

Pancreatic spasmolytic polypeptide: first three-dimensional structure of a member of the mammalian trefoil family of peptides

M Gajhede^{1*}, TN Petersen¹, A Henriksen¹, JFW Petersen¹,
Z Dauter², KS Wilson² and L Thim

¹Department of Chemistry, The H.C. Ørsted Institute, University of Copenhagen, DK-2100 Copenhagen, Denmark, ²European Molecular Biology Laboratory c/o DESY Notkestrasse 85, D-22603 Hamburg, Germany and ³Novo Nordisk A/S, DK-2880 Bagsværd, Denmark

Background: The trefoil peptides are a rapidly growing family of peptides, mainly found in the gastrointestinal tract. There is circumstantial evidence that they stabilize the mucus layer, and may affect the rate of healing of the mucosal epithelium.

Results: We have determined the structure of porcine pancreatic spasmolytic polypeptide (PSP) to 2.5 Å resolution. The polypeptide contains two trefoil domains. The domain structure is compact, and is composed of

a central short antiparallel β -sheet with one short helix above and one below it. This is a novel motif. The two domains are related by two-fold symmetry, and each domain contains a cleft.

Conclusions: The cleft within each domain could accommodate a polysaccharide chain, and may therefore be responsible for binding mucin glycoproteins. We suggest that PSP may cross-link glycoproteins, explaining its ability to stabilize the mucus layer.

Structure 15 December 1993, 1:253–262

Key words: mucus layer, P-domain, pancreatic spasmolytic polypeptide, trefoil family, X-ray structure

Introduction

Trefoil peptides [1] make up a rapidly growing family found mainly in association with the gastrointestinal tract. Circumstantial evidence has accumulated during the last 3–4 years indicating that trefoil peptides have a stabilizing effect on the mucus layer of the gastrointestinal tract, and the peptides have thus been suggested to be naturally-occurring healing factors for peptic ulcers [2], inflammatory bowel disease and other diseases of the gastrointestinal tract involving mucosal injury [3–5]. Mammalian trefoil peptides have one or two characteristic domains comprising 38 or 39 amino acid residues, including six cysteine residues which form disulphide bridges in the configuration 1–5, 2–4 and 3–6. This disulphide bonding pattern results in a distinctive three-leaved structure giving the peptide family its name [1].

Porcine pancreatic spasmolytic polypeptide (PSP) was originally found in a side fraction from the purification of porcine insulin [6]. The amount of PSP in the porcine pancreas is 100 mg kg⁻¹ [7]. PSP contains 106 amino acid residues in a single chain and the peptide is heavily cross-linked by seven disulphide bonds (Fig. 1) [1,8,9]. PSP is secreted in large quantities (50–100 mg l⁻¹) into the pancreatic juice upon stimulation with pancreozymin or secretin [7,10], and it has been found that PSP is resistant to digestion by intraluminal proteases within the gastrointestinal tract [6] explaining how it retains its full range of biological activities following oral administration [11].

In contrast to porcine PSP, the human counterpart, hSP, is expressed in the stomach but not in the pancreas [12]. Human spasmolytic polypeptide shows high sequence homology to PSP [12,13] (Fig. 2) especially within the trefoil domain. This domain, also referred to as the P-domain motif [12], is well conserved in other members of the trefoil family: intestinal trefoil factor, ITF, from rat [14] and man [15]; human breast cancer associated pS2 peptide [16–18]; frog skin peptides xP1, xP2 and xP4 [19,20], frog integumentary mucin C.1 [21] and rabbit *zona pellucida* protein [22].

At least nine different trefoil domains are known in six mammalian peptides (Fig. 2), but so far the three-dimensional structure has not been reported for any trefoil peptide. Here we report the X-ray structure of porcine pancreatic spasmolytic polypeptide (PSP), a member of the family containing two trefoil domains, and describe the atomic model refined at 2.5 Å resolution.

Results and discussion

Structure of the PSP molecule

The approximate size of the PSP molecule is 50 × 20 × 20 Å, giving it a rectangular shape. As expected from the amino acid sequence [8] and disulphide bond assignment [1], PSP is composed of two homologous trefoil domains. If disulphide bridge Cys6–Cys104 is considered part of the molecular skele-

*Corresponding author.

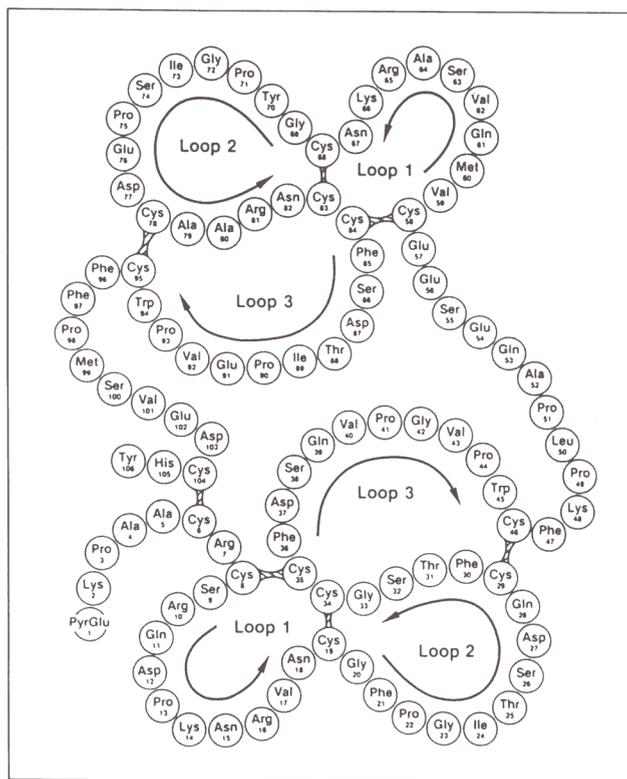


Fig. 1. Primary structure of porcine pancreatic spasmodic polypeptide, PSP [1] showing the two characteristic trefoil domains.

ton, then the two trefoil domains are connected by polypeptide chains formed by residues 47–57 and 96–106 (Fig. 1). The Cys6–Cys104 disulphide bridge closes the molecule as a ring and gives it a very compact shape. It also keeps the interconnecting chains in close contact with the two domains.

In the following description of secondary structure, the assignments are based on the algorithms of Kabsch and Sander [23]. Both the interconnecting chains are in an extended conformation with a short 3_{10} -helix at

residues 55–59 and a hydrogen bonded turn at residues 100–103.

When superimposing the C α coordinates of the four trefoil domains, a root mean standard deviation (rmsd) of 1.07 Å is obtained. However, only residues 5–9, 55–59 and residues 10–51, 59–100 are superimposed due to the extra residue of loop 1 of the amino-terminal domain. This alignment was determined using the LSQ-improved option of the O program [24]. The overall fold of the PSP molecule has approximate two-fold symmetry, as can be clearly seen in Fig. 3.

