FEBS Letters 357 (1995) 50–54 FEBS 14910

Characterisation of the single copy trefoil peptides intestinal trefoil factor and pS2 and their ability to form covalent dimers

Rebecca Chinery[a,**], Paul A. Bates[b,**], Amitabha De[c,***], Paul S. Freemont[c,*]

[a]*Department of Pharmacology, The Royal College of Surgeons of England, 35-43 Lincoln's Inn Fields, London WC2A 3PN, UK*
[b]*Biomolecular Modelling Laboratory, Imperial Cancer Research Fund, P.O. Box 123, Lincoln's Inn Fields, London WC2A 3PX, UK*
[c]*Protein Structure Laboratory, Imperial Cancer Research Fund, P.O. Box 123, Lincoln's Inn Fields, London WC2A 3PX, UK*

Received 18 October 1994; revised version received 11 November 1994

Abstract A bacterial recombinant expression system was established to produce biologically active rat Intestinal Trefoil Factor (rITF). Characterisation of purified rITF shows that both monomers and dimers can be observed under reducing and non-reducing conditions, respectively. Site-directed mutagenesis studies show that Cys57 is necessary for rITF dimer formation. Samples of human gastrointestinal tissue following biopsy also demonstrated the presence of reducible human pS2 and ITF covalent dimers. Three-dimensional models for pS2 and ITF support the hypothesis that both pS2 and ITF can exist as disulphide-linked dimers in vivo and that any proposed function for these peptides must take dimer formation into account.

Key words: Trefoil peptide; Protein modelling

1. Introduction

The trefoil family is a group of extracellular peptides that are highly expressed in the gastrointestinal tract and whose physiological role is largely unknown [1]. Characteristic for these peptides is a conserved pattern of six cysteine residues, which form three intramolecular disulphide bridges, as well as a number of conserved arginine, glycine and tryptophan residues. The spacing between five of the six cysteine residues is invariable, with only the spacing between Cys^1 and Cys^2 subject to minor variation (see Fig. 3a). The proposed primary structure of the motif, with its three disulphide loops, resembles a 'trefoil', which is highly resistant to proteolytic and acid degradation [2].

In mammals, at least two classes of different trefoil proteins have been characterised containing either one (human, mouse pS2; rat, mouse and human intestinal trefoil factor [ITF]) or two (porcine, human, rat and mouse spasmolytic polypeptides [SP]) such motifs [1]. As a general principle, expression of each peptide is confined to specialised cells which secrete various mucin glycoproteins. This has led to the suggestion that trefoil peptides are associated with mucin secretion and/or stabilization, thus explaining their over-expression at sites of chronic inflammatory and ulcerative conditions, where reformation of the mucous gel would be important [3–5].

Since the initial identification of single and multiple trefoil-containing peptides, little information has been published about their structures and/or functional relationships to each other. Recently, however, the NMR and crystal structures for the two trefoil-containing protein, porcine pancreatic (P)SP, have been reported [6–8]. In this peptide, both trefoil-domains adopt similar conformations providing a suitable three-dimensional structural template for a single trefoil motif. As there has been some speculation as to the functions of trefoil proteins and their possible involvement in either mucous-gel stabilization or receptor-mediated responses [1], we have used recombinant rat (r)ITF as a single trefoil-containing peptide to study its ability to self-associate in the presence or absence of its free cysteine residue. In addition, we describe the detection of both mono- and dimeric forms of human pS2 and ITF in human tissue obtained by biopsy. We also present three-dimensional models for pS2 and ITF homodimers based on the crystal structure of the double trefoil-containing protein PSP.

2. Materials and methods

2.1. Expression and purification of recombinant rITF in E. coli

The predicted mature rat ITF cDNA sequence (bases 64–243) [9,10] was cloned in-frame into the bacterial expression vector pGEX-3X. Transformants of *E. coli* were produced and selected by conventional procedures [11]. Induction of transformants, lysis of cells, and the recovery of rITF was carried out as described by Gearing et al. [12]. Purified samples were dialysed against 0.1 M Tris-HCl, pH 7.4, 1 mM EDTA, 1 mM EGTA, and injected onto a Superose 12 gel filtration column (Pharmacia) equilibrated with the same buffer, to confirm the identity of the peptide. Cytochrome *c* (12.4 kDa) and blue dextran (2 kDa) were used for column calibration.

