European Journal of Human Genetics (1998) 6, 121–128

http://www.stockton-press.co.uk/ejhg

ORIGINAL PAPER

A novel 25 bp tandem repeat within the human trefoil peptide gene *TFF2* in 21q22.3: polymorphism and mammalian evolution

Tuncay Kayademir[1], Elisabeth dos Santos Silva[1], Carsten Pusch[1], Stefanie Beck[1], Jose-Carlos Machado[1,2] and Peter Gött[1]

[1]*Division of Molecular Genetics, Department of Anthropology and Human Genetics, University of Tübingen, Germany*
[2]*Institute of Molecular Pathology and Immunology of University of Porto, Medical Faculty, Porto, Portugal*

Trefoil peptides belong to a family of small secretory proteins characterised by three intrachain disulfide bonds forming the trefoil motif (TFF-domain). They serve to maintain or repair the epithelial mucosa, and promote cell migration. They are predominantly found in gastrointestinal tissues, and are upregulated around areas of epithelial damage, and in meta- and neoplasia. The corresponding three genes are clustered in human 21q22.3. *TFF2* is the only human one encoding two TFF-domains on separate exons. In between (intron 2), a novel 25 bp sequence is located that is tandemly repeated approximately 48 times, but is unique in the human genome. A diallelic polymorphism with a second allele comprising approximately 53 repeat units is present among individuals. Both alleles were cloned on BAC recombinants. In Caucasians (*n* = 78) the allele frequencies found are 0.85 and 0.15, respectively, representing a frequency of heterozygosity of 26%. Ape and monkey species exhibit homologous repeats of shorter total array length, whereas none is found in other mammals.

Keywords: TFF-domain; trefoil factor family; bacterial artificial chromosomes; BAC; pS2/BCEI; hP1.B/hITF; hSP/SML1; TFF; spasmolytic protein; minisatellite

Introduction

In recent years a new cysteine-rich structural motif of about 50 amino acid residues, known as the trefoil or TFF-domain (trefoil factor family) 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. TFF-domains have been highly conserved during evolution; in Xenopus, peptides as well as mucins contain multiple TFF-domains (for review, see[3,4]). In mammals, the three TFF members (new nomenclature, see[5]) are coordinately secreted in mucin expressing cells of the normal gastrointestinal tissues and are upregulated in pathological conditions such as damaged epithelia, sites

Correspondence: Dr Peter Gött, Universität Tübingen, Institut für Anthropologie und Humangenetik, Abt. Molekulare Genetik, Wilhelmstr. 27, D-72074 Tübingen, Germany. Fax: (49)7071-296409, Tel: (49)7071-2976401,
Email: peter. goett@uni-tuebingen.de
Received 23 June 1997; revised 15 October 1997; accepted 31 October 1997

of ulceration, and some meta- and neoplasia. TFF1, originally found in breast cancer cell lines (formerly termed pS2 or breast cancer oestrogen-inducible BCEI) is produced mainly in the foveolar part of the stomach, TFF2 (formerly spasmolytic peptide, SP) is expressed in the glandular region of the stomach and duodenal Brunner's glands, while TFF3 is expressed by the goblet cells of the intestine.

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 wound healing by enhancing cell migration.[6–8] Oral and subcutaneous application of trefoil peptides can protect against induced gastric injury in rats.[9,10] *TFF3* knockout mice are sensitive to dextran sulphate-induced colitis.[11] *TFF1* knockout mice lost the expression of gastric mucus, developed gastric adenomas and, to a degree, invasive carcinomas, suggesting a tumor suppressor function.[12] Interestingly, some 30% of these mice display a complete loss of *TFF2* transcription.

At the gene level, we recently mapped the genomic locus of the human *TFF3* to 21q22.3[13] in the region also assigned to *TFF1* and *TFF2*,[14] and found the clustering of all three genes within a region of 50 kb by isolating a contig of genomic fragments cloned into the Bacterial Artificial Chromosome (BAC) system.[15,16] The unidirectional orientation of transcription and the closer localisation of the two stomach specific genes *TFF1* and *TFF2* is reflected by a coordinated cell- and tissue-specific gene expression. In stomach, pancreas and biliary tract carcinomas, the presence of the regular *TFF1* mRNA is strongly correlated with *TFF2* expression.[17–19] This coexpression was also noted in ulceration-associated cell lineage in Crohn's disease.[20] In hyperplastic gastric polyps, *TFF1* and *TFF3* are coexpressed, whereas in some gastric cancer and cell lines all three genes are active.[16]

Among the three *TFF* genes, *TFF2* is located centrally in the cluster and is the only one encoding two TFF-domains. The modular architecture of the genes is reflected by the fact that the TFF-domains are encoded by a single exon in *TFF1*[21] and *TFF3*,[22] whereas two adjacent exons encode the TFF-domains in *TFF2*.[16] This conserved structural organisation may have emerged from gene duplication and exon shuffling during evolution. It may be just a coincidence that the TFF-encoding exons in *TFF2* are separated by an intron containing a new tandem repeat cluster (minisatellite), the characteristics of which are described here. However, it may turn out to be of functional relevance in modulating gene expression due to a binding site for a transcription factor, or interfering with gene expression, or causing genetic instability. Recent examples related to these functions are the diabetes susceptibility locus *IDDM2*,[23] a minisatellite expansion in a non-coding part of a gene also found in 21q22.3 that is responsible for a monogenic form of epilepsy,[24,25] or the fragile X syndrome caused by minisatellite expansion in *FRA16B*.[26]