Arg10 forms a hydrogen bond between the two trefoil domains and also appears to be critical for the folding of the carboxy-terminal tail of PSP (see Table 1). The network of hydrogen bonds in which Arg10 participates is illustrated in Fig. 4. Additional hydrogen bonds involved in the folding of the carboxy-terminal tail are listed in Table 1.

Structure of the trefoil loop domain

Each domain contains three helices, one on each side of a short antiparallel β -sheet and one at the tip of loop 1 (note that loops 1, 2 and 3 are defined by the disulphide bridges, see Fig. 2). In the amino-terminal domain, the helices include residues 5–10 (α -helix A), 25–32 (α -helix B) and 12–16 (3_{10} -helix C), and the β -sheet comprises residues 35–37 and 45–47. In the carboxy-terminal trefoil domain, the helices include residues 55–59 (3_{10} -helix A'), 74–81 (α -helix B') and 61–65 (3_{10} -helix C'), and the β -sheet is composed of residues 84–86 and 94–96 (Fig. 2). Further, there are three hydrogen-bonded turns consisting of residues 21–24, 33–34 and 40–43.

The effect of the deletion in the carboxy-terminal domain after residue 58 is the conversion of the A helix from an α -helix to A', which is a shorter 3_{10} -helix. However, the two halves of the polypeptide chain superimpose very well from the positions of the C and C' helix. From Fig. 3b and Fig. 5 it can be seen that the A and B

Peptide	Domain number	10	20	30	40		
PSP	1	RCSRQDPK	NRVNC	GFPGIT	SDQCFTSGCCFDSQVPGVWCFK		
PSP	2	EC	VMQVSARK	NCGYPGIS	PEDCAARNCCFSDTIPEVWCF		
hSP	1	QCSRLS	PHNR	TNCG	PGITSDQCFDNGCCFDSSVTGVPWCFH		
hSP	2	QC	VMEVSDRR	NCGYPGIS	PEECASRKC	CFSNLIFEVWCF	
mSP	1	RCSRLT	PHNR	KNCG	FGITSEQC	FDLGC	CFDSSVAGVWCFH
mSP	2	QC	VMEVSARK	NCGYPGIS	PEDCASRNCCF	SNLIFEVWCF	
pS2		TC	TVAPRER	QNCG	PGVTPSQ	CANKGCCFDDTVRGVWCFY	
hITF		QC	AVPAKDR	VDCGYPHV	TPKEC	NRNGCCFDSRIPGVWCFK	
rITF		QC	MVPANVR	VDCGYPTVT	SEQC	NNRGCCFDSIPNVWCFK	
Consensus sequence		-C-----R	-CG-P-----C	-----CCF-----	V-----VPWCF-		
			D Y VS E		I		
Disulphide pairing		x	y	z	yx	z	
Loop number		1		2		3	
Kabsch-Sander		tHHHHhtgGGGg		tTTtHHHHHhTtEeE		tTTT eEE	

Fig. 2. Sequence alignment of known mammalian trefoil domains. Sequence numbering is based on the amino-terminal domain of PSP. Key to peptides: PSP, porcine pancreatic spasmodic polypeptide [1,8,9]; hSP, human spasmodic polypeptide [12]; mSP, mouse spasmodic polypeptide [12]; pS2, human breast cancer associated peptide [16,17]; hITF, human intestinal trefoil factor [15]; rITF, rat intestinal trefoil factor [14]. Disulphide pairing and loop numbers are according to Fig. 1. Also included is the Kabsch-Sander assignment [23] of the secondary structural elements found in the amino-terminal domain. (H= α -helix, G= 3_{10} -helix, E=extended strand participating in β -ladder, T=hydrogen-bonded turn. Small letters signify extension of secondary structural elements.

Table 1. Hydrogen bonds within molecule A of PSP.

H-bond donor	H-bond acceptor	Distance (Å)	Location
Arg7 O	Leu50 N	2.79	ATD
Arg10 Nη1	Glu56 O	2.60	inter-domain
Arg10 Nη2	Met99 O	2.87	inter-domain
Arg10 Nη2	Ser100 O	2.64	CTT
Arg10 Nη2	Asp103 Oδ1	2.74	CTT
Arg10 Nζ	Asp103 Oδ2	3.46	CTT
Gln11 Ne2	Lys48 O	2.84	ATD
Asp12 Oδ1	Lys14 Nζ	2.67	ATD
Asp12 Oδ2	Lys14 Nζ	2.50	ATD
Arg16 Nη1	Pro44 O	2.60	ATD
Arg16 Nη2	Asp37 Oδ2	3.35	ATD
Val17 O	Lys48 N	2.81	ATD
Thr25 N	Gln28 Oε1	2.83	ATD
Gln28 O	Ser32 Oγ	2.67	ATD
Val40 O	Trp45 Ne1	2.99	ATD
Arg65 Nη1	Pro93 O	2.92	CTD
Lys66 O	Phe97 N	2.86	CTD
Asn67 Nδ1	Cys68 N	3.19	CTD
Cys68 N	Cys95 O	2.71	CTD
Ser74 Oγ	Asp77 Oδ1	3.22	CTD
Ser74 Oγ	Glu76 Oε2	3.02	CTD
Ser74 Oγ	Asp77 N	2.78	CTD
Asp77 O	Arg81 Ne	3.16	CTD
Asp87 Oδ1	Thr88 N	3.17	CTD
Ile89 O	Trp94 Ne1	2.83	CTD
Met99 N	Glu57 O	2.62	CTT
His105 Nδ1	Tyr106 O	2.93	CTT

ATD, amino-terminal domain; CTD, carboxy-terminal domain; CTT, carboxy-terminal tail.

helices on each side of the β -sheet are approximately orthogonal to each other, with the B helix following the direction of the β -sheet. This assignment of secondary structure elements agrees with that made on the basis of NMR spectroscopic studies, except for the first helix in each domain, which was not assigned by NMR [25].

In addition to the hydrogen bonds associated with secondary structural elements, a number of hydro-

gen bonds are found in the amino-terminal domain (see Table 1). Asp12 and Lys14 clearly form a complex hydrogen-bonding network in this region. Arg16, which makes hydrogen bonds with Pro44 and Asp37 (Table 1), seems to be very important for the folding of the amino-terminal domain since it is conserved in all known members of the family of trefoil loop proteins and is positioned centrally within the domain. Arg16 is 'stacked' with Phe47 which is also a conserved residue. Other hydrogen bonds within this domain are listed in Table 1 along with those found in the carboxy-terminal domain. Arg81 is close to the molecular surface and hence is only involved in one intramolecular hydrogen bond. The water molecules included are all involved in hydrogen bonds to residues on the surface of the peptide (*i.e.* no structural waters are seen in the core of the molecule).

