The EMBO Journal vol.9 no.2 pp.407-414, 1990

hSP, the domain-duplicated homolog of pS2 protein, is co-expressed with pS2 in stomach but not in breast carcinoma

C.Tomasetto, M.-C.Rio, C.Gautier, C.Wolf, M.Hareuveni, P.Chambon and R.Lathe

LGME-CNRS and U184-INSERM, 11 Rue Humann, 67085 Strasbourg, France

Communicated by P.Chambon

Approximately 50% of human breast tumors secrete a small cysteine-rich protein, pS2, of unknown function. pS2 protein was recently found to be homologous to a porcine protein with hormonogastric activity, pancreatic spasmolytic polypeptide (PSP), in which the 5-cysteine domain present in pS2 is tandemly duplicated. We have characterized cDNA species encoding PSP and its human and mouse counterparts, hSP and mSP. We show that hSP and pS2 are separately encoded in the genome, and that the two proteins are co-expressed in normal stomach epithelium. However, expression of hSP was not detected in breast tumors. Computer analysis revealed that the pattern of conserved cysteine residues in hSP and pS2, the P domain, is present at the N termini of two other mammalian proteins, intestinal sucrase-isomaltase and lysosomal α-glucosidase.

Key words: breast cancer/glucosidase/hSP/pS2/spasmolytic/sucrase

Introduction

Breast cancer is a major cause of death in women, and although a large proportion of breast tumors initially respond to anti-estrogen therapy, progression to hormone-independence is a common event. The MCF-7 cell line, derived from a metastatic breast cancer pleural effusion (Soule *et al.*, 1973), was established in order to study the hormone-regulated growth of breast cancer cells. MCF-7 exhibits hormonal regulation *in vitro* (Brown *et al.*, 1984; Vignon and Rochefort, 1987 and references therein), and differential screening of cDNA libraries constructed from MCF-7 cells identified a mRNA species, pS2, whose expression is strongly induced by estrogen (Masiakowski *et al.*, 1982). pS2 protein and mRNA expression is detected in ~50% of human breast tumors, and its expression correlates well with hormono-dependent status (Rio *et al.*, 1987). pS2 mRNA encodes a small 84 amino acid, cysteine-rich secretory protein (Jakowlew *et al.*, 1984) and purified pS2 protein from MCF-7 culture supernatants has a mature size of 60 residues (Rio *et al.*, 1988c). Although pS2 gene expression in MCF-7 cells is principally regulated at the transcriptional level by estrogens (Brown *et al.*, 1984; Berry *et al.*, 1989), pS2 gene expression is also activated by the tumor promoter TPA, by *jun* or *ras* oncogene expression, and by epidermal growth factor (EGF) (Nunez *et al.*, 1989). These and other results have suggested a possible role for pS2 protein in autocrine or paracrine stimulation of breast tumor cell proliferation.

The biological role of pS2 protein is so far unknown. pS2 protein has not been detected in normal breast tissue prior to or during gestation and lactation (Rio *et al.*, 1988b). pS2 gene expression was, however, detected in normal human stomach mucosa, and pS2 protein is secreted in gastric juice (Rio *et al.*, 1988a). Gastric expression of the pS2 gene does not appear to be under estrogen control, because we have failed to detect estrogen receptor in these tissues (Rio *et al.*, 1988a; M.C.Rio, unpublished data); the EGF-responsive element of the pS2 gene regulatory region may instead control pS2 gene expression in stomach (Nunez *et al.*, 1989). To develop an easily accessible model for the involvement of pS2 in breast tumor development we first sought to identify a cognate murine gene. However, cross-hybridizing sequences were not detected in the mouse genome and immunological experiments have so far failed to reveal pS2 protein cross-reacting material in any tissue examined (C. Tomasetto, unpublished). Nevertheless, we and others recently noted a strong sequence homology between pS2 and a porcine pancreatic protein (Baker, 1988; Hoffman, 1988; Rio *et al.*, 1988a; Thim, 1988; see Figure 10). PSP, porcine pancreatic spasmolytic polypeptide, was initially identified as a contaminant in commercial insulin stocks prepared from pancreatic extracts (Jorgensen *et al.*, 1982a). Pharmacologic screening demonstrated that PSP has spasmolytic activity and inhibits gastric acid secretion in rodents (Jorgensen *et al.*, 1982b; Frandsen *et al.*, 1986; Frandsen, 1988); the protein primary sequence (Thim *et al.*, 1985) reveals that PSP harbors a tandem duplication of a 40-50 amino acid cysteine-rich domain with extensive homology to the single central domain present in pS2 protein. We undertook to identify human and mouse sequences homologous to PSP in order to cast light on the biological roles of pS2 and PSP, and to determine whether PSP might be the porcine counterpart of human pS2 protein.

Results

Characterization of PSP coding sequences

To determine the functional and evolutionary relationship between pS2 and PSP we first sought to obtain a molecular clone of the porcine PSP coding sequence. From the published protein sequence of PSP protein (Thim *et al.*, 1985) we used codon utilization data (Lathe, 1985) to design a unique 44mer oligonucleotide probe (see Figure 2). The utility of the probe was assessed by Northern hybridization at low stringency to pancreatic poly(A)$^+$ RNA; a single band at 650 nt was revealed in porcine pancreatic RNA (Figure 1A), whereas mouse RNAs yielded no detectable signal (not shown). The 44mer probe was used to screen a porcine pancreatic cDNA library constructed in λgt10. Three independent but cross-hybridizing recombinants were obtained (Figure 1B) and the nucleotide sequence of the

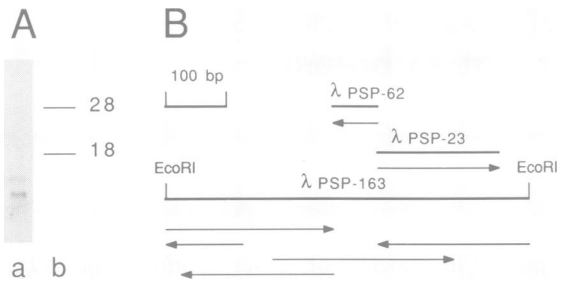


Fig. 1. Characterization of sequences encoding PSP. (**A**) Total porcine pancreatic poly(A)$^+$ RNA (5 μg) was resolved by gel electrophoresis in the presence of HCHO, transferred to nylon, and (**a**) probed with the unique 44mer probe designed for amino acid sequence data; (**b**) gives positions of 18S and 28S ribosomal RNA mol. wt markers. (**B**) PSP cDNA clones obtained and sequencing strategy.

longest insert (PSP-163, ~550 nt) was determined (Figure 2). This clone contains an uninterrupted open-reading frame encoding a protein with extensive homology to PSP, although three differences were detected between the published PSP sequence (Thim *et al.*, 1985; revised by Thim, 1989) and the mature polypeptide encoded by PSP-163 (see Figure 2 and Discussion). At the 5′ terminus the open reading frame extends beyond the N-terminal amino acid of mature PSP into a stretch of hydrophobic amino acids that are presumed to represent a secretion signal sequence. However, PSP-163 is 5′ truncated and does not contain a candidate ATG for PSP translation initiation. The 3′ terminus of PSP-163 appears to be complete, containing an untranslated region of 195 nt followed by a poly(A) tract.

cDNA cloning of the mouse PSP homolog, mSP

Molecular clones of the porcine PSP coding sequence provided hybridization probes to determine whether the mouse harbors a PSP homolog. cDNA probe PSP-163 was first used to screen mouse tissues for cross-hybridizing RNA species. Despite some cross-hybridization with 18S RNA, Northern analysis revealed a specifically hybridizing poly(A)$^+$ RNA species in two different mouse organs: stomach and (very weakly) pancreas (Figure 3, tracks St and Pa). Transcript size appeared to be identical in the two organs.