Materials and Methods

DNA Samples and Probes

Genomic DNA was isolated from various mammalian species by standard methods. Genomic DNA from healthy Caucasians was provided by Dr C Luckenbach, Tübingen. The great ape and monkey DNA (1 μg each) were a kind gift from W Schempp, Freiburg. TFF positive recombinants were obtained[16] from the human genomic BAC library.[27] BAC DNA was prepared by standard alkaline lysis and purified by phenol/chloroform and ethanol precipitation.

Restriction Analysis, Southern Blotting and Hybridisation

*Hind*III restriction digests of BAC DNA were separated on 0.7% agarose-TAE gels. DNA was transferred in alkaline buffer for 2 h to Nytran membrane (Schleicher & Schüll, Dassel, Germany) by downward blotting.[28] For filter hybridisation, the probes were labelled with ^{32}P-dATP by random priming (Boehringer Mannheim Biochemicals, Mannheim, Germany). A 300 bp cDNA fragment of *TFF2* was isolated from pGEM-hsp200,[2] kindly provided by Dr M-C Rio, Strasbourg. The ^{32}P-labelled repeat-probe for hybridisation shown in Figure 4 was generated by PCR on template BAC 90E5 with oligonucleotides SMLF2 and SIN2 Table 1. The probe was hybridised at 65°C overnight in 7% SDS, 1 mM EDTA, 500 mM sodium phosphate buffer, pH = 7.2. Membranes were washed twice at 65°C for 10 min in 0.1% SDS, 40 mM sodium phosphate buffer, pH = 7.2, and exposed to autoradiography (Fuji RX, Düsseldorf, Germany X-ray film) for 2 to 5 days.

PCR Analysis

Oligonucleotides were designed from cDNA or genomic sequences of *TFF2*.[2] PCR primer sets flanking the repeats and listed in Table 1 were used for amplifications from 30 ng BAC DNA or 100–250 ng genomic DNA using Gold Star polymerase (Eurogentech, Belgium, Seraing, Belgium) or Expand Long Template PCR System (Boehringer Mannheim Biochemicals). PCR was processed according to different annealing temperatures using a synthesis time of 3 min (plus extension time of 20 s per cycle) and increased Mg concentration (2.25 mM). Other conditions were applied as recommended by the manufacturers. For blunt end cloning assays *Pwo*-DNA-Polymerase (AGS, Heidelberg, Germany) was used under the same conditions as *Taq*-DNA-Polymerase except for the 2 mM $MgSO_4$ reaction buffer.

Table 1 Primer pairs and PCR amplification products

Name	*Localisation in TFF2*	*Sequence*	*Template (allele)*	*PCR product (bp)*	*Annealing temp. degC*
SMLF	Exon 2, f	CACCAGTGACCAGTGTTTTGAC	90E5 (A)	2400	61
SLM	Exon 3, r	GTTGGAGAAGCAGCACTTCCG	921F2 (a)	2500	61
SMLF	Exon 2, f	CACCAGTGACCAGTGTTTTGAC	90E5 (A)	1600	59
SINTR2	Intron 2, r	GAGGGCTGTCAGCCCTTATG	921F2 (a)	1700	59
SMLF2	Intron 2, f	GAATCTTCCTGGGCCAGCAGC	90E5 (A)	1400	61
SIN2	Intron 2, r	TTCATCTCACTGTTGGCTGACC	921F2 (a)	1500	61
TFF2E2F	Exon 2, f	GTCCCYTGGTGTTTCCACCC	90E5 (A)	2600	57
TFF2E3R	Exon 3, r	CTGATGCCCGGGTASCCAC	921F2 (a)	2700	57

f = forward; r = reverse in respect of transcriptional direction.

Ligations, Sequence Analysis

TA-cloning of *Taq* PCR fragments with a pT 7 Blue Vector (Novagen, Madison WI, USA) and blunt end cloning in pUC 18 (Pharmacia, Freiburg, Germany) was performed using T4 DNA ligase under standard cloning conditions.

Cycle sequencing with ThermoSequenase (Amersham Life Science, Cleveland, OH, USA) using either ^{32}P-primer labelling or ^{35}S-dATP internal labelling was performed according to the manufacturer's protocol. Primers used for sequencing were SMLF2, SIN2, SINTR2 and SIN25-71 (AACCAACCTAACCAACCTGGGCTAC) located in the 3'-end of the repeat. The sequence alignment and evaluation of the degree of homology was performed using GenBank or EMBL databases on the databases of the Genius Workstation of DKFZ, Heidelberg. Sequence data were processed by BLASTN and SIMILARITY algorithms of the GCG Wisconsin package. EMBL database accession numbers of intron 2 of TFF2 are X97791 and X97792.

