Vol. 154, No. 1, 1988 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
July 15, 1988 Pages 130-137

THE HUMAN PREGNANCY-SPECIFIC β_1-GLYCOPROTEIN (PSβG) AND THE CARCINOEMBRYONIC ANTIGEN (CEA)-RELATED PROTEINS ARE MEMBERS OF THE SAME MULTIGENE FAMILY

C. Streydio, K. Lacka*, S. Swillens and G. Vassart

Institut de Recherche Interdisciplinaire and Service de Génétique Médicale,
Université Libre de Bruxelles,
808 route de Lennik, B-1070 Brussels, Belgium

Received May 30, 1988

Pregnancy-specific β_1-glycoprotein (PSβG), a major product of the placenta with unknown function, consists of a set of glycoproteins synthesized by the syncytiotrophoblast. We report here the molecular cloning of 3 cDNA encoding different members of the PSβG family. Two clones (C, D) correspond to a single transcript undergoing differential splicing. The third one (E) originates from a different gene. All three clones have identical (C, D) or similar (E) coding sequences except for the last residues at their carboxyl end. They contain 93 residue motifs related to the ancestral Ig-like domain which makes them new members of this gene superfamily. A striking sequence similarity (50 to 60 %) is observed between PSβG and carcinoembryonic antigen (CEA)-related proteins. The evolutionary relationship between CEA and PSβG points to a possible common function in the control of cell invasion and/or metastasis.

Among placental proteins, pregnancy specific β_1-glycoprotein (PSβG) is a major heterogenous product of the syncytiotrophoblast, with unknown function (1). Like other placental proteins, PSβG is found in neoplasms of various origins (2) and, in some cases, can be used as a marker of malignancy (3). In fact, PSβG is not specific to the placenta (4) and has been found in different cellular types including fibroblasts and granulocytes (5, 6).

We have isolated cDNA clones from a cDNA library of a human term placenta and describe here, the primary structure of 3 PSβG proteins encoded by 2 different genes. Surprisingly, they share a striking homology with the carcinoembryonic (CEA)-related proteins which makes them members of related multigene families.

*Permanent address: Department of Endocrinology,
University School of Medecine of Poznañ, Poznañ, Poland.

Abbreviations: PSβG, pregnancy specific β_1 glycoproteins; CEA, carcinoembryonic antigen; NCA, nonspecific cross-reacting antigen.

0006-291X/88 $1.50

MATERIALS AND METHODS

Construction and screening of the cDNA library. A full-term human placenta was used to construct a cDNA expression library in the λgt 11 vector (7) from a 14 to 19 S mRNA fraction enriched in PSβG mRNA activity as determined by a *in vitro* translation (not shown). The library was screened, first with a rabbit anti PSβG antiserum (RIA Dakoppats-Promega) and secondly, with a probe corresponding to the common 5' portion of the clones (see results)
Subcloning and sequencing of cDNAs. Full-length clones were subcloned in the Eco R1 site of M13mp18 and sequenced (8) using synthetic oligonucleotides (Applied Biosystem 381A) as primers and the exonuclease III digestion method (9).
Sequence analysis. The Protein Identification Resource databank (release 15.0, Dec 87) was searched for similarities using the FASTP programm (10). Homologies in nucleic acids were identified as described (11).

RESULTS

Initial screening of the λgt 11 human placenta cDNA library with anti-PSβG antibody yielded two populations of clones (termed C and D) presenting an identical sequence over 1328 bp from their 5' end but with a different polyadenylated 3' segment (fig. 1). The corresponding transcripts originate most probably from alternative splicing combined with differential use of polyadenylation signals. They contain open reading frames with the potential to encode proteins of 428 (C type, Mr. 48,200) and 419 (D type, Mr. 47,200) amino acids, differing only at their carboxyl-terminal segment (fig. 1). The common primary structure of these polypeptides displays the internal repetition of a 93 residue motif with 72.4% and 50.0 % similarity at the nucleic acid and amino acid levels, respectively (fig. 1, 2, Table I).

