Linkage of two human pregnancy-specific β_1 -glycoprotein genes: One is associated with hydatidiform mole

(protein family/cell adhesion molecule/placenta)

KIMBERLY K. LESLIE^{*}, SHUICHIRO WATANABE[†], KE-JIAN LEI, DANNY Y. CHOU, CATHIE A. PLOUZEK, HWEI-CHUANG DENG, JERRY TORRES^{*}, AND JANICE YANG CHOU[‡]

Human Genetics Branch, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892

Communicated by Roy Hertz, April 27, 1990 (received for review February 1, 1990)

ABSTRACT A genomic clone containing two linked human pregnancy-specific β_1 -glycoprotein (PS β G) genes has been isolated and characterized. The two genes are arranged in the same $5' \rightarrow 3'$ orientation; the 3' region (including the A2 and B-C exons) of the upstream gene, PSGGA, is linked to the 5' region (including the 5'/L and L/N exons) of *PSGGB*, the downstream gene. Depending upon the domains compared, PSGGA and PSGGB share 92-98% nucleotide and 86-95% amino acid sequence identity with PSG93, the most abundant PS β G transcript. The 3' exon (B–C) of PSGGA contains four alternative splice sites and three polyadenylylation sites, which account for the 3' heterogeneity previously reported in the PS β G family. Each of the predicted PSGGA-encoded proteins would have a different carboxyl terminus. PSGGB corresponds to the previously identified cDNA PSG6, which encodes proteins containing a 34-amino acid leader peptide and a 108amino acid N domain, which is one amino acid shorter than the majority of PS β G N domains. Additionally, the PSGGBencoded proteins contain the cell-surface recognition tripeptide Arg-Gly-Asp, shared by several previously reported PSBGs as deduced from cDNA sequences. Northern blot hybridization performed with a PSGGB-specific oligonucleotide probe to the N domain revealed that PSGGB or a PSGGB-like gene encodes a major 1.7-kilobase mRNA in hydatidiform mole tissues and a major 2.0-kilobase mRNA in term placenta tissues. Moreover, the PSGGB-specific probe hybridized most strongly with mRNA from molar trophoblastic tissue, suggesting that the **PSGGB**-like species may be the gene preferentially expressed in gestational trophoblastic disease. Additionally, the sequence of a 2315-base-pair PSBG cDNA (PSG95) that contains an N-A1-A2-B2-C domain arrangement is reported. The coding region of PSG95 is identical to the previously reported cDNA clones PSG1d and FL-NCA, but PSG95 contains an additional 518 and 523 base pairs in the 3' end as compared with PSG1d and FL-NCA, respectively.

Pregnancy-specific β_1 -glycoprotein (PS β G) was the first of a group of pregnancy-associated proteins identified (for reviews see refs. 1 and 2). It is produced by the placenta and detectable in maternal serum as early as 18 days after ovulation (3). Maternal serum concentrations reach 200–400 μ g/ml by the third trimester, making PS β G a major secretory product of the human placenta (3, 4). PS β G has been used clinically to diagnose pregnancy and to predict some pregnancy-related complications. For example, low PS β G values are associated with poor pregnancy outcome in threatened abortions (5–7), intrauterine growth retardation (8), fetal hypoxia (9), and preeclampsia (10). The clinical uses of PS β G are not limited to pregnancy; it is found in the sera of most

patients with hydatidiform mole, invasive mole, and choriocarcinoma (1, 4), and it has been employed as a marker for monitoring the treatment of choriocarcinoma (1).

Although PS β G has been used as a diagnostic marker, the function of PS β G is unknown. To understand more completely this potentially important protein and its function in pregnancy, our laboratory has undertaken a number of studies to characterize human PS β G. We found that placental PS β G represents a family of closely related glycoproteins of 72, 64, and 54 kDa and that placental poly(A)⁺ RNA directed the synthesis of three polypeptides of 50, 48 (major), and 36 kDa as immunoprecipitated by anti-PS β G serum (11). Moreover, near-full-length cDNAs encoding members of the PS β G family were isolated and characterized by our laboratory (11, 12) as well as by others (13–20). Two gene fragments that share strong sequence similarity with the reported PS β G cDNAs have been reported (21, 22). We now report the structural organization of a genomic clone that contains two linked PS β G genes.§

A careful analysis of the structure of PS β G has shown that it is closely related to the carcinoembryonic antigen (CEA) family (23, 24). CEA has recently been shown to be an intercellular adhesion molecule mediating aggregation of cultured human colon adenocarcinoma cells (25). Both $PS\beta G$ and CEA are members of the immunoglobulin superfamily (13-15, 23, 24, 26), which includes proteins that are arranged in domains and have constant and variable regions. A group of proteins, including neural cell adhesion molecule (N-CAM) (for a review see ref. 27), within the immunoglobulin superfamily contain the surface active tripeptide Arg-Gly-Asp, which acts in cell-surface recognition. Many members of the PS β G gene family (including *PSGGB* in this study) are closely related to N-CAM and encode proteins that contain the Arg-Gly-Asp tripeptide (13, 14, 17, 19, 22). In this paper, we explore the possibility that the PSGGB-encoded protein may be involved in the genesis of gestational trophoblastic disease by acting as an adhesion molecule. To support our hypothesis, we present data demonstrating that PSGGB-like mRNA is preferentially expressed in molar trophoblastic tissue. The PSGGB-like product may also serve as a more specific marker for molar pregnancy.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: $PS\beta G$, pregnancy-specific β_1 -glycoprotein; CEA, carcinoembryonic antigen; N-CAM, neural cell adhesion molecule; nt, nucleotide(s).

^{*}Present address: Department of Obstetrics and Gynecology, Division of Maternal-Fetal Medicine, University of Colorado Health Sciences Center, 4200 E. Ninth Avenue, Denver, CO 80262.

[†]Present address: Department of Pediatrics, School of Medicine, Nihon University, Tokyo, Japan.

[‡]To whom reprint requests should be addressed at: Building 10, Room 9S-242, National Institutes of Health, Bethesda, MD 20892. [§]The sequences reported in this paper have been deposited in the GenBank data base [accession nos. M34715 (PSG95), M34716 (*PSGGA*), and M34717 (*PSGGB*)].

