BIOCHEMICAL LETTERS

Oestrogen-induced pS2 protein is similar to pancreatic spasmolytic polypeptide and the kringle domain

There is considerable interest in understanding the mechanism by which oestrogens act as enhancers of gene transcription because of their importance in reproductive function and because about one-third of breast cancers are oestrogen-dependent (Katzenellenbogen, 1980; McGuire, 1980). The MCF-7 cell line (Lippman & Bolan, 1975), a human breast cancer cell line, has been especially useful for understanding the mechanism of action of oestrogen. The pS2 protein ($M_r \sim 6500$) is an abundant oestrogen-induced protein secreted by MCF-7 cells and found in breast tumour tissue (Masiokowski *et al.*, 1982; Jakowlew *et al.*, 1984; Prud'Homme *et al.*, 1985; Nunez *et al.*, 1987), but not in non-malignant breast tissue. pS2's function is unknown. Using computer-based analyses of amino acid sequences, we find that: (1) pS2 is homologous to the pig pancreatic spasmolytic polypeptide, which represses gastrointestinal motility (spasmolytic effect) and inhibits pentagastrin-induced gastric acid secretion (Jørgensen *et al.*, 1982; Thim *et al.*, 1985; Frandsen *et al.*, 1986), and (2) both pS2 and pancreatic spasmolytic polypeptide are similar to the kringle domain, found in the non-catalytic region of plasminogen and several other serine proteases important for blood coagulation and fibrinolysis (Doolittle, 1985; Pathy, 1985; Young *et al.*, 1978). These similarities should provide clues for further studies to understand the role of pS2 in the growth of oestrogen-dependent tumours, as well as the functions of pancreatic spasmolytic polypeptide and the kringle domain.

The similarity between pS2 and pancreatic spasmolytic polypeptide was arrived at as a result of our interest in

the relationship between serine proteases and proteins involved in reproduction (Baker, 1985a,b; Baker *et al.*, 1987). A search of the protein database (over 6400 sequences, 1 550 000 residues) (Dayhoff *et al.*, 1983) with the triplet, Pro-Trp-Cys, a conserved element in the kringle domain, revealed that the pancreatic spasmolytic polypeptide and pS2 contained this triplet. This led to more detailed analysis described below quantifying the similarities among these proteins.

The 106-residue pancreatic spasmolytic polypeptide contains two domains, residues 14–49 and 63–98, which are 50% identical (Thim *et al.*, 1985). The pS2 protein has 58 residues (Nunez *et al.*, 1987). Fig. 1 shows the alignment of residues 1–49 of pS2 with residues 54–102 of pancreatic spasmolytic polypeptide. The RELATE comparison (Dayhoff *et al.*, 1983) of these proteins yields a score that is 22 standard deviation units higher than that obtained with 100 comparisons of randomized sequences of these proteins. A score of this magnitude indicates that the two proteins are homologous, that is, they are derived from a common ancestor. One copy of the ancestral gene is found in pS2. In pancreatic spasmolytic polypeptide this gene has duplicated.

In the original report of the sequence of pancreatic spasmolytic polypeptide, a similarity at the borderline for significance (4.2 standard deviation units) with kringles 4 and 5 of plasminogen was noted (Thim *et al.*, 1985). We compared the entire *N*-terminal non-protease segment of plasminogen, which contains all five kringles, with pancreatic spasmolytic polypeptide and pS2, using the RELATE program and a segment of 20 amino acids. The comparison scores were 10.2 ($P < 10^{-23}$) and 6.5 ($P = 4 \times 10^{-11}$) standard deviations higher, respectively, than that obtained with 100 comparisons of randomized sequences of these proteins. The highest comparison scores were found in kringles 2 and 4 of plasminogen. The *only* proteins in the database containing the sequence Pro-Trp-Cys-Phe are the kringles from the coagulation or fibrinolytic proteases, pancreatic spasmolytic polypeptide, and pS2. We suggest that pancreatic spasmolytic polypeptide, pS2, and the kringles are derived from a common ancestor.