To characterize these messenger species, separate oligo(dT)-primed stomach and pancreas cDNA libraries were constructed and screened in parallel with the PSP-163 probe. In both cases a large number (0.01−0.1% of total clones screened) of independent cross-hybridizing clones were obtained; three stomach cDNA clones and two pancreatic cDNA clones were sequenced. The nucleotide sequences determined (Figure 4) were identical in the overlap regions (~400 nt, not shown) between cDNAs prepared from the two different organs. The encoded protein, referred to as mSP (for murine spasmolytic polypeptide) was found to be closely related to PSP (see Figure 6, below), and contains the tandem duplication of the cysteine-rich domain present as a single repeat in human protein pS2. A mSP-specific probe confirmed that expression in mouse is restricted to the stomach and the pancreas (not presented). The longest mSP cDNA clone (obtained from stomach RNA) is incomplete, lacking a 5′ ATG for translation initiation (Figure 4). The clone is also unusual in that the poly(A) tail is not preceded by the consensus motif AATAAA (see Proudfoot, 1984; Lewitt *et al.*, 1989 and references therein)

```
 1-GAG CCG CAG CGC CCG GCT CCT GGC CAT CCT CCT CCT GCT GGG-42
   Glu Pro Gln Arg Pro Ala Pro Gly His Pro Pro Pro Ala Gly

43-GCT GTG TGC CTG ACG GGG GCC CAG AAG CCC GCC GCC TGC CGG-84
   Ala Val Cys Leu Thr Gly Ala Gln Lys Pro Ala Ala Cys Arg
                              pGlu --- --- --- --- --- ---

85-TGT TCA CGC CAG GAC CCC AAG AAC CGG GTC AAC TGC GGA TTC-126
   Cys Ser Arg gln Asp Pro Lys Asn Arg Val Asn Cys Gly Phe
   --- --- --- --- --- --- --- --- --- --- --- --- --- ---

                                                  G ACG ACG
127-CCG GGC ATC ACC AGC GAC CAG TGC TTC ACC TCG GGG TGC TGC-168
   Pro Gly Ile Thr Ser Asp Gln Cys Phe Thr Ser Gly Cys Cys
   --- --- --- --- --- --- --- --- --- --- --- --- --- ---

     ←————44 mer Oligonucleotide ————————————→
   AAG CTG AGG TGG CAC GGG CCG CAC GGG ACC ACG AAG A
169-TTC GAC TCC CAA GTG CCT GGG GTC CCC TGG TGT TTC AAA CCC-210
   Phe Asp Ser Gln Val Pro Gly Val Pro Trp Cys Phe Lys Pro
   --- --- --- --- --- --- --- --- --- --- --- --- --- ---

211-CTC CCG GCG CAA GAG TCT GAG GAG TGC GTC ATG GAG GTG TCC-252
   Leu Pro Ala Gln Glu Ser Glu Glu Cys Val Met Glu Val Ser
   --- --- --- --- --- --- --- --- --- --- --- Gln --- ---

253-GCC CGC AAG AAC TGT GGC TAC CCG GGC ATC AGC CCC GAG GAC-294
   Ala Arg Lys Asn Cys Gly Tyr Pro Gly Ile Ser Pro Glu Asp
   --- --- --- --- --- --- --- --- --- --- --- --- --- ---

295-TGC GCA CGA CGC AAT TGC TGC TTC TCC GAC ACC ATC CCC GAG-333
   Cys Ala Arg Arg Asn Cys Cys Phe Ser Asp Thr Ile Pro Glu
   --- --- Ala --- --- --- --- --- --- --- --- --- --- ---

334-GTG CCC TGG TGC TTC TTC CCC ATG TCC GTG GAA GAC TGT CAT-375
   Val Pro Trp Cys Phe Phe Pro Met Ser Val Glu Asp Cys His
   --- --- --- --- --- --- --- --- --- --- --- --- --- ---

376-TAC TAA GAGGCCATGGCTTCAGAGAACCCACCAGGCTCCTGGATGTTCTGAAA-428
   Tyr ***
   ---

429-CAAAAGAAGACTATTCATCATCTTATAATTTCAGGTTTTGAGTTTCTGAAATTCC-482

483-TGGGTTTTCTTAATCTGCTTCCCCCTCTGCCCTTTTCAATGGTTTAACACAGAAT-534

535-TTCTTCAAATAAAGAAAACACTGAAAATCTGCTAAAAAAAAAAAAAAAAAAAAAA-586
```

Fig. 2. Sequence of PSP cDNA clone 163 and encoded polypeptide. Flanking linker sequences have been removed. The third line in each row gives the published PSP protein sequence (Thim *et al.*, 1985; Thim, 1989), identical residues are indicated by ---. The sequence of the 44mer probe is indicated (horizontal arrow) and the presumptive signal sequence and the polyadenylation signal are underlined. Despite sequencing both strands, ambiguities remained within a G/C-rich stretch toward the 3′ end of the signal sequence coding region that could possibly change the signal sequence reading frame.

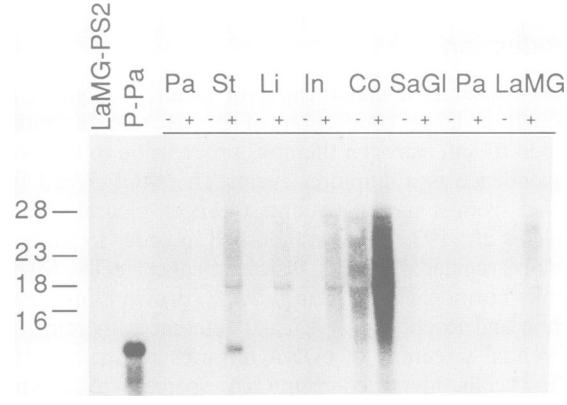


Fig. 3. Northern hybridization of mouse RNA with a PSP probe. Five micrograms of total poly(A)$^+$ (+) or poly(A)$^-$ (−) RNA samples from mouse organs were resolved by gel electrophoresis in the presence of HCHO, transferred to nylon, and probed with internally labeled cDNA PSP-163. Organs were: Pa, pancreas; St, stomach; Li, liver; In, intestine; Co, colon; SaGl, salivary gland; LaMG, 15 day lactating mammary gland. Reference samples were P-Pa, porcine pancreas, and LaMG-pS2, 15 day lactating mammary gland from a transgenic mouse line expressing pS2 protein in mammary gland under the control of the whey acidic protein promoter (Tomasetto *et al.*, 1989). Size standards (S values) are indicated.