MATERIALS AND METHODS

Library Screening and Characterization of cDNA and Genomic Clones. A human placental cDNA library in $\lambda gt11$ (Frank Gonzalez, National Institutes of Health) was screened with a probe containing the 5' EcoRI-BamHI fragment of the PS β G cDNA clone PSG16 (11), and a human leukocyte genomic library in λ EMBL-3 (Clontech) was screened with a probe containing the entire PSG16 cDNA sequence. The cDNA and genomic inserts from positive clones were subcloned into pUC or pGEM vectors (Promega) for further characterization. Sequencing of cDNA and genomic clones was done by the dideoxy chain-termination method (28) using $[\alpha-[^{35}S]$ thio]dATP (400 Ci/mmol; Amersham; 1 Ci = 37 GBq). Both strands of the cDNA and genomic clones were sequenced and the sequences of the genomic clones were compared with cDNA sequences to identify intron-exon junctions.

Oligonucleotide Probes. Probes were synthesized (Cyclone Plus DNA synthesizer, Milligen Biosearch, Navato, CA) from areas of the *PSGGA* and *PSGGB* sequences that contain deletions or substitutions as deduced from comparisons with cDNAs PSG93 and PSG95. Probe 1 is a 30-mer containing nucleotides (nt) 619–648 of *PSGGB*. Probes 2 and 3 contain nt 45–68 and 323–355 of *PSGGA*, respectively. Oligonucleotides were labeled at the 5'-OH end with $[\gamma^{-32}P]$ ATP by T4 polynucleotide kinase (BRL).

Trophoblastic Tissues. Human term placenta was obtained from normal pregnancy. Three hydatidiform mole tissue specimens were obtained from the University of Colorado Health Sciences Center, Denver. Samples 1 and 2 were noninvasive moles and sample 3 was an invasive mole.

Nucleic Acid Hybridization. Total RNA was extracted by the guanidinium thiocyanate method (29), and $poly(A)^+$ RNA was obtained by oligo(dT)-cellulose chromatography. RNA was electrophoresed in 1.2% agarose gels containing 2.2 M formaldehyde (30) and was transferred to Zetabind membranes (AMF Cuno) by electroblotting. The filters were hybridized at 50°C in the presence of an oligonucleotide probe (3×10^6 cpm/ml) in a previously described buffer (11) from which dextran sulfate was removed. The blots were washed twice in 2× SSC with 0.1% SDS for 30 min at room temperature, then three times in 0.2× SSC with 0.1% SDS for 30 min at 50°C. (SSC is 0.15 M NaCl/0.015 M sodium citrate, pH 7.)

RESULTS

Sequence of PSG95. Three PSBG cDNAs, PSG16, PSG93, and PSG95, have been isolated and characterized in this laboratory (11, 12). Sequence analysis indicates that PSG16 and PSG93 are highly similar and that PSG93 contains an additional 86 base pairs (bp) beginning at nt 1309 of PSG16 (11). Moreover, PSG93 is identical to cDNAs PSGD (13), hPSP11 (15), FL-NCA-2 (17), PSG1a (19), and PSBG81 (20), suggesting that PSG93 represents the major $PS\beta G$ species. Like PSG93, the PSG95-encoded protein contains a 34-amino acid leader peptide, an N-A1-A2-B2-C domain arrangement, seven potential glycosylation sites, and six invariant cysteine residues, which are important in disulfide-bond formation (Fig. 1). Sequence analysis of cDNA PSG95 indicates that it is similar to PSGC (13), PSG1d (19), and FL-NCA (16) but is larger than all of them. The 5' sequences of these four cDNAs are identical, except that PSG95 contains the longest 5' untranslated region. The coding regions of PSG95, PSG1d, and FL-NCA are identical and encode a protein of 426 amino acids. However, PSG95 contains an additional 518 and 523 bp in the 3' untranslated region as compared to PSG1d and FL-NCA, respectively. PSGC is identical to PSG95 from nt 45 to 1404 of PSG95; however, PSGC lacks nt 1405-1759 of PSG95. Homology between these two cDNAs again resumes from nt 1760 to the 3' end of PSG95. Because of the 355-bp insertion in PSG95, the proteins encoded by PSG95 and PSGC differ in their sequences at the carboxyl termini. The 3'-end divergence may be generated by alternative splicing events, as discussed later. A complete analysis of the sequences of PSG95 compared to PSG93 and PSG16 reveals that PSG95 and PSG93 are identical until the alternative splicing junction in the carboxyl terminus, at nt 1404 of PSG95 (Fig. 1). However, PSG95 and PSG93 differ from PSG16 at nt 215, 287, 288, and 764, as well as 3' to the alternative splice point (11). Thus, PSG16 is a polymorphic variant.

Isolation and Characterization of PS β G Genes. Nineteen PS β G genomic clones that strongly hybridized with a probe containing the entire PSG16 cDNA were isolated. These genomic clones were characterized by their restriction endonuclease digestion patterns and by their abilities to hybridize with PS β G-specific probes. Clone A, which hybridized strongly with a probe specific to the 5' end of PSG93 (nt 1–142) and gave a restriction endonuclease pattern similar to that of human genomic DNA (data not shown), was extensively analyzed.

A careful analysis of clone A indicates that it contains two linked PS β G genes arranged in the same 5' \rightarrow 3' orientation (Fig. 2). The upstream gene, PSGGA, contains the A2 and B-C exons in its 3' region. The A2 exon is composed of 279 bp encoding 93 amino acids with 98% identity to the A2 protein domain of PSG16/PSG93/PSG95 at the nucleotide level and 95% identity at the amino acid level. The 3' exon, B-C, contains 1950 bp and encodes the PS β G B2 domain and the 3th untranslated regions of mRNAs similar to PSG16, PSG93, PSG95, PSG1d/FL-NCA, and PSGC. These appear to have been generated by alternative splicing. The majority of $PS\beta G$ mRNAs, which differ in their 3' region, diverge precisely at splice site 1 (C1 in Fig. 2). The PSGGA sequence immediately following this point shares 92% identity with PSG95 and contains the entire 3' region of the PSG95-like sequence from nt 1405 to the 3' end of PSG95. PSG1d (19) and FL-NCA (16) are identical to PSG95, except that the FL-NCA-like sequence ends just after splice site 2 (C2 in Fig. 2), and utilizes a polyadenylylation site beginning at PSGGA nt 902. PSGC (13) and PSG95 differ by the presence of an additional 355-bp sequence (nt 1405-1759) in PSG95 (Fig. 1). In PSGGA, this PSG95-like sequence appears to be an intron in the generation of a PSGC-like mRNA. This sequence contains the consensus splicing site GT/AG (31). Within PSGGA, the PSGC-like 3' terminus begins at nt 889 (C2 in Fig. 2). The single polyadenylylation site shared by the putative PSG95- and PSGC-like transcripts begins at PSGGA nt 1409.