In each of the two domains, the six cysteine residues are disulphide linked in the configuration 1–5, 2–4 and 3–6, when the cysteines are numbered from the amino terminus. In the present structure, the assignment of the disulphide bonds is 8–35, 19–34, 29–46 in the amino-terminal domain and 58–84, 68–83, 78–95 in the carboxy-terminal domain. The seventh disulphide bond, between Cys6 and Cys104, confirms the secondary structure previously suggested (Fig. 1). The disulphide bonds in each of the two domains create the three characteristic peptide loops in each domain which give rise to the trefoil nomenclature [1]. The three-dimensional structure reveals that the three loops or 'leaves' are stacked in both domains (see Fig. 5). The secondary structural elements of the trefoil domain are generally short, leading to its very compact form.

Structure of the PSP dimer

In contrast to the solution state [25], PSP forms a dimer in the solid phase, with large solvent channels between the molecules (Fig. 6). There is a cleft between the two domains of PSP. The two molecules pack so that loop 3 of one of the molecules is located in the inter-domain cleft of the other. The two molecules in the

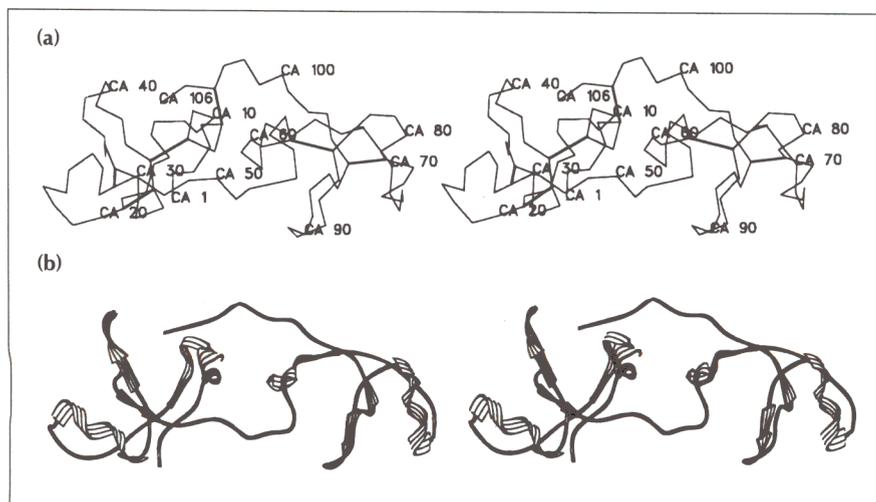


Fig. 3. Stereographic representations of pancreatic spasmodic polypeptide. (a) α backbone trace with every tenth amino acid residue numbered. (b) Richardson-type representation. The orientation clearly shows the pseudo two-fold symmetry of the molecule. (Both figures made with the program SETOR [32].)

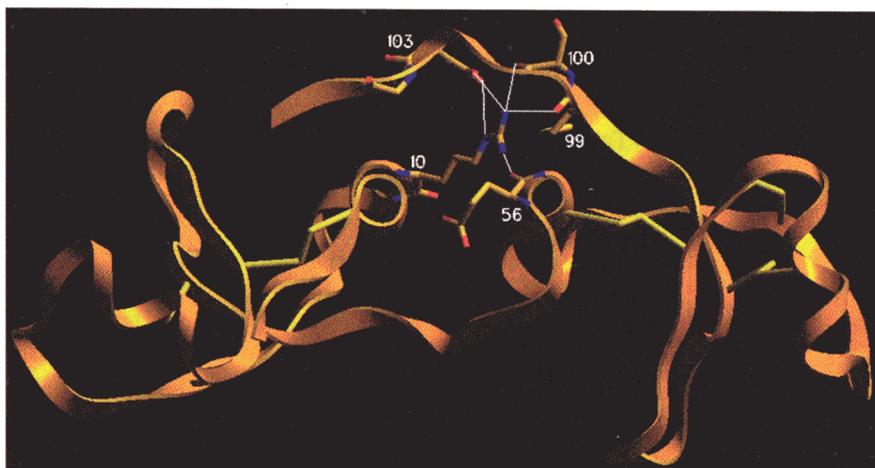


Fig. 4. Ribbon representation of the PSP molecule. The residues involved in hydrogen bonds with Arg10, which is important for the association of the carboxy-terminal tail with the rest of the molecule, are shown with hydrogen bonds in white.

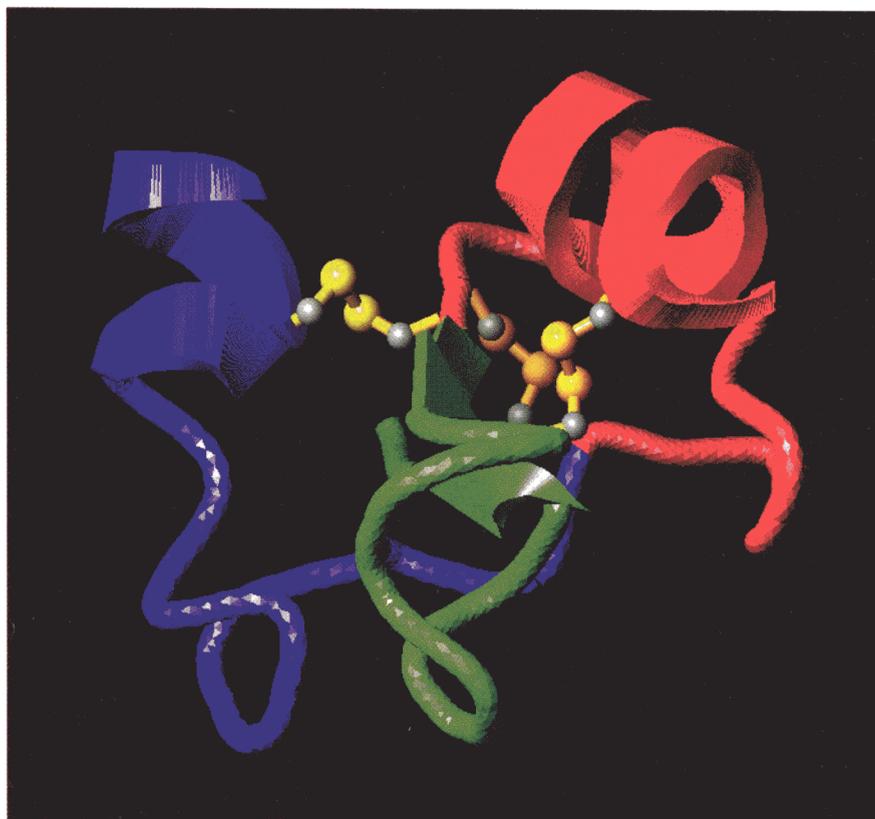


Fig. 5. View of a single (the amino-terminal) domain of PSP showing residues 5–48. Helix A and loop 1 are coloured purple, helix B and loop 2 are red and the β -sheet and loop 3 are green. Disulphide bridges are shown with sulphur coloured yellow and $C\beta$ grey. Note the continuation from loop 3 to the other domain passes through loop 1.

dimer are related by a non-crystallographic two-fold axis. This symmetry axis is in the plane orthogonal to the c-axis, in spherical polar angles at $\phi = 0.2^\circ$, $\psi = 108.5^\circ$ and $\kappa = 180.3^\circ$, using the conventions of the X-PLOR program [26]. This contradicts earlier findings by Gorman *et al.* [27] who found that the non-crystallographic two-fold axis forms an angle of 30° or 60° with the longest axis, based on investigations of the self-rotation function. The reason for this discrepancy is not clear.