for cleavage and polyadenylation of the precursor transcript, and is instead preceded by AAGTAAA at an equivalent position to the AATAAA found in the PSP cDNA clone.

```
1-          TG CGA CCT CGA GGT GCG CCC CTG CTG GCA GTG GTC-35
               Arg Pro Arg Gly Ala Pro Leu Leu Ala Val Leu

36-CTG GTT TTG GGA CTG CAT GCT CTG GTA GAG GGC GAG AAA CCT-77
   Leu Val Leu Gly Leu His Ala Leu Val Glu Gly Glu Lys Pro

78-TCC CCC TGT CGG TGC TCC AGG CTG ACA CCC CAC AAC AGA AAG-119
   Ser Pro Cys Arg Cys Ser Arg Leu Thr Pro His Asn Arg Lys

120-AAC TGT GGC TTC CCG GGC ATC ACC AGT GAG CAG TGC TTT GAT-161
    Asn Cys Gly Phe Pro Gly Ile Thr Ser Glu Gln Cys Phe Asp

162-CTT GGA TGC TGC TTT GAC TCT AGC GTC GCT GGG GTC CCT TGG-203
    Leu Gly Cys Cys Phe Asp Ser Ser Val Ala Gly Val Pro Trp

204-TGT TTC CAC CCA CTT CCA AAC CAA GAA TCG GAG CAG TGT GTC-245
    Cys Phe His Pro Leu Pro Asn Gln Glu Ser Glu Gln Cys Val

246-ATG GAA GTG TCA GCT CGC AAG AAT TGT GGG TAC CCG GGC ATC-287
    Met Glu Val Ser Ala Arg Lys Asn Cys Gly Tyr Pro Gly Ile

289-AGT CCC GAG GAC TGT GCC AGT CGA AAC TGC TGC TTT TCC AAC-329
    Ser Pro Glu Asp Cys Ala Ser Arg Asn Cys Cys Phe Ser Asn

330-CTG ATC TTT GAA GTG CCC TGG TGT TTC TTC CCA CAG TCT GTG-371
    Leu Ile Phe Glu Val Pro Trp Cys Phe Phe Pro Gln Ser Val

372-GAA GAT TGT CAC TAC TGA GAGTTGCTACTGCCGAGCCACCCGTTCCCTG-420
    Glu Asp Cys His Tyr ***

421-GGAGCTGGCAAGCCAGAAGAAAGTTTCAACCAGACTTCATCAATCTCTGGGGTTT-475

476-CTAAAACCATCTTGACCCTTAGCAGTGGCTAGACACAGCATTTTCCAAGTAAAGA-530

531-AAAGTTGAAAAAACCTGAAAAAAAAAAAAAAAAAAAAAAAAAA          -572
```

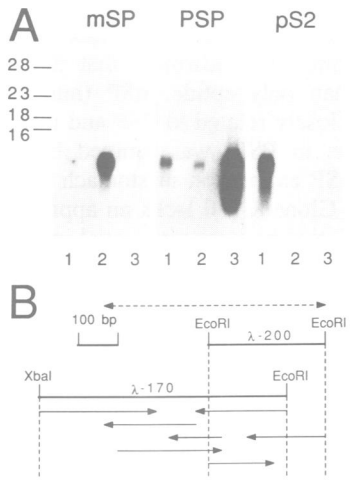
Fig. 4. Sequence of mSP cDNA and the encoded polypeptide. Flanking linker sequences have been removed. The presumptive protein secretion and mRNA polyadenylation signal sequences are underlined. The position of the poly(A) tract given derives from stomach RNA clone 302; in clone 312, produced by $d(T)_{15}$ priming of pancreas RNA, the poly(A) tail (33 nt) continues directly from the $(A)_5$ stretch (arrowed) immediately prior to the site used in clone 302 (not shown).

The same feature was noted in a second independent stomach cDNA clone (not shown).

Cloning of hSP cDNA

The murine (mSP) and porcine (PSP) genes encode proteins in which the human pS2 central domain is tandemly duplicated. PSP expression is pancreatic (Thim *et al.*, 1982), mSP expression is detected in both stomach and in pancreas (above), and pS2 protein has been detected both in breast tumor tissue and in stomach (Rio *et al.*, 1988a,b). To determine whether pS2 protein is the human homolog to PSP/mSP we first studied hybridization of human stomach mRNA to PSP and pS2-specific probes. At reduced hybridization stringency the PSP-specific probe (PSP in Figure 5A) revealed a specific cross-hybridizing species at ~650 nt in human stomach RNA (lane 1), slightly but distinctly larger than the pS2-specific transcript (600 nt) in the same tissue (Figure 5A, probe pS2, lane 1, and data not shown). Because the same probe failed to hybridize at the same stringency to RNA prepared from the pS2-producing breast cancer line MCF-7 (not shown), it was concluded that the PSP cross-hybridizing species in human stomach RNA did not correspond to pS2 RNA. As expected, the mSP probe revealed a very weakly cross-hybridizing RNA species in both human stomach (lane 1, and data not shown) and porcine pancreas (lane 3, and data not shown) and the PSP probe detected a similarly sized RNA species in mouse stomach RNA (lane 2).

To characterize the human stomach RNA species cross-hybridizing to PSP, a PSP-specific probe was used to screen a cDNA library prepared by random priming of total RNA from human stomach. Two independent cross-hybridizing recombinants from the stomach library (λ-170 and λ-200) were analyzed, and sequence analysis (as indicated in Figure 5B) revealed that λ-170 contains both a PSP-hybridizing