Fifty bases downstream from the ends of the PSG95- and PSGC-like sequences is another junction, splice site C3. This 50-bp sequence does not contain the intron consensus splicing site GT/AG; thus it may represent the true 3' end of the PSG95- and PSGC-like transcripts. Beginning at splice site C3, the PSGGA sequence shares 94% sequence identity with the 3' region of PSG93. PSG93 differs from PSG16 by the presence of an additional 86 bp beginning at nt 1309 of PSG16 (11). Splice site C3 marks the beginning of this 86-bp insert and, thus, nt 535-1482 (which contain the C1 and C2 sequences and the consensus splicing site GT/AG) act as an intron to generate a PSG93-like mRNA. The 86-bp region of the PSG93-like sequence is designated as C3 (Fig. 2) and begins at PSGGA nt 1483. Immediately following the 86-bp sequence is splice site C4, which represents where the PSG16-like message joins after splicing at site C1. Sequences in C4 are 94% identical with the shared 3' sequences of PSG16 and PSG93. The shared polyadenylylation site for the putative PSG93- and PSG16-like messages begins at nt 2152 of PSGGA. Although similarity exists between PSGGA and previously reported PS β G cDNAs, PSGGA is not identical to any of them.

							#																							GAA	TTCG(GGCT	GACC	15
CTGCCCATGAGCTTGAGAATTGCTCCTGCCCTGGGAAGAGGCCTCAGCACAGAAAGAGGAAGGA															AGAA	CACA	CAAG	150																
CAG	CAGA	GACC	ATG	GGA Gly	ACC	CTC Leu	TCA Ser	GCC Ala	CCT Pro	CCC Pro	TGC Cys	ACA Thr	CAG Gln	CGC Arg	ATC Ile	AAA Lys	TGG Trp	AAG Lys	GGG Gly	CTC Leu	CTG Leu	CTC Leu	ACA Thr	GCA Ala	TCA Ser	CTT Leu	TTA Leu	AAC Asn	TTC Phe	TGG Trp	AAC Asn	CTG Leu	CCC Pro	254 31
300	3.00	000	+ 1	N dom	ain	3 000	(1)	000	a1.a	003	100					-																		
Thr	Thr	Ala	Gln	Val	Thr	Ile	GAA Glu	Ala	GAG Glu	Pro	Thr	AAA Lys	Val	Ser	GAG Glu	GGG Gly	AAG Lys	GAT Asp	GTT Val	CTT Leu	CTA Leu	CTT Leu	GTC Val	CAC His	AAT Asn	TTG Leu	CCC Pro	CAG Gln	AAT <u>Asn</u>	CTT Leu	ACC	GGC Gly	TAC Tyr	356 65
ATC Ile	TGG Trp	TAC Tyr	AAA Lys	GGG Gly	CAA Gln	ATG Met	AGG Arg	GAC Asp	CTC Leu	TAC Tyr	CAT His	TAC Tyr	ATT Ile	ACA Thr	TCA Ser	TAT Tyr	GTA Val	GTA Val	GAC Asp	GGT Gly	GAA Glu	ATA Ile	ATT Ile	ATA Ile	TAT Tyr	GGG Gly	CCT Pro	GCA Ala	TAT Tyr	AGT Ser	GGA Gly	CGA Arg	GAA Glu	458 99
ACA	GCA Ala	TAT Tyr	TCC Ser	AAT Asn	GCA Ala	TCC Ser	CTG Leu	CTG Leu	ATC	CAG Gln	AAT Asn	GTC Val	ACC	CGG Arg	GAG Glu	GAC	GCA	GGA	TCC	TAC	ACC	TTA	CAC	ATC	ATA	AAG	GGA	GAT	GAT	GGG	ACT	AGA	GGA	560
		-1-						200		+ A	1 Dor	nain			oru	nop	AIG	GLY	Ser	TÀT	THE	Deu	nıs	IIe	IIE	гуз	GTÅ	Asp	Asp	GTÀ	Int	Arg	GTĂ	133
GTA Val	ACT Thr	GGA Gly	CGT Arg	TTC Phe	ACC	TTC Phe	ACC Thr	TTA Leu	CAC His	CTG Leu	GAG Glu	ACT Thr	CCT Pro	AAG Lys	CCC Pro	TCC Ser	ATC Ile	TCC Ser	AGC Ser	AGC Ser	AAC Asn	TTA Leu	AAT Asn	CCC Pro	AGG Arg	GAG Glu	ACC Thr	ATG Met	GAG Glu	GCT Ala	GTG Val	AGC Ser	TTA Leu	662 167
ACC Thr	TGT Cys	GAC Asp	CCT Pro	GAG Glu	ACT Thr	CCA Pro	GAC Asp	GCA Ala	AGC Ser	TAC Tyr	CTG Leu	TGG Trp	TGG Trp	ATG Met	AAT Asn	GGT Gly	CAG Gln	AGC Ser	CTC Leu	CCT Pro	ATG Met	ACT	CAC His	AGC Ser	TTG Leu	AAG Lvs	CTG Leu	TCC Ser	GAA Glu	ACC	AAC Asn	AGG Arg	ACC Thr	764 201
CTTC	արդոր	ста	ጥጥር	CCT	CTTC	202	220	ጥአጥ	3.OT	CC3	CC3	000	mam	C33	*	~		000	110	003	ama	1.00												0//
Leu	Phe	Leu A2 Do	Leu	Gly	Val	Thr	Lys	Tyr	Thr	Ala	Gly	Pro	Tyr	GAA Glu	Cys	GAA Glu	Ile	Arg	AAC	Pro	Val	Ser	Ala	AGC Ser	Arg	AGT Ser	GAC	CCA Pro	GTC Val	ACC	CTG Leu	AAT Asn	CTC Leu	235
CTC	CCG	AAG	CTG	CCC	AAG	CCC	TAC	ATC	ACC	ATC	AAC	AAC	TTA	AAC	ccc	AGG	GAG	AAT	AAG	GAT	GTC	TTA	AAC	TTC	ACC	TGT	GAA	CCT	AAG	AGT	GAG	AAC	TAC	968
