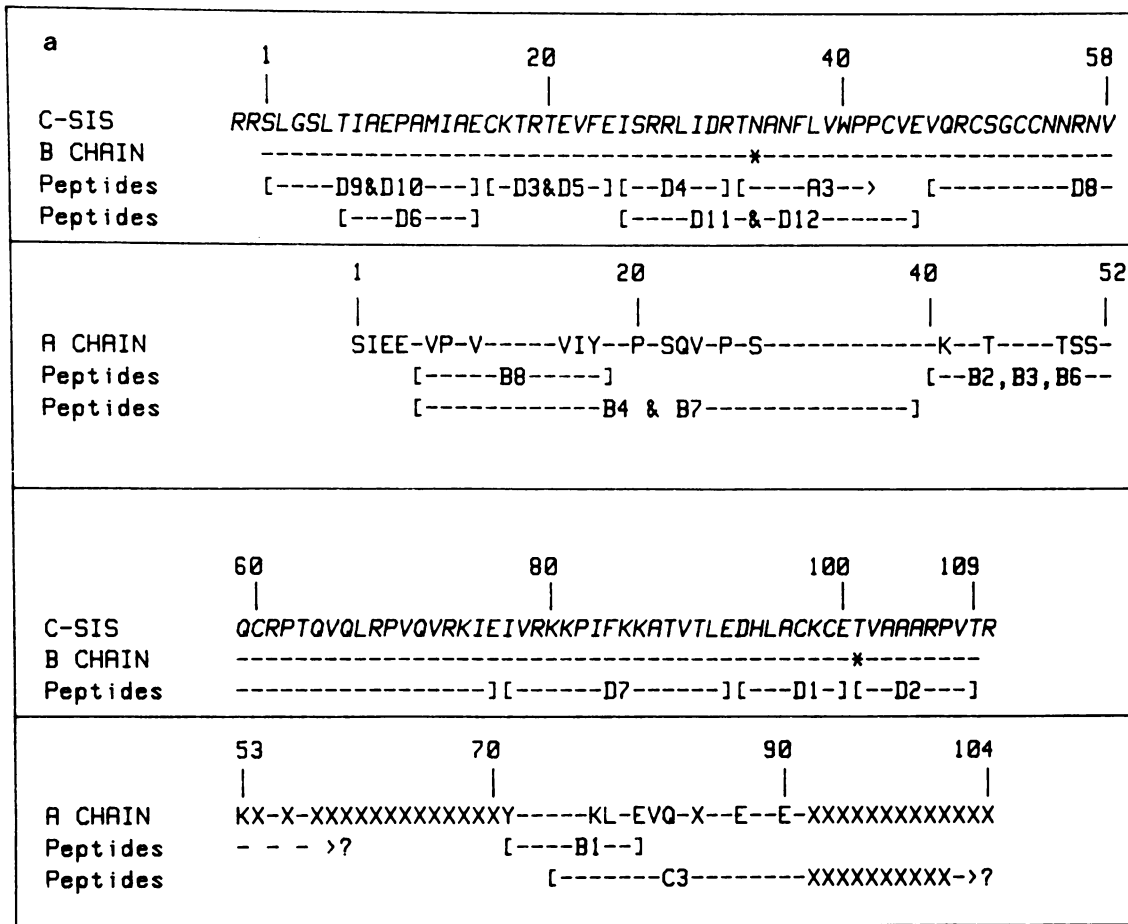


Fig. 5. Reverse phase h.p.l.c. separation of peptides from V8 protease digestion of a pool of material in fractions 11–14 of the chain separation shown in Figure 2 carried out using 0.1% trifluoroacetic acid and acetonitrile gradients. The absorbance at 206 nm is shown.

approximately equal proportions of sequences derived from PDGF A and B chains. This indicates that PDGF I and II are two forms of the same molecule. The mol. wt. difference may then be due to proteolysis, preferentially in the C-terminus, or differences in glycosylation.