The ~80 residue kringle is coded by two exons (Ny *et al.*, 1984). It is the second exon that is found in pancreatic spasmolytic polypeptide (two copies) and pS2 (one copy). The kringle structure is thought to have arisen over 1 billion years ago and is accumulating amino acid changes at a range that is 0.6 times that of the serine protease part of plasminogen and thrombin (Young *et al.*, 1978), suggesting biological importance for this structure, although it is the protease part that has the recognizable biological activity. The existence of proteins consisting primarily of a kringle-related exon in such diverse proteins as pS2 and pancreatic spasmolytic polypeptide suggests that kringles cleaved from the clotting and fibrinolytic factors may have novel biological

pS2 protein	1	Q	T	E	T	C	T	V	A	P	R	E	R	Q	N	C	G	F
		.	.	:	:	:	:	:	:	.	:	:	:	:	:	:	:	.
Spasmolytic protein	54	E	S	E	E	C	V	M	Q	V	K	A	R	K	N	S	G	Y
pS2 protein	18	P	G	V	T	P	S	Q	C	A	N	K	G	C	C	F	D	D
		:	:	.	:	:	:	.	:	:	:	.	:	:	:	:	:	:
Spasmolytic protein	71	P	G	I	C	P	E	D	C	A	A	R	N	C	C	F	S	D
pS2 protein	35	T	V	R	G	V	P	W	C	F	Y	P	N	T	I	D	49	
		.	:	:	:	:	:	:	:	:	:	.	:	:	:	:	:	:
Spasmolytic protein	88	T	I	P	E	V	E	W	C	E	F	P	M	S	V	E	102	

Fig. 1. Comparison of pS2 protein with pancreatic spasmolytic polypeptide

(:) shows identities and (.) shows conservative replacements. There are 21 identities (43%) and 12 conservative replacements (24%). The P-W-C-F segment, unique to pancreatic spasmolytic polypeptide, pS2 and the kringle domain, is underlined.

actions. Interestingly, human apolipoprotein(a), which is homologous to plasminogen, contains 37 kringle domains (McLean *et al.*, 1987).

Evidence from amino acid sequence analyses and analysis of the tertiary structure of prothrombin kringle I indicates that kringles are related to fibronectin type II domain and seminal fluid protein PDC-109 (Baker, 1985b; Patthy *et al.*, 1984; Holland *et al.*, 1987). The fibronectin type II domain is related to a domain on factor XII (McMullen & Fujikawa, 1985), cation-independent mannose 6-phosphate receptor (Lobel *et al.*, 1987), and insulin-like growth factor II receptor (Morgan *et al.*, 1987). The shuffling of a kringle-related exon among these diverse proteins is a good example of the generation of new proteins from common modules (Gilbert, 1978; Blake, 1978, 1983). Interestingly, tissue type plasminogen activator, which contains two kringles, is oestrogen-induced in MCF-7 cells (Butler *et al.*, 1983; Katzenellenbogen *et al.*, 1984; Ryan *et al.*, 1984).

The high comparison score between pancreatic spasmolytic polypeptide and pS2 supports the notion that they have properties in common. Information about each protein should be useful for understanding the other protein. Pancreatic spasmolytic polypeptide clearly formed by a gene duplication of a 36-residue precursor, which appears to be derived from one of the exons in an ancestor of the kringle domain. Pancreatic spasmolytic polypeptide has six disulphide bonds, which, as noted by Thim *et al.* (1985), may account for its stability towards proteolytic digestion by trypsin, which should be important for a protein acting in the gastrointestinal tract, where there are numerous proteases. Pancreatic spasmolytic polypeptide binds to rat intestinal mucosal cell membranes, inhibiting adenylate cyclase (Frandsen *et al.*, 1986). pS2, by analogy, would be expected to have some biological activity involving binding to a membrane receptor. Regarding this possibility, the groups that sequenced pS2 protein (Jakowlew *et al.*, 1984; Prud'Homme *et al.*, 1984), suggested that it had some growth factor activity due to its high number of cysteine residues, even though there was no evidence of amino acid sequence similarity to peptide hormones or growth factors. The similarity reported here provides a stronger basis for seeking a hormonal activity associated with pS2.

In considering hormonal activity possibility for pS2, one potentially relevant known example comes to mind: human epidermal growth factor (EGF), first identified as urogastrone because it inhibits gastric acid secretion (Gregory, 1975; Bower *et al.*, 1975). Epidermal growth factor causes contraction of vascular and uterine smooth muscle (Berk *et al.*, 1985; Muramatsu *et al.*, 1985; Gardner *et al.*, 1987). The multiple actions of urogastrone/epidermal growth factor and of other gastrointestinal peptides (Goodland & Wright, 1987) provide sound basis for investigating pancreatic spasmolytic polypeptide and the kringle domain for diverse biological actions.

I thank Darrell Fanestil for stimulating discussion of these results. I am grateful to Marvin F. Weissberg for encouragement and support.

Michael E. BAKER

Department of Medicine, M-023, University of California, San Diego, La Jolla, CA 92093, U.S.A.

