

Peter Gött<sup>a</sup>  
 Stefanie Beck<sup>a</sup>  
 José Carlos Machado<sup>a,b</sup>  
 Fatima Carneiro<sup>b</sup>  
 Holger Schmitt<sup>a</sup>  
 Nikolaus Blin<sup>a</sup>

<sup>a</sup> Division of Molecular Genetics,  
 Department of Anthropology and Human  
 Genetics, University of Tübingen,  
 Germany;

<sup>b</sup> Institute of Molecular Pathology and  
 Immunology, University of Porto Medical  
 Faculty, Porto, Portugal

# Human Trefoil Peptides: Genomic Structure in 21q22.3 and Coordinated Expression

## Key Words

Trefoil peptide  
 P-domain  
 Bacterial artificial chromosomes  
 pS2/BCEI  
 hITF/TFF3  
 hSP/SML1  
 Transcriptional regulation  
 Gene expression

## Abstract

Trefoil peptides are small secretory proteins characterized by three intrachain disulfide bonds forming the trefoil motif or P-domain. They are abundantly expressed on mucosal surfaces, especially of the gastrointestinal tract. In pathological conditions such as ulcers, metaplasia and neoplasia, their expression is upregulated. Three human trefoil peptides have been described: the estrogen-inducible pS2 protein, the spasmodic protein and the intestinal trefoil factor. Recently, their role in the maintenance of surface integrity and ulcer healing was discussed. We already mapped the corresponding three genes (BCEI, SML1, TFF3) to the same genomic region (21q22.3). In this paper, we show that the three genes are clustered in a tandemly orientated fashion within 50 kb on a bacterial artificial chromosome (BAC) recombinant. This cluster is located adjacent to D21S19 and the locus order is cen-D21S212-TFF3-SML1-BCEI-D21S19-tel, whereas transcription of all three genes is directed towards the centromere. The gene structure of SML1 exhibits four exons, two of which encode the two separate trefoil motifs. TFF3 and BCEI, both containing one trefoil motif, are composed of three exons each, suggesting gene duplication and exon-shuffling events during evolution. The 5'-flanking region of SML1 was compared to the corresponding region of other trefoil genes. Two motifs with identical sequence and positions are shared between SML1 and BCEI, thus presenting possible targets for stomach-specific gene regulation. Two other motifs are shared within all known human and rat trefoil genes, suggesting a coordinated regulation and/or a common locus-controlling region. Using RT-PCR, a change in the pattern of trefoil gene expression is detected in tissue samples from normal gastric mucosa, hyperplastic polyps, gastric cancer, and gastric cancer cell lines, respectively.

## Introduction

In the last few years a new cysteine-rich structural motif of approximately 50 amino acid residues, known as either the trefoil motif or the P-domain, has been described [1, 2]. The conserved six cysteine residues form three intramolecular disulfide bridges resulting in three loops which are responsible for the trefoil-like shape and its resistance against proteolytic degradation. P-domains were discovered in mucins of *Xenopus* skin as well as in porcine, rodent and human gastrointestinal peptides [for a review, see ref. 3]. P-domain-containing peptides display very distinctive expression patterns in normal as well as in pathological gastrointestinal tissues. Although their

precise physiological role is not understood, recent *in vitro* models propose that the peptides are involved in the maintenance of mucosal integrity and may accelerate ulcer healing, presumably by enhancing cell migration after wounding [4, 5]. Transgenic mice that overexpress a human trefoil peptide exhibit increased resistance to intestinal damage [6].

The human pS2 gene, found under estrogen transcriptional control in a subclass of breast cancer cells, was reported to be expressed in normal stomach surface epithelial cells, whereas additional gastrointestinal tissues, such as pancreas and colon, do not produce pS2 at all [7]. Estrogen-independent expression in various tumors of the human gastrointestinal tract, like carcinoma of the stom-

ach, pancreas, colon and biliary tract, was noted [8–11]. The mechanism of pS2 gene activation and its biological role in carcinogenesis remains to be elucidated.

A remarkable sequence homology exists between pS2, a porcine pancreatic spasmodic protein (PSP), and its subsequently isolated human counterpart, hSP [2]. Like the pS2 gene, the hSP gene (also termed human spasmodic; SML1) is expressed in stomach mucosa, but not in any other tissue of the normal gastrointestinal tract. In stomach, pancreas and biliary tract carcinomas, however, the presence of the regular pS2 mRNA is strongly correlated with hSP/SML1 expression [8–10] as well as in ulceration-associated cell lineage in Crohn's disease [12]. Interestingly, both genes were localized in the same chromosomal region 21q22.3 [13].

Moreover, the human cDNA sequence [14], and genomic sequences of human and rat [15, 16] were presented for the third P-domain peptide, the intestinal trefoil factor (hITF). In contrast to stomach specific expression of pS2 and hSP, hITF is found in the human intestine. Most recently, we also mapped the genomic locus for hITF to 21q22.3 [17] and found that all three trefoil peptide genes are clustered within a region of less than 100 kb [18].

With the coordinated cell- and tissue-specific gene expression in mind, we show here the head-to-tail organization of the gene cluster of the three trefoil proteins, compare their 5'-flanking regions and present evidence for a cell-specific and coordinated transcriptional regulation. The structural organization of the genes provides a framework to study gene regulation as a response to gastrointestinal pathological conditions such as defense against pathogens, ulcer healing and carcinogenesis.