2.2. Generation of a mutant form of recombinant rITF

A ^{57}Cys(TGT) to ^{57}Lys(AAA) mutation was made in the mature rITF cDNA sequence using polymerase chain reaction with a mutagenic oligonucleotide 3′ primer. The mutant rITF form encoded by the resulting plasmid was produced and purified as described above for wild-type recombinant rITF.

2.3. Production of antisera specific to rITF

Purified wild-type rITF peptide was used to raise a polyclonal antiserum in two rabbits. Rabbits were injected 4 times at bi-weekly intervals with 50 μg of purified peptide. Serum was obtained 1 month after the fourth injection and was used for immunoblotting at a final dilution of 1:800. The pS2 monoclonal antibody was raised against a synthetic peptide corresponding to the terminal 31 amino acids of the predicted pS2 protein [13] and used at a final dilution of 1:300.

2.4. Homodimerisation of trefoil peptides in vitro and in vivo

The effect of pH on dimerisation was examined in vitro by incubating purified wild-type recombinant rITF peptide overnight in 300 mM potassium phosphate, varying pH values between 4 and 7. To assess pre-existing dimerisation in vivo, small samples of human small intestinal and colonic tissue (1–5 mg) with chronic ulcerative conditions were frozen rapidly following biopsy. All peptide and tissue samples were resuspended in Laemmli sample buffer, either in the presence or

*Corresponding author. Fax: (44) (71) 269 3093.

**Joint first authors.

****Current address:* Khalisani College, Chandannagore, Hooghly, India.

SSDI 0014-5793(94)01297-0

absence of the reducing agent, 2-merceptoethanol. Samples were then boiled for 3 min and the supernatant analysed by SDS-PAGE and Western blotting.

2.5. Western blot analysis of ITF and pS2

10–18% polyacrylamide gels were prepared by the method of Laemmli [14] and used under reducing conditions. Equal amounts of protein (10 μg) were added per lane with molecular mass markers run in parallel lanes. After electrophoresis, proteins were transferred electrophoretically at room temperature for 3 h at 60 V onto nitrocellulose membranes (0.1 mm pore size) in a Bio-Rad Trans-Blot cell [15]. After transfer, membranes were incubated overnight in 5% fetal calf serum in phosphate-buffered saline at 4°C and then incubated successively with the appropriate primary antibody followed by a horseradish peroxidase-conjugated anti-immunoglobulin, both for 2 h at 4°C. The peroxidase reaction was developed using the Amersham's ECL detection system. Negative controls were obtained by preadsorbing the antibody with the appropriate immunising peptide.

2.6. Trefoil structure modelling

A representative set of mammalian trefoil sequences were aligned with the program CLUSTAL V (Fig. 3a) [16]. Small adjustments to the alignment were made manually to better reflect the patches of sequence similarities. The crystal structure of PSP was used as a template to model both ITF and pS2 homodimers. Side chains of PSP domain B (PSP_B in Fig. 3a), and protein backbone fragments with incompatible *phi psi* angles, were substituted for those of ITF and pS2 using the protocol described for the program 3-D JIGSAW [17] and a conformer library [18]. The ITF and pS2 domains were energy refined using 50 steps of steepest decent with the programming package CHARM [19]. The final homodimer models were also energy refined with a further 100 steps of steepest decent and the complete models screened with the Protein Health checks option within the programming package QUANTA [20].

3. Results and discussion

3.1. Isolation and characterisation of bacterial recombinant rITF

Recombinant rITF protein was initially purified according to the method described by Gearing et al. [12]. SDS-PAGE revealed that rITF was synthesised initially as a 33 kDa fusion protein (data not shown). Bound fusion protein was cleaved by Factor X_a and the released moiety recovered in the column supernatant. Eluted material consisted principally of rITF, although there were contaminating levels, as estimated by Coomassie blue staining, of Factor X_a (1%), glutathione-*S*-transferase (10%) and a ~90 kDa host protein (3–5%). The ITF sample was further purified by gel filtration under non-denaturing conditions. The residual contaminants of rITF were found to be less than 1% as estimated by Coomassie blue staining.