While performing hybridization screening of the library to ensure that our "C-D" clones were complete, we came across a series of different clones sharing a strong sequence similarity with them. We report here the sequence of one of these, termed "E" (fig. 1) which encodes a putative polypeptide of 335 residues (predicted Mr. 37,200) containing only one of the 93 amino acid repeats present in C-D clones. Except for this, the overall sequence similarity between C-D and E is 86.7%. Interestingly, the similarity at amino acid level ceases abruptly at the point where the sequences of C and D diverge. Further downstream in 3' untranslated segments, a 126 bp. sequence is found showing a similarity of 85% between C and E clones (fig. 3).

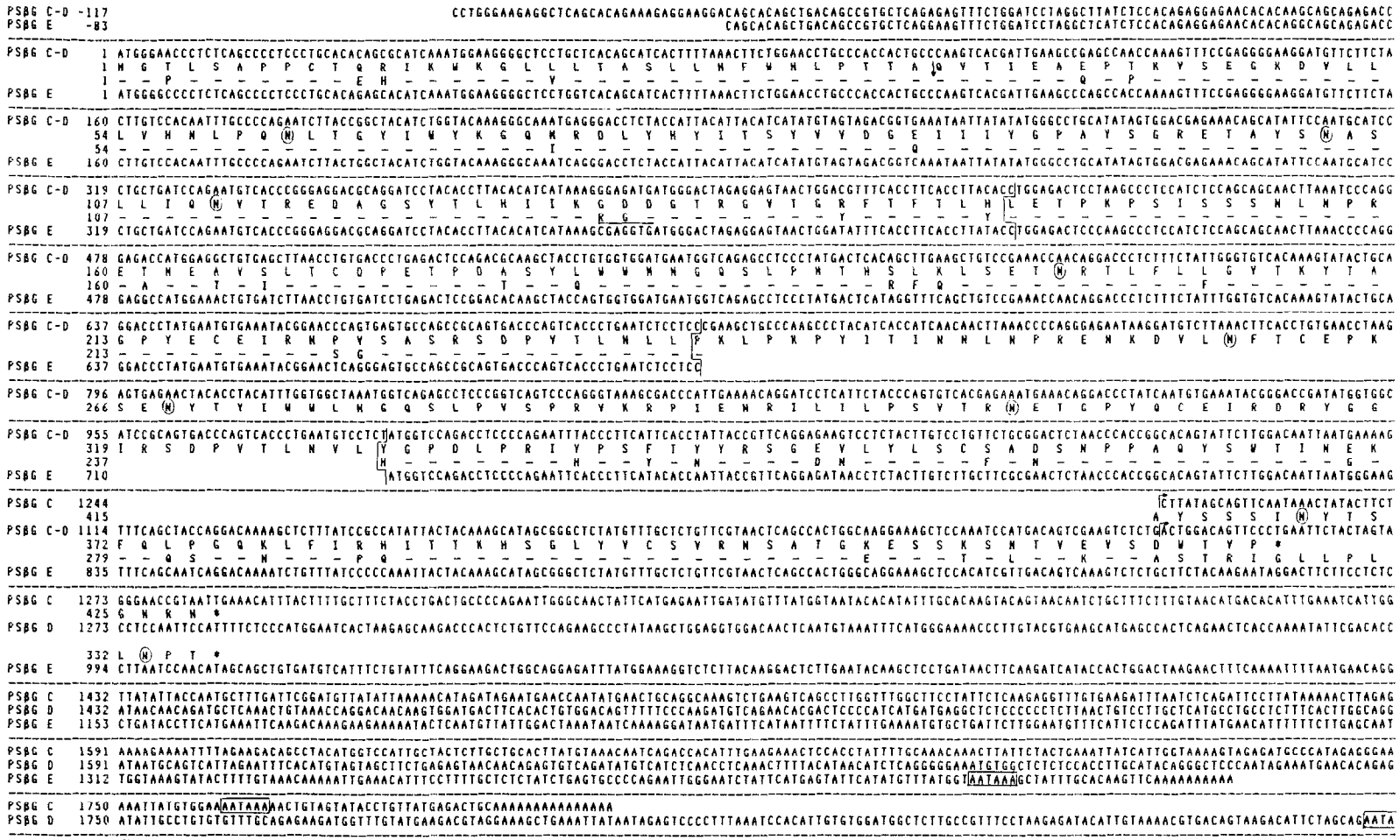

Fig. 1. Nucleotide and amino acid sequences of the PSβG clones. D and C clones begin at positions -117 and -86, respectively. The characteristics of the sequences are indicated as follows: N terminus of