## Materials and Methods

### *Restriction Analysis, Southern Blotting and Hybridization of BAC DNA*

BAC DNA was prepared by standard alkaline lysis and purified by phenol/chloroform and ethanol precipitation. *Hind*III restriction digests were separated on 0.7% agarose-TAE gels. Digests by rare cutting enzymes were separated by PFGE on 1% agarose gels in 0.25 × TBE using an LKB 2015 electrophoresis unit, as described [18]. DNA was transferred in alkaline buffer for 2 h to Nytran membrane (Schleicher & Schüll) by downward blotting. DNA probes were labeled with <sup>32</sup>P-dATP by random priming (Boehringer Mannheim Biochemicals). A 1-kb fragment of the 5'-flanking region of BCEI (–90 to –1100) was isolated from the plasmid pS2-cat. A 300-bp cDNA fragment of SML1 was isolated from pGEM-hsp200 [2]. Both plasmids were kindly provided by Dr. M.-C. Rio, Strasbourg. All other probes for hybridization were generated by PCR with the oligonucleotides listed in table 1. Probes were hybridized at 65 °C overnight in 7% SDS, 0.5 M sodium phosphate buffer.

### *PCR Analysis*

Oligonucleotides were designed from cDNA or genomic sequences of all three trefoil genes [2, 15, 19]. Products listed in table 1 were amplified from 10 ng BAC DNA using Gold Star polymerase (Eurogentech, Belgium) or the Expand Long Template PCR System (Boehringer Mannheim Biochemicals) according to standard conditions recommended by the manufacturers.

### *Sequence Analysis*

Cycle sequencing with ThermoSequenase (Amersham Life Science) using either <sup>32</sup>P-primer labeling or <sup>35</sup>S-dATP internal labeling was performed according to the manufacturer's protocol. Sequence data were processed by BLASTN, FACTOR, and SIMILARITY algorithms of the GCG Wisconsin package.

### *Gene Expression Study by RT-PCR*

For RT-PCR studies we used paraffin-embedded material from two samples of normal mucosa of the stomach, and four gastric hyperplastic polyps. Frozen material of two cases of gastric carcinoma and material from cell lines GP220 and GP202 were also used [20]. RNA was extracted according to the method described by Chomczynski and Sacchi [21]. First-strand synthesis was made by random 6-mer primer using M-MLV-reverse transcriptase at 42 °C for 15 min in the presence of 7 mM MgCl<sub>2</sub>. For PCR we used *Taq* polymerase in the presence of 2 mM MgCl<sub>2</sub>. Specific primers for hSP (5'-GGATCAGTGCTGCATGGAG and 5'-GTTGGAGAAGCAGCACTCC), pS2 (PS1 and PS2, table 1) and hITF (HITFF and HITFR, table 1) were chosen. Cycle conditions were 95 °C, 4 min; 95 °C, 1 min (30 cycles); 64 °C or 55 °C (for hSP) 1 min; 72 °C, 1 min. PCR controls without the addition of reverse transcriptase were routinely performed to exclude genomic DNA contamination.

## Results

### *Orientation of the Contig*

We have previously reported the isolation of clones coding for the three trefoil peptides from a human genomic BAC library [18]. These contiguous genomic fragments were ordered by *Hind*III restriction finger printing, determining overlapping segments after hybridization of BAC 921F4 and BAC 801B4 (fig. 1). The size of each BAC determined by comparison of fragments produced by rare cutters (*Not*I, *Sal*I, *Mlu*I, *Nru*I, *Sgr*AI) after PFGE is concordant with data obtained by summarising *Hind*III fragments. Thus, the contig spans a region of 350 kb. In addition, all seven BACs were probed independently by PCR and hybridization with the well-mapped markers D21S19 and D21S212. Only BACs 843E9 and 43A9 were positive for D21S19 (fig. 2). The known relative map position of BCEI and D21S19 allowed the chromosomal orientation of the whole cluster. BCEI was placed 40–65 kb proximal to D21S19. D21S212 was not localized within the 350-kb BAC contig.

**Table 1.** Oligonucleotides used for PCR

Name	Localisation <sup>1</sup>	Sequence	PCR product, bp	Annealing temp.
PS 1	BCEI exon 1, f	TTTGGAGCAGAGAGGAGGCAATG	3,200	55
PS 2	BCEI exon 2, r	ACCACAATTCTGTCTTTCACGGGG		
R2	BCEI exon 3, f	GGCTTCCTGAGCTGGGATACAAG	12,000	65
T1	SML1 exon 1, r	GGAGCTGGGCGTCTCGCCGT		
R22	BCEI exon 3, f	CTCGCAGAGGTTCTGCATCTTCTG	12,000	65
T11	SML1 exon 1, r	CCAGGACGAGGAGCGCTGCCA		
SLPR2	SML1 prom, f	CAGACTGGCAACCCCTGTC	900	60
T1	SML1 exon 1, r	GGAGCTGGGCGTCTCGCCGT		
SLPR	SML1 prom, f	GTCACAGGCTACTCCGCTGACC	600	60
T1	SML1 exon 1, r	GGAGCTGGGCGTCTCGCCGT		
SLA	SML1 exon 1, f	GCAGACATGGGACGGCGAGA	1,000	60
SLAR	SML1 exon 2, r	ACACTGGTCACTGGTGATTCC		
SMLF	SML1 exon 2, f	CACCAGTGACCAGTGTTTTGAC	2,600	60
SLM	SML1 exon 3, r	GTTGGAGAAGCAGCACTCCG		
SLA	SML1 exon 1, f	GCAGACATGGGACGGCGAGA	3,300	60
SLM	SML1 exon 3, r	GTTGGAGAAGCAGCACTCCG		
SLA	SML1 exon 1, f	GCAGACATGGGACGGCGAGA	4,500	55
SLE	SML1 exon 3, r	CTTAAAGAAATTATATGTAAACCATTG		
SML1	SML1 exon 3, f	TCTCGGAAGTGCTGCTTCTCC	1,000	55
SL2	SML1 exon 4, r	TAAGCGAAGTTTCTTTCTTGG		
HITFA	TFF3 prom, f	ACTTAGGAGGCTGAGGTGGGAG	950	58
H1	TFF3 exon 1, r	TGACCACCGTGGGCTCCGGGA		
XSP66	TFF3 exon 1, f	CTCCAGGGATCCTGGAGTC	1,700	60
XSP67	TFF3 exon 2, r	CAGTCTGAGCTGCGTCCC		
HITFF	TFF3 exon 2, f	GTGGACTGCGGCTACCCCAT	90	65
HITFR	TFF3 exon 2, r	CCAAGGCACTCCTAGGGATCCT		
H2	TFF3 exon 2, f	CATGCTCCACCCGAGGACAGTTC	1,700	47
SP6L	BAC vector	ATTTAGGTGACACTATAG		
H22	TFF3 exon 2, f	CATATCTGGAGCCTGATGTCTTAAC	1,700	47
SP6L	BAC vector	ATTTAGGTGACACTATAG		

