Cell. Mol. Life Sci. 62 (2005) 2956–2973
1420-682X/05/242956-18
DOI 10.1007/s00018-005-5484-6

Cellular and Molecular Life Sciences

Structure of mammalian trefoil factors and functional insights

L. Thim[a,*] and F. E. B. May[b]

[a] Protein Engineering, Novo Nordisk A/S, 2760 Måløv (Denmark), Fax: +45 44663450, e-mail: lth@novonordisk.com
[b] Nothern Institute for Cancer Research, The Medical School, University of Newcastle, Framlington Place, Newcastle Upon Tyne, NE2 4HH (United Kingdom)

Online First 5 December 2005

Abstract. The present review will include the mammalian trefoil factors, TFF1, TFF2 and TFF3. It will summarise the amino acid sequences from different species, their posttranslational modifications and their structures determined by X-ray analysis and nuclear magnetic resonance studies. Trefoil factors all have a well-defined, structurally conserved trefoil domain. The trefoil domain consists of 42 or 43 amino acid residues and contains 6 cysteine residues that form disulphide bonds in a 1–5, 2–4 and 3–6 configuration. By the establishment of an additional intra-molecular disulphide bond at the C-terminal end, TFF1 and TFF3 form homodimers or heterodimers. This dimer formation of TFF1 and TFF3 will be discussed, and the possible implications for biological activity will be reviewed. The physicochemical characteristics including protease stability of trefoil factors will be summarised. The biological implications of different molecular forms of trefoil factors and their interaction with mucins will be discussed together with other functional insights.

Key words. Trefoil Factors; TFF domain; homodimer; heterodimer; spasmolytic polypeptide; pS2, mucins; *Helicobacter pylori*.

Introduction

The mammalian trefoil factor family (TFF) contains three members, TFF1, TFF2 and TFF3. Trefoil factors are expressed and secreted by epithelial cells that line mucus membranes. They are therefore usually expressed in association with mucins, the major protein component of a mucus gel. Frequently referred to as gastrointestinal factors, they are de facto expressed more widely, notably in the bronchial and urogenitary tracts. Individual trefoil factors are expressed by different cells, for instance TFF1 is expressed principally by the superficial cells of the gastric mucosa, whereas TFF2 expression is largely restricted to cells of the basal gastric glands and TFF2, but not TFF1, is expressed in Brunner's glands of the duodenum. Trefoil factors are also expressed, sometimes at high levels and sometimes apparently ectopically, in many human adenocarcinoms [1, 2].

The amino acid sequences of individual mammalian trefoil factors – porcine TFF2; human TFF1 and rat TFF3 – were first reported between 1981 and 1991. The amino acid sequence of porcine TFF2 was the first to be published in 1981 [3, 4]. The nucleotide sequence encoding human TFF1 was reported in 1984–85 [5, 6] followed by the nucleotide sequence encoding rat TFF3 in 1991 [7]. Since these initial sequence determinations, human, porcine, rat and murine TFF1, TFF2 and TFF3 messenger RNAs (mRNAs) have been cloned and sequenced and the translated coding sequences published, with the exception of porcine TFF1, which is not known [8–10]. More recently the sequences of chimpanzee and canine

* Corresponding author.

Figure 1. Sequence alignment of mammalian trefoil factors. Residues completely conserved between the orthologues are shown in green and those with substitutions are shown in purple. Sequences are from the following references: human TFF1 [24], rat TFF1 [63], murine TFF1 [64], human TFF2 [8, 32], rat TFF2 [65], murine TFF2 [8], porcine TFF2 [27, 29], canine TFF2 [66], human TFF3 [10], rat TFF3 [7, 67, 68], murine TFF3 [69, 70], porcine TFF3 [71] and canine TFF3 [72], or from UniProt. α and β indicate α-helix and β-sheet, respectively.

trefoil factors have become available with the release of the corresponding genomes. The orthologues of TFF1, TFF2 and TFF3 are aligned in figure 1.

By definition trefoil factors contain at least one copy of the trefoil domain. A trefoil domain was defined originally as a conserved sequence of 38 or 39 amino acid residues, within which 6 cysteine residues with essentially conserved spacings form disulphide bonds in a 1–5, 2–4 and 3–6 configuration [11]. This definition of a trefoil domain, from the first cysteine to the sixth cysteine, was based on the assignment of disulphide bonds in porcine TFF2 and the linear sequence of human TFF1. The identification of the disulphide bond configuration also led to the suggestion that trefoil domains might form three-leaved structures reminiscent of a trefoil or clover leaf, and thus the trefoil factor family was given its name [11].

TFF1 contains 60 amino acid residues and a single trefoil domain, TFF2 contains 106 amino acid residues and two trefoil domains, and TFF3 contains 59 amino acid residues and a single trefoil domain. The two single-domain trefoil factors, TFF1 and TFF3, contain a seventh cysteine residue, in addition to the six cysteine residues of the trefoil domain, three amino acids from the carboxy-terminus. Both TFF1 and TFF3 form homodimers via a disulphide bond with the extra-trefoil domain cysteine residue and have the potential to form intramolecular disulphide bonds with other proteins [25, 35, 39]. TFF2 has two extra-trefoil domain cysteine residues that form an intermolecular disulphide bond in the TFF2 peptide [11].

The first secondary structure of a trefoil domain was reported in 1992 with the assignment of the nuclear magnetic resonance (NMR) spectra for porcine TFF2 [12] and the first tertiary structure, again for porcine TFF2, was published the following year [13]. Since then the structures of all three trefoil factors have been solved with, most recently, the publication of the structure of dimeric TFF3 [14]. Comparison of the structures indicates striking similarity of both secondary structural elements and overall tertiary structure within the trefoil domain combined with an almost complete absence of

commonality outside the trefoil domains. The divergent structures contribute to the different physicochemical characteristics of the three molecules [39].

From examination of the available three-dimensional structures and comparison of the conservation between mammalian trefoil factors, it has become clear that the conserved trefoil domain extends somewhat beyond the 38–39 amino acid residues suggested originally. From the structural data now available it seems reasonable to redefine a trefoil domain as a sequence constituting 42–43 amino acid residues. Compared with the original definition, this extended definition includes one amino acid residue more at the N-terminal end of the domain and three amino acid residues more in the C-terminal end, as has been indicated in figure 1. Solution of the tertiary structures of all four mammalian trefoil domains confirmed the supposition that they have a three-loop structure. The tertiary structures also showed that the three loops are not orientated in a single plane but are stacked parallel to each other (see below and fig. 2).

The first concrete evidence for a function of trefoil factors came with the demonstration that TFF2 and TFF3 stimulate the migration of intestinal cells grown in culture [15]. This was followed quickly by the demonstration that TFF2 could protect against damage in an in vivo model of gastric ulceration [16]. The compelling results from the TFF-null mice work [17–19] reveal much about the normal roles of trefoil factors and underpin the frequently ignored concept that they have significantly different functions. The phenotypes of TFF1-null mice also emphasised that, at least for TFF1, its interplay with mucins is critical [18]. Parallel studies in vitro [53, 56] and biochemical analyses of human material [38, 58] emphasised further the importance of the interaction with mucins.

The increased structural information combined with knowledge about the roles and sites of expression of trefoil factors in the body allows one to gain some functional insight into how these molecules may work. Such insights include which forms of and which parts of the molecules are important for biological functions, including interactions with molecules such as mucins and other proteins. The present review summarises present knowledge on the primary, secondary, tertiary and quaternary structures of mammalian trefoil factors and their physicochemical characteristics with the aim of gaining some insight into how they may function. It also covers the functional significance of different molecular forms of trefoil factors that have been described.

Structure

Primary structures

Size, amino acid sequence and domain structure of TFFs

The amino acid sequences of the known mammalian TFF1, TFF2 and TFF3 orthologues are shown in figure 1A–C. The majority of the sequences are 'predicted coding sequences'. The notable exception is porcine TFF2, which was first isolated from porcine pancreas as a side fraction during the purification of insulin for therapeutic purposes, and for which the amino acid sequence was determined initially [20–22]. For human TFF1, the amino acid sequence has been confirmed by a combination of amino acid sequencing and mass spectroscopy [23–25]. The numbers of amino acid residues in the mature peptides from different species are the same: TFF1 always contains 60 residues, TFF2 106 residues and TFF3 59 residues. The single trefoil domain in TFF1 consists of 42 amino acid residues (from Thr6 to Pro47), the first trefoil domain of TFF2 consists of 43 residues (from residue Gln7 to Pro48), the second domain of TFF2 consists of 42 residues (from residue Gln57 to Pro98) and the trefoil domain in TFF3 consists of 42 residues (from residue Gln10 to Pro51). The above residues are from the human sequences (see fig. 1). The first domain in TFF2 differs from the other trefoil domains in that it contains an additional amino acid residue.

Homology between orthologues

Between TFF1 orthologues from different species, the overall completely conserved amino acid residues (green residues in fig. 1) constitute 52% (31 out of 60). Within the trefoil domain, the complete conservation is 57% (24 out of 42) compared with 33% (6 out of 18) for the rest of the molecule. In addition to the conservation in the trefoil domain, TFF1 orthologues all have a very acidic carboxy-terminus. In human TFF1, for example, four out of the six C-terminal amino acid residues are glutamic acid residues. For TFF2, the homology between species is also a little higher in the trefoil domains, 60 and 71% for domains 1 and 2, respectively, than it is in the rest of the molecule, 52%. The conservation between TFF3 orthologues is only 22% in the 9-amino acid residue N-terminal region of the molecules. In the trefoil domain, the absolute conservation is 67%, which is similar to the conservation found within the trefoil domains of TFF2 and a little higher than in TFF1. The C-terminal portion of TFF3 is very well conserved with 5 out of 8 residues (63%) being identical. If one ignores canine TFF3 and compares the sequences from the other five species – human, chimpanzee, porcine, rat and murine TFF3 – only a single substitution, Thr55Ala in human TFF3, discriminates between them (fig. 1C). This indicates that in addition to the conservation in the trefoil domain, amino acid conservation in the C-terminal region but not the N-terminal region may be important for the biological function of TFF3.

Conservation between all mammalian trefoil factors

For all three mammalian trefoil factors, there is a tendency to a higher degree of conservation within the trefoil domains than in other parts of the molecules, with the excep-

tion of the carboxy-terminus of TFF3. In addition to the six cysteine residues, some other residues within the trefoil domain are conserved between TFF1, TFF2 and TFF3 from almost all species sequenced thus far: the arginine residue between the first two cysteines; the sequence around the second cysteine, N/DCGF/YP-V/IT/S; the phenylalanine after the third and fourth cysteines; and the VPWCF-P sequence around the sixth cysteine at the C-terminal end of the trefoil domain (fig. 1). This indicates that these amino acids are restricted by important structural or functional constraints common to all trefoil domains.