the mature protein (↓), RNA splice site (↱), stop codon (*), polyadenylation signal (▭), amino acid match (-), RGD peptide (RGD), potential N-linked glycosylation site (Ⓝ), limit between two repeated segments (⊏).

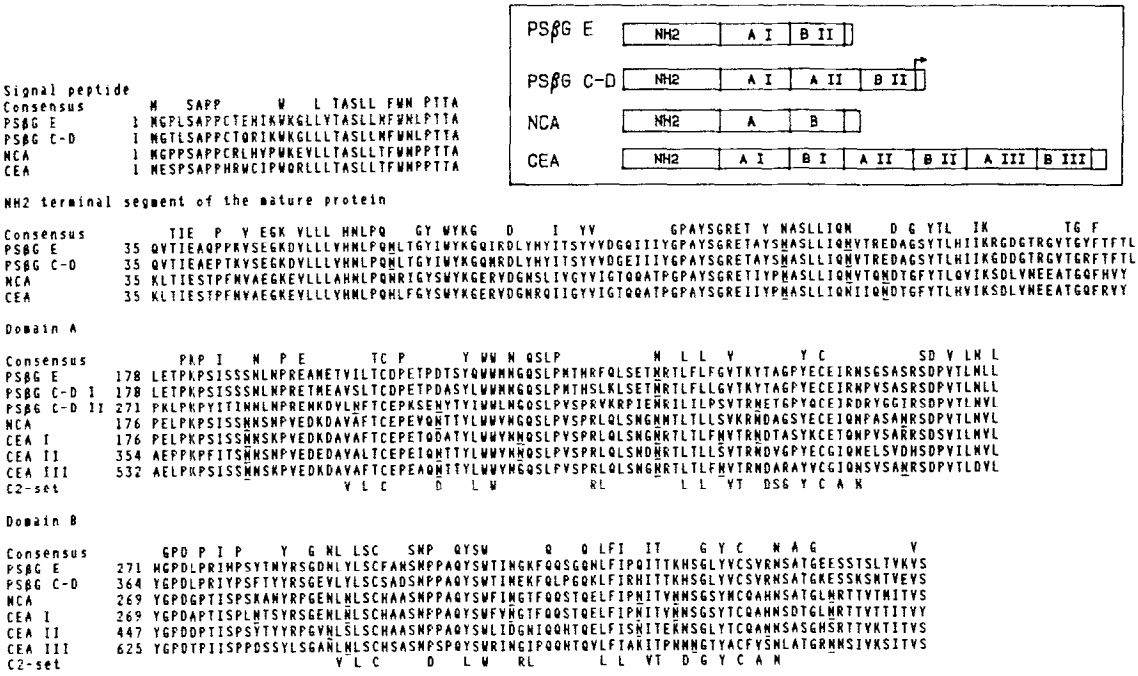

Fig. 2. Comparison of NH2 terminal segments, domains A and domains B between PSβG C-D and E, NCA (18) and CEA (17). The upper sequence indicates the perfect consensus between all the sequences. C2-set is the reported consensus sequence for one subgroup of the Ig superfamily (16). The potential N-linked glycosylation sites are underlined. Amino acid sequences diverge in the COOH terminal end following domain B. Structural protein organization is shown in inset.

Table I. Amino acid similarities between NH_2 terminal segments, domains A and domains B of PSβG C-D and E, CEA (17) and NCA (18). The structural organization of the proteins is showed in fig. 2. The number of matches is expressed as percent of the homology length.