C

```
1-GC AGA CAT GGG ACG GCG AGA CGC CCA GCT CCT GGC AGC GCT-38
     Arg His Gly Thr Ala Arg Arg Pro Ala Pro Gly Ser Ala

39-CCT CGT CCT GGG GCT ATG TGC CTG GCG GGG AGT GAG AAA CCC-80
   Pro Arg Pro Gly Ala Met Cys Leu Ala Gly Ser Glu Lys Pro

81-TCC CCC TGC CAG TGC TCC AGG CTG AGC CCC CAT AAC AGG ACG-122
   Ser Pro Cys Gln Cys Ser Arg Leu Ser Pro His Asn Arg Thr

123-AAC TGC GGC TTC CCT GGA ATC ACC AGT GAC CAG TGT TTT GAC-164
    Asn Cys Gly Phe Pro Gly Ile Thr Ser Asp Gln Cys Phe Asp

165-AAT GGA TGC TGT TTC GAC TCC AGT GTC ACT GGG GTC CCC TGG-206
    Asn Gly Cys Cys Phe Asp Ser Ser Val Thr Gly Val Pro Trp

207-TGT TTC CAC CCC CTC CCA AAG CAA GAG TCG GAT CAG TGC GTC-248
    Cys Phe His Pro Leu Pro Lys Gln Glu Ser Asp Gln Cys Val

249-ATG GAG GTC TCA GAC AGA AGA AAC TGT GGC TAC CCG GGC ATC-290
    Met Glu Val Ser Asp Arg Arg Asn Cys Gly Tyr Pro Gly Ile

291-AGC CCC GAG GAA TGC GCC TCT CGG AAG TGC TGC TTC TCC AAC-332
    Ser Pro Glu Glu Cys Ala Ser Arg Lys Cys Cys Phe Ser Asn

333-TTC ATC TTT GAA GTG CCC TGG TGC TTC TTC CCG AAC TCT GTG-374
    Phe Ile Phe Glu Val Pro Trp Cys Phe Phe Pro Asn Ser Val

375-GAA GAC TGC CAT TAC TAA GAGAGGCTGGTTCCAGAGGATGCATCTGGCT-422
    Glu Asp Cys His Tyr ***

423-CACCGGGTGTTCCGAAACCAAGAAAGAAACTTCGCCTTATCAGCTTCATATTTCA-479

480-TGAAATCCTGGGTTTTCTTCATTTTCTTAACACTCTTTTCAATGGTTTAACATAT-536

537-AATTTCTTTAAG                                          -549
```

Fig. 5. Characterization of hSP coding sequences. (**A**) Northern hybridization of mSP, PSP and pS2 cDNA probes (indicated above) to: **lane 1**, 20 μg total human stomach RNA; **lane 2**, 20 μg total mouse stomach RNA; and **lane 3**, 5 μg of porcine pancreas poly(A)$^+$ RNA. Probes were cDNA clones 302 (mSP), 163 (PSP) and a complete pS2 cDNA (Jakowlew *et al.*, 1984). (**B**) Sequencing strategy for hSP cDNA clones. (**C**) hSP coding sequence (the region presented is indicated by the double-headed arrow in B); the putative signal sequence of the encoded protein is underlined (a sequencing ambiguity similar to that found in PSP cDNA was present in the hSP cDNA clones; see Figure 2 legend).

```
      +1                                                    57
PSP-  QKPAACRCSRQDPKNRVNCGFPGITSDQCFTSGCCFDSQVPGVPWCFKPLPAQESEE
mSP-  E--SP-----LT-H--K---------E---DL------S-A------H---N----Q
hSP-  E--SP-Q---LS-H--T-------------DN------S-T------H---K---DQ
            58                                                106
PSP-        CVMEVSARKNCGYPGISPEDCARRNCCFSDTIPEVPWCFFPMSVEDCHY
mSP-        ----------------------S------NL-F--------Q-------
hSP-        ------D-R----------E--S-K----NF-F--------N-------
```

Fig. 6. Sequence conservation between hSP, mSP and PSP. Amino acid sequences of the three mature polypeptides determined from cDNA sequence data have been aligned to maximize homology between the proteins and between the two tandemly repeated domains. Identical residues are indicated (−), and cysteine residues are underlined.

segment and an unrelated cDNA segment at its 5′ end, possibly introduced during the cloning procedure (not shown). However, the amino acid sequence of the protein

encoded by the unique long open reading frame present in λ-170 and λ-200 (Figure 5C) was homologous to both PSP and mSP (Figure 6), confirming that the cDNA clones encode a human polypeptide, hSP (human spasmolytic polypeptide), closely related to PSP and mSP. (The prefix 'pancreatic', as in PSP, was omitted because we have evidence for hSP expression in stomach but not so far in the pancreas.) Clone λ-170 lacks an appropriate ATG for translation initiation because a potential initiation codon (nt 6–8 in Figure 5C) is out-of-frame with the downstream coding sequence.

pS2 protein and hSP are separately encoded in the genome

pS2 protein and hSP are highly homologous (see Figure 10, below) and we wished to confirm that they are the products of distinct genes. Digests of human genomic DNA were resolved by gel electrophoresis, Southern-blotted to nylon, hybridized with a hSP-specific probe and exposed for autoradiography. The filter was subsequently stripped and rehybridized with a pS2-specific probe. In Figure 7 the two probes identify a non-overlapping pattern of bands, confirming that pS2 protein and hSP are encoded by distinct genes.

Co-expression of pS2 and hSP mRNAs in normal stomach but not in breast tumors

pS2 gene expression was first described in human breast cancer cells (Masiakowski *et al.*, 1982) and ~50% of breast tumor samples produce pS2 protein. We therefore explored whether hSP is similarly expressed in breast tumor cells. Northern hybridization analysis using hSP- and pS2-specific probes was performed upon RNA samples from human breast cancer cell lines MCF-7 and T47-D, cervical cancer HeLa cells, and from a variety of tissue samples. An RNA species hybridizing at high stringency to the hSP-specific probe was only observed in normal stomach (Figure 8A; n-St). RNA from two pS2-negative breast tumor biopsy samples, one pS2-positive breast cancer cell line (MCF-7) and one pS2-positive tumor biopsy (Br-Ca) all failed to hybridize (Figure 8B). Analysis of a further 11 pS2-positive breast tumor samples and two pS2-negative samples failed to produce any evidence of hSP expression (data not shown),

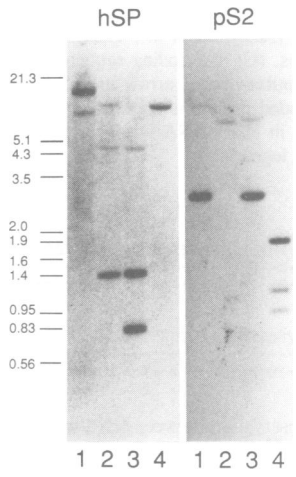


Fig. 7. Southern analysis of human genomic DNA with hSP- and pS2-specific cDNA probes. Restriction digests were: **lane 1**, *Eco*RI; **lane 2**, *Xba*I; **lane 3**, *Eco*RI + *Xba*I; **lane 4**, *Pst*I. Size markers (kilobases) are indicated. The left-hand blot (hSP clone 200) was stripped prior to rehybridization with a pS2-specific probe (complete cDNA).