Post-translational modifications and molecular forms

Removal of signal sequence

TFF1, TFF2 and TFF3 are all synthesised with a signal secretion sequence of between 21 and 27 amino acid residues that is removed by proteolysis during passage through the endoplasmic reticulum.

N-terminal cyclisation

Human TFF1 was purified and characterised first by Rio and co-workers [24] from human gastric juice and medium conditioned by MCF-7 breast cancer cells. The techniques used to characterise TFF1 included direct amino acid sequencing and peptide mapping followed by sequencing of tryptic fragments. The results from this study indicated that human TFF1 isolated from gastric juice was modified at the amino-terminus to pyrrolidone carboxylic acid (pyrGlu), whereas TFF1 from MCF-7 cells was not. The N-terminal sequence of TFF1 isolated from MCF-7 cells was confirmed independently by Mori et al. [26]. More recently Westley et al. [25] purified and characterised human TFF1 from gastric mucosa. N-terminal sequencing of TFF1 indicated that the N-terminal Glu was not modified. It proved possible for the first time to identify all the tryptic peptides of TFF1 by mass spectroscopy. The C-terminal 21 amino acid residue tryptic peptide was identified only in negative ion mode, which may explain why it was not detected previously.

As for TFF1, all the TFF2 orthologues have either a Glu or a Gln as the N-terminal residue, both of which can form pyrGlu. This modification has been described for porcine TFF2 [27] but has not been studied in TFF2 from other species.

There are no published studies that characterise TFF3 isolated from natural sources. However, pyrGlu modification of the N-terminal residue of rat TFF3 produced in yeast has been demonstrated [28].

N-glycosylation

Apart from the N-terminal pyrGlu identified when TFF2 was first isolated from porcine pancreas [20–22], no other posttranslational modifications were described in the porcine molecule [3, 27, 29, 30]. In contrast to the porcine molecule, human TFF2 has a consensus N-glycosylation site, Asn-Arg-Thr at positions 15–17 [8], which indicates that human TFF2 has the potential to be glycosylated on Asn15 (fig. 1B). Both glycosylated and non-glycosylated human TFF2 were produced when it was expressed in yeast, which indicated that the N-glycosylation site was used partially [31]. The question whether or not TFF2 is glycosylated in normal human tissue was addressed in a detailed study by May et al. [32]. In normal human gastric tissue only the glycosylated form of TFF2 could be detected. In gastric juice, the majority of the TFF2 molecules were also glycosylated, although 10–20% could be detected in the non-glycosylated form. Later studies have also shown that both forms are present in normal human gastric juice and that the ratio between these forms as well as the total concentrations may vary considerably during a 24-h period [33].

Intermolecular disulphide bond formation

The six cysteine residues of trefoil domains were shown to form intermolecular disulphide bonds in a 1–5, 2–4 and 3–6 configuration in porcine TFF2 [11]. This configuration of disulphide bond formation contributes to the criteria that define a trefoil domain. In porcine TFF2, purified from the pancreas, the two extra-trefoil domain cysteine residues, Cys5 and Cys104 also form an intermolecular disulphide bond.

Intra-molecular disulphide bond formation

TFF1 has a seventh free cysteine residue outside the trefoil domain at the carboxy-terminus of the molecule that has the potential to form intramolecular disulphide bonds. When recombinant TFF1 was produced in bacteria, it was very heterogeneous [34–35]. The heterogeneity was reduced by treatment with thiol-containing reagents, suggesting that it was caused by the odd number of cysteine residues in mature TFF1. This view was reinforced by substitution of the extra-trefoil domain cysteine residue with a serine residue [34]. Further studies showed that TFF1 Cys58 was able to form homodimers via a disulphide bond involving Cys58, whereas the Ser58 analogue was always produced as a monomer [35]. When recombinant TFF1 was expressed in yeast, it was produced with Cys58 as a free cysteine (–SH form) or coupled to either another cysteine or to glutathione, which reinforces the notion that Cys58 has the potential to form intra-molecular disulphide bonds [36].

Characterisation of TFF1 secreted from MCF-7 and EFM-19 breast cancer cells grown in culture identified three molecular forms: TFF1 monomer, TFF1 homodimer and a heterooligomer of around 65 kDa that was stabilised by a disulphide bond [35, 37]. No further characterisation of the 65-kDa form of TFF1 has been reported.

In later studies three different molecular forms of TFF1 originating from normal human stomach were described

[38]: TFF1 monomer, TFF1 dimer (homodimer) and TFF1 forming a 25-kDa complex (heterodimer). The TFF1 homodimer is indistinguishable from recombinant TFF1 homodimer, and it is most likely that the dimer is formed from a disulphide bond between Cys58 of two monomers. The TFF1 heterodimer was recently purified from human gastric mucosa and the protein partner in the TFF1 heterodimer identified as a previously unknown protein that was called TFIZ1 for *t*refoil *f*actor *i*nteractions(*z*) 1. TFIZ1 is an 18.31-kDa secreted protein that contains a brichos domain and homology with pulmonary surfactant-associated protein C precursor. The TFF1 heterodimer comprises one molecule each of TFF1 and TFIZ1 and is stabilised by a disulphide bond between the two proteins [25].

If TFF2 is to form intramolecular disulphide bonds via the extra-trefoil domain cysteine residues, then the intermolecular bond between Cys5 and Cys104 [11] would need to be broken. There is no evidence that this occurs.

When recombinant rat TFF3 was expressed in yeast, only TFF3 homodimers were produced. When recombinant human TFF3 was expressed, both TFF3 monomers and TFF3 dimers were produced, and in the monomeric form, Cys57 formed a disulphide bond with a cysteine or glutathione [28]. Production of TFF3 Cys57 and the TFF3 Ser57 analogues in bacteria confirmed that TFF3 forms homodimers via a disulphide bond involving Cys57 residues [39].

Physicochemical properties of trefoil factors

The physicochemical properties of trefoil factors have been studied mainly with the human molecules, and this section of the review will therefore focus on them. Clearly the amino acid residue substitutions between different orthologues of the individual trefoil factors will affect the chemical masses, charges and hydrodynamic properties of the molecules, but it is anticipated that the principal findings of the studies with the human trefoil factors will be applicable to trefoil factors from other species.

Hydrodynamic properties

TFF2 varies from the dimeric forms of TFF1 and TFF3 in that the two domains are not only linked by a disulphide bond, between Cys6 to Cys104, but also by a peptide chain from Leu50 to Asp56 (see fig. 1). From this extra link connecting the two domains, one might expect TFF2 to be somewhat more compact than the dimer forms of TFF1 and TFF3. Early gel-filtration studies on Sepharose G50 showed that porcine TFF2 behaves as a very compact molecule with an apparent molecular mass of only 8,600 Da [22], although the theoretical molecular mass is 37% larger (11,801 Da).

Analytical centrifugation of purified recombinant TFF1 and TFF3 monomers and dimers has been used to confirm their monomeric and dimeric nature under native conditions in solution [35,39]. Both the TFF1 monomer and TFF3 monomer behaved as ideal solutes with anhydrous molecular masses close to the predicted chemical masses. This is consistent with there being no aggregation of the molecules at any concentration. Analysis of the TFF1 and TFF3 dimers showed that they also behaved as ideal solutes with anhydrous molecular masses that were close to the predicted chemical masses. This confirmed the dimeric nature of the two molecules [35, 39].

The molecular masses of the TFF1 and TFF3 monomers differ by less than 100 Da, and those of the two homodimers by less than 200 Da. The Stokes radii of the TFF1 monomer and dimer and of the TFF3 monomer and dimer were measured by gel filtration on Superdex 75. The TFF1 monomer eluted much earlier than would be expected if it were a globular protein almost in the elution volume of a globular protein of twice the size of TFF1, whereas the TFF3 monomer eluted somewhat later than the TFF1 monomer. The Stokes radii of the TFF1 and TFF3 monomers calculated from the elution volumes were 15.5 and 13.7 Å, respectively. The frictional coefficients were calculated from the Stokes radii and the molecular masses, and were 1.25 for the TFF1 monomer and 1.12 for the TFF3 monomer. This shows that TFF1 and TFF3 are somewhat asymmetric and that TFF1 is more asymmetric than TFF3 [35, 39].

Gel filtration of the TFF1 and TFF3 dimers showed that both molecules eluted earlier than would have been expected if they were globular proteins and that this was particularly evident for the TFF1 dimer. The Stokes radii of the TFF1 dimer and the TFF3 dimer were 20.8 and 17.8 Å, respectively and the frictional coefficients were 1.34 and 1.14. The TFF1 dimer is therefore more asymmetric than the TFF3 dimer, and the difference in asymmetry between the TFF1 dimer and the TFF1 monomer is larger than for the corresponding forms of TFF3 [35, 39]. The asymmetry of the molecules means that caution must be exercised when using gel filtration columns calibrated with globular standards to determine if a trefoil factor is dimeric.

Overall charge

Comparison of the theoretical isoelectric points of human TFF1 and TFF3 monomers shows that they are both acidic, with pI values of 3.94 and 4.75, respectively. However, empirical studies suggested that the effective overall charges of TFF1 and TFF3 were more dissimilar than is apparent from their isoelectric points [39]. Construction of theoretical titration curves indicted that TFF1 might be considerably more acidic than TFF3 within a pH range of 4.0–10.0. These predicted differences in charge were confirmed experimentally by the very different interaction of TFF1 and TFF3 with cationic and anionic exchange resins during ion-exchange chromatography at different pHs [39]. Thus over a fairly wide pH range around neutral, TFF1 is a much more acidic molecule than TFF3.

With dimerisation, the predicted overall negative charges at pH values around neutral of TFF1 and TFF3 effectively double, and this is reflected in their interaction with greater avidity than the corresponding monomers with an anion exchange resin [39]. Human TFF2 has a predicted isoelectric point of 4.91, and its theoretical titration curve is more akin to that of the TFF3 dimer than the TFF1 dimer, but no detailed studies of its overall charge characteristics have been published.

Charge distribution

Both human TFF1 and TFF3 have exclusively acidic charged residues in the N- and C-termini (fig. 1), but the surface charge distribution is much more polarised in TFF1 than in TFF3 [39]. TFF1 has a greater number of acidic residues in the termini than in the trefoil domain, whereas TFF3 has the same number of acidic residues in the termini and in the trefoil domain. TFF3 contains more charged residues in the trefoil domain than TFF1, and a greater proportion of them are basic. Thus TFF1 has a greater number of acidic residues than TFF3 and the additional acidic residues are outside the trefoil domain. In contrast, TFF3 has more basic residues than TFF1, and the additional basic residues are in the trefoil domain [39]. The different orientations of the side chains of the charged amino acid residues in the molecules exacerbate the greater polarisation of TFF1 [39].

Human TFF2 has not been studied in detail, but there is at least one basic residue in each of the amino-terminus, the link region and the carboxy-terminus, which suggests that the uneven distribution of charged residues characteristic of human TFF1 and to a lesser extent human TFF3 is not an important feature of human TFF2 function.