Results

Identification and Amplification of the Repeat

During cloning and sequencing the exon-intron boundaries of *TFF2*,[16] we discovered a repetitive element in the intron 2 (Figure 1). Among several overlapping BAC recombinants, BAC 921F2 was chosen to determine a *Hind*III restriction map and use it as a template to amplify intron 2 by primers SMLF and SML (Figure 1 and Table 1). A 2.5 kb PCR product was obtained that was analysed by cycle sequencing, indicating a tandemly repeated motif in the 5' half of intron 2. TA-cloning of the 2.5 kb PCR fragment into a high copy number plasmid gave rise to various deletion inserts in

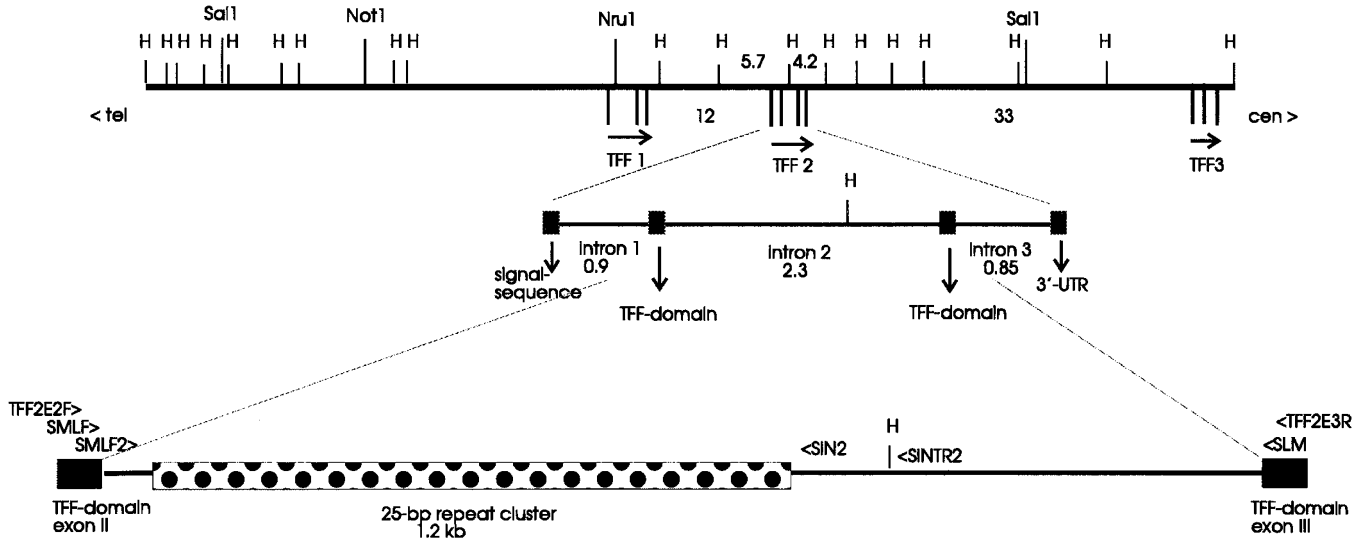

Figure 1 *Genomic organization of the human TFF gene cluster in 21q22.3* TFF2 *and the 25 bp tandem repeats. Bars and black boxes refer to* TFF *exons. Top:* Hind*III (H) restriction map of BAC 921F2 recombinant. Bottom: Map of the 25 bp repeat cluster in the intron 2 of* TFF2 *and orientation of flanking primers. Numbers refer to distances in kilobases.*

the recombinant plasmids (data not shown). Moreover, attempts to subclone the corresponding *Hind*III restriction fragment of 5.7 kb known to contain intron 2 from BAC 921F2 and BAC 90E5 into pUC-derived vectors failed, although the adjacent restriction fragment of 4.2 kb as well as several other *Hind*III fragments of this size were successfully cloned. This indicates that the repeat is unclonable by standard cloning methods, as also observed for other tandem repeats.[29] Cloning and sequencing of the adjacent 4.2 kb *Hind*III restriction fragment allowed us to design new primers (SIN2, SINTR2, Table 1, Figure 1) for amplification of the entire tandem repeat cluster, and to determine the primary structure of intron 2 (EMBL database acc.no. X97791 and X97792). The intron sequence downstream to the *Hind*III site matched perfectly the one published elsewhere, ([30] EMBL database accession no. U47290 and U47291).