<sup>1</sup> 'f' and 'r' refer to the 'forward' and 'reverse' orientation with respect to the transcriptional direction.

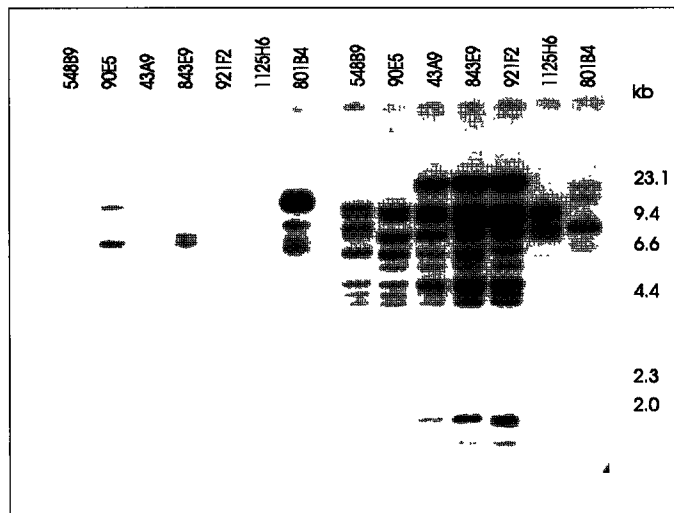
### Distances and Orientation of Genes

The three trefoil genes BCEI, SML1 and TFF3, encoding pS2, hSP, and hITF, respectively, were mapped in detail by PCR with appropriate primers (table 1) using the six BACs shown in figure 2. The genes' positioning was verified by hybridizing gene probes to restriction fragments generated by rare cutting enzymes (fig. 2). Whereas BAC 1125H6 was positive for TFF3, exclusively, BACs 548B9 and 90E5 in addition contained SML1. BACs 921F4, 843E9 and 43A9 yielded signals with all three gene probes (fig. 2).

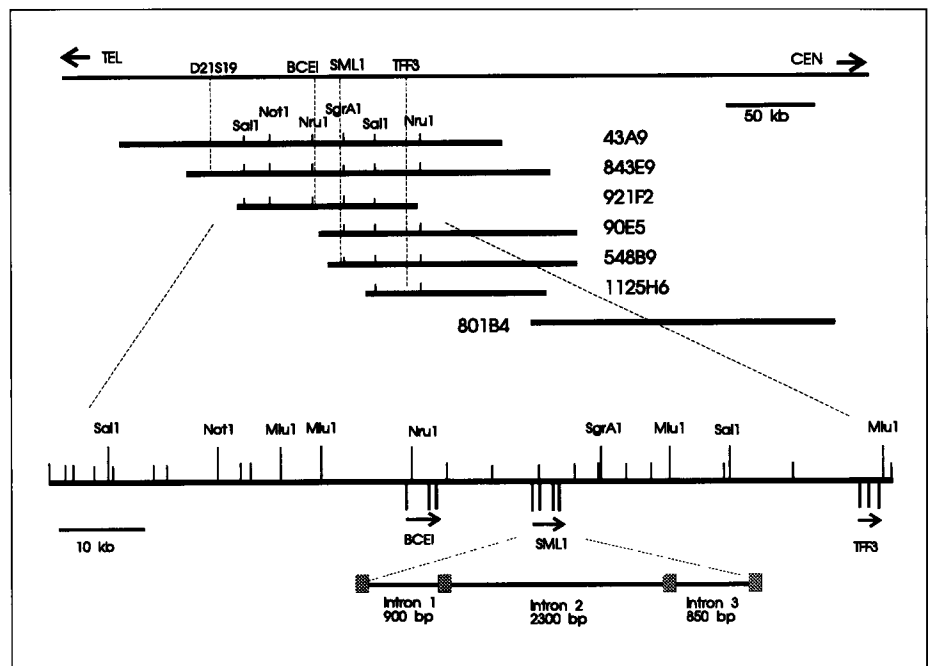
Using outward-directed primers derived from the known cDNA sequences, long range PCR was performed to determine the intergenic distances of the three genes. A 12-kb product was specifically generated using primers