#### The relationship between PDGF and the *c-sis* locus

The sequence studies described here extend the original observations which suggested that one polypeptide chain of PDGF may be encoded by the *c-sis* locus (Waterfield *et al.*, 1983a; Doolittle *et al.*, 1983). The amino acid sequence analysis of PDGF presented here supports the concept that PDGF contains two different prototype polypeptide chains. The B chain sequence has been shown to cover a stretch of 109 amino



**Fig. 6.** The amino acid sequences of the putative A and B chains of PDGF. (a) The amino acid sequence of the proteins predicted from the c-sis nucleotide sequences is shown (see Figure 1) for the region homologous to the sequences of the A and B chains (residues 66–175 of p28<sup>sis</sup>). The sequences of the A and B chains deduced from structural studies are shown together with the location of the peptides used to obtain these sequences. The peptide numbering refers to h.p.l.c. fractionated V8 peptides derived from polypeptides a<sub>1</sub>, a<sub>2</sub> and a<sub>3</sub> from fractions 7 and 8 in Figure 2 (see Figure 3), from subfractionation of peptides in fractions 3 and 4 from Figure 2 (see Figure 4) and from fractionated V8 cleavage products of material in fractions 11–14 from Figure 2 (see Figure 5). – indicates identity to the amino acid predicted from the c-sis gene, \* that polymorphic variants exist, and X unknown amino acid. (b) Two alternative models to account for the structure of PDGF. It is proposed that two chains, A and B, B being a cleavage product of a c-sis encoded precursor polypeptide, can form the PDGF protein. Three alternative dimeric species could be formed AA, BB and AB.

**Table I.** The relative proportions of amino terminal sequences derived from polypeptides of the A and B type in preparations of PDGF I and PDGF II.

Chain type	Amino acid sequence residue number	Relative contribution of polypeptides (%)	
		PDGF I	PDGF II
A	1 2 3 4 5 S I E E A	27	27
	A/B } T S A N F T T A N F T N A N F	17	9
		A	20
B	K K P K L	4	8
	S L G S L	17	21
B	S V A E P	5	6
B	K K P I F	6	2
B	K P I F K	2	—

acids which is identical to part of the predicted amino acid sequence of c-sis (see Figures 1B and 6). Results show that two polymorphic residues may exist (at residues 34 and 101).

The PDGF A chain cannot be encoded by the sequences of c-sis established here. v-sis hybridizes to a large transcript (~4 kb in size) in certain human tumor cell lines (Eva *et al.*, 1982). Assuming that this transcript is representative of that of the corresponding gene in normal cells, the coding sequences of the A chain may be located 5' to those of the B chain, and transcribed to a common mRNA. A single chain PDGF precursor may be processed to a two chain heterodimer structure. Alternatively, the A chain may be encoded at a separate locus, and the A and B chains translated from separate transcripts. PDGF would then be assembled as A-A and B-B homodimers, or A-B heterodimers. PDGF may be composed of one A chain and one B chain (see above), but SSV has acquired sequences corresponding only to the B chain. In such a case the acquisition of the gene for only one chain of PDGF by the virus, may have been accompanied by a mutation introducing an extra cysteine residue (residue 145, see Figure 1B). The resulting odd number of cysteine residues in p28<sup>sis</sup> may then contribute to the dimerisation of the molecule that occurs in SSV-transformed cells (Robbins *et al.*, 1983). Further processing of the dimer by N-terminal and C-terminal cleavage yields p24<sup>sis</sup> (Robbins *et al.*, 1983). This product resembles a PDGF B-B homodimer, which may have PDGF agonist activity.

## Materials and methods

### Isolation and characterization of recombinant lambda clones

A human fibroblast DNA partial Sau3A Library in lambda Charon 30 (kindly provided by W.I.Wood) was screened following standard procedures (Maniatis *et al.*, 1982). The probe used consisted of two chemically synthesised 40 residue long oligonucleotides (5'-CCATTCTGAGGAGCTC-TATAAGATGCTGAGTGGCCACT-3', 5'-CAGCAGGCGTTGGAGATCATCAAAGGAGCGGATCGAGTGG-3') which were complementary over a five base stretch at their 3' ends. *Escherichia coli* polymerase I (Klenow fragment), was used under standard conditions to fill in single-stranded regions with alpha [<sup>32</sup>P]-deoxynucleotides resulting in a 72-bp long double-stranded radioactive probe. The duplex was melted at 100°C (3 mins) and hybridized to nitrocellulose filters (Maniatis *et al.*, 1982) as described (Ullrich *et al.*, 1984). Hybridizing lambda clones were plaque purified, grown and characterized using standard methods (Maniatis *et al.*, 1982; Sanger *et al.*, 1977).