- Baker, M. E. (1985a) FEBS Lett. **182**, 47–52
 Baker, M. E. (1985b) Biochem. Biophys. Res. Commun. **130**, 1010–1014
 Baker, M. E., French, F. S. & Joseph, D. R. (1987) Biochem. J. **243**, 293–296
 Berk, B. C., Brock, T. A., Webb, R. C., Taubman, M. B., Atkinson, W. J., Gimbrone, M. A., Jr. & Alexander, R. W. (1985) J. Clin. Invest. **75**, 1083–1086
 Blake, C. C. F. (1978) Nature (London) **273**, 267
 Blake, C. C. F. (1983) Trends Biochem. Sci. **8**, 11–13
 Bower, J. M., Camble, R., Gregory, H., Gerring, E. L. & Willshire, I. R. (1975) Experientia **31**, 825–826
 Butler, W. B., Kirkland, W. L., Gargala, T. L., Goren, N., Kelsey, W. H. & Berlinski, P. (1983) Cancer Res. **43**, 1637–1641
 Dayhoff, M. O., Barker, W. C. & Hunt, L. T. (1983) Methods Enzymol. **91**, 524–545
 Doolittle, R. F. (1985) Trends Biochem. Sci. **10**, 233–237
 Frandsen, E. K., Jørgensen, K. H. & Thim, L. (1986) Regul. Peptides **16**, 291–297
 Gardner, R. M., Lingham, R. B. & Stancel, G. M. (1987) FASEB J. **1**, 224–228
 Gilbert, W. (1978) Nature (London) **271**, 501
 Goodland, R. A. & Wright, N. A. (1987) Experientia **43**, 780–784
 Gregory, H. (1975) Nature (London) **257**, 325–327
 Holland, S. K., Harlow, K. & Blake, C. C. F. (1987) EMBO J. **6**, 1875–1880
 Jakowlew, S. B., Breathnach, R., Jeltsch, J. M., Maisakowski, P. & Chambon, P. (1984) Nucleic Acids Res. **12**, 2861–2878
 Jørgensen, K. H., Diamant, B., Jørgensen, K. D. & Thim, L. (1982) Regul. Peptides **3**, 231–243
 Katzenellenbogen, B. S. (1980) Annu. Rev. Physiol. **42**, 17–35
 Katzenellenbogen, B. S., Norman, M. J., Eckert, R. E., Peltz, S. W. & Mangel, W. F. (1984) Cancer Res. **42**, 112–119
 Lippman, M. E. & Bolan, G. (1975) Nature (London) **256**, 592–593
 Lobel, P., Dahms, N. M., Breitmeyer, J., Chirgwin, J. M. & Kornfeld, S. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**, 2233–2237
 Masiakowski, P., Breathnach, R., Bloch, J., Gannon, F., Krust, A. & Chambon, P. (1982) Nucleic Acids Res. **10**, 7895–7903
 McGuire, W. L. (1980) Rec. Prog. Horm. Res. **36**, 135–156
 McLean, J. W., Tomlinson, J. E., Kuang, W.-J., Eaton, D. L., Chen, E. Y., Fless, G. M., Scanu, A. M. & Lawn, R. M. (1987) Nature (London) **330**, 132–137
 McMullen, B. A. & Fujikawa, K. (1985) J. Biol. Chem. **260**, 5328–5341
 Morgan, D. O., Edman, J. C., Standring, D. N., Fried, V. A., Smith, M. C., Roth, R. A. & Rutter, W. J. (1987) Nature (London) **329**, 301–307
 Muramatsu, I., Hollenberg, M. D. & Lederis, K. (1985) Can. J. Physiol. Pharmacol. **63**, 994–999
 Nunez, A.-M., Jakowlew, S., Briand, J.-P., Gaire, M., Krust, A., Rio, M.-C. & Chambon, P. (1987) Endocrinology (Baltimore) **121**, 1759–1765
 Ny, T., Elgh, F. & Lund, B. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 5355–5359
 Patthy, L. (1985) Cell **41**, 657–663
 Patthy, L., Trexler, M., Váli, Z., Bányai, L. & Váradi, A. (1984) FEBS Lett. **171**, 131–136
 Prud'Homme, J. F., Fridlansky, F., LeCunff, M., Atager, M., Mercier-Bodard, C., Pichon, M.-F. & Milgrom, E. (1985) DNA **4**, 11–21
 Ryan, T. J., Seegar, J. I., Kumar, S. A. & Dickerman, H. W. (1984) J. Biol. Chem. **259**, 14324–14327
 Thim, L., Thompson, J., Christmen, M. & Jørgensen, K. H. (1985) Biochim. Biophys. Acta **827**, 410–418

Young, C. L., Barker, W. C., Tomaselli, C. M. & Dayhoff, M. O. (1978) in *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., ed.), vol. 5, suppl. 3, pp. 73–93, National Biomedical Research Foundation, Washington, DC