Leu	Pro	гÀг	Leu	Pro	LYS	Pro	Tyr	IIe	Thr	IIe	Asn	Asn	Leu	Asn	Pro	Arg	Glu	Asn	Lys	Asp	Val	Leu	<u>Asn</u>	Phe	Thr	Cys	Glu	Pro	Lys	Ser	Glu	Asn	Tyr	269
ACC Thr	TAC Tyr	ATT Ile	TGG Trp	TGG Trp	CTA Leu	AAT Asn	GGT Gly	CAG Gln	AGC Ser	CTC Leu	CCG Pro	GTC Val	AGT Ser	CCC Pro	AGG Arg	GTA Val	AAG Lys	CGA Arg	CCC Pro	ATT Ile	GAA Glu	AAC Asn	AGG Arg	ATC Ile	CTC Leu	ATT Ile	CTA Leu	CCC Pro	AGT Ser	GTC Val	ACG Thr	AGA Arg	AAT Asn	1070 303
GAA	ACA	GGA	ccc	TAT	CAA	TGT	GAA	ATA	CGG	GAC	CGA	TAT	GGT	GGC	GTC	CGC	AGT	GAC	CCA	GTC	ACC	CTG	AAT	GTC	CTC	TAT	GGT	main CCA	GAC	CTC	CCC	AGA	ATT	1172
Glu	Thr	Gly	Pro	Tyr	Gln	Cys	Glu	Ile	Arg	Asp	Arg	Tyr	Gly	Gly	Val	Arg	Ser	Asp	Pro	Val	Thr	Leu	Asn	Val	Leu	Tyr	Gly	Pro	Asp	Leu	Pro	Arg	Ile	337
TAC	CCT	TCA	TTC	ACC	TAT	TAC	CGT	TCA	GGA	GAA	GTC	CTC	TAC	TTG	TCC	TGT	TCT	GCG	GAC	TCT	AAC	CCA	CCG	GCA	CAG	TAT	TCT	TGG	ACA	ATT	AAT	GAA	AAG	1274
Tyr	Pro	Ser	Phe	Thr	Tyr	Tyr	Arg	Ser	Gly	Glu	Val	Leu	Tyr	Leu	Ser	Cys	Ser	Ala	Asp	Ser	Asn	Pro	Pro	Ala	Gln	Tyr	Ser	Trp	Thr	Ile	Asn	Glu	Lys	371
TTT Phe	CAG Gln	CTA Leu	CCA Pro	GGA Gly	CAA Gln	AAG Lys	CTC Leu	TTT Phe	ATC Ile	CGC Arg	CAT His	ATT Ile	ACT Thr	ACA Thr	AAG Lys	CAT His	AGC Ser	GGG Gly	CTC Leu	TAT Tyr	GTT Val	TGC Cys	TCT Ser	GTT Val	CGT Arg	AAC Asn	TCA Ser	GCC Ala	ACT Thr	GGC Gly	AAG Lys	GAA Glu	AGC Ser	1376 405
TCC Ser	AAA Lys	TCC Ser	ATG Met	ACA Thr	GTC Val	GAA Glu	GTC Val	TCT Ser	G GT Gly	AAG Lys	TGG	ATC Ile	CCA Pro	GCA Ala	TCG	TTG	GCA	ATA	GGG Glv	TTT	TAG	ey ve et	GTC	1110	IGGC)	177(2)	GNGI	AGA	FTCA	SGAN	AACA	ATTG	PATT	1490 426
ccc	AGCC	GTG	rece	ATGG	CAC	AAGC	AAATY	SCC M	ATT	CPCCT	CCR	-		-	-		ACRA(MARK			PAAC		1000	N 3 19 1	PER-WENT					2003-02				1625
(33)	- A THINK	- Filala	No B da	PC33		THIN A	PAAT		ACA			-	- 2 (1997)													Sec.14	MIN			C 1 PRG	ACCA.			1740
Sarara			LINCLA	A GIFAFAI	CI A PARS.	A T CHAT	10011	MILOR		Innio	****			LT TWI	LIGI	1000	FIIGA	TGA	FFCF		ATGG:	1.1.1.0.	A 4 4 7	FIAC	VITE	aree	VALUE -	4419	Residia	FTAA	TCTA	ICTA	CAGC	1/00
TTA	FAGC/	AGTT	CAAT	AAAC	TATA	CTTC	TGGG	ACCO	STAA!	ITGA	ACAI	TTAC	TTT	GCTI	TCT	ACCT	GACTO	GCCC	CAGA	ATTGO	GGCA	ACTA	TCAS	rgag/	ATTO	ATAT	GTT	FATGO	GTAA?	FACA	CATA	TTTG	CACA	1895
AGT	ACAG	PAAC/	ATC	IGCT	TTCT	FTGT	AACA	rgac/	ACAT	FTGA	ATC	TTG	STTAT	CATT?	CCA	ATGC	PTTG/	ATTCO	GGAT	GTTAT	TATT	AAAA	ACATZ	AGATZ	GAAJ	GAA	GAAG	CCAAS	TATG	AACT	GCAG	GCAA	AGTC	2030
TGA	AGTCI	AGCC	rTGG?	TTTG (GCTT	CCTA	TTCT	CAAGA	AGGT	FTGTO	GAAGA	TTT	ATCI	CAGA	TTC	CTTA	raaa/	ACT	FAGAG	GAAAJ	AGAA	AATT	TAG	AAGAG	CAGCO	TAC	TGG	FCCAT	TGC	FACT	CTTG	CTGC	ACTT	2165
ATG	TAAA	CAAT	CAGA	CCAC	ATTT	GAAG	AAAC	CCAO	CCTA	FTTTC	GCAAA	CAA	ACTTZ	ATTCI	ACTO	GAAA	TATO	CATTO	GGTA	AAAG	FAGA	GATG	CCA	ragao	GGAA	AAA	TATO	GTGG/	AAA	<u>FAAA</u>	AACT	GTAG	FATA	2300
CCT	GTAC	CC <u>GA</u>	ATTC																								9							2315

FIG. 1. Nucleotide and deduced amino acid sequences of cDNA PSG95, as compared to cDNA PSGC. The 5' sequence of cDNA PSGC (13) begins at nucleotide 45 (indicated as #) of PSG95. Arrows indicate the leader (L) peptide and the N, A1, A2, and B2 protein domains of the PSG95-encoded protein. Potential asparagine-linked glycosylation sites and poly(A)-addition signals are underlined. Asterisks indicate cysteine residues involved in disulfide-bridge formation. The potential termination codon (TGA) for PSGC is boxed. The shaded area represents the sequence that is absent from PSGC (13). The nucleotide that differs from PSGC is shown below the substituted position, and dashes at nt 2003–2006 indicate deletions.