Susceptibility to proteolytic degradation

The protease stability of porcine TFF2 has been investigated by incubation with trypsin, which cleaves after basic residues (Arg and Lys residues), and chymotrypsin, which has a broader specificity (cleavage after Tyr, Phe, Trp, Met, Leu, Gln and Asn) [22]. Using 2% (w/w) trypsin or 0.5% (w/w) chymotrypsin and incubation at neutral pH for up to 1 h, no degradation was observed, as judged by isoelectric focusing, immunoreactivity and N-terminal sequence analysis [22]. This is despite the presence of 10 basic residues in porcine TFF2. These results indicate that the seven disulphide bonds in porcine TFF2 hold the molecule in the very compact configuration and thus protect it from digestion at least from pancreatic enzymes. The two trefoil domains of TFF2 are connected by a peptide chain as well as a disulphide bond, and this may make it compact and therefore more resistant to gastrointestinal proteases. When porcine TFF2 was reduced and S-carboxymethylated before incubation, both trypsin and chymotrypsin were able to digest the molecule into several small fragments [30]. TFF1 was also only digestible by typsin after it had been reduced [25].

Similar results have been obtained for rat and human TFF3 incubated with carboxypeptidase, trypsin, chymotrypsin and pepsin [H. Kindon, personal communication]. Interestingly, rat but not human TFF3 seems to be degraded by a cocktail of mixed bacterial proteases [H. Kindon, personal communication]. Poulsen et al. [40] demonstrated that porcine TFF2 administered orally to rats is stable in the stomach and small intestine but is degraded when it enters the caecum. Taken together, these results indicate that although trefoil factors are generally considered to be protease resistant, this is true for gastric (pepsin) and pancreatic/duodenal proteases (trypsin, chymotrypsin and carboxypetptidases) but not for the bacterial proteases present in the caecum and colon.

Secondary structure

The secondary and tertiary structures of mammalian trefoil factors have been studied by the use of crystallisation followed by X-ray analysis and by NMR-based studies of the molecules in solution. Porcine TFF2 has been studied by both X-ray crystallography of orthorhombic and trigonal crystal forms, and NMR analysis, and human TFF1 and TFF3 by NMR analysis.

Common secondary structure within the trefoil domain

The identities of the three disulphide bonds within the trefoil domain have been determined unequivocally for all four mammalian trefoil domains, TFF1, TFF2 domain 1, TFF2 domain 2 and TFF3, from structural studies. The cysteine residues are always paired 1–5, 2–4 and 3–6 (fig. 2A) [41–43].

A short α-helix that includes the third cysteine residues is present in all four trefoil domains (figs. 1D, 2B, 2D) [13, 42, 62]. The C-terminal amino acid of the α-helix is always the residue before the second conserved glycine residues in TFF1, TFF2 D1 and TFF3 and the equivalent residue in TFF2 D2. The α-helix identified in the NMR-based studies comprises six residues, but it contains eight residues in the crystal structures [12, 13, 42, 62].

There is a short irregular four-to-five-amino-acid-long two-stranded antiparallel β-sheet, the first strand of which is separated from the α-helix by one residue (figs. 1D, 2B, 2D). The first strand comprises the completely conserved CCF motif and the second strand the conserved WCF-P motif [13, 41–43].

Other secondary structure within the trefoil domain

In the TFF1 trefoil domain and both trefoil domains of TFF2 there is a five-amino acid residue 3_{10}-helix towards the amino-terminus of the domain [43, 44]. The C-terminal residue of this helix is the completely conserved arginine residue (fig. 1A, D). In TFF1, these residues Ala10

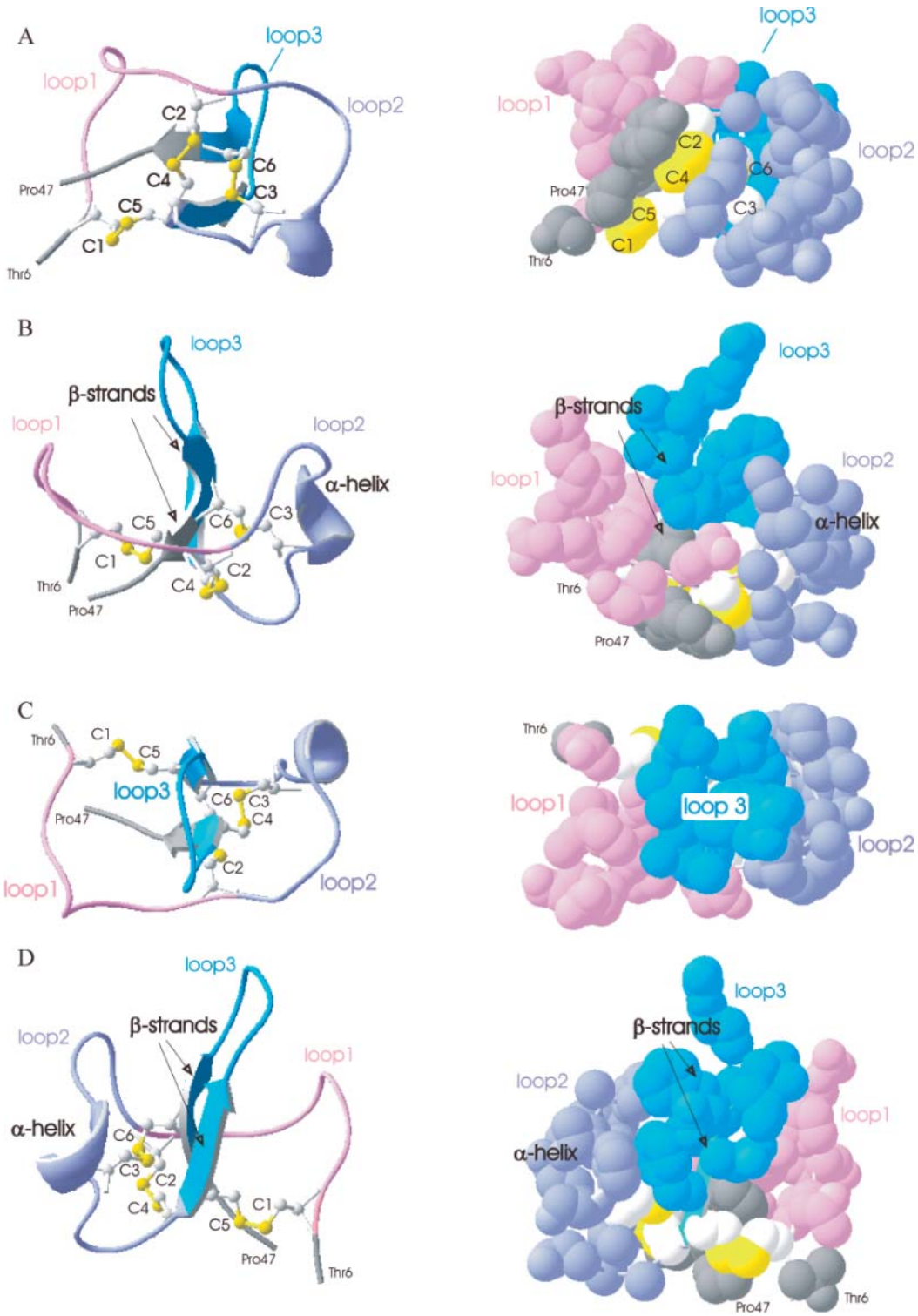

Figure 2. Elements of secondary and tertiary structure shared between the four mammalian trefoil domains. The trefoil domain of human TFF1 [43], residues Thr6–Pro47, is shown as a ribbon (LHS) and with the side chains shown in a space filling view (RHS). The cysteine residues are shown in ball-and-stick view (LHS) or space filling view (RHS) and with the sulphur atoms coloured yellow. The other residues in loop 1, loop 2 and loop 3 are coloured pink, purple and blue, respectively. The remaining residues in the trefoil domain are coloured grey. (*A*) The trefoil domain is viewed from below to show the disulphide bond formation, 1–5, 2–4, 3–6. (*B–D*): The trefoil domain is viewed from the side (*B*) from the top (*C*) and from the opposite side (*D*) to show the conserved α-helix and β-strands (*C* and *D*) and the positioning of the third loop between the first and second loops (*B–D*).

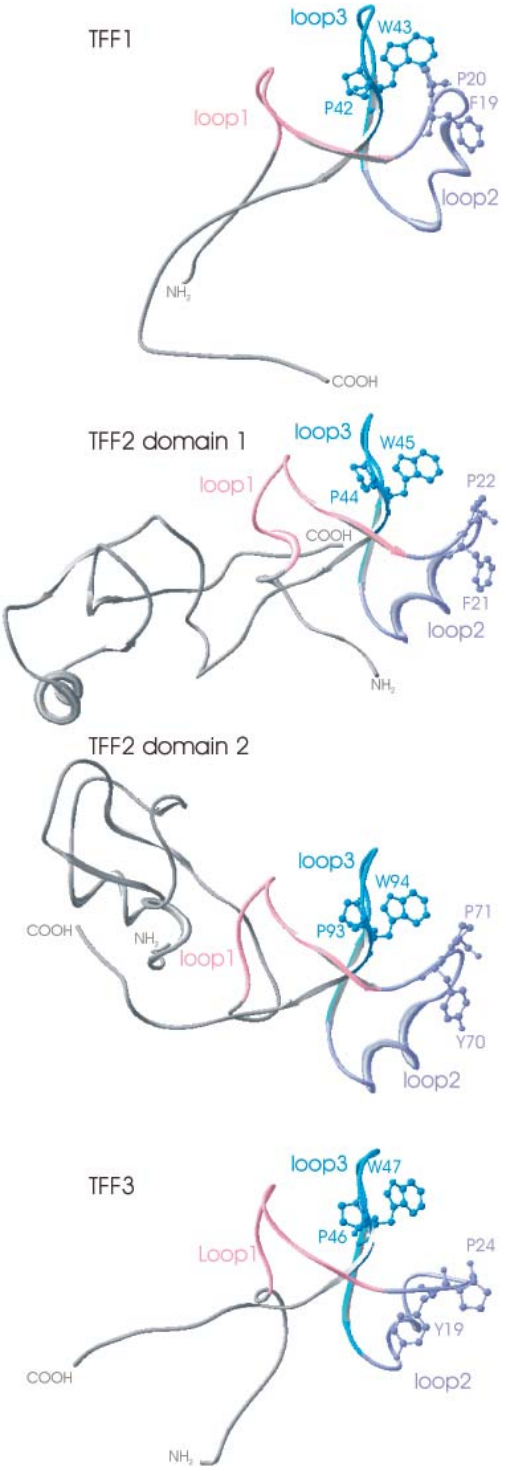

Figure 3. Comparison of the backbone conformations of, and the orientations of the side chains of the four conserved and semi-conserved solvent-accessible hydrophobic amino acid residues in the four mammalian trefoil domains in TFF1 [43, 62], TFF2 [44] and TFF3 [42]. The molecules are viewed from the side of each trefoil domain as in figure 2B. All the amino acid residues in loop 1, loop 2 and loop 3 are coloured pink, purple and blue, respectively, and remaining residues in the molecules are coloured grey. The side chains of the four hydrophobic amino acid residues are shown in ball-and-stick view.