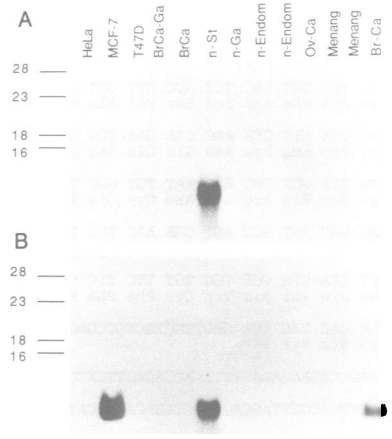


Fig. 8. Northern analysis of human normal or tumor tissues for transcripts hybridizing with probes specific for (**A**) hSP or (**B**) pS2. RNA samples (20 μg, total RNA) were: HeLa, established human cervical cancer cell line; MCF-7 and T47D, established cell lines from human breast cancer metastases; BrCa-Ga, metastatic ganglion from a breast cancer patient; BrCa, primary breast tumor biopsy; n-St, normal stomach; n-Ga, normal breast ganglion; n-Endom, normal endometrium; Ov-Ca, ovarian carcinoma biopsy; Menang, two independent primary menangioma biopsy samples. Probes were hSP clone 200 and a complete pS2 cDNA. Size markers are given as S values.

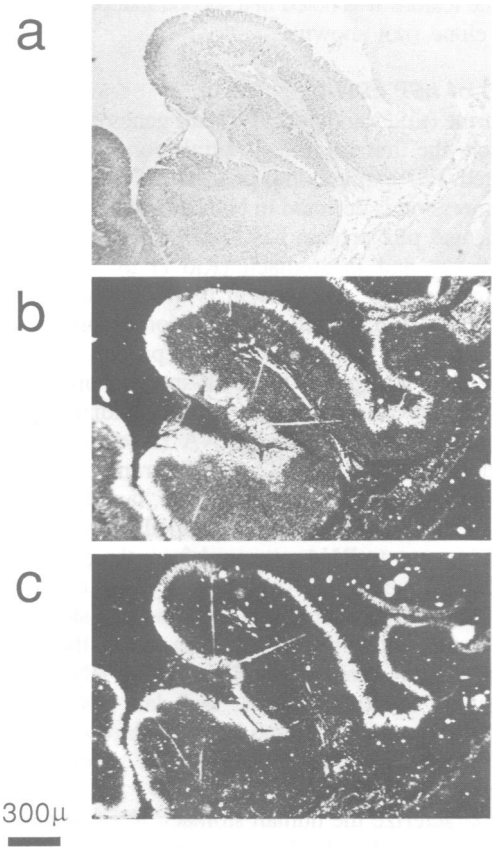


Fig. 9. *In situ* hybridization of serial sections of human stomach antrum epithelium, (**a**) hematoxylin-stained section; hybridization to antisense probes specific for (**b**) hSP, or (**c**) pS2 RNAs; probe sequences were hSP clone 200 or a complete pS2 cDNA.

indicating that hSP is not co-expressed with pS2 in breast tumor cells.

Because both the hSP and pS2 genes are transcribed in normal human stomach we determined whether the same cells are responsible for the synthesis of both proteins. Previous results using immunological staining localized pS2 to the stomach epithelium (Rio *et al.*, 1988a). *In situ* hybridization employing pS2 or hSP-specific 'antisense' RNA probes was used to analyze paraffin sections of normal human stomach for expression. Strong expression of both pS2 and hSP mRNAs was detected in the same epithelial mucosa layer of the stomach antrum (Figure 9); no hybridization was obtained using control 'sense' probes (not shown). The pattern of expression was identical for both pS2 and hSP, expression being restricted to the superficial epithelium regardless of type of mucous cell (parietal, chief or mucous neck cells). No signal was detected in the underlying cells of the submucosa, the muscularis or the serosa, and no hybridization was observed in other sections of the normal human digestive tract (duodenum, ileum and colon; not shown). Because suitable human pancreas RNA samples were not available, we were unable to determine whether hSP is expressed, as in mouse, in pancreas as well as in stomach.

The P domain, a protein motif related to the toxin–hemagglutinin fold

pS2 protein (Jakowlew *et al.*, 1984), the three members of the PSP group (PSP, Thim *et al.*, 1985; Thim, 1989; mSP and hSP, this work), and a *Xenopus* protein xSL (Hoffmann, 1988), all share a common core of five conserved cysteine residues in the arrangement C-C-CC-C. Computer searching (not described) of the NBRF database was performed to identify other proteins containing a similarly spaced five-cysteine motif. In this searching only two further proteins, sucrase-isomaltase (SIM) and α-glucosidase (AG), were found that contain the motif (Figure 10A).

SIM is a membrane-associated enzyme of the intestinal brush border that appears to have undergone internal duplication during the course of evolution (Semenza, 1986; Hunziker *et al.*, 1986). SIM is synthesized as a large precursor molecule that is proteolytically cleaved at the cell surface to yield a N-terminal membrane-bound subunit with sucrase activity; the homologous C-terminal subunit possesses isomaltase activity and remains non-covalently associated to the membrane-bound sucrase (reviewed by Semenza, 1986). α-Glucosidase (acid maltase) is a lysosomal enzyme that is homologous to both SIM subunits (Hoefsloot *et al.*, 1988). In SIM and AG the region homologous to the