The nucleotide sequence of the downstream gene, *PSGGB*, is identical to cDNA PSG6 reported by Zimmermann *et al.* (19). We have obtained \approx 180 bp of the *PSGGB* sequence upstream from the transcriptional start site of the PSG93-like sequence (the beginning of exon 5'/L). The 5'/L exon (\approx 250 bp) encodes the 5' untranslated region and the first 21 amino acids of the leader peptide. The L/N exon (363 bp) encodes the remaining leader codon and the N protein domain (108 amino acid residues). Amino acids 126–128 of *PSGGB* are Arg-Gly-Asp. This tripeptide is the signal-surface active component of cellular adhesion molecules (27) and gives a possible clue to the function of this member of the PS β G family.

Northern Analysis. Fig. 2 illustrates the nucleotide sequences of *PSGGA* and *PSGGB* and the proposed exon and protein sequences compared to PSG16, PSG93, PSG95, and PSGC. To demonstrate that mRNAs encoded by *PSGGA* and *PSGGB* are actually expressed in placental tissue, we synthesized three oligonucleotides (probes 1–3) that would hybridize to the *PSGGA* and *PSGGB* sequences in areas that contain deletions or substitutions as compared with cDNA PSG93/PSG95. Probe 1 (30-mer) is located in the N domain of *PSGGB* and includes the 3-bp deletion present in this region as well as four base-pair substitutions when compared with PSG95/PSG93. Probe 2 (24-mer) is located in the A2 domain of *PSGGA*, where it contains three single base-pair substitutions when compared with PSG95/PSG93. Probe 3 (33-mer) is located in the B2 domain of *PSGGA*, where it contains four single base-pair substitutions and one double substitution when compared with PSG95/PSG93. Northern blots revealed that the PSGGB-specific probe, probe 1, hybridized with a major 1.7-kb mRNA and a minor 1.5-kb mRNA from the three molar tissues, but with a 2.0-kb mRNA (major) and a 1.5-kb mRNA from term placental tissue (Fig. 3). Moreover, the transcript recognized by probe 1 appears to be preferentially expressed in hydatidiform mole when compared with normal trophoblastic tissue (Fig. 3). Furthermore, the level of the transcript that hybridized with probe 1 was higher in an invasive mole than in two noninvasive moles. The PSGGA probe, probe 2, is the least specific probe and hybridized with two mRNAs of 2.2- and 1.5-kb in term placenta and with a 1.7-kb message in molar tissue. However, the relative amounts of probe 2-hybridizable message were higher in placenta than in hydatidiform moles. Probe 3, which is more specific to PSGGA, hybridized poorly with either placenta or molar RNA (data not shown).

DISCUSSION

We report the isolation and characterization of two linked genes (*PSGGA* and *PSGGB*) encoding members of the PS β G protein family and present the complete sequence of another PS β G cDNA, PSG95. PS β G is a family of proteins with many closely related individual members encoded by more than one gene on chromosome 19 (ref. 32; K.-J.L. and J.Y.C., unpublished results). In addition to this genomic clone, we