Primary Structure and Detection of Different Alleles

Cycle sequencing from both ends of the amplified PCR products of 1.7 and 1.6 kb (Table 1) revealed a repetitive consensus sequence of 25 bp. Ten repeat units were determined on the 5'-end and eight repeat units on the 3'-end (Figure 2). At both sites at least twelve additional repetitive units were detected on autoradiographs, although their primary sequence was not unambiguously determined, indicating a total array size of at least 40 repeat units. As determined from the length of the amplified PCR product (1.7 kb, SMLF2-SIN2, on BAC 921F2, Table 1) the total array size was estimated to be 1.32 kb equalling a total copy number of 53 tandem repeats. PCR amplification from four additional BAC clones (43A9, 90E5, 548B9, 843E9,[16]) yielded total array sizes that are 120 bp shorter than the corresponding one of clone 921F2. Sequence analysis of both ends of the repetitive cluster from template 90E5 exhibited no differences with respect to clone 921F2. Since also at least 40 repetitive units were detected, the size of the 90E5 allele corresponds to a total array size of 48 repetitive units (Figure 2).

At the 5' flanking site two cryptic repeat units were evident 50 bp downstream to the start site of intron 2 (Figure 2). One hundred bp downstream to the 5' site of intron 2, the following repetitive consensus sequence was found: AGTCT(A/G)GGGCTAGCC-CAGGTTGGTT. A palindromic sequence able to form a perfect GC-rich 5 base pair hairpin element (GGGCTAGCCC) is located in the centre of each unit. The fine structure of the cluster revealed variation in the 25 bp unit. After 6 conserved repetitive units two of three adjacent cytosin bases are deleted in unit no.7 (Figure 2).

Even more variation of the consensus sequence was obvious at the 3' end of the repeat, located 208 bp upstream to the *Hind*III site, where the consensus sequence was 'fading out' by two cryptic copies. Every second repetitive copy showed a 24% variation. It included a deletion of one of the three adjacent cytosine bases. The other copies showed variability within the adjacent three guanosine bases to an overall extent of 8%. These data indicate a high degree of inter-repeat variability, which suggests a low frequency of heterozygosity.

```
  C  F  H  P  L  P  K  Q
TGTTTCCACCCCCTCCCAAAGCAAGGTAATCTTCCAGGGAATCTTCCTGG
GCCAGCAGCTGGCAACCCAGGACCCAGCTTCACAGGCGGAGCCCAGAGCA
GGGGCTGGAGGAGGCCCAGTTGCTAGTCTAGGGTTAGCCTGGGTGGGTTA
GTCTCGAGCTAGCCCCGGTTGGTTAGTCTGGGGCTAGCCCAGGTTGGTTA
GTCTAGAGCTAGCCCAGGTTGGTTAGTCTGGGGCTAGCCCAGGTTGGTTA
GTCTGGGGCTAGCCCAGGTTGGTTAGTCTAGGGCTATCTAGGCTAGTTAG
TCTAGAGCTACCCAGGTTGGTTAGTCTGGGGCTAGCCCAGGTTGGTTAGT
CTGGGGCTAGCCCAGGTTGGTTAGTCT.....
.....590 bp (refer to allele A).....
.......TAGCCTTGGACTGNTAGTCTAGAGGTTAGCCTAGAGGACTGCT
AGTCTAGAGGTAGTCTAGGGCTAGCCCAGGTTGGTTAGTCTGGGGTAGTC
TGGGTAGCCNATGTTGGTTAGTNTTAGACTAGCCTGGA[C/G]TGCTAGTCTAG
AGGTAGCCCAGGTTGGTTAGTCTGGTACTAGCCTGGACTGTNAGTCTAGA
GGTAGCCCAGGTTGGTTAGGTTGGTTAGTCTGGGACTAGTCTGGACTGTT
AGTCTAGAGGTAGCCCAGGTTGGTTAGTTTGGGACTAGCCTGGACTGTTA
GTCTAGAGGTAGCCCAGATTGGTTAGTCTGGGACTAGTCTGGACTGCTAG
TCTAGAGGTAGCCCAGGTTGGTTAGCCTGGGGCCAGCCTGGACTGTTAGT
CTAGAGGTAACCCAGGTCAGCCAACAGTGCAGATGAAAATTTCCCACCTA
CCCTGTTTCTACACTGTTAGTTCTTTCAACAGACATGTGTGTGTGGAGCC
ATCAGTTTTACTTTAGTTGAGAAAAAAATATATATATATAGTAGGTCT
CCTCTAGTTTTTGAAGTGTGACTTCTGAAGAAGCTTCCATGGGGAAATGA
AGGTATTTAATAGGACAGCAGTAACATAAGGGCTGACAGCCCTCAAATGT
```

Figure 2 *Flanking sequences of the 25 bp repeat cluster. Exon 2 and the* Hind*III restriction site – bold type; primers SMLF2, SIN2, SINTR2 and SIN25-71 – underlined. (EMBL database acc.no.X97791 and X97792); 25 bp consensus motif (AGTC-*T^{A}_{G}*GGGCTAGCCCAGGTTGGTT) –. The 590 bp sequence not determined by nucleotide sequencing refers to allele A.*

A New Single Copy-Minisatellite in the Human Genome

To determine the abundance of this repeat cluster within the human genome, genomic DNA of three human individuals was digested by *Hind*III and subjected to Southern blot analysis using the total array of the repeat cluster (1.5 kb PCR product) as a probe. Beside the 5.8 kb band no other autoradiographic signal was detected in the genomic DNA as well as in BAC 843E9 and 921F2 recombinants (data not shown). Sequence data library searching by the 25 bp consensus sequence exhibited no significantly homologous

sequence. These results indicate that the 25 bp minisatellite is exclusively present in the intron 2 of *TFF2* in human 21q22.3.