NH_2 terminal

	NCA	PSβG C-D	PSβG E
CEA	83.6	58	57.2
NCA		58.7	59.4
PSβG C-D			92.3

Domain A

	CEA II	CEA III	NCA	PSβG C-D I	PSβG C-D II	PSβG E
CEA I	73.1	82.8	81.7	58.7	53.8	55.3
CEA II		76.3	78.5	56.5	56.7	55.4
CEA III			84.9	57.6	54.8	58.7
NCA				62	59.1	59.8
PSβG C-D I					50	88.2
PSβG C-D II						50

Domain B

	CEA II	CEA III	NCA	PSβG C-D	PSβG E
CEA I	73.8	60.7	86.9	54.8	60.2
CEA II		58.8	74.1	55.3	59.5
CEA III			60	48.2	51.2
NCA				54.1	60.7
PSβG C-D					79.8

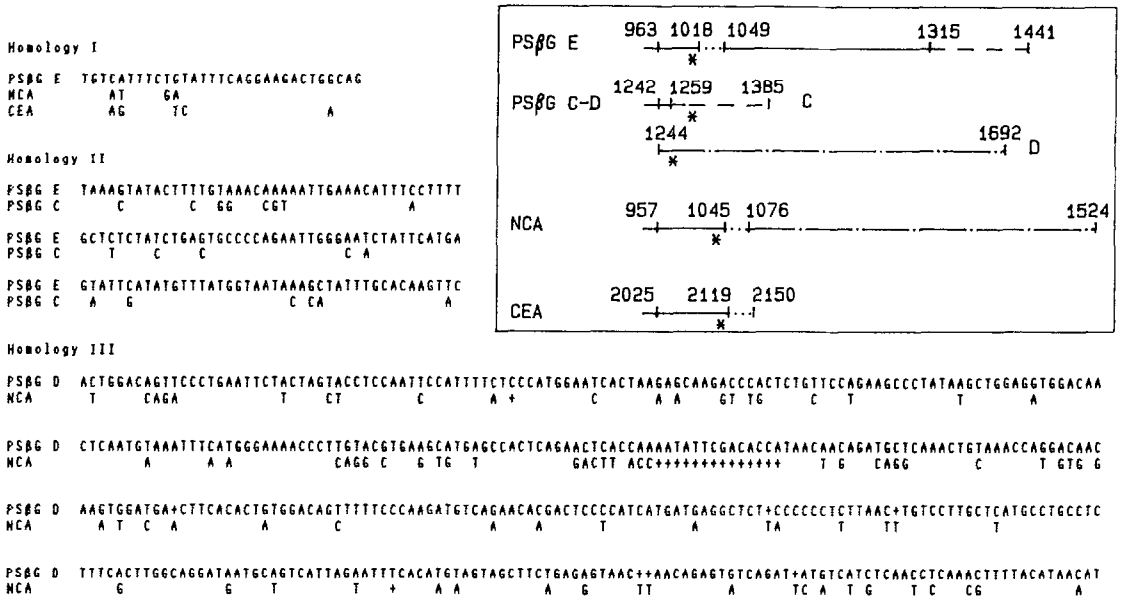

Fig. 3. Nucleotide alignements of similarities in the 3' end between PSβG C, D and E, CEA (17) and NCA (18) clones. All nucleotides are identical, except printed nucleotides and "+" representing mismatches and gaps, respectively. The inset shows the localisation of homologies I (....), II (_ _) and III (_._). The first number of each clone corresponds to the last base of the similarity between domains B. Stars indicate the stop codons.

In its NH_2 terminal region, the E sequence contains an Arg-Gly-Asp (RGD) tripeptide which has been implicated in recognition phenomena between cell surface receptors and extracellular proteins (12). Two others type of clones under current study have a similarly located RGD (not shown).