(A) PSGGA

PSG93/95	→ Exon A2 C G	_												
PSGGA	ttcacacag bg Ang CTG CCC Ang CTC ATC ATC ATC AAT AAC TTA AAC CCC AGG GAC Lys Lew Pro Lys Pro Tyr 11e Arn Asn Asn Lew Asn Pro Art Asp	50 17												
	Pro Glu													
PSG93/95 PSGGA	ANT ANG GAT GTC TCA ACC TTC ACC TTCT GAA CCT ANG AGT GAG AAC TAC ACC TAC ATT TGG TGG CTA ANT GGT CAG AGC CTC CCG GTC AGT CCC AGG GTA AAG	152												
	Asn Lys Asp Val Ser Thr Phe Thr Cys Glu Pro Lys Ser Glu <u>Asn Tyr Thr</u> Tyr Ile Trp Trp Leu Asn Gly Gln Ser Leu Pro Val Ser Pro Arg Val Lys	51												
PSG93/95	Leti Astri C G G													
PSGGA	GA OGC ATT GAA AAC AGG ATC CTC ATT CTA COC AGT GTC ACG AGA AAT GAA ACA GGA CCC TAT CAA TOT GAA ATA CGG GAC CGA TAT GGT GGC ATC CGC AGT	254												
	Pro													
PSG93/95	G G													
PSGGA	GAC CCA GTC ACC CTG AAT GTC CTC T gtgagtatc-350 bp-ttcctccag AT GGT CCA GAC CTC CCC AGA ATT TAC CCT TCA ACC TAT TAC CAT TCA GAC CAA	335												
	Asp rio var mi beu ash var beu far and the sing rio asp													
PSG93/95	GT T C T G													
PSGGA	AAC CTC TAC TTG TCC TGC TTT GCG GAC TCT AAC CCA CCG GCA CAG TAT TCT TGG ACA ATT AAT GGG AAG TTT CAG CTA TCA GGA AAA GC TT TCT AC CCC	437												
	As in the first bet ber (ys) me ala as ber as from to and on the set of the ber as bet of on by a bet ber as ber as the from the ber as the from the ber as the from the ber as	140												
PSG93/95														
PSGGA	CAG ATT ACT ACA ANG CAT AGC GGG CTC TAT GCT TGC TCC GTT CGT AAC TCA GCC ACT GGC ANG GAA AGC TCC AAA TCC GTG ACA GTC AGA GTC TCT GGT AAG	539												
	Gin lie Thr Thr Lys His Ser Giy Leu Tyr Ala Cys Ser Val Arg Asn Ser Ala Thr Giy Lys Giu Ser Ser Lys Ser Val Thr Val Arg Val Ser Giy Lys	180												
PSG95														
PSGGA	TEG ATC CCA GCA TCC TTE GCA ATA GGE TTT TAG GTGCAGTCTATCTCACTTTCAGAGAAAAGTCAGGAAAACATTTGTATTCCCCAGGCCTGTGTCCCCATGGGCACAAGCAAATCCCAAATTTTTCC	663												
	Trp Ile Pro Ala Ser Leu Ala Ile Gly Phe End	190												
PSG95 PSCCA		798												
PSG95/C	$T \qquad A T T \qquad C G T T C \qquad A C \longrightarrow C^2 C A C C G$													
PSGGA	стттття салтттталалататестваттстттатетесаластттстасалттатесала тот-етстттаатстатетствестватасттс <u>алтала</u> статастстесссалтсалтсалтся	932												
AlaTyrSerGlySerfleAsnTyrThrSerGlyAsnAspAsnEnd														
PSC95/C														
PSGGA	ANTATTTACTTTECCCCTEATACTCEATATCCCCCAGAATTGGGCAACTATTCATGATATTGATATGTTTATGGTAATACAGATATTTGCACAAAACAGGATACCACCTGCTCTTTTGTAACAGGACACATTTCA	1067												
PSG95/C	T T C A A CAT G AAT A G G G													
PSGGA	ANTCATTGGTTATATTACCAAGGCTTTGACTGGGATGTTATATTTAAGAATATAGCTAGAATGAACTGCAGGCAAGCTCTGAAGTCAGCCTTGGTTTGGCTTCCTATTCTCAAGAGTTTTGT	1199												
PSGGA	AAAGGTTTAATCTEAGATTCCTTATAAAAACTTACAGCAAAGAAAATTTTAAAAGGAGCCTACAATAGTCCAATGCTACTATGCTGCACTTATGTAAACAATCAGACCAAGTTTGAAGAAAACTCAAGATTGAAGAAAACTCAAGATTGAAGAAAACTCAAGTTTGAAGAAAACTCAAGTTAGACGAAGTTAGAAGAAACTCAAGTTAGAAGAAAACTCAAGATCAAGACCAAGTTAGAAGAAACTCAAGACCAAGTTAGAAGAAACTCAAGACCAAGTTAGAAGAAACTCAAGACCAAGATGAAGAAACTCAAGACCAAGATGCAATAGTCAAGACCAAGATGCAATAGTCAAGACCAAGATGCAAGACCAAGATGCAAGACCAAGATGCAAGACCAAGATGCAAGACCAAGATGCAAGACCAAGAAAAGTAGACAAGACCAAGACCAAGATGAAGAACAAGACCAAGATGAAGAAAACTCAAGACCAAGATGCAAGAAGAAGAAGAAGACCAAGATGAAGAAAACTCAAGACCAAGATGAAGAAAGTCAAGACCAAGATGAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAA	1334												
PSG95/C	G G AT T C													
PSGGA PSG95/C/93	GCAAACAAACTTATTCTACTGAAAATTTATTCTACGAAAACTACGAGAGCCCCATACGAGAGAAAATTATCTGCGAAAATTAAACCGTAGTACACCTCTATGAGATCCACCTCTTTTTTTCTGCGAAAATTTATCTGCGAAAATTATCTGCGAAAATTATCTGCGAAGATCCACCTCTTATGAGATCCACCTCTTATGAGATCCACCTCTTATGAGATCCACCTCTTATGAGATCCACCTCTTATGAGATCCACCTCTTATGAGATCCACCTCTTATGAGATCCACCTCTTTTTTTT	1400												
PSGGA	ATTATCCACTOTAGACTOGACATTACCCTGAATTCCACTAGTTCCTCCAATTCCACTCCCCATGGAACCTCAAAGAGCAAGACCCACTCTATAAGTCAGAGAGTGGACAACTCAATGTA	1603												
	AspTrpSerLeuProEnd GluSerLeuEnd													
DCC02/16														
PSGGA	AATTTCATGGGAAAATCCTTGTACCTGATGTCTGAGCCACTCAGAACTCACCAAAATGTTCAACACCCATAACAACGGCTGCTCAAACTGTAACAACGGAAAACAAGTTGATGACACCTGTGGACAGCTTTT	1738												
PSG93/16		1973												
PSGGA PSG93/16		10/3												
PSGGA	TGAGGGTAACAACAGAGTGTCAGATATGTCATCTCAACCTCAGACTTTTACATAACATCTCAGGAGGAAATGTGGCTCTCTCCATCATGCATACAGGGCTCCCAATAGAAATGAAACACAGAGATATTGCCTGTGT	2008												
PSG93/16 PSCGA	G G G G G G G G G G	2142												
PSG93/16														
PSGGA	TTCTAGCAGAATAAAACATGTACTACAATTTGCTAATACTGTTCTTCTTAAAAAAAA	2229												
(8) PSGG	2													
PSCCB	GETA TETETEACTERICATERICATERICATERICATERICATERICATERICATERICATERICATERICATERICATERICATERIA A A A A CONSTRUCTOR A	112												
	Exon 5'/L													
PSG93/95														
PSGGB	CTCCA <mark>EAATTEE</mark> CAGGGAUTGAAACAUAGAUAAAAGGAAAAAGGAAGGACAAGGAGGAUAGGAGGAUAGAGGAUGAAGAGGAUGAAUGA	247												
PSG93/95	A G G AGT T A AG A													
PECCH	CCCTTECCAAGAGGCTCACCACAGAAGGAAGGAAGGACAGCACCACCTGACAGCCCTGCTCAGGAAGTCTCTGGATCCTAGGCTCATCTCCACAGGGGAGAACACAGAGAGACGAGAGACC ATG GGA CCC CTC	378												