There is a long intermolecular hydrogen bond between Asn15 N δ 2 in molecule A and Asn67 O δ 1 in molecule B (3.55 Å) and one between Asn67 O δ 1 in

molecule A and Asn15 N δ 2 in molecule B (3.18 Å). The side chains of residues involved in intermolecular hydrogen bonding are shown in Fig. 6. The number of hydrogen bonds (only two bonds) between the two molecules does not seem significant. However, the hydrophobic regions defined by Val62, Ile89 and Val92 of both molecules are in close contact in the dimer, forming a closed hydrophobic core. These residues could be important for dimerization. The positions of these hydrophobic residues are also shown in Fig. 6. The rmsd of alignment of the two molecules in the asymmetric unit is 0.51 Å for the $C\alpha$ atoms. Hence the two molecules in the asymmetric unit are essentially identical.

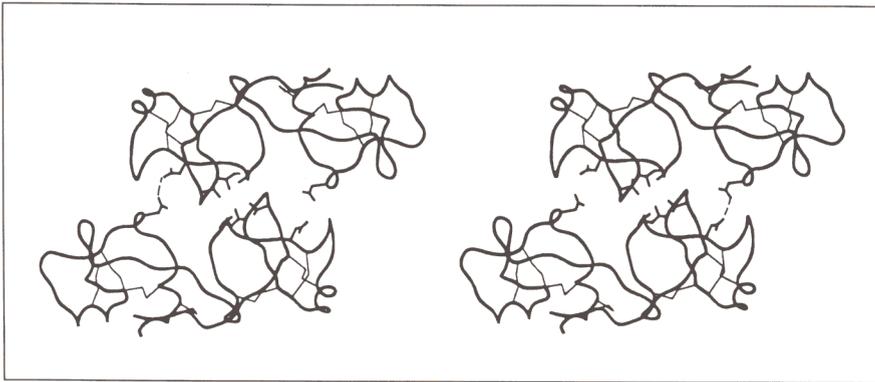


Fig. 6. A stereographic view of the dimer formed in the solid phase, showing the non-crystallographic two-fold symmetry. The side chains of residues forming a hydrophobic core and involved in intermolecular hydrogen bonding (with hydrogen bonds indicated by dashed lines) are shown. Disulphide bonds are shown as thinner lines. (Drawing made with SETOR [32].)

Resistance to degradation by proteases

PSP has been shown to be highly resistant to the proteases trypsin and chymotrypsin [6]. While trypsin normally cleaves peptides after arginine or lysine residues, chymotrypsin is less specific, and generally cleaves after hydrophobic residues. Since stability towards proteases appears to be a feature of all the members of the trefoil loop family of peptides, it can be assumed that the stability is related to the overall folding of the molecule. As can be seen from Fig. 3, the molecule is compact due to the short secondary structural elements and the large number of disulphide bridges. The Cys6–Cys104 disulphide bridge in particular, ensures that the molecule does not have easily accessible terminal ends. This combination of structural features probably accounts for the pronounced resistance towards proteases.

Inspection of the main chain B-factor plot (Fig. 7) reveals that the most flexible region of the PSP molecule occurs around residues 99–102, and Val101 seems to be an obvious site for hydrolysis by chymotrypsin. However, because the carboxy-terminal end is kept close to the two domains, this part of the PSP molecule cannot fit into the binding pockets of the proteases.

Possible interactions of trefoil peptides with mucins

Peptides and proteins containing trefoil domains have been isolated or cloned from a series of different organs and tissues from different species. In mammals, they appear to function mainly by association with the mucus layer of the gastrointestinal tract. In man, three trefoil peptides are known at present: hSP, ITF and pS2 (as described in the Introduction). Under normal physiological conditions, all three peptides are expressed in the gastrointestinal tract — hSP and pS2 in the epithelial mucosal layer of the stomach [12,18], and ITF in the epithelial mucosal layer of the small intestine and colon [15].

It is tempting to speculate that trefoil peptides may act to stabilize the mucus gel layer since they are co-expressed with mucin-type glycoproteins [28] in the human gastrointestinal tract. Such a stabilizing effect could be mediated by formation of complexes between the highly glycosylated mucin glycoproteins and trefoil peptides.

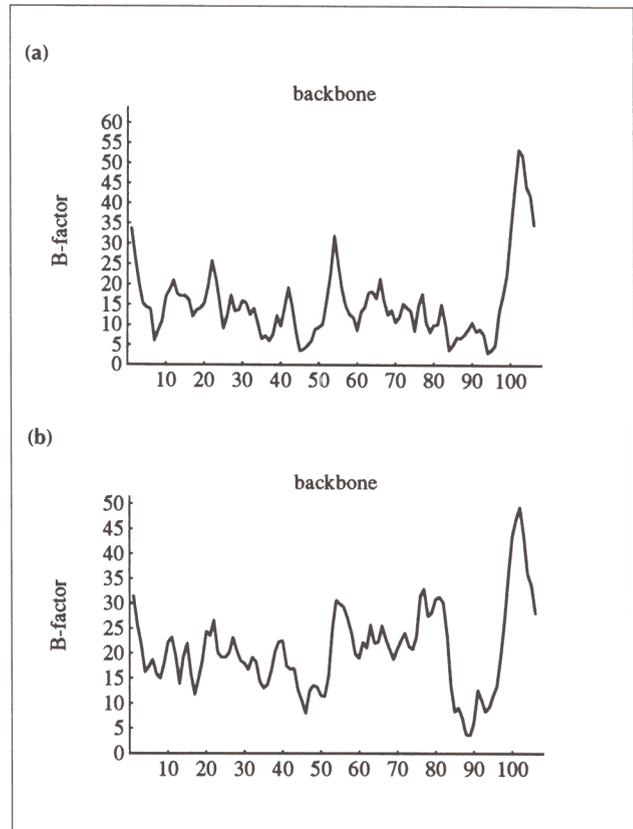


Fig. 7. Plot of average B-values for main-chain atoms in all residues. (a) Molecule A in the asymmetric unit. (b) Molecule B in the asymmetric unit.