– Arg14 are found in a composite turn or a 3_{10}-helix in roughly equal proportions.

There is a type II β-turn in TFF1 and both domains of TFF2 immediately N-terminal to the α-helix [43, 45]. The turn is just after the second cysteine residue and absolutely conserved glycine and encompasses the absolutely conserved proline residue that follows.

Secondary structure outside the trefoil domain

An additional short 3_{10}-helix of three amino acid residues is located at the extreme amino-terminus of the TFF3 trefoil domain around the first cysteine residue [42]. The same region of the second domain of TFF2 is also in a 3_{10}-helix, but it extends further amino-terminal to include some of the link region between the two domains [13]. There is an α-helix of seven amino acid residues in the same region of the first domain of TFF2. The extra residue in the first domain of TFF2 is located at the C-terminal end of the helix and allows the formation of an α-helix in lieu of the 3_{10}-helix in domain 2.

There is an additional 3_{10}-helix from residues 100–103 in the carboxy-terminal region of TFF2 [44].

In TFF1, an additional antiparallel β-sheet is formed by part of the N- and C-termini, residues 3–7 and 47–51 [43, 46]. This encompasses the region that forms the N-terminal 3_{10}-helix of TFF2 and TFF3.

Tertiary structure

Common tertiary structure

In all trefoil domains studied, the three disulphide bonds 'lock' the secondary structural elements to form a stable core from which three striking loops protrude (fig. 2). The central core of the loop-like regions consists of the short, irregular two-stranded β-sheet, which forms the lower region of the third loop, the upper region of which is an irregular peptide chain (fig. 2B, D) [13]. The short α-helix of the second loop packs against the N-terminal strand of the β-sheet of the third loop (fig. 2D). The three loops can be considered as running from the first cysteine to the second cysteine, from the conserved glycine to the fourth cysteine and from the fifth cysteine to the sixth cysteine. Thus the second loop is slightly longer than the first and third. The three loops stack together with their faces parallel to each other and with the third loop sandwiched between the first and second loops (fig. 2) [43]. The three closely packed loops with only a small proportion of short elements of secondary structure give the trefoil domain a compact structure, which may protect it from digestion by proteases [13].

Comparison of the main chain positions in the trefoil domains

The high degree of sequence conservation between different trefoil domains is reflected in the very similar folds of

the four domains and is confirmed by the pairwise rmsd values for the backbone atoms. This is illustrated in figure 3, which shows a visual comparison of the backbone conformations of the four trefoil domains, viewed as in figure 2B, in TFF1, TFF2 and TFF3. After the structure of TFF1 was solved, the backbone for residues 9–47 was superimposed on domains 1 and 2 of TFF2, and the rms differences were calculated to be 1.5 and 0.9 Å, respectively [43]. This showed that the overall conformation of the TFF1 trefoil domain is more similar to that of domain 2 of TFF2 (fig. 3) [43, 45]. The principal reason for this difference is that the α-helix and its preceding turn are shifted by approximately 4 Å in domain 1 of TFF2 with respect to the position of the corresponding residues of TFF1. The separation of the second and third loops is consequently greater in domain 1 of TFF2 than in TFF1, producing a more open structure (fig. 3).

Comparison of the backbone conformations of TFF1 for residues 7–47 with TFF3 gave an rmsd value of 1.98 Å and between TFF3 and TFF2 D2 the value was 1.44 Å [42]. The major differences are in the conformation of the second loop. In TFF3 the backbone of residues Tyr23–Pro28 forms an essentially flat, extended loop, whereas the equivalent regions of TFF1 and both domains of TFF2 form an uplifted β-turn, which gives the appearance of a pronounced cleft between the second and third loops (fig. 3). The substitution of a histidine at position 25 in human TFF3 compared with a glycine in the trefoil domains of TFF1 and TFF2 is clearly not conservative and is the only significant sequence difference in this region (fig. 1), which suggests strongly that this substitution accounts for the unique conformation of TFF3 in this region.

Conservation of structurally important residues

There are striking similarities between the positions and interactions made by the side chains of many highly conserved residues, such as those corresponding to Cys7, Arg14, Cys17, Cys27, Cys32, Cys33, Phe34, Asp35, Pro42, Trp43, Cys44 and Phe45 in TFF1. A number of long-range side chain-to-side chain interactions are found in all four trefoil domains. These include the disulphide bonds and the salt bridge between the conserved residues corresponding to Arg14 and Asp35 in TFF1 [41–43].

Several of the conserved and semi-conserved residues are buried in the interfaces of the loops and are responsible for holding them in the observed compact three-dimensional arrangement. The three residues equivalent to Arg14, Val22 and Phe45 in TFF1 are buried in all of the known trefoil domain structures. The low degree of solvent accessibility and high degree of conservation indicates that they are particularly important to the integrity of the trefoil fold [43].

The conserved arginine is positioned centrally and interacts with the equivalent of Asp35 in TFF1 [43]. The hydrogen bonding interactions between Arg14 and its neighbours appear to contribute to create a very rigid pocket for the binding of the aromatic ring of Phe45, which prevents the transient fluctuations in the protein structure required for aromatic ring flipping. The aliphatic part of the side chain of Arg14 is packed against the aromatic ring of Phe45. The importance structurally of the conserved arginine was also noted for TFF2. It makes hydrogen bonds with Pro44 and Asp37 and seems to be important for the folding of the amino-terminus of the domain. The stacking of Arg16 with Phe47 was also noted in TFF2. Thus the buried Phe45 residue of TFF1 with its aromatic ring in a well-packed environment together with the interacting Arg14 and Asp35 residues is an important contributor to the stabilisation of the common core structures found for the trefoil domains between the first and third loops. There are also many conserved residues in the interface between the second and third loops; however, only the equivalent of Val22 in TFF1 is buried in this region and can be regarded with confidence as having a structural role. The conserved phenylalanine at position 34 in TFF1 is in the interface between the second and third loops and is probably important structurally, as it is the only other residue to have a solvent accessibility less than 25% [43].

Potential binding motifs identified from structural work.

Inspection of the structures of the trefoil domains has identified clusters of conserved residues on the surfaces of loops 2 and 3 that are solvent accessible and have therefore been proposed as potential binding sites with so far unknown ligands or receptors [13, 41, 45].

In particular, a pronounced groove between the ends of the second and third loops that encompasses a patch of solvent accessible conserved hydrophobic residues has drawn attention as a potential binding site capable of accommodating either an oligosaccharide or an aromatic amino acid side chain (fig. 3) [13, 43, 44]. Conserved residues are either placed around this groove or form the core of the domain. The groove is formed primarily from peptide backbone atoms. The only side chains that point into the groove are of conserved residues. In addition to the fully conserved hydrophobic residues equivalent to Pro20, Pro42 and Trp43 in TFF1, Phe19, which is replaced conservatively with a tyrosine residue in some domains, forms a contiguous hydrophobic patch on the surface of the trefoil domain in this location (fig. 3). Several of the surface positions surrounding this patch have features that are fully conserved: hydrophobic residues always replace Pro11, Val38, and Val41; a hydrogen bond donor, Arg12; and hydrogen bond acceptors, Asn16, Thr23, Gln26 and Asp36 [43].

The high degree of conservation of the physical properties of this region suggested strongly that the mammalian

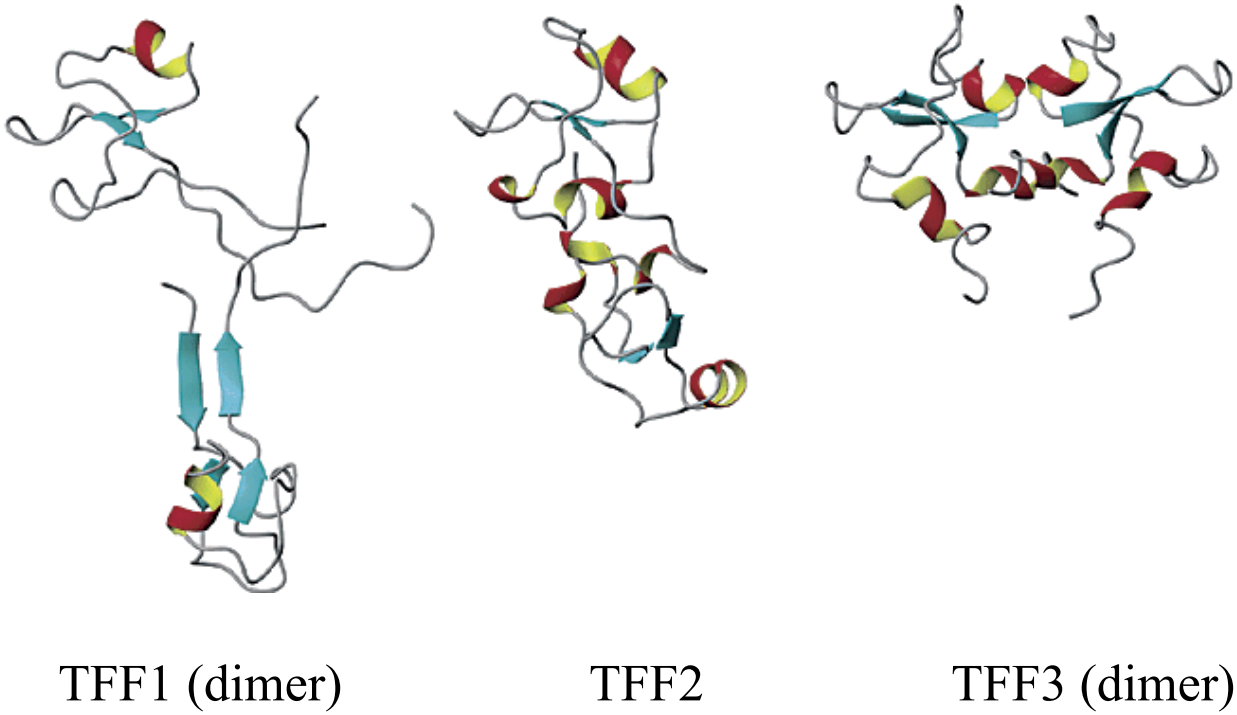

Figure 4. Ribbon representation of TFF1 dimer, TFF2 and TFF3 dimer. The figure is adapted from Muskett et al. [14].

trefoil domains bind similar molecules in this region. However, the widths of the clefts in the four mammalian trefoil domains as measured between the Cαs of the first conserved proline and the tryptophan residues in the trefoil domains [13] are very different: in TFF1 the cleft is narrow, 6 Å; whereas in TFF2 the clefts are somewhat wider, 10 Å for D1 and 8 Å for D2; and in TFF3 the cleft is even more open and has a shelf like structure that measures 11–12 Å accross (fig. 3) [13, 14, 42, 43].