T1 and R2 derived from the 3'-end of BCEI and the 5'-end of SML1. The product was verified by a second PCR with nested primers T11 and R22. Thus the orientation of both genes is organized in tandem (head to tail) and transcription is directed towards the centromere. The orientation and location of BCEI were further verified using Southern hybridization. A single *NruI* site, localized in intron I, created restriction fragments of 45 and 55 kb, which hybridized specifically with a probe harboring the promoter (-1100 to -90) and a probe harboring exon II, respectively. Attempts to generate a PCR fragment between SML1 and TFF3 failed, presumably because of the large distance (30-35 kb). To determine the localization and orientation of TFF3, BAC 921F4 was probed by PCR with vector primers and primers corresponding to the 3'-

**Fig. 1.** Fingerprint analysis of a contig of 7 BACs. 1  $\mu$ g of BAC DNA was digested with *Hind*III, Southern blotted to duplicate filters and hybridized with labeled BAC 801B9 (left) and BAC 921F2 (right).



**Fig. 2.** Genomic organization of the gene cluster of trefoil peptides within a BAC contig located on 21q22.3 (upper part), and restriction enzyme mapping, gene order, orientation, and exon-intron structure of trefoil peptides on BAC 921F2 (lower part). Small upper bars indicate *Hind*III sites as determined by Southern hybridization with various probes, lower bars and hatched boxes indicate gene exons.



as well as the 5'-end of hITF cDNA. A 1.7-kb PCR product, specifically amplified with the vector primer (SP6L) and the 3'-end-specific primers (H2 and H22, table 1), verified that TFF3 is orientated in the same direction as SML1 and BCEI (head to tail), and is localized 30-35 kb downstream from SML1, as judged by PFGE using rare cutters (data not shown). To confirm these results, the 20-kb *Sal*I restriction fragment of BAC 921F4 was isolated, further digested with *Bam*HI, *Eco*RI, and *Hind*III, and hybridized with cDNA probes of hSP and hITF, respectively (not shown). The intergenic region between SML1

and TFF3 was estimated to be 32 kb, as calculated from the sizes of the intergenic *Hind*III fragments (fig. 1).

#### Exon-Intron Mapping of SML1 and Translational Initiation

The genomic organization of SML1 was of special interest, since this gene codes for a protein with two P-domains, in contrast to the genes BCEI and TFF3 that code for only one P-domain. Using the cDNA-derived primer pairs (SLA-SLAR, SLMF-SLM, and SML1-SL2), we amplified introns of 900, 2,300 and 850 bp, respec-

SML1 CAGACTGG -880  
SML1 CAACCCCTGTCATTTCCCTGGCGTGGGAACTTCGGGTCCCCTCTGTCCTTCCACCAC -820  
SML1 ACTTTTCCCTCTTTCTTTCCGGGTGCTACTCTCTGGCTTCTGTCTTCTCTGTGAGGTCC -760

SML1 ACAGAATCCTTCTCCAGCACATCCTACCCAGGAAGGCCATGGGCTGGGTCCCAGGTGCC -700  
BCEI AAGTGATTCTCCTGACTTAACCTCCAGAGTAGCTAGGATTAC -701

SML1 ATCTTTCAGAAGATGTAGAGCATTCCCATGGAACAAAAATAACCCATTTTCAGGGGTGGC -640  
BCEI AGGCACCCGCACCATGCCTGGCTAATTTTTGTATTTTTTTTTTTTTTTGTAGAGACGGGTTT -641

motif I

SML1 TGAAAAATGAACTTATTAAAACCTGCCTGTACAGGCTACTCCGCTGACCCTGTGACGCTC -580  
BCEI CGGCCATGTGGCCAGGCTAGTCTCAAACCTCGACTTTAAGTGATCCGCTGCTTTGGC -581  
TFF3 ATTTCACCACTTATTAAAATGTAGTACTAATTTTTAGGGGAGAAAGCAAAAAGGAAAG -534  
RITF TTAATATGTTTTATTAAAATTACATTTTCAGCCATTAAGTGATAGTAAACGATAAAAAATG -415

SML1 ATCTCCATGGAGAGCAGCCCTCCTGTGAAGATGGGACAAAGGGCATCGTGTGCGGTT -520  
BCEI CTCCAAAGTGTTGGGATTACAGCGTGAGCCACTGCGCCAGGCCTACAATTTCATATTAA -521

motif II

SML1 GGGGAGGCTCTAACACAGCCCTGGGAGCAGTCTTACCTCCTCTGAGATGCTTCCCTT -460  
BCEI AACCAATTCCACTGTAAAAGAATTAGCTTAGGCCTAGACGAATGGGCTTCATGAGCTCC -461

**ERE**

SML1 **CTC**AGGGAGGGGACTTTTTCCATGCTATCTGTGGCCTGTACATTTTCCCCAGTAAACTT -400  
BCEI **TTCCCTT**CCCCCTGCAAGGTCAGGGTGGCCACCCCGTGAGCCACTGTTGTGACGGCAAGC -401

SML1 GGCCCTAATATTTCTAAATTCCTGTGGTCCCTGCCACTCTATCAATAGAAATGCATAG -340  
BCEI CTTTTCCGGCCATCTCTCACTATGAATCACTTCTGCAGTGAGTACAGTATTTACCCCTGG -341

**MYC**

SML1 CTTATCCCTTCCTGGGTGTGACCCTGTGTGTGCCAGCCCCAGACCTGCACGTGGCCCGGT -280  
BCEI CGGGAGGGCCTCTCAGATATGAGTAGGACCTGGATTAAGGTCAGGTTGGAGGAGACTCCC -281