Two different representations of the trefoil domain are shown in Fig. 8. They reveal a cleft in each domain between loops 2 and 3. The width of this cleft, as judged by the distance between C α s of Trp45 and Pro22, is 10.12 Å in molecule A and 9.10 Å in molecule B. This cleft is significantly narrower in the carboxy-terminal domain, where it is found to be only 8.41 Å in molecule A and 8.25 Å in molecule B. This means that the two grooves are not identical in size, but they are primarily built of the same residues. Further, the observed differences can be related to packing effects or to a different number of water molecules bound in the two grooves. The differences in cleft size observed in the four trefoil loop domains found in this crystal structure

indicate that the volume is flexible, and can adapt to a specific substrate. Fig. 8 also shows that all the conserved residues (see Fig. 2) are either placed around this groove, or form the core of the domain (Arg16, Val40, Val43 and Phe47). It is tempting to propose a role for this groove that is related to the function of the molecule. The groove is primarily formed from peptide backbone atoms. The only side chains pointing into the groove are the conserved residues Ile24, Phe36, Pro44 and Trp45. In contrast, all of the non-conserved residues in this region of the molecule point away from the groove.

As judged from a preliminary docking experiment, both clefts are deep enough to accommodate an oligosaccharide, or part of a longer polysaccharide chain. Pro44 and Trp45 could interact with the hydrophobic part of the carbohydrates and the potential also exists for hydrogen-bond formation to the hydroxyl groups in the cleft. A function of the trefoil domains could be to act as 'cross-linkers' of the polysaccharide chains of the mucin glycoproteins, and in this way increase the viscosity of the mucus layer. This proposed function of course requires at least two trefoil domains in a cross-linking unit. If PSP does dimerize *in vivo*, this would lead to four potential bonding sites (see Fig. 6) all oriented in different directions. We are currently investigating the validity of this cross-linking hypothesis by crystal soaking experiments using a number of different substrates.

Biological implications

Trefoil peptides are believed to contribute to the healing of peptic ulcers and other mucosal injuries by stabilizing the mucus layer of the intestinal tract. The mechanism of this stabilization is unknown. The X-ray structure of porcine pancreatic spasmolytic polypeptide (PSP), which

contains two trefoil domains, shows that the molecule is compact, due to extensive cross-linking, and has short secondary structural elements that are not sufficiently flexible to be accessible to protease active sites. This may explain the observed resistance of trefoil peptides to proteases, which is important for their survival in the gastric tract.

The trefoil fold is a new motif, distinct from the folds found in all other known protein structures, for example the 'kringle' domain of blood coagulation factors and the epidermal growth factor domain. The conserved residues Arg16 and Phe47 appear particularly important for the folding of the domain, stacking together to form the centre of the hydrophobic core.

The majority of the conserved residues contribute to a cleft, 8–10 Å wide, found in each of the trefoil domains. Preliminary docking experiments show that the cleft could accommodate part of an oligosaccharide chain, for example the carbohydrate attached to a mucin glycoprotein. If this is the case, since PSP contains two such clefts, it may cross-link mucins, helping them to form a protective gel over the mucosal epithelium.

Three trefoil peptides are known in man, which are specifically expressed either in the stomach (hSP and ITF) or in the small intestine and colon (pS2). hSP is the human counterpart of PSP, and, like PSP, has two trefoil domains. ITF and pS2, however, have only one trefoil domain; it seems plausible that they bind to mucins, but less likely that they cross-link them unless they dimerize *in vivo*. The mechanism of action of the single-domain trefoil peptides may therefore be distinct from that of the two-domain trefoil peptides.

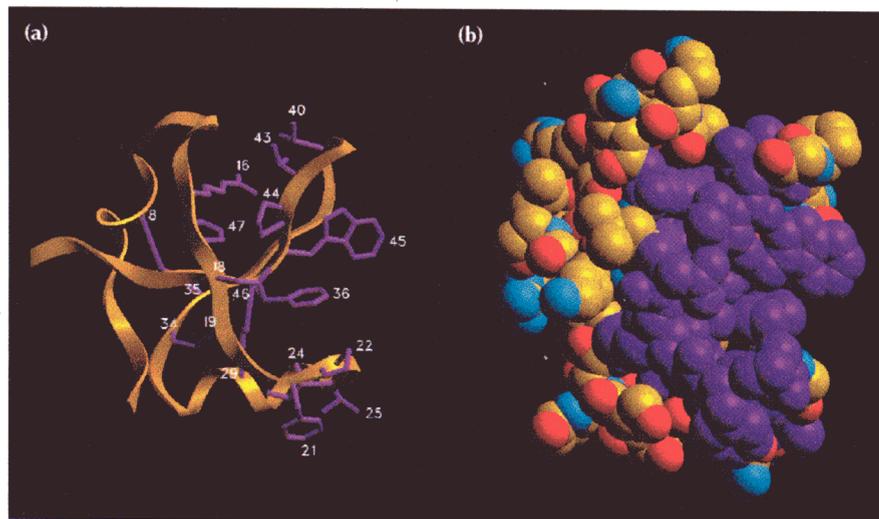


Fig. 8. The amino-terminal trefoil loop domain. (a) Ribbon representation. (b) Space-filling representation. (Atom colours: carbon, yellow; nitrogen, blue; oxygen, red.) All residues which are conserved in mammalian trefoil loop peptides are displayed and coloured purple.

Materials and methods

Purification, crystallization and data collection

Preliminary structure analysis by X-ray [27,29] as well as by NMR [25] techniques has previously been carried out on PSP. In the present study, PSP was purified [6] and crystallized [29] as previously described. Pancreatic spasmodic polypeptide crystallizes in space group $I2_12_12_1$. The crystal contains two molecules per asymmetric unit (referred to as molecule A and molecule B below) and 62% water.

All experimental data were collected (using monochromatic synchrotron radiation) at the EMBL outstation at DESY, Hamburg. The Hendrix Lentfer imaging plate scanner, constructed at EMBL Hamburg, was used as a detector. The native data were collected in two batches: (1) low order data ($>8\text{ \AA}$ resolution), 23 frames at a crystal-to-plate distance of 400 mm with 4° oscillations and (2) higher order data, 46 frames at a distance 230 mm with 2° oscillations. Only one crystal was used. Wavelength of synchrotron radiation was 1.009 \AA . All data were processed with the DENZO program (An oscillation data processing program for macromolecular crystallography, Yale University, CT, USA, 1986). Further data processing was performed with the CCP4 [SERC (UK), Collaborative Computer Project No.4, Daresbury Laboratory, UK, 1979] and PHASES [W Furey (1991). A program package for phasing in macromolecular crystallography, University of Pittsburg, PA, USA] program packages. The program AGROVATA (from the CCP4 package) was used to analyze the data and to perform the averaging of equivalent measurements. In the averaging, individual reflections with a deviation of more than 3 standard deviations from the mean were rejected until at least 3 measurements remained. The unit cell parameters, determined from the data processing, were $a=54.7\text{ \AA}$, $b=72.6\text{ \AA}$ and $c=181.5\text{ \AA}$. The standard deviations derived from counting statistics were modified according to the following scheme:

$$\sigma_{\text{new}} = k\sigma_{\text{counting}}(I) + pI^2$$

where $k=1.7$ and $p=0.0$ since the standard deviations found from counting statistics were generally found to be too low compared with the statistical standard deviations.