Tertiary structure outside the trefoil domain

Outside the trefoil domains, there is little evidence for commonality of tertiary structure between different molecules (fig. 3). In TFF1, the N- and C-terminal strands are closely associated and form an extended structure that has some β-sheet character for part of its length and which becomes more disordered towards the termini [43].

In contrast, the structural positions of the extra-trefoil domain regions of TFF2 are constrained. The formation of the extra-trefoil domain disulphide bond has been noted already, and both the C-terminal region and the linking region lie alongside the two helices, residues 4–10 and 55–59, at the interface between the two TFF2 trefoil domains [13].

The N- and C-termini of TFF3 appear to be relatively mobile, as was found for TFF1, but there is no evidence for the formation of a β-sheet between the termini (fig. 3) [42].

Quaternary structure

TFF1 homo-dimer

The two trefoil domains in the TFF1 dimer are linked covalently by a disulphide bond between the Cys58 residues [35]. Analysis of the structure of the TFF1 dimer using NMR techniques indicates that the trefoil domains have very similar structures in the monomer and the dimer with no evidence for any contact between the two domains in the dimer (fig. 4) [46]. The two trefoil domains are well separated and are held apart on opposite ends of a flexible linker. This implies that the overall structure is constrained by the steric/packing properties of the peptide sequence near the disulphide junction. The adjacent proline residues, Pro53 and Pro54, kink the chain and are relatively conformationally restricted (fig. 3) [43], and the disulphide bond itself forces the C-termini to cross each other at nearly 90° at Cys58. The two monomer units in the human TFF1 dimer have no fixed orientation with respect to each other and are present in a mixture of many conformations [46].

The location of four glutamic acid residues within the last six amino acid residues of TFF1 (fig. 1A) means that in the dimeric form of TFF1, 10 negatively charged residues, 8 Glu and 2 C-terminal Phe, are located within a rather short distance, giving this molecule a very acid region at the centre of the flexible peptide linker. It may be that this region of localised charge contributes to the constraints on the steric properties of the peptide sequence near the disulphide junction.

TFF2, a natural dimer

The dimeric structure of porcine TFF2 is of necessity an asymmetric structure formed from a continuous stretch of peptide and containing seven disulphide bonds. The N- and C-termini of TFF2 are linked by the seventh disulphide bond, which helps to fix the relative orientation of the two trefoil domains. The structures of both domains are homologous (1.2 rms deviation on main-chain atoms) as expected from their sequence homology and are related to each other by a rotation of 135° [45]. TFF2 is a relatively compact, elongated molecule consisting of the two compact globular trefoil domains joined via a short interface (fig. 4). The interface between the two domains is not extensive (surface area ~250 Å) and consists of the N-terminal α-helix from domain 1 packing loosely against the corresponding helical region in domain 2. These two helices are bounded on each side by the C-terminal extended chain, residues 99–106, and the linking region, residues 49–54, which are perpendicular to them and form a number of hydrogen bonds [45].

Although porcine TFF2 contains 10 basic residues, it is very resistant to digestion with trypsin [22]. None of the lysine and arginine residues, except perhaps Lys14, are located on the surface of the molecule, and they are thereby inaccessible to trypsin digestion. More surprisingly, porcine TFF2 is also resistant to digestion with chymotrypsin [22], which cleaves after Tyr, Phe, Trp, Met, Leu, Gln and Asn residues, because several of these residues are located at the surface of the molecule. Again, the compact structure of TFF2 seems to protect it from digestion.

TFF3 homodimer

The TFF3 dimer is like the TFF1 dimer formed via a disulphide bind involving the cysteine residue located near the C-terminus [39], but the two symmetrical dimers have very different structures [14, 46]. The TFF3 dimer has a well-defined dimer structure with a twofold axis of rotational symmetry (fig. 4). There is an additional 3_{10}-helix involving residues 53–55 in the TFF3 dimer that was not found in the monomer. The two additional 3_{10}-helices lie antiparallel to each other at the interface between the monomer units. The orientation of the N-terminal 3_{10}-helix is changed in the dimer compared with the monomer so that the poorly defined N-terminal region is positioned further underneath the trefoil domain to accommodate the interaction of the two domains. The C-terminal region is also different in the dimer where it loops back on itself to position the additional 3_{10}-helix beneath the α-helix and at the tip of the β-sheet (fig. 4). The interface between the TFF3 monomer units in the TFF3 dimer involves residues in the additional C-terminal 3_{10}-helices and nearby residues. The C-terminal 3_{10}-helix is an integral part of the interface between the two monomer units, and this probably contributes to its stabilisation and may explain its presence only in the dimeric form of TFF3 [14].

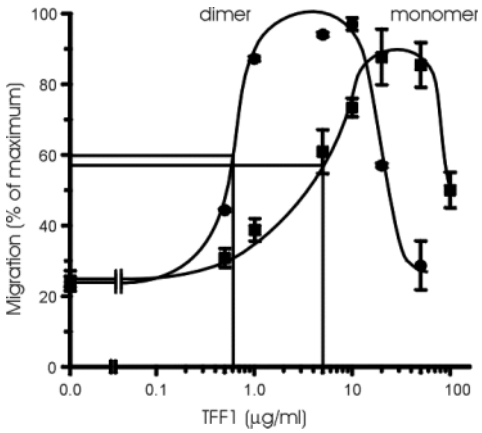

Figure 5. Stimulation of the movement of breast cancer cell migration. MCF-7 cell migration was stimulated with varying concentrations of TFF1 monomer (filled square) or TFF1 dimer (filled circle). The figure is adapted from Prest et al. [37].

The high conservation in the C-terminal but not the N-terminal regions of TFF3 orthologues suggests that there are constraints on the sequence in this region that may be important for biological function of TFF3 (fig. 1C). The formation of a 3_{10}-helix in this region upon dimerisation and the contribution of this region of TFF3 to the interface between the two monomer units in the TFF3 homodimer support this supposition and indicate that the conservation is for structural reasons (fig. 4) [14]. This underpins the importance of the induction of this structural element in the function of the TFF3 dimer and thereby probably of TFF3.

Separation and orientation of the monomer units in trefoil factor dimers

The structural studies show that the TFF3 dimer is much more compact than the TFF1 dimer and somewhat less mobile [14, 46]. The juxtaposition of the two 3_{10}-helices induced upon dimer formation, together with the intramolecular interactions around the α-helix at the base of loop 2, give the TFF3 dimer a compact structure that is almost as constrained as that of TFF2. TFF3 has a globular discoid structure (26 x 19 x 10 $Å^3$; [14]) whereas TFF2 is more elongated (29 x 12 x 11 $Å^3$; [F. W. Muskett, personal communication]).

The mobilities in the core regions of the TFF1 and TFF3 dimers are close to the values expected for the rigid part of peptides/proteins of this size, but even in these core regions the mobility is greater in the TFF1 dimer than in the TFF3 dimer [14, 46]. The mobility in the C-terminal linker regions, residues 48–60 and 51–59 for TFF1 and TFF3, respectively, appears to be restricted rather than free. The residues close to the intra-molecular disulphide bonds are more mobile in the TFF1 dimer than in the TFF3 dimer. This is fully

compatible with the increased compactness of TFF3 compared with TFF1 found in the structural studies. The ultracentrifugation and gel filtration experiments had also indicated that the TFF3 dimer is more compact than the TFF1 dimer [59].

Structural constraints and dimensions of the trefoil factor dimers

The distances and relative angles between corresponding structural elements in each trefoil factor dimer are very different [14]. This results in the positioning and orientations of the two patches of conserved hydrophobic residues being very different in the three dimers (fig. 4). For the TFF1 dimer [46], where the linkage between the two domains is very flexible, the distances between the hydrophobic clusters and their relative orientations are fairly random. The distances between the α-carbons of the Pro20 residues are in the range ~36–73 Å with the separation in the majority of the structures being greater than 50 Å. In the case of the rigid TFF2 structure, the hydrophobic clusters are positioned at the opposite ends of the elongated molecule resulting in a 44 Å separation between the α-carbons of the Pro22 and Pro71. TFF3 is again very different, with the conserved residues on the same face of the dimer and the distances between the α-carbons of the Pro24 residues being only 13–19 Å [14].

Functional insights

Functional consequences of different molecular forms and posttranslational modifications

This review will not cover all studies published on the biological activities of trefoil factors but will rather focus on those studies that address the functional consequences of structural differences using mature correctly-folded trefoil factors.

Activities in vitro

Three studies have compared the biological potencies of the TFF1 monomer and dimer. In a wounding model of restitution with the human colorectal cell line HT29, the extent of migration was increased in the presence of TFF1 dimer but not monomer [47]. Both forms of TFF1 inhibited the growth of the human gastric cell line AGS, but the dimer was more potent and effected a greater inhibition in cell growth [48]. Prest et al. compared the relative activities of TFF1 monomer and dimer as motogens in breast cancer cells [37], utilising the wounding assay and a modified Boyden assay. Both molecules stimulated the migration of MCF-7 and MDA-MB-231 breast cancer cells. The dimer was consistently more potent than the monomer and was, for instance, eightfold more potent in stimulating migration of MCF-7 cells (fig. 5).

Protection against damage

The activities of the TFF1 monomer and dimer were compared in an animal model of gastric damage. Dimeric TFF1 reduced gastric damage by 70%, whereas monomeric TFF1 was significantly less effective. Similar results were obtained in an animal model of colitis in which TFF1 dimer protected against the induction of colitis but TFF1 monomer was ineffective [47].

The relative activities of TFF3 monomer and dimer have been compared in a detailed study in a rodent model of ischaemic necrotising enterocolitis [49]. The TFF3 dimer was sometimes more active than the TFF3 monomer both when the molecules were administered prior to the induction of ischaemia and when the healing effects were compared. However, these differences were not observed consistently. It is not known if this is because the monomer and dimer have similar potencies or whether they have different potencies that were not apparent at the concentrations tested.

There is one unconfirmed report that glycosylated human TFF2 is more effective than the non-glycosylated TFF2 in protection against gastric damage in an animal model [50]. The glycosylation of recombinant TFF2 used in this study will be high mannose because it was expressed and secreted by *Saccharomyces cerevisiae*, whereas in vivo, TFF2 is glycosylated with a complex glycan [unpublished data]. It will be interesting to compare the activities of non-glycosylated TFF2 with appropriately glycosylated human TFF2.

Interactions with mucins

Mucins are large, heavily glycosylated proteins expressed in a tissue specific manner in several organs of the body including the gastrointestinal tract. Mucins are normally classified into two categories: secreted and membrane-associated mucins. Seventeen human epithelial mucin genes have been cloned, of which five have been classified as secretory: MUC2, MUC5AC, MUC5B, MUC6 and MUC7 [51]. Four of these, MUC2, MUC5AC, MUC5B, and MUC6, are large gel-forming gastrointestinal mucins responsible for the rheological properties of the mucus layer [51].