3.2. Dimerisation of wild-type rITF in vitro

In Fig. 1, recombinant rITF is identified by SDS-PAGE (lanes 2–3, 6–9). Under non-reducing conditions, two entities of approximately 7 and 14 kDa can be observed (lane 2). When the larger protein was extracted from the gel and reduced, this molecule was found to yield a single protein of 7 kDa (lane 3), consistent with the shift from dimeric to monomeric rITF. Both proteins showed immunological cross-reactivity with the polyclonal antisera (data not shown), demonstrating that recombinant rITF proteins occur as both monomeric and homodimeric forms. Densitometric scanning of Coomassie blue-stained gels suggest that recombinant rITF as purified exists predominantly in the dimeric form, comprising about 80% of the total peptide sample. It is possible that the monomeric recombinant rITF observed under non-reducing conditions (Fig. 1, lane 2), has lost the ability to dimerise, due to either oxidation of the free

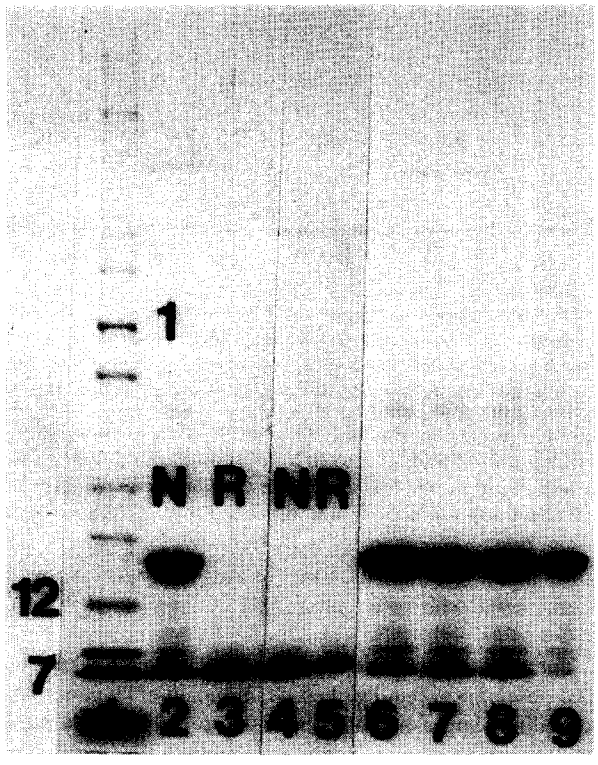


Fig. 1. Separation and identification of the recombinant wild-type and mutant rITF by SDS-PAGE. Coomassie blue-stained gel showing electrophoretic separation of recombinant rITF. Lanes 2 and 3 show the presence of dimeric and monomeric forms of recombinant rITF when run under non-reducing conditions (lane 2, N) and reducing conditions (lane 3, R). Lanes 4 and 5 shows the presence of C58K mutant rITF run under non-reducing (N) and reducing (R) conditions, the position of which coincides precisely with that of the monomeric form of rITF. Lanes 6–9 show rITF following incubation with potassium phosphate at varying pH values (pH 7–4: lanes 6–9, respectively). Molecular-mass markers are shown in lane 1 and include from bottom to top aprotinin (6.5 kDa), myoglobin (8.2 kDa), lysozyme (14.3 kDa) and trypsin inhibitor (20.1 kDa).

sulphydryl or by binding of the free sulphydryl to glutathione at the affinity column step during the purification scheme. Interestingly, a small band is observed running slightly higher than the monomer protein (Fig. 1, lane 2) which disappears under reducing conditions (Fig. 1, lane 3) which would be consistent with the presence of an ITF-glutathione complex.