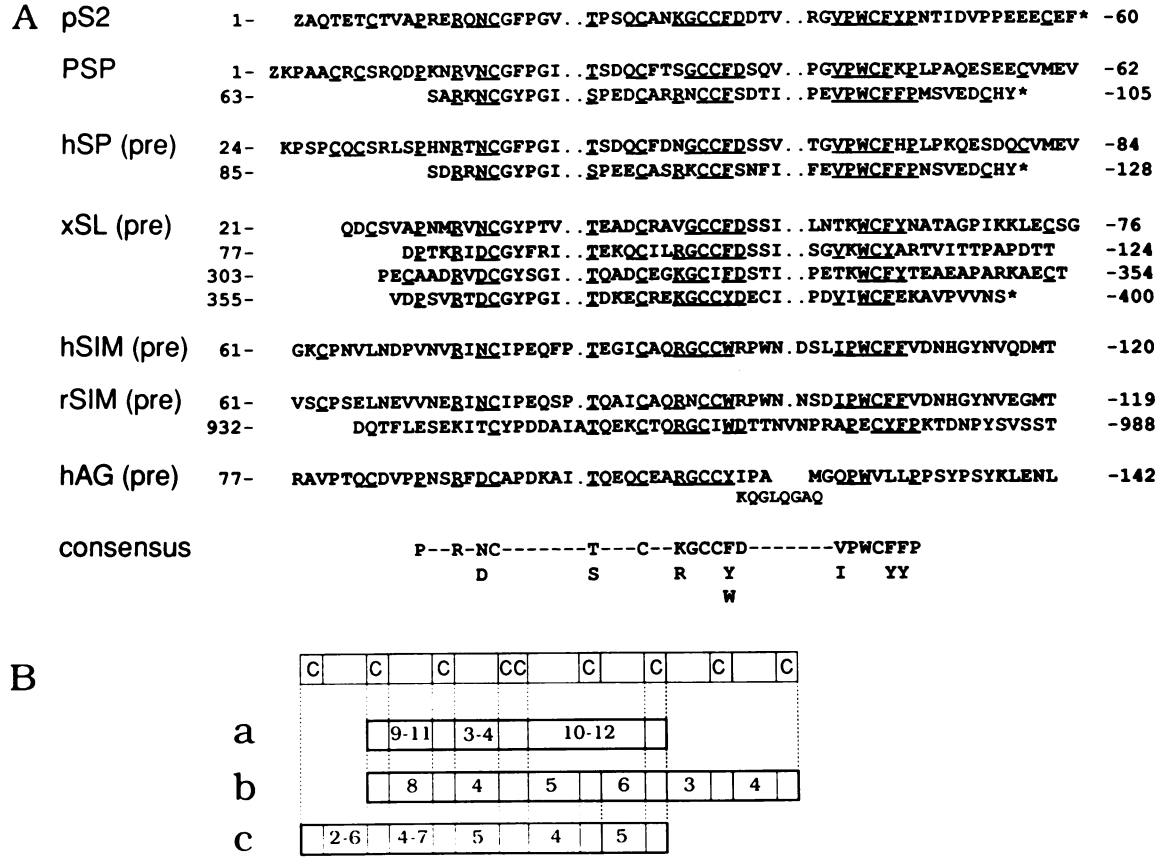


Fig. 10. (**A**) The P domain. Proteins have been aligned to maximize homologies, a period (.) indicates a gap introduced for alignment. Sequences are pS2, human pS2 protein; PSP, porcine pancreatic spasmolytic polypeptide sequence deduced from the cDNA nucleotide sequence; hSP, human spasmolytic polypeptide; xSL, *Xenopus laevis* spasmolysin; hSIM, human sucrase-isomaltase; rSIM, rabbit sucrase-isomaltase; hAG, human α-glucosidase. For literature references see text. Numbers give amino acid positions in the mature protein or in the precursor (pre) deduced from the coding sequence. Z, pyroglutamic acid; *translation stop. Only the regions homologous to the P domain are presented. Amino acids conforming to the consensus, and all cysteine residues including those outwith the consensus, are underlined. (**B**) Similarity to the toxin–agglutinin fold. Distribution of cysteine residues in (**a**) the P domain; (**b**) four repeat domains of wheat agglutinin isolectin (NBRF sequence AEWT2, Wright *et al.*, 1984); (**c**) two repeat domains of the bovine vasopressin-neurophysin precursor (sequence NVBO2, Land *et al.*, 1982).

pS2/hSP motif is located towards the N terminus of each enzyme subunit, and is present once in AG and twice in SIM (at the N termini of the sucrase and isomaltase subunits).

Because the 35–45 amino acid, five-cysteine motif is highly conserved in pS2 and in the PSP family (PSP, hSP and mSP) we have termed this motif the P domain. The homology is not restricted to the conserved cysteine residues, and certain other residues are rigorously conserved between the identified proteins containing the P domain (Figure 10A).

The arrangement of cysteine residues within the P domain also resembles that found in a number of other proteins, particularly those adopting a protein structure referred to as the toxin–agglutinin fold (Drenth *et al.*, 1980) (Figure 10B). The similarity may, however, reflect convergence upon a favorable pattern of disulfide bonds rather than a common evolutionary origin.

Discussion

Differential screening of a cDNA library prepared from hormone-treated MCF-7 cell lines identified a mRNA species, pS2, whose transcription is strongly induced by estrogen (Masiakowski *et al.*, 1982; Jakowlew *et al.*, 1984). pS2 mRNA encodes a small 60 amino acid cysteine-rich secreted protein, and the possibility that pS2 protein might play an autocrine growth regulatory role in breast tumor development has been discussed (Jakowlew *et al.*, 1984; Jeltsch *et al.*, 1987). Although pS2 gene expression is observed in ~50% of human breast tumors (Rio *et al.*, 1987), we have failed to detect pS2 protein expression in normal breast tissue (Rio *et al.*, 1988a) and the biological role of pS2 is so far unknown.

To explore the function of pS2 we first attempted to identify a cognate murine gene; however, no pS2-hybridizing sequence has so far been detected in the mouse genome or in mammary gland RNA and no pS2 protein cross-reactive material was detected in any mouse tissue sample so far examined (unpublished). Computer searching, however, revealed a strong sequence homology between pS2 protein and a porcine pancreatic spasmolytic polypeptide (PSP) in which the cysteine-rich domain of pS2 is tandemly duplicated (Rio *et al.*, 1988a).

To determine if PSP is the porcine counterpart of pS2 protein, we first used a probe based upon the PSP protein coding sequence to identify molecular clones encoding PSP. Sequence analysis revealed six differences between the deduced sequence of the encoded protein and the published (Thim *et al.*, 1985) PSP amino acid sequence. Although four of these differences have been eliminated by more recent protein sequence data (Thim, 1989), the sequence reported here differs from the published PSP sequence at two positions —a glutamine to glutamic acid (Q-E) change at position 61 of the mature protein, and an alanine to arginine (A-R) change at position 80. These differences may represent allelic variants; however, the second alteration (A-R) is encoded by a small region that was not sequenced on both DNA strands. Although PSP, like pS2 protein (Rio *et al.*, 1988c), is thought to harbor a pyro-glutamic acid (pyro-Glu; pyrollidone carboxylic acid) residue at its mature N terminus (Thim, 1989), the PSP cDNA sequence predicts an N-terminal glutamine residue instead of glutamic acid, the common pyro-Glu precursor. However, both glutamic acid and glutamine can be cyclized to generate pyro-Glu (see Wold, 1981).

The molecular clone encoding PSP was used to probe mouse tissues for expression, and specific cross-hybridizing mRNA species were observed in both pancreas and stomach. Sequencing of cDNA clones revealed that the same messenger RNA species appears to be expressed in both stomach and pancreas, and the encoded stomach and pancreatic polypeptides have identical primary amino acid sequences. The mouse protein species is referred to here as murine spasmolytic polypeptide, mSP. Because suitable human pancreatic RNA samples are not readily obtainable, we used the PSP molecular clone to probe human stomach RNA. A major cross-hybridizing RNA species was detected, and cDNA cloning was performed. Nucleotide sequence analysis indicated that the cross-hybridizing RNA species from human stomach encodes a close relative of PSP, and this protein is referred to here as hSP (human spasmolytic polypeptide).

Both hSP and pS2 protein are homologous to PSP (see Figure 10); however, hSP presents the domain duplication observed in PSP. Further, the degree of homology of hSP with PSP (76% in the mature protein) is rather greater than with pS2 protein (65 and 54% of the first and second domains respectively), and hSP is presumed to be the human equivalent of PSP. Genomic Southern blotting confirmed that pS2 protein and hSP are separately encoded in the human genome.

The availability of pS2- and hSP-specific probes permitted comparative analysis of their expression. We previously reported pS2 gene expression in normal human stomach, and *in situ* hybridization with hSP-specific probes revealed that hSP is co-expressed with pS2 in the surface epithelium of the stomach antrum. We also examined the expression of hSP in breast tumor samples. RNA hybridizing to an hSP probe was not detected in a range of tumor samples and analysis of 12 independent pS2-positive breast cancer biopsy samples failed to provide evidence for co-expression of hSP with pS2.