Coordinated localisation of mucins and trefoil factors in the normal gastrointestinal tract has been reported: TFF1 is co-localised with MUC5AC, TFF2 is co-localised with MUC6 and TFF3 is co-localised with MUC2 [52]. These results seem to indicate that specific trefoil factors have specific mucin partners. Previous results utilising an in vitro model of epithelial barrier function have suggested that mucins and trefoil factors act in a cooperative manner to protect the mucosa [16]. In this study, rat TFF3 in combination with rat colonic mucin, which should consist mainly of MUC2, was more effective than rat TFF3 in combination with rat gastric mucin, which should consist mainly of MUC5AC and MUC6. This is consistent

with individual trefoil factors being more effective with specific mucins. However, TFF2 and TFF3 were equally effective in combination with human colonic mucins. In another study an increase in viscosity was demonstrated when trefoil factors were mixed with mucin in vitro [53]. The combination of TFF2 with porcine gastric mucin resulted in a larger increase than the combination of TFF3 with gastric mucin.

Taken together these results suggest that optimal functional effectiveness may depend on structural interactions between trefoil factors and mucins that are specific to individual molecules. The colocalisation of different trefoil factors with different mucins probably reflects this optimal functional effectiveness but different trefoil factors may to a certain degree be able to act as a substitute for each other. Co-ordinated trefoil factor expression with mucins may to some extent be altered in diseased conditions such as Barrett's oesophagus [54] and inflammatory bowel disease [55].

Tomassetto and co-workers used yeast two-hybrid analysis to demonstrate a direct interaction between murine TFF1 and both murine MUC5AC and MUC2 [56]. Inside the yeast cells, proteins are in a reducing environment and will not be glycosylated, which means that the interaction did not occur between TFF1 and the oligosaccharide side chains of the mucins but must have been a direct protein-protein interaction between TFF1 and the two mucins. The interaction was with the vWFC1 and vWFC2 (von Willebrand factor cysteine-rich domain 1 and 2) domains that are located towards the C-terminus of MUC5AC and MUC2 and are generally believed to be responsible for joining mucin monomers into large multimers with high intrinsic viscosity [57].

There is some evidence that human TFF1 interacts with gastric mucins and in particular MUC5AC in vivo [38, 58]. A direct interaction between TFF1 and MUC5AC has been demonstrated by gel filtration of human gastric mucosal proteins followed by Western transfer analysis and by co-immunoprecipitation using specific antibodies towards the two proteins [58]. This complements the in vitro work because it demonstrates that an interaction between the secreted correctly folded posttranslationally modified proteins occurs in vivo. The interaction between TFF1 and MUC5AC can be abolished by high salt concentrations or the presence of chelating agents such as EDTA, indicating that the two proteins are not linked covalently (e.g. by disulphide bond) but may be stabilised by the presence of divalent metal ions such as Ca^{++} [58].

Of particular relevance to the present review are the observations that different forms of TFF1 have different avidities for gastric mucins. While all three forms of TFF1 identified are present in the adherent mucus gel layer, the TFF1 dimer and to a lesser extent the TFF1 monomer remain associated with gastric mucins after CsCl gradient centrifugation, whereas the TFF1 heterodimer does not [59]. Likewise, a greater proportion of the TFF1 dimer than the more prominent TFF1 heterodimer is associated with MUC5AC after gel filtration under physiological conditions, and the TFF1 dimer is preferentially co-immunoprecipitated by a MUC5AC specific antibody [58]. This may imply that the disulphide bond that connects the two TFF1 monomer units (Cys58–Cys58) is important and that if this bond is broken the entire TFF1-MUC5AC complex may dissociate.

Binding to H. pylori

A recent study has raised the possibility that colonisation of the human stomach with *Helicobacter pylori* is mediated by direct interaction between the bacteria and TFF1 [60]. An interaction was demonstrated by flow cytometry between *H. pylori* and TFF1-coated latex beads and by plasmon resonance between *H. pylori* and TFF1-coated dextran chips. *H. pylori* bound to the TFF1 dimer but not the TFF1 monomer. The binding was rapid and seems to be specific and of high affinity. This suggests that the surfaces of *H. pylori* present a specific protein, which may be a receptor-like protein, a binding protein or a MUC5AC-like protein, that interacts directly with the dimer form of TFF1. *H. pylori* was able to bind gastric mucin only if it had been preincubated with TFF1, which indicates that, when it is bound to *H. pylori*, TFF1 retains its ability to interact with mucins and suggests that TFF1 binds to *H. pylori* with a different binding surface than the one that binds to MUC5AC [60].

Theoretical effects of mutations and analogues

Another type of functional insight comes from comparison of substitutions between different mammalian trefoil domains or between trefoil factor orthologues and from the study of somatic mutations of the *TFF1* gene in gastric cancer.

Non-conservation between the four mammalian trefoil domains

The presence of an additional residue in TFF2 D1 compared with the other mammalian trefoil domains results in the formation of an α-helix in place of a 3_{10}-helix at the beginning of loop 1 [13].

Polshakov et al. [43] remarked that the probable cause of the less 'open' structure of TFF1 compared with domain 1 of TFF2 (fig. 3) is the substitution of a phenylalanine residue in the position equivalent to Ala28 in TFF1 (fig. 1). The more open conformation allows the aromatic ring of this phenylalanine ring to interact with residues of the N- and C-termini to bury a substantial amount of hydrophobic surface. Amongst the TFF1 orthologues, this residue can be a Lys or a Thr, and the presence of the

more bulky side chains on these residues may affect the relative separations of loop 2 and loop 3.

TFF1 contains four glycine residues within the trefoil domain. The first, Gly8, is conserved in all mammalian trefoil factors sequenced thus far. The second, Gly21, is conserved in all orthologues of TFF1 and both domains of TFF2 but is replaced by a His, Gln, Glu, Thr or Ser in different orthologues of TFF3. The substitution of the glycine residue after the second cysteine probably prevents the formation of the type II β-turn in the second loop of TFF3. It was suggested that the positioning of this conserved glycine residue on the edge of the conserved hydrophobic surface patch might indicate that it was necessary to allow easy access of a ligand to the hydrophobic cleft. The substitution of a histidine at this position in human TFF3 may be responsible for the distinctive conformation of loop 2 in human TFF3 (fig. 3) [42]. In human TFF3 the backbone of the residues in loop 2 forms an essentially flat, extended loop, whereas the equivalent region in TFF1 and TFF2 D1 and TFF2 D2 adopts an uplifted turnlike conformation. This gives the appearance of the aforementioned cleft, while in human TFF3 the same region of the molecule is described as having a shelflike appearance [42]. The substitutions at this position in different TFF3 orthologues suggest that the requirement is for a charged or polar side chain rather than for a histidine per se. The third glycine is conserved in TFF1, TFF2 D1 and TFF3 but is replaced by a lysine, asparagine or histidine in D2 of TFF2. This residue lies between the α-helix and the first strand of the β-sheet at the junction between the second and third loops and the bulkier, polar or charged side chains will reduce the sharpness of the bend between these two elements of secondary structure. The fourth glycine, equivalent to Gly 40 in TFF1 is almost always a glycine in TFF2 domain 1 and TFF3, and is a glutamic acid in domain 2 of TFF2 but can be replaced by an asparagine. The replacement of the glycine means that the hairpin at the tip of loop 3 has a more rounded peptide backbone and that there is an additional charged or polar residue on the edge of the hydrophobic patch.

Non-conservation between orthologous trefoil factors

There are some notable amino acid sequence differences between individual trefoil factors from different species that hint to the existence of significant species-specific characteristics of trefoil factors.

Most mammalian trefoil factors have either a Glu or a Gln at position 1 which is frequently cyclicised to a pyrGlu. In TFF3, the equivalent residue can also be a simple hydrophobic residue: Gly, Val or Ala. The substitutions will reduce the polarisation of the TFF3 molecules compared with human TFF3 [39] and suggest that the pyrGlu formation is not critical to the function of at least TFF3.

The Asn15 that is glycosylated in human TFF2 [32] is on the surface of loop 1 of domain 1 but is not in a consensus glycosylation site in any other known TFF2 orthologues other than chimpanzee TFF2. Human TFF2 is predominantly glycosylated, and it seems probable that this confers some functional advantage to human TFF2 that is not shared by TFF2 from other species.

The double proline at positions 53 and 54 in human TFF1 has been commented on previously and accounts for the pronounced kink in the peptide chain of the otherwise extended C-terminus of the TFF1 monomer (fig. 3) and probably contributes to the lack of contact between the monomer units in the TFF1 dimer (fig. 4) [43, 46]. The absence of both these proline residues in rodent TFF1s suggests that the structure of rodent TFF1 monomers and dimers will not be similarly constrained (fig. 1A).

The location of an arginine residue at position 58 in canine TFF3 is unprecedented, as the carboxy-termini of TFF1 and TFF3 are characterised by a high concentration of charged but exclusively acidic residues. Caution should be exercised, though, in overinterpreting this substitution until it has been confirmed independently.

In human and chimpanzee TFF3, there is an alanine residue at position 55 that is a threonine in all other TFF3 sequences (fig. 1C). This residue forms part of the additional 3_{10}-helix induced upon formation of the TFF3 dimer (fig. 4). Recombinant rat TFF3 produced in yeast always forms homodimers, whereas human TFF3 is produced as a monomer and a dimer. It may be that the substitution of a threonine residue at the interface between the two monomer units facilitates dimer formation.

Mutations in gastric cancer

In 61 human gastric tumours studied, eight mutations were detected in the *TFF1* gene [61]. These mutations occurred in five specific positions (Thr8, Ala10, Glu13, Val22 and Gly31) of the TFF1 sequence (see fig. 1A). One substitution was detected in DNA from 1 of 18 adenomas studied. The replacement of Val22, which is located in the β-turn at the end of loop 1, with an isoleucine is likely to be conservative. Indeed, an isoleucine is present at this position in canine TFF1 and TFF3, and in all molecules of TFF2 in both domains 1 and 2. DNA from 43 carcinomas was sequenced, and a Thr8Ile substitution was found in two cases. This substitution occurs in murine TFF1 and therefore may not be very deleterious to TFF1 function. The same residue which lies in the core of the trefoil domain at the base of loop 1 was replaced by a lysine in another case. Canine TFF3 is also reported to have a charged amino acid residue in this position, but it seems probable that this substitution will affect the folding of TFF1. The substitution Glu13Lys was identified in two cases. This residue is immediately N-terminal to the absolutely conserved structurally essential arginine. While nearly always a glutamic acid in TFF1, apart from cannine TFF1 in which it is a histidine, in other mammalian trefoil domains it can be substituted by residues with

charged, polar or hydrophobic side chains. Finally, the conserved Gly31 was replaced by a valine in one case. As noted previously, substitutions in this position such as occur in domain 2 of TFF2 will affect the junction at the base of loops 2 and 3.