**PEA3**

SML1 TTTCCACGCTGGCAGCCTGGCATGACCCAACTCTCTGTCCAGGGCAGGAAGAGGTATCAC -220  
BCEI ATGGGAAAGAGGGACTTTCTGAATCTCAGATCCCTCAGCCAAGATGACCTCACACATGT -221

**PEA3**

SML1 CGAGCAGGGAGAGAGTACCCCTGGCCCGGAAGCCTCGCCTGCACAGGGCAGAGCTGCCTC -160  
BCEI CGTCTCTGTCTATCAGCAAATCCTCCATGTAGCTTGACCATGTCTAGGAAACACCTTTG -161

motif III

SML1 TTGCCTCCTCTTCGCCTCCACGGTGGAAAGGGCTGGGGCCAGGGGAGAGAAGAAGGTT -100  
BCEI ATAAAAATCAGTGGAGATTATTTGTCTCAGAGGATCCCCGGGCCTCCTTAGGCAAAATGTTA -101  
TFF3 AAAGGCATGCAGGAGAGAAACAGGAGCAGCCACAGCCAGGAGGGAGAGCCTTCCCAAGCA -95

motif IV

SML1 **ATCT**CTGCTTGTGGCAAAACAGAGGGGAGATTTATAAAACATACCCGGCAGTGGACACCA -40  
BCEI **TCTA**ACGCTCTTTAAGCAAAACAGAGCCTGCCCTTATAAAATCCGGGGCTCGGGCGGCCTCT -41  
TFF3 **AACA**ATCCAGAGCAGCTGTGCAAAACAACGGTGCCATAATGAGGCCTCCTGGACCTGAAGC -35  
RITF CAGAGTCCGCAGAACCCAGCCAAACACAATGGCTTATAAAAGGGCTTTTCCTTGGCACCCAC -33

+1 M G R R D A

SML1 TGCATTTCTGC -AAGCCACCCTGGGGTGCAGCTGAGCTAGACATGGGACGGCGAGACGCCC +19  
BCEI CATCCCTGACTCGGGTTCGCTTTGGAGCAGAGAGGAGCAATGGCCACCATGGAGAACA +19  
TFF3 -----GAGTCTGAGCTGCGTCCCGGAGCCACGGTGGTCTATGGCTGCCAGAGCGCTCT +19  
RITF -----GCATCTCGGGATACCGAAGTTTGCTGTGCTGCCATGGAGACCAGAGCCTTCT +19

mSP P L L A V L L V L G L H A L V E G E K P  
Q L L A A L L V L G L C A L A G S E K P

SML1 AGCTCCTGGCAGCGCTCCTCGTCCCTGGGGCTATGTGCCCTGGCGGGAGTGAGAAACCCCT +79  
style="text-align: center;">intron I S P C Q C S R L S P

SML1 gtaagtgaagga-900bp-tctttccacagCCCCCTGCCAGTGCTCCAGGCTGAGCCCCC +129  
H N R T N C G F P G I T S D Q C F D N G

SML1 ATAACAGGACGAACTGCGGCTTCCCTGGAATCACCAGTGACCAGTGTTTTGACAATGGAT +189  
C C F D S S V T G V P W C F H P L P K Q

SML1 GCTGTTTCGACTCCAGTGTCACTGGGGTCCCTGGTGTTCACCCCTCCCAAAGCAAG +249  
style="text-align: center;">intron II E S D Q C V M E V S

SML1 gtaatcttcag-2400bp-tcccatggcagAGTCGGATCAGTCGGCTCATGGAGGCTCTCA +298  
D R R N C G Y P G I S P E C A S R K C

SML1 GACCGAAGAACTGTGGCTACCCGGGCATCAGCCCCGAGGAATGCGCCTCTCGGAAGTGC +359  
C F S N F I F E V P W C F F P K S V E

SML1 TGCTTCTCCAACCTCATCTTTGAAGTGCCCTGGTGTCTTCTCCCGAAGTCTGTGGAAGgt +419  
style="text-align: center;">intron III D C H Y \*\*\*

SML1 aacgtcgctg-850bp-taaacttctagACTGCCATTACTAAGAGAGGCTGGTTCCAGAG +469  
SML1 GATGCATCTGGCTCACCCGGGTGTTCCGAAACCAAGAAGAAACTTCGCCTTATCAGCTTC +529  
SML1 ATATTTCAATGAAATCCTGGGTTTTCTTAACCATCTTTTCTCATTTTCAATGGTTTAAACA +589

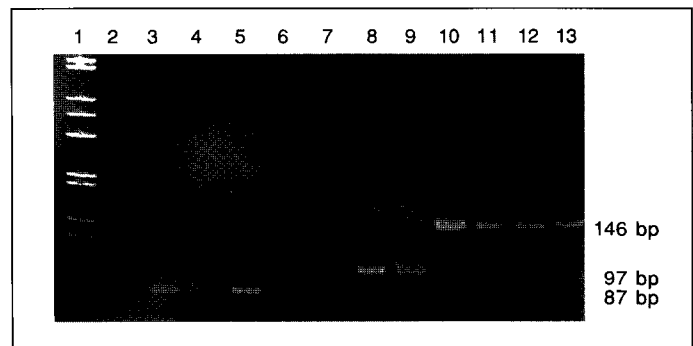
tively (fig. 2, 3). The exon-intron junctions were determined by sequence analysis of the corresponding PCR products (EMBL accession No. X97790, X97791, X97792, X97793) revealing that the two trefoil motifs of SML1 are encoded by exons 2 and 3 (fig. 3). Thus, each P-domain is encoded by a single exon. An additional cytosin base at position 56 was found in contrast to the previously published sequence of hSP [2]. As a consequence, the corrected human sequence exhibits high amino acid homology to the mouse signal peptide (22 identical residues out of 30, see fig. 3). Thus, both signal peptides comprise basic amino acid residues followed by a cluster of hydrophobic residues.