								~																		Exc	n L/I	N			Met	GIY	Pro Thr	Leu	
PSG93/95														~	-	~		~ .	-			b			- L.	-	-			_	maa		-	000	
PSGGB	TCA	GCC	CCT	CCC	TGC	ACT	CAG	CAC	ATC	ACC	TGG	AAG	GGG	CIU	CIG	Cre	ACA	G	jegag	gag-	1100	-qa	accu	cetaç	J CA	TCA	CIT	114	AAC	TIC	166	AAC	CIG	CCC	455
	Ser	Ala	Pro	Pro	Cys	Thr	Gln	His	Ile	Thr	Trp	Lys	Gly	Leu	Leu	Leu	Thr								Ala	Ser	Leu	Leu	Asn	Phe	Trp	Asn	Leu	Pro	31
								λrg		Lys																									
PSG93/95					С	CG				G		λ																				с			
PSCCB	ACC	ACT	GCC	CAA	GTA	λτλ	ATT	GAA	GCC	AAG	CCA	CCC	XXX	GTT	TCC	GAG	GGG	AAG	GAT	GTT	CTT	СТА	CTT	GTC	CYC	λλT	TTG	ccc	CAG	AAT	CTT	ACT	GGC	TAC	561
	Thr	Thr	Ala	Gln	Val	Ile	Ile	Glu	Ala	LVS	Pro	Pro	Lvs	Val	Ser	Glu	Gly	Lys	s Asp	Val	Leu	Leu	Leu	Val	His	λsn	Leu	Pro	Gln	Asn	Leu	Thr	Gly	Tyr	65
						Thr				Gĺu		Thr					-		-														-	•	
D0000 /05								G												G		G	ATA						λ	т					
P5G93/95	3.000	TCC	TAC		CCC	CAA	ATG	ACG	GAC	CTC	TAC	CAT	TAC	ATT	ACA	TCA	TAT	GT	GTA	CAC	GGT	CAA		ATT	ATA	TAT	GGG	CCT	GCC	TAC	AGT	GGA	CGA	GAA	660
PSGGB	710	100	Three	Tyre	Cly	Cla	Not	Thr	Aen.	Len	Tur	Hie	Tvr	TIO	Thr	Ser	Tyr	Val	Val	His	GIV	Gln		Tle	Tle	Tvr	Glv	Pro	Ala	Tvr	Ser	Glv	Ara	Glu	91
	TIE	пp	1.1.1	Lys	913	GIU	Het		пор	L.C.u	- 3 -		- , -			001	- 1 -			Acn		610	TIA			- 7 -				- 4 -					
		-						MY						~	~		~			пор		014	110					c	37						
PSG93/95		С									~ ~		_					~~~		-		100		~~~	100		336	000		C. 10			202	CC1	
PSGGB	усу	GTA	TAT	TCC	AAT	GCA	TCC	CTG	CTG	ATC	CAG	AAT	GTC	ACA	CAG	GAG	GAT	GCA	GGA	TUU	TAC	ACC	TTA	CAC	AIC	AIA	AAG	CGA	GGC	GAI	GGG	ACT.	GGA	GGA	/62
	Thr	Val	Tyr	Ser	<u>Asn</u>	Ala	Ser	Leu	Leu	Ile	Gin	Asn	Val	Thr	Gin	Glu	Asp	A15	a GIY	ser	Tyr	Thr	Leu	HIS	116	116	LYS	Arg	GIY	ASP	GIY	Thr	GIY	GIY	132
		Ala													Arg													GIÅ	Asp				Arg		
PSG93/95				CG		С	T			с	с																								
PSGGB	GTA	ACT	GGA	TAT	TTC	ACT	GTC	ACC	TTA	TAC	Tg	tgag	tga																						793
	Val	Thr	Glv	Tvr	Phe	Thr	Val	Thr	Leu	Tyr																									142
				Ara			Phe			His																									

FIG. 2. Restriction map (B, BamHI; R, EcoRI; H, HindIII) and the nucleotide and deduced amino acid sequences of two linked PS_βG genes, PSGGA and PSGGB, as compared to cDNAs PSG16, PSG93, PSG95, and PSGC. (A) The upstream gene, PSGGA, contains the A2 and B-C exons. The B-C exon contains four alternative splicing sites (C1, C2, C3, and C4) that may be involved in the generation of four PSBG mRNAs, each encoding a protein with a different carboxyl terminus. The PSG1d-like 3' sequence is composed of C1 (19), the PSG95-like 3' sequence is composed of C1 and C2, the PSGC-like 3' sequence is composed of C2 (13), the PSG93-like 3' sequence is composed of C3 and C4 (11), and the PSG16-like 3' sequence is composed of C4 (11). Cysteine residues are boxed, and potential asparagine-linked glycosylation sites and the poly(A) addition sequences are underlined. (B) The downstream gene, PSGGB, contains the 5'/L and L/N exons. The CAAT box and Arg-Gly-Asp tripeptide are boxed.

have identified another clone containing two different linked PS β G genes (clone B, unpublished results). Our results suggest that all of the PS β G genes may be linked. The B-C exon of PSGGA encodes the B2 protein domain and four alternative splicing points within the carboxyl terminus that could result in the generation of multiple transcripts. The



FIG. 3. Northern blot hybridization of placental and hydatidiform mole RNAs with PSGGA- and PSGGB-specific probes. Total (20 µg per lane, T) or poly(A)⁺ (2 μ g per lane, A⁺) RNAs from human term placenta and hydatidiform moles were electrophoresed in formaldehyde/agarose gels. RNA was hybridized to ³²P-labeled oligonucleotide probes containing nt 619-648 (oligo 1) of PSGGB or nt 45-68 (oligo 2) of PSGGA. Samples 1 and 2 are total RNAs from two noninvasive moles, and sample 3 is total RNA from an invasive mole. Position of 18S rRNA is indicated.

similarity between the PSGGA 3' sequences and the 3' sequences of PSG95, PSGC (13), PSG1d (19)/FL-NCA (16), PSG93 (11), and PSG16 (11) suggests that these five $PS\beta G$ mRNAs may be encoded by a gene with a similar B-C exon.

Oikawa et al. (21, 22) have isolated two other $PS\beta G$ genomic fragments. The gene fragment CGM35 (21) contains the A1, B1, A2, and B-C exons. The B-C exon also contains several splice points, and alternative splicing is likely in this gene fragment as a means to generate different mRNA species; however, CGM35 does not encode any of the reported PSBG mRNAs. The second gene fragment, PSG-HL 12-2 (22), contains the 5'/L, L/N, A1, B1, and A2 exons with part of the B2 domain of the B-C exon. A PSBG cDNA (PS β G-HL clone 22) has been identified that has the N-A2-B2-C domain arrangement and shares complete sequence homology with genomic fragment PSG-HL 12-2 (22). This suggests that the A1 and B1 exons have been spliced out to generate this mRNA. A comparison of the intron sequences of PSGGA and PSGGB with the corresponding intron sequences of PSG-HL 12-2 reveals that they are very similar (data not shown). The similarity in intron sequences among the different PS β G genes indicates that divergence within this family is a relatively recent evolutionary event.

Although a putative CAAT box is located at nt 118-123 of **PSGGB**, there is no consensus TATA box-like sequence in the 5' flanking region of this gene. The linkage of PSGGA and PSGGB indicates that an additional 5' exon does not exist in **PSGGB.** This is supported by previous primer extension experiments, which showed that PSG93 may be a full-length cDNA clone (11), and the entire 5'-untranslated region of PSG93 is represented in PSGGB.