The biological roles of pS2 and hSP are so far unknown. The strong conservation of primary structure between PSP, mSP and hSP suggests that the three proteins fulfil similar biological functions. PSP is reported to act as a gastric hormone, inhibiting gastric acid secretion and gastrointestinal motility in rodents (Jorgensen *et al.*, 1982b; Frandsen *et al.*, 1986; Frandsen, 1988). These authors have also reported recently that PSP may have a weak growth-stimulatory effect upon H-119 colon tumor cells and upon cultured breast cancer MCF-7 cells, and an interaction with a putative pS2 receptor was discussed (Hoosein *et al.*, 1989). In contrast, we have failed to observe growth-promoting activity of pS2 protein upon MCF-7 cells (M-.C.Rio, unpublished data) and transgenic mice expressing pS2 in lactating mammary gland (and their suckling progeny) fail to exhibit detectable pathology (Tomasetto *et al.*, 1989). These observations argue against an autocrine role for human pS2 protein in growth stimulation of the mammary epithelium.

Although the primary structures of pS2 protein and hSP diverge, PSP, hSP, mSP and pS2 all exhibit striking conservation of a pattern of cysteine residues. This motif is also present in a protein ('spasmolysin', SL) of unknown function from *Xenopus* skin (Hoffmann, 1988). We report here that a very similar motif is present at the N terminus

of α-glucosidase (acid maltase) (Hoefsloot *et al.*, 1988) as well as at the N termini of both subunits of mammalian sucrase-isomaltase (Semenza, 1986; Hunziker *et al.*, 1986). We refer to this conserved five-cysteine motif as the P domain.

The disposition of cysteine residues within the P domain also resembles that found in the toxin–agglutinin fold (Drenth *et al.*, 1980). However, structural similarities may reflect either a common evolutionary origin or convergence upon a favorable disposition of cysteine residues and disulfide bonding (required to stabilize small protein domains that lack a large hydrophobic core; Drenth *et al.*, 1980). Conservation of cysteine residues does not necessarily reflect a present-day common functionality—for instance, the epidermal growth factor, pancreatic secretory trypsin inhibitor and prothrombin possess structural similarities with one another (see Strydom, 1977).

The primary sequence homologies between pS2 protein, the PSP group (PSP, mSP and hSP), xSL, SIM and AG are not, however, restricted to the cysteine residues: a number of other residues are strongly conserved throughout the P domain family (see Figure 10A). It is particularly noteworthy that a similar alteration of a single cysteine residue within the central P domain core sequence K/R,G,C,C,F/W/Y found in the *Xenopus* protein (K,G,C,I,F) is repeated in SIM (R,G,C,I,W), because a codon determining isoleucine (AUU, AUC, AUA) cannot be generated in one step from a codon determining cysteine (UGC, UGU). This result implies that similar evolutionary pressures have operated on both xSL and SIM, possibly indicating similar functionality.

The observed sequence homologies between hSP, pS2 protein and the carbohydrate-degrading enzymes SIM and AG argue that hSP and pS2 might play a role in sugar breakdown, raising the possibility that pS2 protein secreted by breast tumor cells might contribute to degradation of the extracellular glycoprotein matrix. Further experiments employing purified pS2 and hSP will be required to test this hypothesis.

Materials and methods

Isolation of RNA and Northern hybridization

Tissue samples were rapidly frozen in liquid nitrogen, and RNA was prepared by the LiCl procedure (Auffray and Rougeon, 1980). Poly(A)-containing RNA was purified by oligo(dT)–cellulose chromatography. RNAs were fractionated by agarose gel electrophoresis in the presence of formaldehyde (Lehrach *et al.*, 1977) and transferred to nylon membranes (Hybond-N, Amersham). Filters were acidified (10 min, 10% CH_3COOH) and stained (10 min, 0.004% methylene blue, 0.5 M CH_3COONa, pH 5.0) prior to hybridization. Synthetic oligonucleotides were radiolabeled at their 5′ termini using T4 polynucleotide kinase and [γ-^{32}P]ATP, and double-stranded DNA was labeled by random-priming of DNA polymerase I (Feinberg and Vogelstein, 1983) with synthetic hexamers in the presence of [α-^{32}P]deoxynucleotides. Northern blots were hybridized with the 44mer oligonucleotide probe (2×10^7 c.p.m./ml) at 42°C in 4 × SET, 0.1% SDS, 0.1% sodium pyrophosphate, 100 μg/ml denatured salmon sperm DNA, 2 mg/ml heparin (Figure 1) (1 × SET is 0.15 M NaCl, 1 mM Na_2EDTA, 10 mM Tris–HCl, pH 7.5). For double-stranded probes (2×10^6 c.p.m./ml) the hybridization medium was supplemented with 30% formamide (Figures 3 and 5) or 50% formamide (Figure 8). Washes were performed at 50°C in 2 × SSC, 0.1% SDS (Figures 3 and 5) or in 0.1 × SSC, 0.1% SDS (Figure 8) (1 × SSC is 0.15 M NaCl, 0.015 M sodium citrate).

cDNA library construction and screening

cDNA synthesis was performed essentially as described (Huynh *et al.*, 1985). First-strand cDNA synthesis using AMV reverse transcriptase in the presence of [α-^{32}P]dCTP was primed with $(dT)_{15}$ (porcine pancreas, mouse pancreas and stomach RNA) or with random hexamers (human stomach RNA); second-strand synthesis used DNA polymerase I. Double-stranded cDNA molecules were blunted by S1 nuclease and T4 DNA polymerase treatment. Internal *Eco*RI sites were protected by methylation prior to terminal addition of phosphorylated *Eco*RI linkers. After *Eco*RI digestion excess linkers were removed by Biogel A50M chromatography. cDNA peak fractions were ligated to *Eco*RI-digested λgt10. Recombinant genomes were packaged *in vitro* and selected by plating on *Escherichia coli* C600 *hfl hsdR*. 2×10^5, 1×10^7, 1×10^7 and 6×10^5 recombinants were obtained for the porcine pancreas, mouse pancreas, mouse stomach and the human stomach RNAs respectively. In each case $\sim 2 \times 10^5$ primary plaques were transferred to nylon filters (Pall, Biodyne, NY) and hybridized as described for Northern analysis, either with end-labeled oligonucleotide or with randomly primed nick-translated DNA.

Subcloning and sequencing

cDNA inserts of lambda recombinants were transferred to pGEM1 (Promega Biotec, Madison) or to M13tg130 or tg131 (Kieny *et al.*, 1983) for sequencing by the dideoxy protocol employing Sequenase (Tabor and Richardson, 1987) and dGTP analogs.

In situ hybridization

Histological staining and *in situ* hybridization with ^{35}S-labeled *in vitro* transcript probes was performed as described previously (Tomasetto *et al.*, 1989).