A Diallelic Polymorphism in the Caucasian Population

The genotype distribution and allele frequencies in 78 unrelated white Caucasian subjects is summarised in Table 2. Two alleles were found that are identical in size to those already cloned in BAC recombinants (Figure 3A). A 120 bp difference in size accounts for the differences of alleles termed 'A' and 'a', corresponding to a repeat copy number of 48 and 53, respectively (Table 2). The allelic distribution agreed well with the Hardy–Weinberg equilibrium. From size analysis of DNA molecular markers the resolution of gel electrophoresis used in this study was able to detect a size difference of 25 bp, thus excluding the possibility of additional alleles differing in one unit within the panel of 78 unrelated individuals. Inheritance analysis in five families, each including father, mother and one child, showed a Mendelian type of inheritance (Figure 3B). Two families (nos 2 and 3) were genotyped to be heterozygous, the other three were homozygous for the allele A.

Table 2 Genotype and allele frequency in 78 unrelated individuals

	AA	*Aa*	*aa*	*A*	*a*
n=78	56 (71.8%)	20 (25.6%)	2 (2.6%)	0.85	0.15

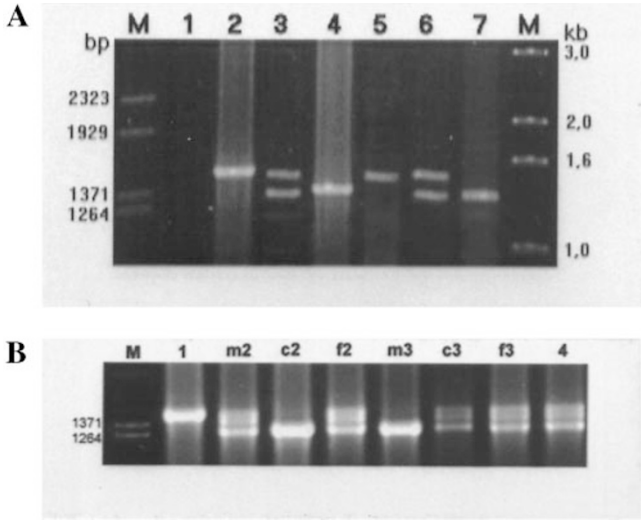

Figure 3 *Diallelic polymorphism of the 25 bp repeats shown by PCR amplification with primers SMLF2 and SIN2.* ***a****) Total array size in BAC recombinants and in unrelated individuals. Slot 1: negative control; 2: BAC 921F2; 3 and 6: heterozygous individuals; 4: BAC 90E5; 5: homozygous individual displaying allel 'a'; 7: homozygous individual displaying allel 'A'; M: DNA size markers.* ***b****) Inheritance analysis in two families. Family no. 2 (m2, c2, f2) presents a heterozygous mother (m) and father (f) and a homozygous child (c), whereas in family no. 3 (m3, c3, f3) only the mother is homozygous. Nos 1 and 4 refer to unrelated homozygous and heterozygous individuals, respectively. The upper band in the heterozygous individuals appeared doubled probably due to a PCR artefact.*

Evolutionary Origin of the Repeated Sequence

The presence of the 25 bp consensus motif was tested in several mammalian species to reveal its evolutionary origin. Classical zooblot analysis was hampered by lack of enough material from some monkey species. Using PCR with human intron primers flanking the repeats (SMLF2, SIN2), amplified products from chimpanzee (*Pan troglodytes*) and gorilla (*Gorilla gorilla*) were probed to be positive for the tandem repeat by Southern hybridisation (Figure 4a and Figure 4b). In the gorilla, a double band was amplified, indicating heterozygosity in great apes. In other monkeys either no PCR product was found or the hybridisation revealed no signal, presumably because of sequence variation to the human specific intron primers. To circumvent this problem, cross-species primers TFF2E2F, and TFF2E3R were designed from conserved human and murine TFF-domain sequences in exon 2 and 3.[2] Since the exon–intron boundaries are conserved in human and mouse (O Lefebvre, M-C Rio, personal communication), the size of the intron can be deduced, and results shown in Figures 4c and 4d are summarised in Table 3. Amplified products were obtained from genomic DNA of mouse (0.85 kb), rat (1.0 kb), *Oryctolagus spec* (rabbit; 1.11 kb), *Lemur catta* (1.15 kb), *Theropithecus gelada* (1.58 kb), chimpanzee (2.2 kb) and man (2.4 kb and 2.5 kb), indicating a continuous enlargement of intron 2 during mammalian evolution (Figure 4c, Table 3). When probed for the presence of the tandem repeat, man, chimpanzee and the simian *Theropithecus gelada* were positive, whereas the lemur species and the other mammalian species were negative (Figure 4d, Table 3), indicating that the minisatellite was acquired during simian evolution. Additional bands amplified by the cross-species primer are nonspecific, since they are probed to be negative for the tandem-repeat by hybridisation. During the process of reviewing this paper the mouse *TFF2* gene was sequenced (GenBank database acc. No. U78770). Intron 2 has a size of 773 bp, which agreed well with the