Moreover, the translational initiation sequences of all trefoil peptides are homologous. In positions -2 and +4, a common G is obvious, although the sites exhibit marked differences to the Kozak consensus sequence site (ACCATGG; fig. 3). Nevertheless, this consensus site is present three codons downstream of the presumed translational initiation site of BCEI and therefore might contribute to an enhanced translation of the pS2 protein.

#### *Promoter Sequence of SML1 and Comparison of Gene Regulatory Sequences*

Since the 5'-flanking regions of BCEI and TFF3 were already published, only the corresponding region of SML1 was analyzed. Using primers T1 and T11, the nucleotide sequence of this region was determined up to position -887 by sequencing of BAC 921F2 DNA using a primer walking strategy (EMBL accession No. X97790). At this position, the downstream primer designated SPLR2 (table 1) was designed to amplify the promoter region. The nucleotide sequence was confirmed by sequencing the PCR product. A TATAA box is present 67 bases upstream of the ATG codon, a feature shared with the BCEI but not with the TFF3 promoter. We found potential binding sites for several transcription factors (Myc; PEA3, a Ets-like factor; fig. 3) whose involvement in regulation remains to be determined. A search for

**Fig. 3.** Sequence of the SML1 gene and partial comparison with other trefoil peptide genes. The 5'-flanking region of SML1 is displayed up to position -887 with respect to the translational start codon (+1). Introns sequences are indicated by small letters. Translated codons are denoted by single code amino acid residues. The resulting signal sequence of hSP is presented in italics and aligned with the homologous sequence of mSP. Possible regulatory targets are either underlined or printed in bold letters. Canonical target sequences of Myc, Pea3 (Ets-like transcription factor) and the ERE are indicated in bold and italics. rITF denotes the corresponding gene sequences of rITF.



**Fig. 4.** Example of trefoil gene expression by RT-PCR analyzed on a 5% PAA-TBE stained with ethidium bromide using primers specific for TFF3 (hITF; lanes 2-5), SML1 (hSP; lanes 6-9), and BCEI (pS2; lanes 10-13), respectively. RNA samples were from normal gastric mucosa (lanes 2, 6, 10), hyperplastic polyp (lanes 3, 7, 11), gastric tumour (lanes 4, 8, 12) and cell line GP220 (lanes 5, 9, 13). Lane 1 = molecular DNA marker (GibcoBRL No. 15615-024).

homologous motifs shared by the promoters of the trefoil genes revealed some interesting features which underline the concept of coordinated gene expression. Although there is no overall homology among the promoters, some common signals with almost identical spacing with respect to the TATAA box were found (fig. 3). Motifs II and III are shared only by SML1 and BCEI. Motif II, a pyrimidin-rich region TGAGA/CTG/CCTTCCCTTCC) is located adjacent to the estrogen-responsive element (ERE) of the pS2 gene. Remarkably, SML1 exhibited no ERE. Another consensus sequence, AAAG/TGTTATCT is located at almost identical positions (-106, -107) just upstream to the TATAA box. A consensus sequence (CAAACA, motif IV) located 12-16 bases upstream of the TATAA box is found in all human trefoil genes as well as the rat ITF gene. This motif is extended to CAACAGAG in SLM1 and BCEI. Finally, a consensus sequence (motif I; TTATTA AAA; fig. 3) located at position No. -628 is found in all trefoil gene promoters at similar positions. This consensus sequence partially overlaps with a 14-bp homeodomain-like sequence found also at a comparable position in the 5'-flanking region of the rat intestinal fatty acid binding protein [22] with partial homology to the Pit-1 homeodomain [23].

#### *Expression Pattern in Gastric Tissue*

To study the expression of trefoil peptides on the level of transcription, we used RT-PCR as a sensitive tool (fig. 4). In two paraffin-embedded samples from normal superficial mucosa we detected pS2 (BCEI) mRNA, but

not hITF (TFF3) or hSP (SML1) mRNA. In four paraffin samples from hyperplastic polyps of stomach mucosa pS2 and, surprisingly, hITF was present. Frozen material from two different gastric carcinomas of the diffuse type exhibited expression of all three trefoil peptides, also noted in two gastric tumor cell lines, GP220 and GP202.

## Discussion

The data presented provide some interesting factors that shed light mainly on the regulation of the genes coding for all yet known human trefoil peptides, and may eventually elucidate their physiological function. The order and distances of the corresponding genes are now mapped within a physically defined region on 21q22.3 (cen-D21S212-TFF3-(32 kb)-SML1-(12 kb)-BCEI-(40–65 kb)-D21S19-tel). The location of BCEI and D21S19 was previously assigned to adjacent *NotI* fragments 400 kb and 440 kb in size, respectively [24]. Since D21S212, the BCEI proximal marker located on the 400 kb *NotI* fragment, is not present on our genomic contig, it is likely to be located 270–400 kb adjacent to BCEI, towards the centromere. Recently, the clustering of genes encoding the trefoil peptide family was also reported by others [25], without presenting the gene order, distances and chromosomal fine mapping. This gene cluster in 21q22.3 is positioned within a CpG-rich region.