As ITF is found within the gastrointestinal tract, the effect of pH on the stability of the recombinant ITF homodimer was therefore investigated. Purified rITF protein samples were incubated overnight in phosphate buffer of varying pH (pH 4–7) and subsequently analysed by non-reducing SDS-PAGE (Fig. 1, lanes 6–9) and size exclusion chromatography. This treatment had little effect on the concentration and/or ratio of ITF homodimers.

3.3. Characterisation of the ^{57}Cys mutant of recombinant rITF

In order to determine whether the free cysteine residue at the C-terminus is involved in the dimerisation of rITF, site-directed mutagenesis was carried out changing Cys^{57} to Lys. As shown in Fig. 1 (lane 4), the loss of the Cys^{57} prevents the interaction of recombinant rITF to form homodimers. Interestingly, the inability to form dimeric forms of rITF inhibits changes in ion transport across confluent cultured epithelial layers, using

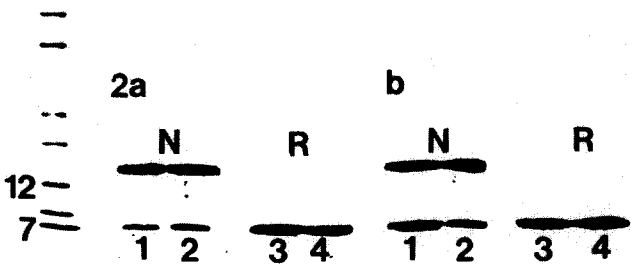


Fig. 2. Identification of pS2 and ITF in human tissue by Western blot analysis. Human intestinal tissue was boiled in Laemmli loading buffer with (lanes 1,2 in duplicate) or without (lane 3,4 in duplicate) the presence of 2-mercaptoethanol and proteins resolved under non-reducing conditions (see section 2). (a) Autograph derived from the above gel showing the presence of monomeric and dimeric forms of pS2, the positions of which coincide precisely with that of aprotinin (6.5 kDa) and lysozyme (14.3 kDa), respectively. (b) Autograph showing a similar arrangement for human ITF. Molecular mass markers are indicated at the left hand side (kDa) as in Fig. 1a.

voltage-clamp techniques, which have been previously described for the wild-type peptide [21].

3.4. Potential dimerisation of human ITF and pS2 in vivo

The observation that purified recombinant rITF can exist as a homodimer suggests that either itself, or other single trefoil-containing proteins, may be associated in a similar way in vivo. To try and test this hypothesis, homogenised samples of human small intestinal tissue obtained from a patient with Crohn's disease (a chronic inflammatory bowel disorder associated with the over-expression of pS2, SP, and ITF) were analysed by Western blotting using a monoclonal antibody to pS2 or the polyclonal antisera raised against recombinant rITF (Fig. 2). Under non-reducing conditions, two proteins of approximately 7 and 14 kDa could be observed following pS2 (Fig. 2a: lanes 1,2) and ITF (Fig. 2b: lanes 1,2) immunostaining. When the same sample was electrophoresed under reducing conditions, the larger immunoreactive band disappeared (lanes 3,4), suggesting that both human pS2 and ITF exist as disulphide-linked complexes in vivo. The presence of the monomeric form of human ITF and pS2 under non-reducing conditions may be due in some way to the extraction and/or storage procedures of the tissue samples after biopsy, which resulted in an unknown modification of the free sulfhydryl. A similar result was seen for the purified recombinant rITF (see above). Due to the similar sizes of pS2 and ITF, it is possible that the 14 kDa immunoreactive covalent complexes observed in Fig. 2 are pS2/ITF heterodimers. Although, neither antibody demonstrated cross-reactivity with the opposing peptide, it is not possible at this stage to discriminate between pS2/ITF homo- or heterodimers as both peptides are present in the same tissue sample. It is also possible that the 14 kDa immunoreactive covalent complexes of human pS2 and ITF are disulphide-linked complexes of pS2 or ITF with unknown proteins of similar sizes to pS2 and ITF. However, the molecular modelling studies of the pS2 and ITF homodimers (see below) combined with the in vitro analysis of recombinant rITF support the conclusion that pS2 and ITF can exist as covalent dimers in vivo.