### Purification and structural analysis of PDGF

PDGF used for structural studies of component polypeptides was purified

from platelet lysates by the method of Heldin *et al.* (1981) and Johnsson *et al.* (1982). Polypeptides were fully reduced and alkylated (Skehel and Waterfield, 1975) and separated by reverse phase h.p.l.c. in 2 M guanidine HCl, 1 M acetic acid (Johnsson *et al.*, 1982) or by gel permeation h.p.l.c. in 6 M guanidine, 0.1 M phosphate pH 4.5 (Ui, 1979; Waterfield *et al.*, 1983a). Human PDGF I and II were purified by the method of Deuel *et al.* (1981b) from outdated platelets.

Amino acid compositions were determined following hydrolysis in 6 N HCl *in vacuo* using a Beckman 6300 analyser and amino-terminal sequence analysis using a gas phase sequencer constructed and operated as described by Hewick *et al.* (1981). PTH amino acids were analysed by reverse phase h.p.l.c. (Waterfield *et al.*, 1983a; 1983b).

### Enzymatic digestion and h.p.l.c. purification of peptides

Purified peptides were digested with 3% by weight of *Staphylococcus* V8 protease (Boehringer) in 10 mM ammonium bicarbonate containing 2 M urea. Digests were injected directly onto a RPP-C8 reverse phase column (5 x 0.45 cm, Synchrom, Linden, IN, USA) equilibrated in 0.1% trifluoroacetic acid and the column was developed with a 0–60% gradient of acetonitrile at a flow rate of 1 ml/min (Bennett *et al.*, 1980). Effluent was monitored at 280 and 206 nm using two Uvicord monitors (LKB, Sweden) and radioactivity in aliquots of fractions was measured by scintillation counting. Amino acid analysis of appropriate fractions was carried out to locate peptides.

## Acknowledgements

We thank the Finnish Red Cross (Dr G.Myllylä), Helsinki, the Department of Virology (Dr G.Alm), University of Uppsala and the American Red Cross Blood Banks in Chicago (R.Gilbert), Toledo (F.Courtwright and P.Lan), Tulsa (D.Kasprisin), Waterloo (J.Bender and T.Brown) and Fort Wayne (G.Drew and D.Dunfee) for generous supplies of platelets. We also thank N.F.Totty, R.Philp, S.Wennergren, B.Kennedy and D.Chang for technical assistance. The work was supported by Swedish MRC grant 4486, Swedish Cancer Society grants 689, 786 and 1794 (to Å.W., B.W. and C.-H.H.), and NIH grants CA22409, HL14147 and HL22119 (to T.F.D.).

## References

- Antoniades, H.N. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 7314-7317.
- Antoniades, H.N. and Hunkapiller, M.W. (1983) *Science (Wash.)*, **220**, 963-965.
- Bennett, H.P.J., Browne, C.A. and Solomon, S. (1980) *J. Liq. Chromatogr.*, **3**, 1353-1365.
- Betsholtz, C., Heldin, C.-H., Nister, M., Ek, B., Wasteson, Å. and Westermark, B. (1983) *Biochem. Biophys. Res. Commun.*, **117**, 176-182.
- Blobel, G. and Dobberstein, B. (1975) *J. Cell Biol.*, **67**, 835-851.
- Bourne, H. and Rozenfurt, E. (1976) *Proc. Natl. Acad. Sci. USA*, **73**, 4555-4559.
- Burk, R.R. (1976) *Exp. Cell Res.*, **101**, 293-298.
- Dalla Favera, R., Gelmann, E.P., Gallo, R.C. and Wong-Staal, F. (1981) *Nature*, **292**, 31-35.
- Dalla Favera, R., Gallo, R.C., Giallongo, A. and Croce, C.M. (1982) *Science (Wash.)*, **218**, 686-688.
- Deuel, T.F. and Huang, J.S. (1983) *Prog. Hematol.*, **13**, 202-221.
- Deuel, T.F., Huang, J.S., Proffitt, R.T., Chang, D. and Kennedy, B.B. (1981a) *J. Supramol. Struct. Cell Biochem.*, Suppl., **5**, 128.
- Deuel, T.F., Huang, J.S., Proffitt, R.T., Baenziger, J.U., Chang, D. and Kennedy, B.B. (1981b) *J. Biol. Chem.*, **256**, 8896-8899.
- Deuel, T.F., Huang, J.S., Huang, S.S., Stroobant, P. and Waterfield, M.D. (1983) *Science (Wash.)*, **221**, 1348-1350.
- 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.
- Dicker, P., Pohjanpelto, P., Pettican, P. and Rozenfurt, E. (1981) *Exp. Cell Res.*, **135**, 221-227.
- Doolittle, R.F., Hunkapiller, M.W., Hood, L., Devare, S.G., Robbins, K.C., Aaronson, S.A. and Antoniades, H.N. (1983) *Science (Wash.)*, **221**, 275-277.
- Eva, A., Robbins, K.C., Anderson, P.R., Srinivasan, A., Tronick, S.R., Reddy, E.P., Ellmore, N.W., Galen, A.T., Lautenberger, S.A., Papas, T.S., Weston, E.H., Wong-Staal, F., Gallo, R.C. and Aaronson, S.A. (1982) *Nature*, **295**, 116-119.
- Heldin, C.-H., Westermark, B. and Wasteson, Å. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 3722-3726.
- Heldin, C.-H., Westermark, B. and Wasteson, Å. (1980) *J. Cell. Physiol.*, **105**, 235-246.
- Heldin, C.-H., Westermark, B. and Wasteson, Å. (1981) *Biochem. J.*, **193**, 907-913.