Analysis of the exon-intron boundaries of SML1 revealed that the two P domains are encoded by two different exons. The exon structure of the three trefoil peptides are very similarly organized. The first exon encodes the secretion signal sequence, the second exon (and third exon of SML1) encodes the P-domain or trefoil motif, and the third exon (fourth exon of SML1) encodes three to four residues of the carboxy terminus. This conserved structural organization may have evolved by gene duplication and exon shuffling.

The main data presented here support the idea of a coordinated regulation of gene expression of the trefoil peptides. Besides gene clustering, the transcriptional orientation of all three genes is identical. More interestingly, the genes' 5'-flanking regions share several motifs with almost identical sequences and spacings. This is reminiscent of a situation found for  $\beta$ -globin genes, which are directed by a cis-located locus-controlling region. In this case, a developmental switch of gene expression is mediated by consensus sequences shared by the genes' 5'-flanking regions that compete for binding to the distant locus-controlling region [26]. For the presently known

promoters of genes coding for mammalian trefoil peptides, two motifs are identified which share similar sequences and spacings. They may present targets for unknown regulatory DNA-binding proteins. In fact, motif I overlaps with a 14-bp homeodomain-like sequence (ATTAAAATACATTT) present at a corresponding position in the 5'-flanking region of the rat intestinal fatty acid binding protein [22] with partial homology to the Pit-1 homeodomain [23].

Transcriptional activity of the trefoil genes is known to be tissue specific. In the normal gastrointestinal mucosa, BCEI and SML1 are expressed in the superficial and glandular area of the stomach, respectively, whereas TFF3 is expressed in goblet cells of the intestine. Aligning of the genes' 5'-flanking regions reveals two motifs exclusively shared by the stomach-specific genes BCEI and SML1. They provide a first hint for testing stomach-specific gene regulation by reporter gene technology. The transcriptional regulation of the pS2 gene has been studied in greater detail. A variety of factors are known to be involved in the transcriptional regulation [27], but up to now, studies have been focussed on the estrogen-responsive MCF7 breast cancer cells. In this respect, we find it interesting to assess the influence of the pyrimidin-rich motif II (TGA-GA/CTG/CCTTCCCTCC), localized close to the ERE, for both SML1 and BCEI transcription in gastrointestinal tumor cell lines.

Finally, the expression of trefoil peptides is associated with changes of the physiological status of the cell, and is upregulated in pathological conditions of the gastrointestinal tract, like damage of the mucosa, ulcerative lesions and cancer [12, 28]. In several cancers of the gastrointestinal tract, we have previously identified coordinated regulation of pS2 and hSP at the mRNA and protein level [8–10]. Here we show that different pathological and morphological stages of the stomach mucosa, are associated with alterations in the pattern of trefoil peptide expression. In normal mucosa, only pS2 is found, whereas in hyperplastic polyps, a premalignant stage of gastric mucosa, pS2 and, surprisingly, hITF mRNA are present. All three trefoil peptides are expressed in tumors of this tissue as well as tumor cell lines. Additionally, a role for the intestinal trefoil factor has been recently proposed in the development of the mouse brain [29].

Although the physiological relevance of this switch of expression pattern is not known, our results suggest a coordinated gene regulation of trefoil peptides, which is also supported by the genomic structure and sequence comparison of the 5'-flanking regions of the corresponding genes. In this respect, probing promoters by transient

transfection in different cell lines should elucidate regulation steps in inflammatory and preneoplastic processes, and contribute to understanding the yet unclear role of trefoil peptides.

## Acknowledgements

S.B. is a recipient of a Baden-Württemberg predoctoral training grant. The authors acknowledge a DFG research grant to N.B. (BL166/11-1).