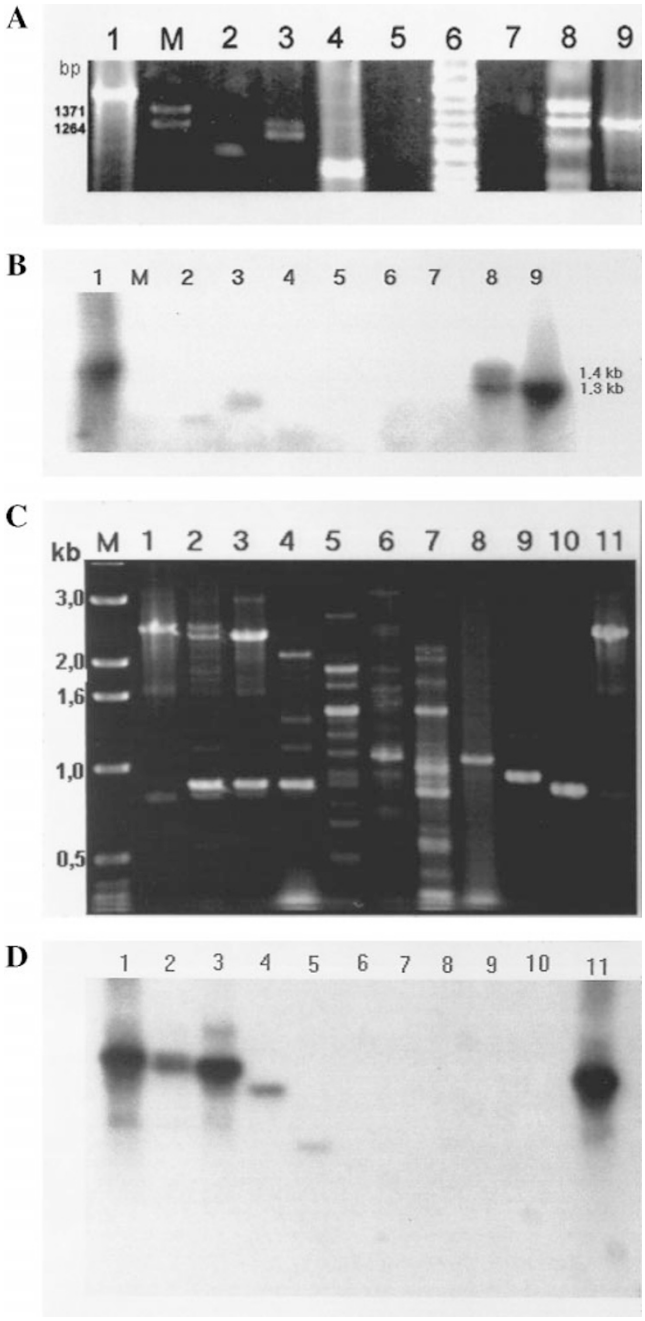

Figure 4 *Evolutionary origin of the 25 bp minisatellite.* ***a****) PCR amplification with primers SMLF2 and SIN2 flanking the minisatellite, using DNA of various sources: 1: BAC 921F2; 2:* Pan troglodytes*; 3:* Gorilla gorilla*; 4:* Theropithecus gelada *(old world monkey); 5:* Ateles geoffreyi *(new world monkey); 6:* Lemur catta*; 7: negative control; 8: heterozygous individual; 9: BAC 90E5.* ***b****) corresponding Southern blot of amplified products hybridised with the 25 bp repeat probe.* ***c****) Amplification of the intron 2 of* TFF2 *in mammalian species by conserved cross-species primers.* ***d****) corresponding Southern blot hybridisation with the 25 bp minisatellite probe 1: BAC 921F2; 2: heterozygous individual; 3: homozygous individual; 4:* Pan troglodytes *(chimpanzee); 5:* Theropithecus gelada*; 6:* Lemur catta*; 7:* Sus suis *(pig); 8:* Oryctolagus spec. *(rabbit); 9:* Rattus norvegicus*; 10:* Mus musculus*; 11: BAC 843E9.*

size of 750 bp determined by our analysis. Also, the data from the GenBank database revealed no sequences homologous to the 25 bp tandem repeat.

Discussion

In the following we describe the characterisation of a minisatellite in intron 2 of *TFF2* coding for the human trefoil factor 2. Several factors point to the relevance of trefoil peptides in mucosal restitution, eg wound healing and cell motility, differentiation, and association with cancerogenesis.[5,10,12,16] The application of trefoil peptides as therapeutic agents in several ulcerative diseases has been suggested.[9,31] Trefoil peptides are coordinately secreted from mucin expressing cells of the normal gastrointestinal tissues and are upregulated in pathological conditions such as damaged epithelia, sites of ulceration, and some meta- and neoplasia. *TFF2* has very restricted expression sites in normal gastrointestinal tissue, since it is limited to the glandular region of the stomach and duodenal Brunner's glands.