3.5. The pS2 and ITF homodimer models

The crystal structure of PSP consists of two covalently linked trefoil domains (Fig. 3) [6–8]. Inspection of the structure suggests that single copy members of the trefoil family, such as ITF and pS2, could pack together in a similar orientation to the two trefoils in PSP, thus forming homodimers. A sequence alignment of the trefoil family shows that human, rITF and pS2 do not have the first cysteine of the analogous N-terminal trefoil of PSP but an extra cysteine at the C-terminus (Fig. 3a). As all other cysteines are conserved it is these terminal cysteines that are predicted to link two molecules of either ITF or pS2 together. The N and C-terminal trefoil domains of PSP have high sequence similarity to ITF and pS2 (Fig. 3a). However, the C-terminal PSP trefoil domain (PSP_B in Fig. 3a) was thought to be the best template for modelling as no insertions or deletions were necessary within the core of the motif. We modelled the hITF homodimer rather than the rITF homodimer, as pS2 is a human sequences and there is high sequence identity (76%) between the mature human and rat ITF proteins, including all of the important hydrophobic core residues and disulphide bridges (Fig. 3a).

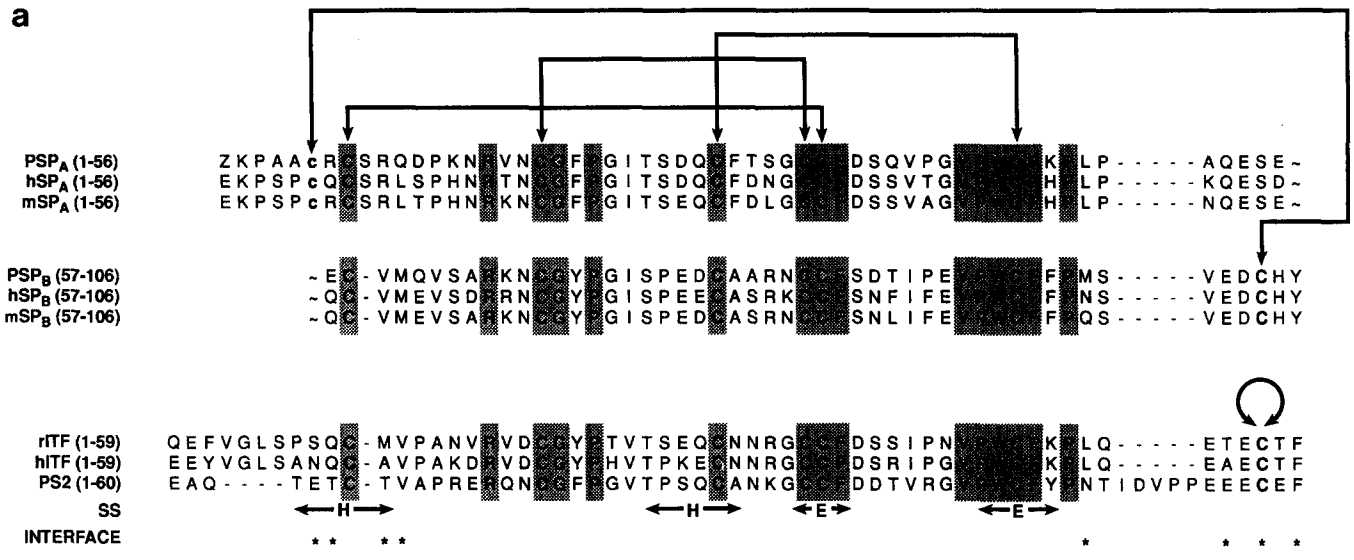
Inspection of the sequences and PSP structure indicated that the last conserved residue before the C-termini is a proline, six residues back from the free cysteine of human ITF (Fig. 3a). The remaining residues between this proline and the C-termini do not contribute to the structure of the PSP trefoils but some do contribute to the interface between trefoils both in the crystal structure and predicted homodimer interfaces. Other interface residues common to both the PSP X-ray crystal structure and predicted homodimer interfaces are on the whole conserved but there are some differences between the crystal structure and potential homodimers (Fig. 3a). The most notable is the disulphide link at the interface of the homodimers and this alone suggests that small adjustments are needed between the relative orientations of the trefoil domains.