PSGGB encodes the PSG6 transcript reported by Zimmermann et al. (19). This PS β G species contains the Arg-Gly-Asp tripeptide (amino acids 126-128) in a position similar to that found in some other PS β G species (13, 14, 17, 19, 22). This tripeptide is also present in the well known cellular adhesion molecule N-CAM and in other proteins that are known to be important in the cellular recognition process (27, 33). The Arg-Gly-Asp-containing PSBG proteins could be important in maintaining the functional and structural integrity of the placenta, interacting between the trophoblastic and decidual cells, and mediating trophoblastic invasion. Although the association of PSSGB-like mRNA with hydatidiform moles was demonstrated with only three molar samples, our data support the theory that proteins encoded by PSGGB or a PSGGB-like gene may be acting as molar trophoblastic adhesion molecules in a manner analogous to CEA in adenocarcinoma of the colon (25). Additionally, the specificity of the PSGGB-like protein for gestational trophoblastic disease may be of use clinically in the diagnosis and monitoring of molar pregnancies.

In a broader sense, the concept that certain members of the $PS\beta G$ family may be associated with specific disease entities could be of great importance. For instance, if a protein encoded by the PSGGB-like gene represents a specific marker for invasive mole or choriocarcinoma, does a specific $PS\beta G$ marker exist for pregnancies that are destined to abort? Now that a variety of PSBG cDNAs have been identified, differential expression of these species should be investigated.

This work was done while C.A.P. held a National Research Council (Human Genetics Branch, National Institute of Child Health and Human Development) research associateship.

- Tatarinov, Y. S. (1978) Gynecol. Obstet. Invest. 9, 65-97.
- Sorensen, S. (1984) Tumor Biol. 5, 275-302. 2.
- Tatarinov, Y. S. & Masyukevich, V. N. (1970) Byull. Eksp. Biol. Med. 3. USSR 69, 66-68.
- 4. Lin, T. M., Halbert, S. P. & Spellacy, W. N. (1974) J. Clin. Invest. 54, 576-582.
- 5. Wurz, H., Geiger, W., Kunzig, H. J., Jabs-Lehmann, A., Bohn, H. & Luben, G. (1981) J. Perinat. Med. 2, 67-78.
- Hertz, J. B. & Schultz-Larsen, P. (1983) Int. J. Gynaecol. Obstet. 21, 111-117. 7.
- Masson, G. M., Anthony, F. & Wilson, M. S. (1983) Br. J. Obstet. Gynaecol. 90, 146-149.
- 8. Tamsen, L., Johansson, S. G. O. & Axelsson, O. (1983) J. Perinat. Med. 11. 19-25.
- MacDonald, D. J., Scott, J. M., Gemmel, R. S. & Mack, D. S. (1983) 9. Am. J. Obstet. Gynecol. 147, 430-436.
- 10. Grudzinskas, J. G., Gordon, Y. B., Menabawey, M., Lee, J. N., Wadsworth, J. & Chard, T. (1983) Am. J. Obstet. Gynecol. 147, 10-12.
- Watanabe, S. & Chou, J. Y. (1988) J. Biol. Chem. 263, 2049-2054. 11.
- 12. Watanabe, S. & Chou, J. Y. (1988) in Placental and Endometrial Proteins, ed. Mizutani, S. (VNU Sci. Press, Japan), pp. 155-158. 13. Streydio, C., Lacka, K., Swillens, S. & Vassart, G. (1988) Biochem.
- Biophys. Res. Commun. 154, 130-137. Rooney, B. C., Horne, C. H. W. & Hardman, N. (1988) Gene 71, 14.
- 439-449 Chan, W.-Y., Borjigin, J., Zheng, Q.-X. & Shupert, W. L. (1988) DNA 15.
- 7, 545-555 Khan, W. N., Osterman, A. & Hammarstrom, S. (1989) Proc. Natl. 16. Acad. Sci. USA 86, 3332-3336.
- 17. Khan, W. N. & Hammarstrom, S. (1989) Biochem. Biophys. Res. Commun. 161, 525-535.
- 18. McLenachan, T. & Mansfield, B. (1989) Biochem. Biophys. Res. Commun. 162, 1486-1493.
- 19. Zimmermann, W., Weiss, M. & Thompson, J. A. (1989) Biochem. Biophys. Res. Commun. 163, 1197-1209.
- 20. Niemann, S. C., Flake, A., Bohn, H. & Bartels, I. (1989) Hum. Genet. 82, 239-243.
- 21. Oikawa, S., Inuzuka, C., Kosaki, G. & Nakazato, H. (1988) Biochem. Biophys. Res. Commun. 156, 68-77.
- 22. Oikawa, S., Inuzuka, C., Kuroki, M., Matsuoka, Y., Kosaki, G. & Nakazato, H. (1989) Biochem. Biophys. Res. Commun. 163, 1021-1031.
- 23 Oikawa, S., Nakazato, H. & Kosaki, G. (1987) Biochem. Biophys. Res. Commun. 142, 511-518.
- Beauchemin, N., Benchimol, S., Cournoyer, D., Furks, A. & Stanners, 24. C. P. (1987) Mol. Cell. Biol. 7, 3221-3230.
- Benchimol, S., Fuks, A., Jothy, S., Beauchemin, N., Shirota, K. & Stanners, C. P. (1989) Cell 57, 327–334. Watanabe, S. & Chou, J. Y. (1988) Biochem. Biophys. Res. Commun. 25.
- 26. 152, 762-768.
- Cunningham, B. A., Hemperly, J. J., Murray, B. A., Predoger, E. A., 27. Brackenbury, R. & Edelman, G. M. (1987) Science 236, 799-806
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. 28. USA 74, 5463-5467.
- Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. 29. (1979) Biochemistry 18, 5294–5299.
- 30. Lehrach, H., Diamond, D., Wozney, J. M. & Boedtker, H. (1977) Biochemistry 16, 4743-4751.
- Breathnach, R., Benoist, C., O'Hare, K. & Chambon, P. (1978) Proc. 31. Natl. Acad. Sci. USA 75, 4853-4857.
- Barnett, T. R., Pickle, W., Rae, P. M. M., Hart, J., Kamarck, M. & Elting, J. (1989) Am. J. Hum. Genet. 44, 890-893. 32.
- Ruoslahti, E. & Pierschbacher, M. D. (1986) Cell 44, 517-518. 33.