The crystals for a Pt^{2+} derivative were obtained by adding 1 mM K_2PtCl_4 to the mother liquor and soaking the crystals for 1 week. A wavelength of 0.9901 \AA was used for collecting data on heavy-atom derivatives. The diffraction data were collected in 46 frames, with oscillations of 2° and a crystal-to-film distance of 230 mm. The cell dimensions of the derivatized crystals were altered to $a=53.5\text{ \AA}$, $b=72.9\text{ \AA}$ and $c=181.7\text{ \AA}$, indicating non-isomorphism with the native crystal. In addition, the crystals showed a rapid decay in the beam; the scale factor of the last frame was determined as 0.55 with a relative B-factor of -1.77 \AA^2 . A scaled $\langle F_{\text{PH}} \rangle / \langle F_{\text{P}} \rangle$ plot against resolution indicated that the Pt^{2+} derivative was only isomorphous to 4 \AA since the slope became significantly different from zero at this resolution.

An Ag^+ derivative was made by soaking crystals for 2 days in the mother liquor, made 5 mM with respect to AgNO_3 . The same data collection parameters were used as for the Pt^{2+} derivative. These crystals showed no signs of non-isomorphism.

Table 2 contains a summary of the data collection statistics. Only isomorphous contributions were used from both derivatives. The resolution cutoff criterion for all three data sets was that the internal R-factor should not exceed 0.20 in the outermost resolution shell.

Structure determination by Multiple Isomorphous Replacement (MIR)

Phases were derived by MIR using the two heavy-atom derivatives. Scaling each of the derivatives to the native data was ac-

Table 2. Summary of data collection statistics.

	Native	Ag	Pt
Resolution (\AA)	2.5	2.8	3.0
Number of reflections	46 452	30 195	25 628
Number rejected	232	285	184
Number of unique reflections	12 661	9 190	4 718
Overall completeness (%)	98.3	99.5	99.5
Outermost shell (\AA)	2.5–2.63	2.80–2.95	3.00–3.16
Outermost shell completeness (%)	98.6	100	100
$1/\sigma(I) > 2$ (%)	89.3	91.0	89.9
$1/\sigma(I) > 2$ outermost shell (%)	76.0	80.4	76.8
R_i (%)	5.9	5.5	6.1

Resolution is determined as the highest order shell where the internal R-factor, $R_i = \sum_{\text{hkl}} \sum_i |I_i(\text{hkl}) - \langle I(\text{hkl}) \rangle| / \sum_{\text{hkl}} \sum_i I_i(\text{hkl})$ is <0.20 . Number of reflections is the total number of reflections within the resolution limit and with a partiality above 0.5. Number rejected is the number of reflections rejected in the averaging procedure. Number of unique reflections is the number of unique reflections after averaging. Overall completeness is the percentage of theoretically obtainable data measured. Outermost shell is the highest order data referred to below. $1/\sigma(I) > 2$ is the number of reflections with intensity greater than 2 standard deviations. These data were used in subsequent refinements.

complished with the program CMBISO and the derived difference Patterson maps were analyzed with the program MAPVIEW (both from the phasing package by W Furey).

The Ag^+ derivative showed peaks in the Harker sections corresponding to two heavy atom sites. These peaks could not be seen in the anomalous Patterson map probably due to the low f'' at the applied wavelength. The two sites were related to a common origin by standard difference Fourier techniques. Phase information to 3.0 \AA resolution was extracted from this derivative.

For the Pt^{2+} derivative a difference Patterson map was calculated to 3.0 \AA resolution. This map showed only one site. A cross difference Fourier, using the phases determined from the Ag^+ derivative, using data up to 4 \AA resolution, verified the site. Subsequent processing was done with programs from the CCP4 suite. The positions of the heavy atoms were refined with the program HEAVY. Further phase refinement and phase calculation was done with the program MLPHARE. Table 3 lists the heavy atom positions determined. In fact, only one binding site was found for both derivatives. However, the site was occupied by Ag^+ in both molecules comprising the asymmetric unit, but Pt^{2+} only complexed with molecule A. Both metal ions form a complex with the Met60 and Met99 side chains which are in close proximity, and located on the surface of the peptide.

Table 3. Heavy atom positions.

Derivative	Position			B-factor (\AA^2)	Occupancy (%)	Binding site
	x	y	z			
Ag 1	0.236	0.069	0.139	27	0.39	Met A60, Met A99
2	0.147	0.281	0.199	41	0.45	Met B60, Met B99
Pt 1	0.222	0.057	0.138	33	0.40	Met A60, Met A99

Later analysis of the crystal packing clearly showed that the Met60 and Met99 positions were much more readily accessible from the solvent channels in molecule A than in molecule B. The refinement and phasing procedure is summarized in Table 4.

Table 4. Heavy atom refinement and phasing statistics.

	Ag derivative	Pt derivative
Resolution limit	3.0	3.0
R_{cullis} (centric reflections)	0.64	0.83
R_{cullis} (acentric reflections)	0.39	0.85
Phasing power (centric reflections)	1.0	0.7
Phasing power (acentric reflections)	1.4	1.0
FOM (centric)		0.67
FOM (acentric)		0.44

Resolution limit is the highest order of data used for each derivative.
 $R_{\text{cullis}} = \sum_{\text{hkl}} | |F_{\text{PH}} \pm F_{\text{P}}| - F_{\text{H}(\text{calc})} | / \sum_{\text{hkl}} |F_{\text{PH}} - F_{\text{H}}|$. The phasing power is defined as: $\sum_{\text{hkl}} F_{\text{H}} / \sum_{\text{hkl}} |F_{\text{PH}(\text{obs})} - F_{\text{PH}(\text{calc})}|$. FOM is the phasing figure of merit.

The electron density map produced from the MIR data showed well defined large solvent channels in the structure, but it was not possible to trace the polypeptide chain. This is probably related to the low phasing power of the Pt^{2+} derivative.

The program SQUASH [30] was used for solvent flattening and phase extension from 4 to 3 Å. Histogram matching, automatic Wang solvent mask generation and Sayre equation techniques were employed. This improved the overall figure of merit to 0.75. The resulting map was of high quality with clear continuity in the electron density (see Fig. 9). The remarkable success of the procedure is probably related to the large and well defined solvent content of the crystals.

The initial model was built with the program O [24]. Both molecules in the asymmetric unit could be traced and no averaging procedures were used.