Twelve backbone *phi/psi* angles, six from each domain of human ITF starting from the conserved proline and ending with the terminal cysteine, were adjusted manually to form the inter-trefoil disulphide link. Equal adjustments were made to equivalent angles in each trefoil and only allowed regions of the Ramachandran *phi/psi* map [22] were sampled with the best packing of residues into the core and between each trefoil maintained. Under these constraints a disulphide link could be formed. For pS2, there is a five residue insertion in the pS2 sequence (see Fig. 3a), thus a fragment database search [17] was carried out to model this insertion with the endpoints of the search fixed. The best fitting fragment chosen was for that of a random coil as might be expected due to the occurrence of two consecutive prolines within the insert region (see Fig. 3b). The final energy refined ITF and pS2 homodimer models were assessed for correct geometry, with both models passing all protein health checks set by the program QUANTA [20].

The possibility that ITF and pS2 homodimers link via a disulphide but with no stabilising homodimer interface, cannot be discounted. Indeed, even free rotation of both domains around the disulphide bond is possible. However, there is enough conservation in both sequence and secondary structure at the trefoil interfaces, relative to the known crystal structure of PSP, to suggest that the homodimers are stable as modelled. Moreover, the modelled intermolecular disulphide link is of good geometry and is energetically stable.

3.5. Functional implications of pS2 and ITF dimerisation

Does the formation of homo- or heterodimers by ITF and



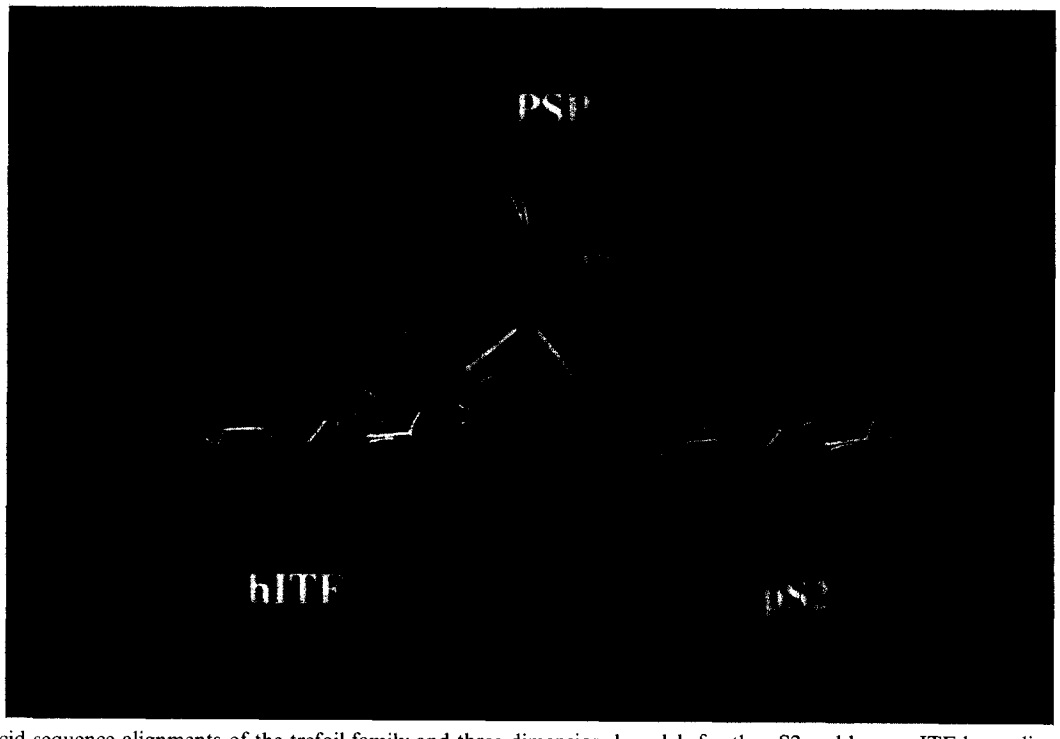


Fig. 3. Amino acid sequence alignments of the trefoil family and three-dimensional models for the pS2 and human ITF homodimers. (a) Multiple sequence alignments of trefoil domains (reviewed in [1]). The top sequence of each sequence block PSP$_A$ and PSP$_B$ are from the X-ray crystal structure of PSP. The bottom two sequences of the lower block are the trefoils modelled, human ITF [4], and human breast cancer associated peptide pS2. Residues that are 100% conserved in all trefoils are boxed. Disulphide links within and between trefoils are indicated by arrows. The homodimer disulphide is indicated by a circular arrow. The symbol ~, indicates the peptide break made between the two trefoil units of PSP. The row labelled SS, indicate the position of secondary structure elements as determined from the crystal structure of PSP. The row labelled **INTERFACE** indicate those residues that are part of the interface between trefoils both in the crystal structure and predicted homodimers. (b) Model for the homodimers of hITF and pS2 and how they were modelled from PSP. The protein chain of each structure is illustrated as a coloured cylinder. Each trefoil domain