Until more is known about the interactions that TFF1 makes with other molecules, it is very difficult to assess the import of the mutations identified in gastric tumour DNA. These point mutations are all located in the trefoil domain of TFF1, and one could speculate that by altering the structure of the trefoil domain they result in a TFF1 molecule that is no longer able to form complexes with MUC5AC and that this leaves the gastric tissue less protected and increases the probability of development of gastric carcinoma.

In conclusion

How much redundancy is there between the three mammalian trefoil factors?

The three mammalian trefoil factors each have clearly defined, specific sites of expression both within and outside the gastrointestinal tract. Our knowledge about the biological activities of trefoil factors has increased. However, so far there is no compelling evidence from studies using assays of biological activity that there is any difference in activity of the different trefoil factors. It is generally accepted that all three increase motility of cells in culture and protect against mucosal damage in vivo. There is, however, compelling evidence from the TFF1-, TFF2- and TFF3-null mice studies for differences in biological function between individual trefoil factors [17–19]. This is unlikely to be due solely to differential sites of expression, as the observed phenotypes were so different. For instance while TFF1-null mice manifest a complete absence of antral mucus [18], no loss of intestinal mucus formation or function was observed in TFF3-null mice [19]. It therefore seems probable that different trefoil factors are in different locations because they do different things rather than because they do the same things in those different places.

Do structural studies support different functions for individual trefoil factors?

Clearly there is significant sequence homology between the four mammalian trefoil domains, which gave rise to the initial concept of a trefoil factor family and which accounts for the conserved trefoil motif structure. Clearly there are also significant sequence differences between the domains. For instance, the human TFF1 trefoil domain shares 60% sequence identity with domain 1 of TFF2, but only 45% identity with domain 2 and 55% identity with the trefoil domain of TFF3 (fig. 1D). These sequence differences produce notable differences in the structure of the different trefoil domains. The region on which most comment has been passed, ironically drew attention originally partly because of sequence conservation of surface residues. That is the cleft or patch of hydrophobic residues with solvent-exposed side chains that is found at the interface between loops 2 and 3 (fig. 3). The width and shape of, and nature of the side chains of residues that surround, this proposed binding site differ between the different mammalian trefoil factors. Given these significant structural similarities, it seems improbable that, if this is a binding surface, all four trefoil domains have evolved to interact with the same ligand.

Does the existence of different forms of trefoil factors contribute to a diversity of function?

This review has covered the differences in physicochemial properties and structures of the three mammalian trefoil factor dimers. Clearly dimerisation of the trefoil factors results in increased structural divergence. If we take the supposedly conserved hydrophobic region as our example once again, comparison of the three homodimer structures indicates that the two regions in each dimer can be either relatively far apart with no fixed orientation to each other; relatively close together on the same face of a constrained dimer structure; or twice as far apart in a different orientation on a second constrained dimer structure (fig. 4) [13, 14, 46]. Again, it seems improbable that these significant structural differences exist for interaction with a single ligand.

Does a different form indicate a different activity?

Different naturally occurring forms of individual trefoil factors are being identified and evidence is emerging that the different forms have differing activities both in vitro and in vivo and that they interact with ligands with differing avidity.

Although the TFF1 monomer and dimer both stimulate migration, the dimer is more potent. The concentrations of TFF1 required to stimulate migration of MCF-7 breast cancer cells are compatible with interaction with a high-affinity cell surface receptor, as occurs with other small mitogenic or motogenic factors (fig. 5) [37]. Several peptide growth factors and hormones bind to and activate their receptors by cross-linking receptor monomers. The emerging consensus that the TFF1 dimer is more potent is consistent with TFF1 stimulating migration via a bivalent dimeric receptor [37, 46].

The studies reviewed herein underpin the importance of being clear which molecular form of a trefoil factor is being measured in a particular assay. Until it is known which are the biologically most active or relevant forms, it is best to measure all forms and ideally to distinguish between different forms.

For TFF1, which has been the most studied to date, the dimer has increased activity compared with the monomer in its ability to stimulate migration [37], afford protection

against damage in vivo [47] and interact with the class I carcinogen *H. pylori* [60]. This illustrates the importance of understanding the different forms that exist and of studying their individual activities. For instance, the interaction between *H. pylori* and TFF1 would not have been discovered if only the TFF1 monomer had been analysed. Until now it has not been possible to assess the biological activity of the TFF1 heterodimer, which appears to be the most prominent form present in human gastric mucosa. The identification of TFIZ1 as the protein partner in the TFF1 heterodimer means that this is now a realistic possibility [25].

It seems unlikely that all molecular forms of mammalian trefoil factors have been identified, and until this is accomplished and pure sources of them become available, we cannot hope to understand the biology of this enigmatic family of molecules.

1 Henry J. A., Bennett M. K., Piggott N. H., Levett D. L., May F. E. B. and Westley B. R. (1991) Expression of the pNR-2/pS2 protein in diverse human epithelial tumours. Br. J. Cancer **64:** 677–682

2 May F. E. B. and Westley B. R. (1997) Expression of human intestinal trefoil factor in malignant-cells and its regulation by estrogen in breast-cancer cells. J. Pathol. **182:** 404–413

3 Jørgensen K. H., Thim L. and Jacobsen H. E. (1981) Preparation and characterization of PSP, a new polypeptide from porcine pancreas. Diabetalogia **21:** 288

4 Jørgensen K. H., Jørgensen K. D. and Thim L. (1983) Pancreatic spasmolytic polypeptide, US Patent 4,370,317 (Filed Mar. 9, 1981)

5 Jakowlew S. B., Breathnach R., Jeltsch J. M., Masiakowski P. and Chambon P. (1984) Sequence of the pS2 mRNA induced by estrogen in the human breast cancer cell line MCF-7. Nucleic Acids Res. **12:** 2861–2878

6 Prud'homme J. F., Fridlansky F., Le Cunff M., Atger M., Mercier-Bodart C., Pichon M. F. et al. (1985) Cloning of a gene expressed in human breast cancer and regulated by estrogen in MCF-7 cells. DNA **4:** 11–21

7 Suemori S., Lynch-Devaney K. and Podolsky D. K. (1991) Identification and characterization of rat intestinal trefoil factor: tissue- and cell-specific menber of the trefoil protein family. Proc. Natl. Acad. Sci. USA **88:** 11017–11021

8 Tomasetto C., Rio M., Gautier C., Wolf C., Hareuveni M., Chambon P. et al. (1990) hSP, the domain-dublicated homolog of pS2 protein, is co-expressed with pS2 in the stomach but not in breast carcinoma. EMBO J. **9:** 407–414

9 Podolsky D. K., Lynch-Devaney K., Stow J. L., Oates P., Murgue B., Debeaumont M. et al. (1993) Identification of human intestinal trefoil factor. Goblet cell- specific expression of a peptide targeted for apical secretion. J. Biol. Chem. **268:** 6694–6702

10 Hauser F., Poulsom R., Chinery R., Rogers L. A., Hanby A. M., Wright N. A. et al. (1993) hP1.B, a human P-domain peptide homologous with rat intestinal trefoil factor, is expressed also in the ulcer-associated cell lineage and the uterus. Proc. Natl. Acad. Sci. USA **90:** 6961–6965

11 Thim L. (1989) 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. **250:** 85–90

12 Carr M. D. (1992) 1H NMR-based determination of the secondary structure of porcine pancreatic spasmolytic polypeptide: one of a new family 'trefoil' motif containing cell growth factors. Biochemistry **31:** 1998–2004

13 Gajhede M., Petersen T. N., Henriksen A., Petersen J. F., Dauter Z., Wilson K. S. et al. (1993) Pancreatic spasmolytic polypeptide: first three dimensional structure of a member of the mammalian trefoil family of peptides. Structure **1:** 253–262

14 Muskett F. W., May F. E. B., Westley B. R. and Feeney J. (2003) Solution structure of the disulfide-linked dimer of human intestinal trefoil factor (TFF3): The intermolecular orientation and interactions are markedly different from those of other dimeric trefoil protein. Biochemistry **42:** 15139–15147

15 Dignass A., Lynch-Devaney K., Kindon H., Thim L. and Podolsky D. K. (1994) Trefoil peptides promote epithelial migration through a transforming growth factor beta-independent pathway. J. Clin. Invest. **94:** 376–383

16 Kindon H., Pothoulakis C., Thim L., Lynch-Devaney G. and Podolsky D. K. (1995) Trefoil peptide protection of intestinal epithelial barrier function – cooperative interaction with mucin glycoprotein. Gastroenterology **109:** 516–523

17 Farrell J. J., Taupin D., Koh T. J., Chen D., Zhao C. M., Podolsky D. K. et al. (2002) TFF2/SP-deficient mice show decreased gastric proliferation, increased acid secretion, and increased susceptibility to NSAID injury. J. Clin. Invest. **109:** 193–204

18 Lefebvre O., Chenard M. P., Masson R., Linares J., Dierich A., Lemeur M. et al. (1996) Gastric-mucosa abnormalities and tumorigenesis in mice lacking the ps2 trefoil protein. Science **274:** 259–262

19 Mashimo H., Wu D. C., Podolsky D. K. and Fishman M. C. (1996) Impaired defense of intestinal-mucosa in mice lacking intestinal trefoil factor. Science **274:** 262–265

20 Jørgensen K. H., Jørgensen K. D. and Thim L. (1979) Spasmolytic polypeptide, UK Patent Application 2059970

21 Jørgensen K. H., Jørgensen K. D. and Thim L. (1980) Pancreatic spasmolytic polypeptide, US Patent Application 185,796

22 Jørgensen K. H., Thim L. and Jacobsen H. E. (1982) Pancreatic spasmolytic polypeptide (PSP): I. Preparation and initial chemical characterization of a new polypeptide from porcine pancreas. Regul. Peptides **3:** 207–219

23 Mori K., Fujii R., Kida N., Takahashi H., Ohkubo S., Fujino M. et al. (1990) Complete primary structure of the human estrogen-responsive gene (pS2) product. J. Biochem. **107:** 73–76

24 Rio M. C., Lepage P., Diemunsch P., Roitsch C. and Chambon P. (1988) Primary structure of human protein pS2. CR. Acad. Sci. III **307:** 825–831

25 Westley B. R., Griffin S. M. and May F. E. B. (2005) Interaction between TFF1, a gastric tumor suppressor trefoil protein, and TFIZ1, a brichos domain-containing protein with homology to SP-C. Biochemistry **44:** 7967–7975

26 Mori K., Fujii R., Kida N., Ohta M. and Hayashi K. (1988) Identification of a polypeptide secreted by human breast cancer cells (MCF-7) as the human estrogen-responsive gene (pS2) product. Biochem. Biophys. Res. Commun. **155:** 366–372

27 Thim L., Thomsen J., Christensen M. and Jørgensen K. H. (1985) The amino acid sequence of pancreatic spasmolytic polypeptide. Biochim. Biophys. Acta **827:** 410–418

28 Thim L., Wöldike H. F., Nielsen P. F., Christensen M., Lynch-Devaney K. and Podolsky D. K. (1995) Characterization of human and rat intestinal trefoil factor produced in yeast. Biochemistry **34:** 4757–4764

29 Rose K., Savoy L. A., Thim L., Christensen M. and Jørgensen K. H. (1989) Revised amino acid sequence of pancreatic spasmolytic polypeptide exhibits greater similarity with an inducible pS2 peptide found in a human breast cancer cell line. Biochim. Biophys. Acta **998:** 297–300

30 Thim L., Jørgensen K. H., Thomsen J., Christensen M. and Jacobsen H. E. (1982) The amino acid sequence of pancreatic spasmolytic polypeptide. 4th International Symposium on Gastrointestinal Hormones, Stockholm, Sweden 20–23 June 1982 (Abstr.)