- Hewick,R.M., Hunkapiller,M.W., Hood,L.E. and Dreyer,W.J. (1981) *J. Biol. Chem.*, **256**, 7990-7997.
- Johnsson,A., Heldin,C.-H., Westermark,B. and Wasteson,Å. (1982) *Biochem. Biophys. Res. Commun.*, **104**, 66-74.
- Josephs,S.F., Dalla Favera,R., Gelmann,E.P., Gallo,R.C. and Wong-Staal, F. (1983) *Science (Wash.)*, **219**, 503-505.
- Josephs,S.F., Guo,C., Ratner,L. and Wong-Staal,F. (1984) *Science (Wash.)*, **223**, 487-490.
- Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) in *Molecular Cloning: A Laboratory Manual*, published by Cold Spring Harbor Laboratory, NY.
- Morrissey,J.H. (1981) *Anal. Biochem.*, **117**, 307-310.
- Nistér,M., Heldin,C.-H., Wasteson,Å. and Westermark,B. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 926-930.
- Raines,E.W. and Ross,R. (1982) *J. Biol. Chem.*, **257**, 5154-5160.
- Robbins,K.C., Antoniades,H.N., Devare,S.G., Hunkapiller,M.W. and Aaronson,S.A. (1983) *Nature*, **305**, 605-608.
- Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.
- Skehel,J.J. and Waterfield,M.D. (1975) *Proc. Natl. Acad. Sci. USA*, **72**, 93-97.
- Swan,D.C., McBride,O.W., Robbins,K.C., Keithley,D.A., Reddy,E.P. and Aaronson,S.A. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 4691-4695.
- Ui,N. (1979) *Anal. Biochem.*, **97**, 65-71.
- Ullrich,A., Berman,C.H., Dull,T.J., Gray,A. and Lee,J.M. (1984) *EMBO J.*, **3**, 361-364.
- Waterfield,M.D., Scrace,G.T., Whittle,N., Stroobant,P., Johnsson,A., Wasteson,Å., Westermark,B., Heldin,C.-H., Huang,J.S. and Deuel,T.F. (1983a) *Nature*, **304**, 35-39.
- Waterfield,M.D., Scrace,G. and Totty,N. (1983b) in Darbre,A. and Waterfield,M.D. (eds.), *Practical Protein Biochemistry*, Wiley, NY, in press.
- Westermark,B. and Wasteson,Å. (1975) *Adv. Metab. Disord.*, **8**, 85-100.
- Westermark,B., Heldin,C.-H., Ek,B., Johnsson,A., Mellstrom,K. and Wasteson,Å. (1983) in Guroff,G. (ed.), *Growth and Maturation Factors*, Wiley and Sons, vol. 1, pp. 73-115.
- Wong-Staal,F., Dalla Favera,R., Franchini,G., Gelmann,E.P. and Gallo, R.C. (1981) *Science (Wash.)*, **213**, 226-228.

Received on 28 February 1984

**Note added in proof**

After this work was completed the nucleotide sequences of the six regions within c-sis that correspond to v-sis were published (Josephs et al., 1984).