for the homodimers is in a different colour, green and light blue. Disulphide bonds are indicated by yellow bars. The cysteine that has been lost by the homodimers relative to the PSP structure is labelled, **C→N/E**, as **N** is the mutation for hITF and **E** for pS2 at this position. The cysteine gained by the homodimers is labelled by, **S→C**. The position of the five residue insert of pS2 relative to hITF is also labelled. Figure produced with PREPI (S. Islam and M. Sternberg, ICRF).

pS2 in vivo indicate a functional mechanism? It has been suggested that trefoil peptides in general, including ITF and pS2, may act to stabilise the mucous gel layer since they are co-expressed with mucin-type glycoproteins in the human gastrointestinal tract [23]. These mucins are covered by O-linked carbohydrates and preliminary modelling studies by us as well as by others [6] indicate that some of the surface pockets on each trefoil may bind carbohydrate units. Hence, trefoil domains may bind carbohydrate chains attached to mucins which under the right conditions, such as mucosal damage (ulceration, injury), the trefoils could form disulphide-linked dimers. The cross-linking of mucin chains could effect the viscosity of the mucous layer possibly facilitating cellular restitution across the damaged area. Interestingly, for secreted mucins, free cysteines on separate mucin molecules cross-link the cores of the mucin chains through intermolecular disulphide bonds forming the mucous layer [24]. Trefoil peptides could bind to the polysaccharide part of the mucin molecules and provide further cross-linking by forming covalent homo- or heterodimers.

Alternatively, the existence of homo- or heterodimeric forms of pS2 and ITF could enable these single trefoil proteins to compete with double trefoil-containing proteins for receptor sites. High affinity binding sites for PSP have already been demonstrated in the rat small intestine [25], and putative binding sites for rITF were demonstrated histochemically in the rat stomach, kidney, and colon [26], regions of the gut rich in both single (pS2, ITF) and double (SP) trefoil-containing proteins. It has been suggested previously that the growth stimulatory effects of PSP on a human breast carcinoma cell line may be mediated by its cross-recognition of a pS2-binding site [27], reflecting the related tertiary structures of these trefoil proteins. The functional implications of either homo- or heterodimeric pS2 and ITF, both physiological and pathological, will have to await further in vitro and in vivo studies.

Acknowledgements: R.C. is indebted to the Wellcome Trust for her postdoctorate funding. We thank Dr Suhail Islam for the use of his protein presentation program PREPI. We also thank Drs Kathy Borden and Richard Poulsom for comments on the manuscript and M.J.E. Sternberg for discussion about the modelling procedures.