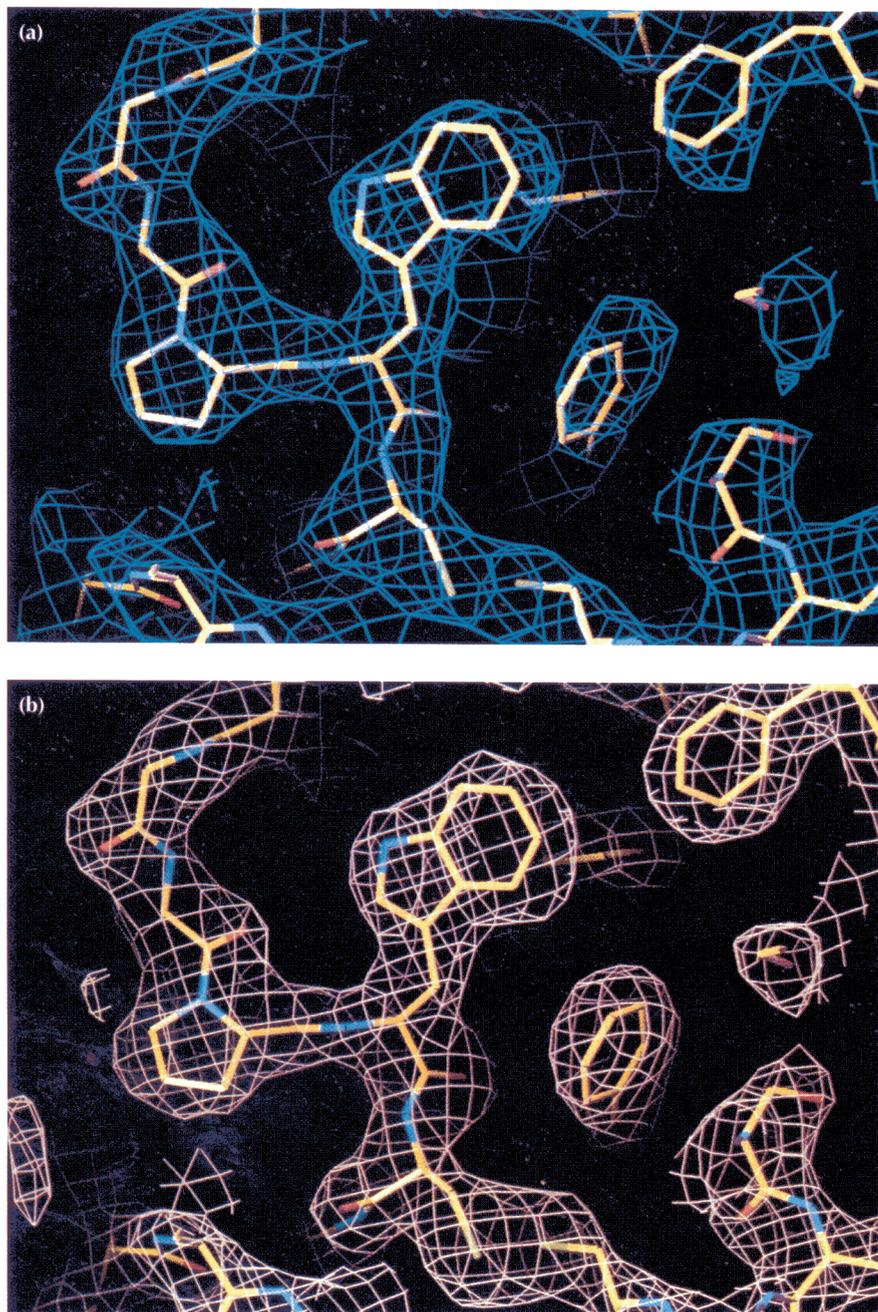


Fig. 9. Views of the electron density of the calculated maps in the region around Trp45. (a) The SQUASH-treated MIR map at 3 Å resolution contoured at the 1.0 rmsd level. (b) $2F_o - F_c$ map contoured at 1.3 rmsd level.

Refinement of the structure

The initial model was refined by simulated annealing using the program X-PLOR [26]. Data in the range 8–2.5 Å were used in the refinement, only including reflections for which $I > 2\sigma(I)$. This excluded 1449 out of a total of 12 348 reflections. The standard slow cool protocol was used in the refinement, starting initially at 3000 K and performing 50 steps of molecular dynamics (time step 0.5 fs) then reducing the temperature by 25 K and repeating until the temperature reached 300 K. In later refinement calculations, the initial temperature was set at 800 K.

A total of 23 refinement cycles were conducted with alternating X-PLOR calculations and manual density fitting. The final restricted R-value is 0.20 with inclusion of 88 water molecules, all with B-factors below 60 Å², returning $2F_o - F_c$ density and in proximity of the protein molecules. The corresponding free R-value is 0.30 [26] using 10 % of the data as a test set. When all measured reflections are included in the calculation of the restricted R-value, a value of 0.22 is obtained. Table 5 shows a breakdown of R-values on resolution. The average deviation from ideality in bond lengths is 0.023 Å and in bond angles is 2.3°.

Table 5. R-factors as a function of resolution.

Resolution (Å)	NREF free	R-value free	R-value free cum.	NREF rest.	R-value rest.	R-value rest. cum.
4.69–8.00	164	0.3170	0.3170	1331	0.2183	0.2183
3.85–4.69	143	0.2808	0.2988	1339	0.1663	0.1908
3.41–3.85	146	0.2717	0.2901	1332	0.1769	0.1862
3.12–3.41	146	0.2806	0.2883	1311	0.1904	0.1871
2.91–3.12	139	0.3231	0.2926	1208	0.2180	0.1910
2.74–2.91	149	0.2924	0.2926	1179	0.2271	0.1946
2.61–2.74	118	0.3644	0.2977	1080	0.2544	0.1990
2.50–2.61	118	0.3575	0.3012	996	0.2678	0.2030

NREF free is the number of reflections included in the calculation of R-free. $R = \sum_{hkl} |F_o - F_c| / \sum F_o$. R-value free cum. is the cumulative R_{free} . NREF rest. is the number of reflections included in the calculation of the restricted R-value.

The Ramachandran plot (Fig. 10) shows no residues, other than glycines or prolines, outside favoured or allowed regions. The plot includes both molecules in the asymmetric unit. 82 % of the residues are in the most favoured regions, the rest are in favoured regions.

The real space correlation coefficients were examined as a function of residue number, as implemented in program O [31]. A plot of these correlation coefficients is shown in Fig. 11. The function is computed by calculating grid sums in a series of density maps calculated from the model coordinates of a selected group of atoms from each residue. This function can reveal problem areas in a structure. The plot shows no signs of errors in the model and values are as expected for a structure at this resolution. However, it does reveal that the density around residues 99–105 is not as well defined as the rest of the density in either of the molecules. As can be seen from Fig. 1, this part of the molecule is not part of the domain structure and is probably quite flexible.