31 Thim L., Norris K., Norris F., Nielsen P. F., Bjørn S. E., Christensen M. et al. (1993) Purification and characterization of the

trefoil peptide human spasmolytic polypeptide (hSP) produced in yeast. FEBS Lett. **318:** 345–352
32 May F. E. B., Semple J. I., Newton J. L. and Westley B. R. (2000) The human two domain trefoil protein, TFF2, is glycosylated in vivo in the stomach. Gut **46:** 454–459
33 Semple J. I., Newton J. L., Westley B. R. and May F. E. B. (2001) Dramatic diurnal variation in the concentration of the human trefoil peptide TFF2 in gastric juice. Gut **48:** 648–655
34 Chadwick M. P., May F. E. B. and Westley B. R. (1995) Production and comparison of mature single-domain trefoil peptides pnr-2/ps2 cys(58) and pnr-2/ps2 ser(58). Biochem. J. **308:** 1001–1007
35 Chadwick M. P., Westley B. R. and May F. B. (1997) Homodimerization and hetero-oligomerization of the single-domain trefoil protein pnr-2/ps2 through cysteine-58. Biochem. J. **327:** 117–123
36 Kannan R., Tomasetto C., Staub A., Bossenmeyer-Pourie C., Thim L., Nielsen P. F. et al. (2001) Human pS2/trefoil factor 1: production and characterization in Pichia pastoris. Protein Expr. Purif. **21:** 92–98
37 Prest S. J., May F. E. B. and Westley B. R. (2002) The estrogen-regulated protein, TFF1, stimulates migration of human breast cancer cells. FASEB J. **16:** 592–594
38 Newton J. L., Allen A., Westley B. R. and May F. E. B. (2000) The human trefoil peptide, TFF1, is present in different molecular forms that are intimately associated with mucus in normal stomach. Gut **46:** 312–320
39 May F. E. B., Church S. T., Major S. and Westley B. R. (2003) The closely related estrogen-regulated trefoil proteins TFF1 and TFF3 have markedly different hydrodynamic properties, overall charge and distribution of surface charge. Biochemistry **42:** 8250–8259
40 Poulsen S. S., Thulesen J., Christensen L., Nexø E. and Thim L. (1999) Metabolism of oral trefoil factor 2 (TFF2) and the effect of oral and parenteral TFF2 on gastric and duodenal ulcer healing in the rat. Gut **45:** 516–522
41 Carr M. D., Bauer C. J., Gradwell M. J. and Feeney J. (1994) Solution structure of a trefoil-motif-containing cell growth factor, porcine spasmolytic polypeptide. Proc. Natl. Acad. Sci. USA **91:** 2206–2210
42 Lemercinier X., Muskett F. W., Cheeseman B., McIntosh P. B., Thim L. and Carr M. D. (2001) High-resolution solution structure of human intestinal trefoil factor and functional insights from detailed structural comparisons with the other members of the trefoil family of mammalian cell motility factors. Biochemistry **40:** 9552–9559
43 Polshakov V. I., Williams M. A., Gargaro A. R., Frenkiel T. A., Westley B. R., Chadwick M. P. et al. (1997) High-resolution solution structure of human pnr-2/ps2 – a single trefoil motif protein. J. Mol. Biol. **267:** 418–432
44 Petersen T. N., Henriksen A. and Gajhede M. (1996) Structure of porcine spasmolytic polypeptide at 1.95 Å resulution. Acta Cryst. **D52:** 730–737
45 De A., Brown D. G., Gorman M. A., Carr M. D., Sanderson M. R. and Freemont P. S. (1994) Crystal structure of a disulphide-linked 'trefoil' motif found in a large family of putative growth factors. Proc. Natl. Acad. Sci. USA **91:** 1084–1088
46 Williams M. A., Westley B. R., May F. E. B. and Feeney J. (2001) The solution structure of the disulphide-linked homodimer of the human trefoil protein TFF1. FEBS Lett. **493:** 70–74
47 Marchbank T., Westley B. R., May F. E. B., Calnan D. P. and Playford R. J. (1998) Dimerization of human ps2 (tff1) plays a key role in its protective/healing effects. J. Pathol. **185:** 153–158
48 Calnan D. P., Westley B. R., May F. E. B., Floyd D. N., Marchbank T. and Playford R. J. (1999) The trefoil peptide TFF1 inhibits the growth of the human gastric adenocarcinoma cell line AGS. J. Pathol. **188:** 312–317
49 Carrasco R., Pera M., May F. E. B., Westley B. R., Martinez A. and Morales L. (2004) Trefoil factor family peptide 3 prevents the development and promotes healing of ischemia-reperfusion injury in weanling rats. J. Pediatr. Surg. **39:** 1693–1700
50 Playford R. J., Marchbank T., Chinery R., Evison R., Pignatelli M., Boulton R. A. et al. (1995) Human spasmolytic polypeptide is a cytoprotective agent that stimulates cell migration. Gastroenterology **108:** 108–116
51 Gipson I. K. and Argueso P. (2003) Role of mucins in the function of the corneal and conjunctival epithelia. Int. Rev. Cytol. Survey Cell Biol. **231:** 1–49
52 Longman R. J., Douthwaite J., Sylvester P. A., Poulsom R., Corfield A. P., Thomas M. G. et al. (2000) Coordinated localisation of mucins and trefoil peptides in the ulcer associated cell lineage and the gastrointestinal mucosa. Gut **47:** 792–800
53 Thim L., Madsen F. and Poulsen S. S. (2002) Effect of trefoil factors on the viscoelastic properties of mucus gels. Eur. J. Clin. Invest. **32:** 519–527
54 Warson C., van de Bovenkamp J. H. B., Korteland-Van Male A. M., Buller H. A., Einerhand A. W. C., Ectors N. L. E. Y. et al. (2002) Barrett's esophagus is characterized by expression of gastric-type mucins (MUC5AC, MUC6) and TFF peptides (TFF1 and TFF2), but the risk of carcinoma development may be indicated by the intestinal-type mucin, MUC2. Hum. Pathol. **33:** 660–668
55 Shaoul R., Okada Y., Cutz E. and Marcon M. A. (2004) Colonic expression of MUC2, MUC5AC, and TFF1 in inflammatory bowel disease in children. J. Pediatr. Gastroenterol. Nutr. **38:** 488–493
56 Tomasetto C., Masson R., Linares J. L., Wendling C., Lefebvre O., Chenard M. P. et al. (2000) pS2/TFF1 interacts directly with the VWFC cysteine-rich domains of mucins. Gastroenterology **118:** 70–80
57 Gum J. R., Hicks J. W., Toribara N. W., Siddiki B. and Kim Y. S. (1994) Molecular-cloning of human intestinal mucin (muc2) cdna – identification of the amino-terminus and overall sequence similarity to prepro-von-Willebrand factor. J. Biol. Chem. **269:** 2440–2446
58 Ruchaud-Sparagano M. H., Westley B. R. and May F. E. B. (2004) The trefoil protein TFF1 is bound to MUC5AC in humangastric mucosa. Cell. Mol. Life Sci. **61:** 1946–1954
59 Newton J. L., Allen A., Westley B. R. and May F. E. B. (2000) The human trefoil peptide, TFF1, is present in different molecular forms that are intimately associated with mucus in normal stomach. Gut **46:** 312–320
60 Clyne M., Dillon P., Daly S., O'Kennedy R., May F. E. B., Westley B. R. et al. (2004) Helicobacter pylori interacts with the human single-domain trefoil protein TFF1. Proc. Natl. Acad. Sci. USA **101:** 7409–7414
61 Park W. S., Oh R. R., Park J. Y., Lee J. H., Shin M. S., Kim H. S. et al. (2000) Somatic mutations of the trefoil factor family 1 gene in gastric cancer. Gastroenterology **119:** 691–698
62 Polshakov V. I., Frenkiel T. A., Westley B. R., Chadwick M., May F. Carr M. D. et al. (1995) NMR-based structural studies of the pNR-2/pS2 single domain trefoil peptide. Similarities to porcine spasmolytic peptide and evidence for a monomeric structure. Eur. J. Biochem. **233:** 847–855
63 Itoh H., Tomita M., Uchino H., Kobayashi T., Kataoka H., Sekiya R. et al. (1996) cDNA cloning of rat ps2 peptide and expression of trefoil peptides in acetic acid-induced colitis. Biochem. J. **318:** 939–944
64 Lefebvre O., Wolf C., Kedinger M., Chenard M. P., Tomasetto C., Chambon P. et al. (1993) The mouse one P-domain (pS2) and two P-domain (mSP) genes exhibit distinct patterns of expression. J. Cell Biol. **122:** 191–198
65 Jeffrey G. P., Oates P. S., Wang T. C., Babyatsky M. W. and Brand S. J. (1994) Spasmolytic polypeptide: a trefoil peptide secreted by rat gastric mucuos cells. Gastroenterology **106:** 336–345
66 Campbell B. G. and Jabbes M. (2003) Canine TFF2 sequence, EMBL Database AY264843

67 Probst J. C., Skutella T., Mullerschmid A., Jirikowski G. F. and Hoffmann W. (1995) 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. **33:** 269–276
68 Sands B. E., Ogata H., Lynch-Devaney K., DeBeaumont M., Ezzell R. M. and Podolsky D. K. (1995) Molecular-cloning of the rat intestinal trefoil factor gene – characterization of an intestinal goblet cell-associated promoter. J. Biol. Chem. **270:** 9353–9361
69 Mashimo H., Podolsky D. K. and Fishman M. C. (1995) Structure and expression of murine intestinal trefoil factor – high evolutionary conservation and postnatal expression. Biochem. Biophys. Res. Commun. **210:** 31–37
70 Tomita M., Itoh H., Ishikawa N., Higa A., Ide H., Murakumo Y. et al. (1995) Molecular cloning of mouse intestinal trefoil factor and its expression during goblet cell changes. Biochem. J. **311:** 293–297
71 Winteroe A. K., Fredholm M. and Davies W. (1996) Evaluation and characterization of a porcine small intestine cDNA library. Mamm. Genome **7:** 509
72 Campbell B. G. and Jabbes M. (2003) Canine TFF3, EMBL Database AY264844