References

[1] Poulsom, R. and Wright, N.A. (1993) Am. J. Physiol. 265, 6205–6213.
[2] Thim, L. (1989) FEBS Lett. 250, 85–90.
[3] Ashwood, P., Chinery, R., Alison, M., Jeffery, R.E., Poulsom, R. and Wright, N.A. (1993) Gut 34, F199.
[4] Hauser, F., Poulsom, R., Chinery, R., Rogers, L.A., Hanby, A.M., Wright, N.A. and Hoffmann, W. (1993) Proc. Natl. Acad. Sci. USA 90, 6961–6965.
[5] Wright, N.A., Poulsom, R., Stamp, G.W.H., Hall, P.A., Jeffery, R.E., Longcroft, J.M., Rio, M.-C., Tomasetto, C. and Chambon, P. (1990) J. Pathol. 162, 279–284.
[6] Gajhede, M., Petersen, T.N., Henriksen, A., Petersen, J.F.W., Dauter, Z., Wilson, K.S. and Thim, L. (1993) Structure 1, 253–262.
[7] De, A., Brown, D.G., Gorman, M.A., Carr, M., Sanderson, M.R. and Freemont, P.S. (1994) Proc. Natl. Acad. Sci. USA 91, 1084–1088.
[8] Carr, M.D., Bauer, C.J., Gradwell, M.J. and Fenney, J. (1994) Proc. Natl. Acad. Sci. USA 91, 2206–2210.
[9] Suemori, S., Lynch-Devaney, K. and Podolsky, D.K. (1991) Proc. Natl. Acad. Sci. USA 88, 11017–11021.
[10] Chinery, R., Poulsom, R., Rogers, L.A., Jeffery, R.E., Longcroft, J.M., Hanby, A.M. and Wright, N.A. (1992) Biochem. J. 285, 5–8.
[11] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[12] Gearing, D.P., Nicola, N.A., Metcalf, D., Foote, S., Willson, T.A., Gough, N.M. and Williams, R.L. (1989) Biotechnology 7, 1157–1161.
[13] Poulsom, R., Chinery, R., Sarraf, C., Lalani, E.-N., Stamp, G.W.H., Elia, G. and Wright, N.A. (1992) Scand. J. Gastroenterol. 27, 17–21.
[14] Laemmli, U.K. (1970) Nature 227, 680–685.
[15] Towbin, H., Staehelin, T. and Gordon, J. (1979) Proc. Natl. Acad. Sci. USA 76, 4350–4354.
[16] Higgins, D.G., Bleasby, A.J. and Fuchs, R. (1992) Comput. Applic. Biosci. 8, 189–191.
[17] Bates, P.A. and Sternberg, M.J.E. (1992) in: Protein Engineering. A Practical Approach (Rees, A.R., Sternberg, M.J.E. and Wetzel, R., Eds.) pp. 117–141, IRL Press.
[18] McGregor, M., Islam, S. and Sternberg, M.J.E. (1987) J. Mol. Biol. 198, 295–310.
[19] CHARM software version 3.3 (1992) Molecular Simulations Inc. 200 Fifth Avenue, Waltham, MA 02154.
[20] QUANTA software version 3.3 (1992) Molecular Simulations Inc. 200 Fifth Avenue, Waltham, MA 02154.
[21] Cox, H.M., Chinery, R. and Wright, N.A. (1993) Regul. Pept. 47, 90.
[22] Ramachandran, G.N., Ramakrishnan, C. and Sasisekharan, V. (1963) J. Mol. Biol. 7, 95–99.
[23] Strous, G.J. and Dekker, J. (1992) Crit. Rev. Biochem. Mol. Biochem. 27, 57–92.
[24] Gum, J.R. (1992) Am. J. Respir. Cell Mol. Biol. 7, 557–564.
[25] Frandsen, E.K. (1988) Regul. Peptides 3, 231–243.
[26] Chinery, R., Poulsom, R., Elia, G., Hanby, A.M. and Wright, N.A. (1993) Eur. J. Biochem. 212, 557–563.
[27] Hoosein, N.M., Thim, L., Jorgensen, K.H. and Brattain, M.G. (1989) FEBS Lett. 247, 303–306.