Southern hybridization

Total peripheral blood lymphocyte DNA was digested with restriction enzymes, resolved by agarose gel electrophoresis, and transferred to nylon (Hybond-N, Amersham, UK) according to standard procedures. Probe labeling was performed by random-priming of DNA polymerase I (Feinberg and Vogelstein, 1983) and hybridization was performed in 50% formamide, 5 × SSC, 0.1% SDS, 0.1% sodium pyrophosphate, 0.1% polyvinylpyrollidone, 0.1% Ficoll 400, 0.1% bovine serum albumin at 42°C. Washing was in 0.2 × SSC, 0.1% SDS, at 60°C.

Acknowledgements

We would like to thank R.Boukamel, K.Haffen, O.Bchini, N.Rouyer, J.N.Freund, F.Raul and C.Roitsch for helpful discussions, S.Ali for critical reading of the manuscript, B.Schuhbaur and I.Kuhn for technical assistance, and A.Staub and F.Ruffenach for oligonucleotide synthesis. We are particularly indebted to G.Zint for assistance in procuring porcine tissue samples. This work was supported by the INSERM, the CNRS, the Association pour la Recherche contre le Cancer, and the Fondation pour la Recherche Médicale. M.H. is the recipient of a long-term EMBO fellowship and C.G. is supported by a fellowship from the Agence Nationale de Recherches sur le SIDA.

References

Auffray,C. and Rougeon,F. (1980) *Eur. J. Biochem.*, **107**, 303–314.
Baker,M.E. (1988) *Biochem. J.*, **253**, 307–311.
Berry,M., Nunez,A.M. and Chambon,P. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 1218–1222.
Brown,A.M.C., Jeltsch,J.M., Roberts,M. and Chambon,P. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 6344–6348.
Drenth,J., Low,B.W., Richardson,J.S. and Wright,C.S. (1980) *J. Biol. Chem.*, **255**, 2652–2655.
Feinberg,A.P. and Vogelstein,B. (1983) *Anal. Biochem.*, **132**, 6–13.
Frandsen,E.K. (1988) *Regul. Peptides*, **20**, 45–52.
Frandsen,E.K., Jorgensen,K.H. and Thim,L. (1986) *Regul. Peptides*, **16**, 291–297.
Hoefsloot,L.H., Hoogeveen-Westerveld,M., Kroos,M.A., van Beeumen,J., Reuser,A.J.J. and Oostra,B.A. (1988) *EMBO J.*, **7**, 1697–1704.
Hoffmann,W. (1988) *J. Biol. Chem.*, **263**, 7686–7690.
Hoosein,N.M., Thim,L., Jorgensen,K.H. and Brattain,M.G. (1989) *FEBS Lett.*, **247**, 303–306.
Hunziker,W., Spiess,M., Semenza,G. and Lodish,H.F. (1986) *Cell*, **46**, 227–234.
Huynh,T.V., Young,R.A. and Davis,R.W. (1985) In Glover,D.M. (ed.), *DNA Cloning: A Practical Approach*. IRL Press, Oxford, Vol. 1, pp. 49–78.
Jakowlew,S.B., Breathnach,R., Jeltsch,J.M., Masiakowski,P. and

Chambon,P. (1984) *Nucleic Acids Res.*, **12**, 2861–2878.
Jeltsch,J.M., Roberts,M., Schatz,C., Garnier,J.M., Brown,A.M.C. and Chambon,P. (1987) *Nucleic Acids Res.*, **15**, 1401–1414.
Jorgensen,K.H., Thim,L. and Jacobsen,H.E. (1982a) *Regul. Peptides*, **3**, 207–219.
Jorgensen,K.D., Diamant,B., Jorgensen,K.H. and Thim,L. (1982b) *Regul. Peptides*, **3**, 231–243.
Kieny,M.P., Lathe,R. and Lecocq,J.P. (1984) *Gene*, **26**, 91–99.
Land,H., Schütz,G., Schmale,H. and Richter,D. (1982) *Nature*, **295**, 299–303.
Lathe,R. (1985) *J. Mol. Biol.*, **183**, 1–12.
Lehrach,H., Diamond,D., Wozney,J.M. and Boedtker,H. (1977) *Biochemistry*, **16**, 4743–4751.
Lewitt,N., Briggs,D., Gil,A. and Proudfoot,N.J. (1989) *Genes Dev.*, **3**, 1019–1025.
Masiakowski,P., Breathnach,R., Bloch,J., Gannon,F., Krust,A. and Chambon,P. (1982) *Nucleic Acids Res.*, **10**, 7895–7903.
Nunez,A.M., Berry,M., Imler,J.L. and Chambon,P. (1989) *EMBO J.*, **8**, 823–829.
Proudfoot,N. (1984) *Nature*, **307**, 412–413.
Rio,M-.C., Bellocq,J.P., Gairard,B., Rasmussen,U., Krust,A., Koehl,C., Calderoli,H., Schiff,V., Renaud,R. and Chambon,P. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 9243–9247.
Rio,M-.C., Bellocq,J.P., Daniel,J-.Y., Tomasetto,C., Lathe,R., Chenard,M.P., Batzenschlager,A. and Chambon,P. (1988a) *Science*, **241**, 705–708.
Rio,M-.C., Bellocq,J.P., Gairard,B., Koehl,C., Renaud,R. and Chambon,P. (1988b) *Biochimie*, **70**, 961–968.
Rio,M-.C., Lepage,P., Diemunsch,P., Roitsch,C. and Chambon,P. (1988c) *C.R. Acad. Sci. Paris*, **307-iii**, 825–831.
Semenza,G. (1986) *Annu. Rev. Cell. Biol.*, **2**, 255–313.
Soule,H.D., Vasquez,J., Long,A., Albert,S. and Brennan,M.J. (1973) *J. Natl. Cancer Inst.*, **51**, 1409–1413.
Strydom,D.J. (1977) *J. Mol. Evol.*, **9**, 349–361.
Tabor,S. and Richardson,C.C. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 4767–4771.
Thim,L. (1988) *Bichem. J.*, **253**, 309.
Thim,L. (1989) *FEBS Lett.*, **250**, 85–90.
Thim,L., Jorgensen,K.H. and Jorgensen,K.D. (1982) *Regul. Peptides*, **3**, 221–230.
Thim,L., Thomsen,J.K., Christensen,M. and Jorgensen,K.H. (1985) *Biochim. Biophys. Acta*, **827**, 410–418.
Tomasetto,C., Wolf,C., Rio,M-.C., Mehtali,M., LeMeur,M., Gerlinger,P., Chambon,P. and Lathe,R. (1989) *Mol. Endocrinol.*, **3**, 1579–1584.
Vignon,F. and Rochefort,H. (1987) In Moudgil,V.K. (ed.), *Recent Advances in Cancer Research*. de Gruyter, Berlin, pp. 405–425.
Wold,F. (1981) *Annu. Rev. Biochem.*, **50**, 783–814.
Wright,C.S., Gavilanes,F. and Peterson,D.L. (1984) *Biochemistry*, **23**, 280–287.

Received on November 9, 1989