Received 23 December 1987

A surprising sequence homology

A few years ago we isolated a new peptide from porcine pancreatic extracts (Jørgensen *et al.*, 1982b). An initial characterization revealed that the peptide had some peculiar characteristics, e.g. seven disulphide bridges were present in the peptide consisting of only 106 amino acid residues. Further studies showed that the peptide was probably related to the exocrine function of the pancreas (Thim *et al.*, 1982), and a pharmacological screening indicated that the peptide had a spasmolytic effect (Jørgensen *et al.*, 1982a). Although we were not completely certain that this effect was actually the physiological function, we decided to name the peptide Pancreatic Spasmolytic Polypeptide (PSP). Following the amino acid sequence determination of PSP, the Dayhoff protein sequence database (National Biomedical Research Foundation, Georgetown, Washington, DC, U.S.A.) was searched (Thim *et al.*, 1985). This search did not pick up any proteins with significant homology to PSP (Thim *et al.*, 1985).

Recently, we received a new, updated version of the database (version 13, updated 30 June 1987), and to our surprise a search picked up a significantly high alignment score between PSP and a pS2 peptide. The amino acid sequence of the pS2 peptide had been derived from an mRNA which is induced by oestrogen in the human breast cancer cell line MCF-7 (Jakowlew *et al.*, 1984; Prud'homme *et al.*, 1985). The nucleotide sequence of a pS2 full-length cDNA predicted a peptide of 84 amino acid residues, out of which 26 or 21 residues represented a putative signal peptide corresponding to a mature pS2 peptide of 58 or 63 residues (Jakowlew *et al.*, 1984). PSP contains two homologous domains (Thim *et al.*, 1985) and both of these show a high degree of homology to the mature pS2 peptide (Fig. 1).

The implications of this high degree of homology cannot be established at present. However, it is our hope that the identified structural similarity may contribute to the elucidation of the physiological functions of the two peptides.

Lars THIM

Nova Research Institute, DK-2880 Bagsvaerd,
Denmark

Jakowlew, S. B., Breathnach, R., Jeltsch, J.-M., Masiakowski, P. & Chambon, P. (1984) *Nucleic Acids Res.* **12**, 2861–2878
Jørgensen, K. D., Diamant, B., Jørgensen, K. H. & Thim, L. (1982a) *Regul. Peptides* **3**, 231–243
Jørgensen, K. H., Thim, L. & Jacobsen, H. E. (1982b) *Regul. Peptides* **3**, 207–219
Prud'homme, J.-F., Fridlansky, F., Le Cunff, M., Atger, M., Mercier-Bodart, C., Pichon, M.-F. & Milgrom, E. (1985) *DNA* **4**, 11–21
Thim, L., Jørgensen, K. H. & Jørgensen, K. D. (1982) *Regul. Peptides* **3**, 221–230
Thim, L., Thomsen, J., Christensen, M. & Jørgensen, K. H. (1985) *Biochim. Biophys. Acta* **827**, 410–418

Received 22 January 1988

Homology of human pancreatic stone protein with animal lectins

The main protein constituent of human pancreatic stones has been shown to be derived from a glycoprotein precursor that accounts for about 10% of the secretory proteins of human pancreatic juice (Multigner *et al.*, 1985; Montalto *et al.*, 1986). In pancreatitis, premature zymogen activation leads to proteolysis of this precursor within the pancreas, yielding form(s) of stone protein that are practically insoluble at neutral pH. It has been suggested that precipitation of degraded forms of stone protein may be responsible for initiating the formation of pancreatic stones in chronic calcifying pancreatitis (Figarella, 1988). The normal physiological function of this secretory protein, however, is unclear. De Caro *et al.* (1984) have proposed that stone protein may act as a stabilizer of supersaturated pancreatic secretion by inhibiting the growth of calcium carbonate crystals. This view, however, was challenged recently (Figarella, 1988) since the inhibitor of crystal nucleation was shown to be distinct from stone protein.

Recently the complete amino acid sequence of human pancreatic stone protein has been published (Montalto *et al.*, 1986; Rouimi *et al.*, 1987; De Caro *et al.*, 1987). Computer-based surveys of protein sequences failed to detect homology of this protein with any other sequence except with bovine pancreatic thread protein (Gross *et al.*, 1985), the bovine equivalent of human pancreatic stone protein.

We have recently developed a sensitive procedure that can detect homologies not recognized by conventional programs of sequence comparison (Patthy, 1987a). In this procedure, similarity to a consensus sequence characteristic of a protein family is used to decide whether a new sequence has the features typical of that family. Using

Fig. 1. Homology between PSP domain 2 (residues 65–96) (a), PSP domain 1 (residues 16–47) (b) and pS2 peptide (residues 38–69) (c)