The finding of a minisatellite raises at least three possible functional aspects. First, the minisatellite could regulate transcription of *TFF2* or the coordinated expression all *TFF* genes through binding of a transcription factor. Examples are the 14-bp repeats 365 bp upstream of the transcription start site of the human insulin gene[23] and the 28-bp repeats 1 kb downstream of the polyadenylation site of the human *HRAS1* oncogene.[32] These examples are associated with an increased risk of insulin-dependent diabetes and several cancers, respectively. The palindromic sequence GGGCTAGCCC present in each repeat of the *TFF2* intron suggests the existence of a functional DNA

Table 3 Occurrence of the 25 bp tandem repeat and sizes of intron II in *TFF2*

Mammalian species	*Presence of the tandem repeat*[a]	*No. of 25 bp units*[a]	*Size of intron 2 (kb)*[a]
Homo sapiens	+(polymorphic)	48 (A) 53 (a)	2.3–2.4
Gorilla gorilla	+(polymorphic)	44; 47	n.d.
Pan troglodytes	+	41	2.1
Theropithecus gelada	+	<40	1.5
Lemur catta	–	–	1.05
Oryctolagus spec.	–	–	1.0
Rattus norvegicus	–	–	0.9
Mus musculus	–	–	0.75

[a]Data from Figure 4, intron sizes are calculated using the conserved sizes of exons 2 and 3.

binding protein. The presence of the 25 bp repeats in human and simian primates, but not in lower mammals, would argue for a different regulation of *TFF2* acquired during evolution. In fact. *TFF2* is most prominent in porcine pancreas, and is present in mouse and rat acinar cells of the pancreas, whereas it is not found in the human pancreas.[2,33] Although no direct relationship of *TFF2* with a human disease is yet proven, *TFF2* and *TFF1* transcription is coordinately upregulated in some adenomas of stomach, pancreas, and biliary tract.[17,18,19] On the other hand, *TFF1*-knockout mice that are also defective in *TFF2* expression develop adenocarcinomas in stomach and duodenum,[12] which may indicate different regulation in lower mammals.

Secondly, minisatellites can disturb gene expression by instability leading to genomic rearrangements or by expansion processes interfering with transcription on the chromatin level. For example, an expanded 33-bp, AT-rich minisatellite results in the common distamycin A-sensitive fragile site, *FRA16B* causing fragile X syndrome.[26] The small alleles containing 7-12 sequence-heterogeneous repeats are enlarged up to 2000 copies of one repeat type in *FRA16B*. Thirdly, recent examples show that minisatellite expansion disrupt gene expression by yet unknown mechanism. The expansion of a GC-rich 12-bp repeat unit is responsible for a monogenic form of epilepsy termed EPM1.[25,34] This repeat, normally present in 2–3 copies in normal alleles of the promoter region of the gene encoding cystatin B is expanded to a large insertion (600–900 bp) leading to inhibition of transcription in affected patients.

The use of the 25 bp minisatellite in *TFF2* for LOH studies is somewhat hampered by the low frequency of heterozygosity (26%). However, it may be of great interest to study the stability of this repeat cluster, especially in clinical collectives displaying chronic gastritis, gastric dys- and neoplasia, or inflammatory bowel disease (Crohn's disease, ulcerative colitis), conditions that are likely to be linked to TFF expression. Two commonly deleted regions on 21q in differentiated gastric adenocarcinomas have been recently suggested.[35] Studies to determine the role of this repeat in transcriptional regulation and in gastric carcinogenesis are currently being undertaken.

Acknowledgements

We would like to thank J Kömpf, C Luckenbach and R Pöltl for providing human and mammalian DNA samples. We are indebted to N Blin for helpful discussions. The study was supported by a grant from the Deutsche Krebshilfe (10-1120-Gö1) and the BMBF (01KS9602). S Beck is a recipient of a Baden-Württemberg predoctoral training grant.