Computer screening of a protein data bank (see methods) revealed a significant sequence similarity between the 93 residue repeat and repeated motifs found in myelin-associated-glycoprotein (MAG) (13). MAG is a member of the Ig-gene superfamily (14) which includes a large set of proteins having evolved from a common 100 to 110 residue Ig-like domain (15). The 93 amino acid repeat of C-D and E contains several conserved residues (including the invariant cysteines) characteristic of one subgroup of the Ig superfamily (c2-set in ref 16) (fig. 2). This clearly makes PSβG a new member of the Ig superfamily. Comparaison of PSβG with individual proteins of the family, revealed a striking sequence similarity with the recently sequenced carcinoembryonic antigen (CEA)(17) and other related proteins NCA-55 (18), NCA-95 (19) (fig. 2). PSβG and CEA families share a very close structural

organisation with homologies involving signal peptides (34 residues), amino terminal domains (110-108 residues) and tandemly arranged doublets of Ig-like domains (A and B in fig. 2). The proteins differ mainly in the number of doublets they contain: one in NCA (18) and in PGβG E, three in CEA (17) and one-and-a-half in PSβG C-D (fig. 2). In spite of their identical organisation, PSβG E and NCA doublets share a higher similarity with PSβG C-D and CEA, respectively, which clearly defines two separate gene families (see details in Table I). Interestingly, the similarity extends to the 3' untranslated regions of the different mRNA's: a 31 bp. segment constituting the 3' border of the NCA-CEA homology, 13 bp. 3' from their stop codon, is found at a similar location in PSβG-E (fig. 3). More surprising, a 448 bp. segment constituting the 3' extremity of PSβG D is found with 76% similarity within the 3' untranslated segment of NCA, beginning precisely where NCA-CEA sequences become dissimilar (fig. 3)

DISCUSSION

Our results demonstrate that PSβG proteins are encoded by a multigene family comprising a minimum of four members of which two, C-D type and E type, are described here. At least one gene transcript (C-D type, fig. 1) is subjected to alternative splicing which helps explaining the structural heterogeneity of the protein (2). Watanabe and Chou recently reported the sequence of two PSβG cDNAs (20). One is virtually identical to our D clone, showing only three amino acid differences at position 41, 43 and 319, respectively, which could correspond to polymorphism. Our E clone, encoding a putative 37.2 kD protein, is a good candidate to account for the minor 36 kD PSβG-related peptide observed by these authors in *in vitro* translation experiments (20).

The PSβG multigene family clearly belongs to the Ig-gene superfamily (fig. 2). In particular, with an overal 50 to 60 % sequence similarity, it is very close to the CEA-NCA multigene family, suggesting a relatively recent common ancestor. Sequence similarity, on one hand, between PSβG and CEA-NCA polypeptides and between individual internal repeats (Table 1) and, on the

other hand, between the 3' non translated end of the differents clones (fig. 3) will allow the definition of an evolutionary tree (in preparation).

In the absence of any solid knowledge on the role of PSβG during pregnancy, this evolutionary relationship provides important clues regarding possible functions of the protein and may explain some of its structural features. The known tendancy of PSβG to form multimers or aggregates (21) might be explained by the aptitude of Ig-like domains to engage in homophilic or heterophilic interactions (15). Most members of this superfamily encode membrane proteins with demonstrated (poly Ig receptor, macrophage Fc receptor, T-cell receptor, MHC) or suspected (MAG, CEA, NCAM, Thy-1) roles in recognition phenomena (14). While there is no indication from presently available sequences that PSβG could have membrane-bound forms, it is tempting to speculate that the various members of PSβG family could play a role in the recognition phenomena involved in the invasion of the mother tissue by the trophoblast. The presence in three members of PSβG of the RGD tripeptide gives further support to this hypothesis. RGD, in extracellular matrix proteins, has been demonstrated as a signal recognized by specific cellular receptors and controlling cell adhesion and migration on subtrates (12). Considering the "tumoral" nature of placenta, a potential role for the high concentration of soluble PSβG during pregnancy might be searched in the light of experiments showing that RGD-containing peptides can inhibit tumor cell invasion (22).