## References

- Thim L: A new family of growth factor-like peptides. 'Trefoil' disulphide loop structures as a common feature in breast cancer associated peptide (pS2), pancreatic spasmodic polypeptide (PSP), and frog skin peptides (spasmolysins). *FEBS Lett* 1989;250:85-90.
- Tomasetto C, Rio MC, Gautier C, Wolf C, Hareuveni M, Chambon P, Lathe R: hSP, the domain-duplicated homolog of pS2 protein, is co-expressed with pS2 in stomach but not in breast carcinoma. *EMBO J* 1990;9:407-414.
- Hoffmann W, Hauser F: The P-domain or trefoil motif: a role in renewal and pathology of mucous epithelia? *Trends Biochem Sci* 1993;18:239-243.
- Kindon H, Pothoulakis C, Thim L, Lynch Devaney K, Podolsky DK: Trefoil peptide protection of intestinal epithelial barrier function: Cooperative interaction with mucin glycoprotein. *Gastroenterology* 1995;109:516-523.
- Alison MR, Chinery R, Poulosom R, Ashwood P, Longcroft JM, Wright NA: Experimental ulceration leads to sequential expression of spasmodic polypeptide, intestinal trefoil factor, epidermal growth factor and transforming growth factor alpha mRNAs in rat stomach. *J Pathol* 1995;175:405-414.
- Playford RJ, Marchbank T, Goodlad RA, Chinery RA, Poulosom R, Hanby AM, Wright NA: Transgenic mice that overexpress the human trefoil peptide pS2 have an increased resistance to intestinal damage. *Proc Natl Acad Sci USA* 1996;93:2137-2142.
- Rio MC, Bellocq JP, Daniel JY, Tomasetto C, Lathe R, Chenard MP, Batzenschlager A, Chambon P: Breast cancer-associated pS2 protein: Synthesis and secretion by normal stomach mucosa. *Science* 1988;241:705-708.
- Seitz G, Theisinger B, Tomasetto G, Rio MC, Chambon P, Blin N, Welter G: Breast cancer-associated protein pS2 expression in tumors of the biliary tract. *Am J Gastroenterol* 1991;86:1491-1494.
- Theisinger B, Welter C, Seitz G, Rio MC, Lathe R, Chambon P, Blin N: Expression of the breast cancer associated gene pS2 and the pancreatic spasmodic polypeptide gene (hSP) in diffuse type of stomach carcinoma. *Eur J Cancer* 1991;27:770-773.
- Welter C, Theisinger B, Seitz G, Tomasetto C, Rio MC, Chambon P, Blin N: Association of the human spasmodic polypeptide and an estrogen-induced breast cancer protein (pS2) with human pancreatic carcinoma. *Lab Invest* 1992;66:187-192.
- Welter C, Theisinger B, Rio MC, Seitz G, Schuder G, Blin N: Expression pattern of breast-cancer-associated protein pS2/BCEI in colorectal tumors. *Int J Cancer* 1994;56:52-55.
- Wright NA, Poulosom R, Stamp G, Van Noorden S, Sarraf C, Elia G, Ahnen D, Jeffery R, Longcroft J, Pike C, et al: Trefoil peptide gene expression in gastrointestinal epithelial cells in inflammatory bowel disease. *Gastroenterology* 1993;104:12-20.
- Tomasetto C, Rockel N, Mattei MG, Fujita R, Rio MC: The gene encoding the human spasmodic protein (SML1/hSP) is in 21q 22.3, physically linked to the homologous breast cancer marker gene BCEI/pS2. *Genomics* 1992;13:1328-1330.
- Podolsky DK, Lynch Devaney K, Stow JL, Oates P, Murgue B, De Beaumont M, Sands BE, Mahida YR: Identification of human intestinal trefoil factor. Goblet cell-specific expression of a peptide targeted for apical secretion [published erratum appears in *J Biol Chem* 1993;268:12230]. *J Biol Chem* 1993;268:6694-6702.
- Seib T, Dooley S, Welter C: Characterization of the genomic structure and the promoter region of the human intestinal trefoil factor. *Biochem Biophys Res Commun* 1995;214:195-199.
- Sands BE, Ogata H, Lynch Devaney K, De Beaumont M, Ezzell RM, Podolsky DK: Molecular cloning of the rat intestinal trefoil factor gene. Characterization of an intestinal goblet cell-associated promoter. *J Biol Chem* 1995;270:9353-9361.
- Schmitt H, Wundrack I, Beck S, Gött P, Welter C, Shizuya H, Simon M, Blin N: A third P-domain peptide gene, human intestinal trefoil factor (TFF3), maps to 21q22.3. *Cytogenet Cell Genet* 1996;72:299-302.
- Beck S, Schmitt H, Shizuya H, Blin N, Gött P: Cloning of contiguous fragments from human chromosome 21 harbouring three trefoil peptide genes. *Hum Genet*, in press.
- Jeltsch JM, Roberts M, Schatz C, Garnier JM, Brown AM, Chambon P: Structure of the human oestrogen-responsive gene pS2. *Nucleic Acids Res* 1987;15:1401-1414.
- Gärtner F, David L, Seruca R, Machado JC, Sobrinho-Simoes M: Establishment and characterization of two cell lines derived from human diffuse gastric carcinomas xenografted in nude mice. *Virchows Arch*, in press.
- Chomczynski P, Sacchi N: Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 1987;162:156-159.
- Sweetser DA, Birkenmeier EH, Klsak IJ, Zollman S, Sparkes RS, Mohandas T, Lulis AJ, Gordon JI: The human and rodent intestinal fatty acid binding protein genes. A comparative analysis of their structure, expression, and linkage relationships. *J Biol Chem* 1987;262:16060-16071.
- Ingraham HA, Chen RP, Mangalam HJ, Elsholtz HP, Flynn SE, Lin CR, Simmons DM, Swanson L, Rosenfeld MG: A tissue-specific transcription factor containing a homeodomain specifies a pituitary phenotype. *Cell* 1988;55:519-529.
- Wang D, Fang H, Cantor CR, Smith CL: A continuous *NotI* restriction map of band q22.3 of human chromosome 21. *Proc Natl Acad Sci USA* 1992;89:3222-3226.
- Chinery R, Williamson J, Poulosom R: The gene encoding human intestinal trefoil factor (TFF3) is located on chromosome 21q22.3 clustered with other members of the trefoil peptide family. *Genomics* 1996;32:281-284.
- Wijgerde M, Grosveld F, Fraser P: Transcription complex stability and chromatin dynamics in vivo. *Nature* 1996;377:209-213.
- Nunez AM, Berry M, Imler JL, Chambon P: The 5' flanking region of the pS2 gene contains a complex enhancer region responsive to oestrogens, epidermal growth factor, a tumour promoter (TPA), the c-Ha-ras oncoprotein and the c-jun protein. *EMBO J* 1989;8:823-829.
- Hanby AM, Jankowski JA, Elia G, Poulosom R, Wright NA: Expression of the trefoil peptides pS2 and human spasmodic polypeptide (hSP) in Barrett's metaplasia and the native oesophageal epithelium: Delineation of epithelial phenotype. *J Pathol* 1994;173:213-219.
- Probst JC, Skutella T, Müller-Schmid A, Jirikowski GF, Hoffmann W: Molecular and cellular analysis of rP1.B in the rat hypothalamus: In situ hybridization and immunohistochemistry of a new P-domain neuropeptide. *Mol Brain Res* 1995;33:269-276.

## Note Added in Proof

The SML1 gene sequences are also available under the accession numbers U47289-47292 (T. Seib, pers. communication).