References

1 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 spasmolytic polypeptide (PSP), and frog skin peptides (spasmolysins). *FEBS Lett* 1989; **250**: 85–90.
2 Tomasetto C, Rio MC, Gautier C *et al*: 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.
3 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.
4 Sands BE, Podolsky DK: The trefoil peptide family. *Annu Rev Physiol* 1996; **58**: 253–273.
5 Wright NA, Hoffmann W, Otto WR, Rio M-C, Thim L: Rolling in the clover: Trefoil factor family (TFF) domain peptides, cell migration and cancer. *FEBS Lett* 1997; **408**: 121–123.
6 Dignass A, Lynch Devaney K, Kindon H, Thim L, Podolsky DK: Trefoil peptides promote epithelial migration through a transforming growth factor beta-independent pathway. *J Clin Invest* 1994; **94**: 376–383.
7 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.
8 Alison MR, Chinery R, Poulsom R, Ashwood P, Longcroft JM, Wright NA: Experimental ulceration leads to sequential expression of spasmolytic polypeptide, intestinal trefoil factor, epidermal growth factor and transforming growth factor alpha mRNAs in rat stomach. *J Pathol* 1995 **175**: 405–414.
9 Babyatsky MW, DeBeaumont M, Thim L, Podolsky DK: Oral trefoil peptides protect against ethanol- and indomethacin-induced gastric injury in rats. *Gastroenterology* 1996; **110**: 489–497.
10 Playford RJ, Marchbank T, Chinery R *et al*: Human spasmolytic polypeptide is a cytoprotective agent that stimulates cell migration. *Gastroenterology* 1995; **108**: 108–116.
11 Mashimo H, Wu DC, Podolsky DK, Fishman MC: Impaired defense of intestinal mucosa in mice lacking intestinal trefoil factor. *Science*; 1996; **274**: 262–265.
12 Lefebvre O, Chenard MP, Masson R *et al*: Gastric mucosa abnormalities and tumorigenesis in mice lacking the pS2 trefoil protein. *Science* 1996; **274**: 259–262.
13 Schmitt H, Wundrack I, Beck S *et al*: A third P-domain peptide gene, human intestinal trefoil factor (TFF3), maps to 21q22.3. *Cytogenet Cell Genet* 1996; **72**: 299–302.
14 Theisinger B, Welter C, Grzeschik KH, Blin N: Assignment of the gene for human spasmolytic protein (hSP/SML1) to chromosome 21. *Hum Genet* 1992; **89**: 681–682.

15 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* 1996; **98**: 233–235.
16 Gött P, Beck S, Machado JC, Carneiro F, Schmitt H, Blin N: Human trefoil peptides: Genomic structure in 21q22.3 and coordinated expression. *Eur J Hum Genet* 1996; **4**: 308–315.
17 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 spasmolytic polypeptide gene (hSP) in diffuse type of stomach carcinoma. *Eur J Cancer* 1991; **27**: 770–773.
18 Welter C, Theisinger B, Seitz G, Tomasetto C, Rio MC, Chambon P, Blin N: Association of the human spasmolytic polypeptide and an estrogen-induced breast cancer protein (pS2) with human pancreatic carcinoma. *Lab Invest* 1992; **66**: 187–192.
19 Seitz G, Thelsinger 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.
20 Wright NA, Poulsom R, Stamp G *et al*: Trefoil peptide gene expression in gastrointestinal epithelial cells in inflammatory bowel disease. *Gastroenterology* 1993; **104**: 12–20.
21 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.
22 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.
23 Kennedy GC, German MS, Rutter WJ: The minisatellite in the diabetes susceptibility locus IDDM2 regulates insulin transcription. *Nature Genet* 1995; **9**: 293–298.
.24 Virtaneva K, D'Amato E, Miao IM *et al*: Unstable minisatellite expansion causing recessively inherited myoclonus epilepsy, EPM1. *Nature Genet* 1997; **15**: 393–396.
25 Lalioti MD, Scott HS, Buresi C *et al*: Dodecamer repeat expansion in cystatin B gene in progressive myoclonus epilepsy. *Nature* 1997; **386**: 847–851.
26 Yu S, Mangelsdorf M, Hewett D *et al*: Human chromosomal fragile site *FRA16B* is an amplified AT-rich minisatellite repeat. *Cell*1997; **88**: 367–374.
27 Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M: Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc Natl Acad Sci USA* 1992; **89**: 8794–8797.
28 Chomczynski P: One-hour downward alkaline capillary transfer for blotting of DNA and RNA. *Analytical Biochemistry* 1992; **201**: 134–139.
29 Kang HK, Cox DW: Tandem repeats 3' of the IGHA genes in the human immunoglobulin heavy chain gene cluster. *Genomics* 1996 **35**: 189–195.
30 Seib T, Blin N, Hilgert K *et al*: The three human trefoil genes *TFF1, TFF2* and *TFF3* are located within a region of 55 kb on chromosome 21q22.3. *Genomics* 1997; **40**: 200–202.
31 Featherstone C: Soothing the troubled gut with trefoil factors. *Lancet* 1997; **349**: 706.
32 Green M, Krontiris TG: Allelic variation of reporter gene activation by the *HRAS1* minisatellite. *Genomics* 1993 **17**: 429–434.
33 Taupin DR, Pang KC, Green SP, Giraud AS: The trefoil peptides spasmolyitc polypeptide and intestinal trefoil factor are major secretory product of the rat gut. *Peptides* 1995; **16**: 1001–1005.
34 Lafrenière RG, Rochefort DL, Chrétien N *et al*: Unstable insertion in the 5' flanking region of the cystatin B gene is the most common mutation in progressive myoclonus epilepsy type 1, EPM1. *Nature Genet* 1997; **15**: 298–302.
35 Sakata K, Tamura G, Nishizuka S *et al*: Commonly deleted regions on the long arm of chromosome 21 in differentiated adenocarcinoma of the stomach. *Genes, chromosomes & cancer* 1997; **18**: 318–321.