From a more practical viewpoint, the high similarity between PSβG and CEA-NCA may have implications in the assay of these proteins by immunological methods. The possibility of immunological cross-reaction should be considered in analysing reports on the extratrophoblastic production of PSβG (4, 5). In granulocytes, two "NCA", with the same molecular weight as two PSβG-like proteins, have indeed been identified (6, 23).

The demonstration of close evolutionary and structural relationships between PSβG and CEA will undoubtedly orient future research on the role of both proteins. The availability of PSβG cDNA clones provides the tools to

test the various hypotheses regarding the functions of the protein in stably transfected cells and in transgenic mice.

ACKNOWLEDGEMENTS

We are grateful to Dr J.E. Dumont for continual support and interest. We greatly thank M. Georges for helpful discussion, Dr Ph. Simon for his aid in the initial steps, F. Libert and Dr Ch. Dinsart for their help in the preparation of the library, Ch. Christophe for the synthesis of oligo primers. This work was supported by PREST contract and the Bourse Hoguet, FRSM, CGER and Ministère de la Politique Scientifique (Sciences de la Vie).

REFERENCES

1. Bohn, D. and Dati, F. (1984) Proteins in Body Fluids, Amino Acids and Tumor Markers: Diagnostic and Clinical Aspects, Alan R. Liss, New York, pp. 333-374.
2. Sorensen, S. (1984) Tumour Biology, 5, 275-302.
3. Sakuragi, N. (1982) Gyn. Onc., 13, 393-398.
4. Heikinheimo, A., Walstrom, T., Lehto, V. and Stenman U. (1985) J. of Clin. Endocr. and Metabolism, 61, 188-191.
5. Rosen, S., Kaminska J., Calvert, I. and Aaronson, S. (1979) Am. J. Obstet. Gynecol., 134, 734-738.
6. Heikinheimo, M., Gahmberg, C.G., Bohn, H. and Andersson, L.C. (1987) Blood, 70, 1279-1283.
7. Young, R.A. and Davis, R.W. (1985) in :"DNA cloning". vol I, ed. D.M. Gloser, IRL Press, Oxford, pp. 49-78.
8. Sanger, F., Nickmem, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA, 74, 5463-5467.
9. Henikoff, S. (1984) Gene, 28, 351-359.
10. Lipman, D.J. and Pearson, W.R. (1985) Science, 227, 1435-1441.
11. Reich, J.G., Drabsch, H. and Däumler, A. (1984) Nucl. Acids. Res., 12, 5529-5543.
12. Ruoslahti, E. and Pierschbacher, M.D. (1987) Science, 238, 491-49.
13. Salzer, J.L., Holmes, W.P. and Colman, D.R. (1987) J. Cell Biol., 104, 957-965.
14. Hunkapiller, T. and Hood, L. (1986) Nature, 323, 15-16.
15. Amzel, L.M. and Poljak, R.J. (1979) Ann. Rev. Biochem., 48, 961-997.
16. Williams, A.F. (1987) Immun. Today, 8, 298-303.
17. Oikawa, S., Nakazato, H. and Kosaki, G. (1987) Bioch. Bioph. Res. Comm. 142, 511-518.
18. Neumaier, M., Zimmermann, W., Shively, L., Hinoda, Y., Riggs, A.D. and Shively, J.E. (1988) J. Biol. Chem., 263, 3202-3207.
19. Paxton, R.J., Mooser, G., Pande, H., Lee, T.D. and Shively, J.E. (1987) Proc. Natl. Acad. Sci. USA, 84, 920-924.
20. Watanabe, S. and Chou, J.Y. (1988) J. Biol. Chem., 263, 2049-2054.
21. Rosen, S.W., Clavert, I., Lee, N., Bohn H., Papadopoulos, N. and Osborne, J.C. (1986) Clin. Chem. Acta, 157, 65-72.
22. Gehlsen, K.R., Argraves, W.S., Pierschbacher, M.D. and Ruoslahti, E. (1988) J. Cell Biol., 106, 925-930.
23. Buchegger, F., Schreyer, M., Carrel, S. and Mach, J.-P. (1984) Int. J. Cancer, 33, 643-649.