The average B-value, including both molecules in the asymmetric unit, was 20.3 Å² after the final cycle of refinement. A plot of the average B-values of the main-chain atoms for all residues is shown in Fig. 7. This plot also reveals that the region around 99–105 is probably disordered, and indicates some disorder around the carboxy-terminal end and around Glu54 which is in

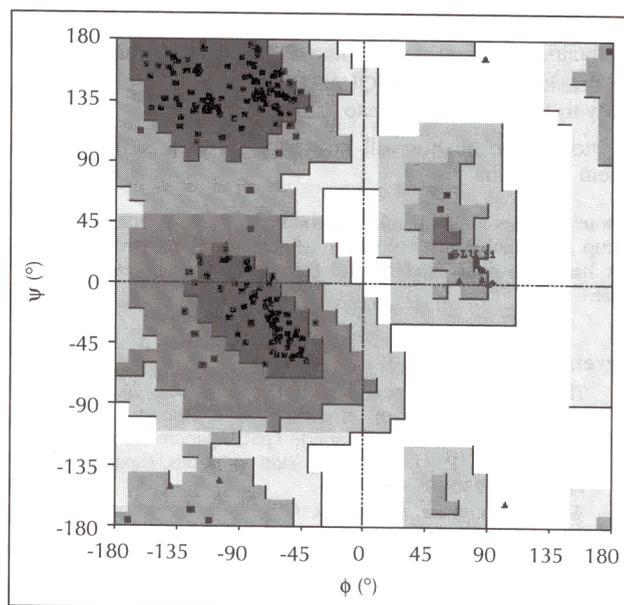


Fig. 10. Ramachandran plot including both molecules in the asymmetric unit. The plot is produced with program PROCHECK [33]. Levels, indicated by black, dark grey, medium grey and light grey shading, represent regions which are most favoured, additional allowed, generously allowed and disallowed, respectively. This classification is based on analysis of 118 high quality X-ray structures.

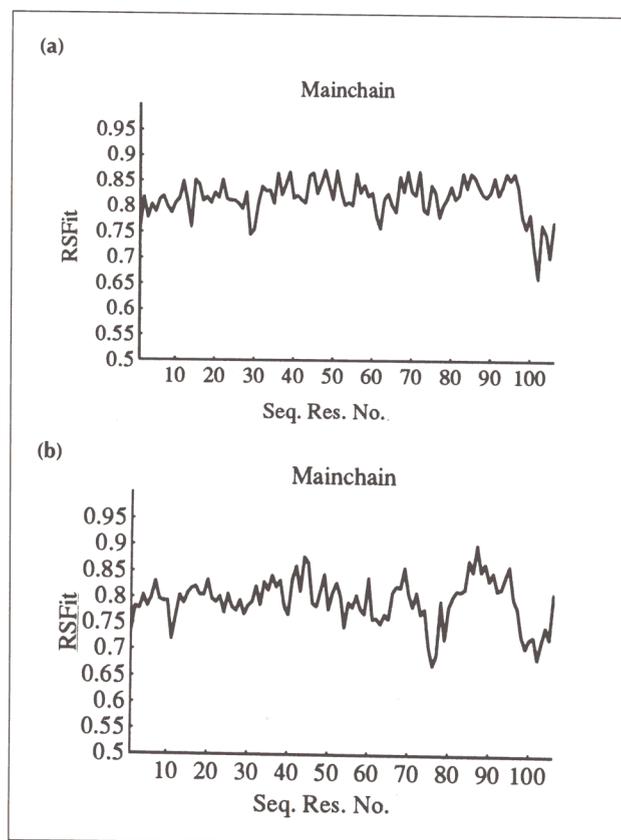


Fig. 11. Plot of the real space R-value versus residue number (a) for molecule A and (b) for molecule B in the asymmetric unit [31].

the region linking the two domains. Fig. 11 shows the quality of the obtained electron density maps. The quality of the MIR electron density map after the SQUASH procedure is of comparable quality to the final $2F_o - F_c$ map.

The atomic coordinates will be deposited in the Brookhaven protein data bank.

Acknowledgements We thank Anders Kadziola and Jette Sandholm Kastrup for valuable discussion on refinement of the structure. This work has been supported by the Danish Signale Peptide Research Centre.

References

1. 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.
2. Wright, N.A., Pike, C. & Elia, G. (1990). Induction of a novel epidermal growth factor-secreting cell lineage by mucosal ulceration in human gastrointestinal stem cells. *Nature*, **343**, 82–85.
3. Wright, N.A., *et al.*, & Chambon, P. (1990). Epidermal growth factor (EFG/URO) induces expression of regulatory peptides in damaged human gastrointestinal tissues. *J. Pathol.* **162**, 279–284.
4. Rio, M.-C., *et al.*, & Chambon, P. (1991). Induction of pS2 and hSP genes as markers of mucosal ulceration of the digestive tract. *Gastroenterology*, **100**, 375–379.
5. Wright, N.A., *et al.*, & Chambon, P. (1993). Trefoil peptide gene expression in gastrointestinal epithelial cells in inflammatory bowel disease. *Gastroenterology*, **104**, 12–20.
6. Jørgensen, K.H., Thim, L. & Jacobsen, H.E. (1982). Pancreatic spasmolytic polypeptide (PSP): I. Preparation and initial chemical characterization of a new polypeptide from porcine pancreas. *Regul. Pept.* **3**, 207–219.
7. Thim, L., Jørgensen, K.H. & Jørgensen, K.D. (1983). Pancreatic spasmolytic polypeptide (PSP): II. Radioimmunological determination of PSP in porcine tissues, plasma and pancreatic juice. *Regul. Pept.* **3**, 221–230.
8. Thim, L., Thomsen, J., Christensen, M. & Jørgensen, K.H. (1985). The amino acid sequence of pancreatic spasmolytic polypeptide. *Biochem. Biophys. Acta* **827**, 410–418.
9. Rose, K., Savoy, L.-A., Thim, L., Christensen, M. & 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. *Biochem. Biophys. Acta*, **998**, 297–300.
10. Rasmussen, T.N., Harling, H., Thim, L., Pierznowsky, S., Westström, B.R. & Host, J.J. (1993). Regulation of secretion of pancreatic spasmolytic polypeptide from the porcine pancreas. *Am. J. Physiol.* **264**, G22–G29.
11. Jørgensen, K.D., Diamant, B., Jørgensen, K.H. & Thim, L. (1982). Pancreatic spasmolytic polypeptide (PSP): III. Pharmacology of a new porcine pancreatic polypeptide with spasmolytic and gastric acid secretion inhibitory effects. *Regul. Pept.* **3**, 231–243.
12. Tomasetto, C., *et al.*, & Lathe, R. (1990). hSP, the domain-duplicated homolog of pS2 protein, is co-expressed with pS2 in stomach but not in breast carcinoma. *EMBO J.* **9**, 407–414.
13. Thim, L., *et al.*, & Petersen, J. (1993). Purification and characterization of the trefoil peptide human spasmolytic polypeptide (hSP) produced in yeast. *FEBS Lett.* **318**, 345–352.
14. Suemori, S., Lynch-Devaney, K. & Podolsky, D.K. (1991). Identification and characterization of rat intestinal trefoil factor: tissue and cell-specific member of the trefoil protein family. *Proc. Natl. Acad. Sci. USA*, **88**, 11017–11021.
15. Podolsky, D.K., *et al.*, & Mahida, Y.R. (1993). Identification of human intestinal trefoil factor. Goblet cell-specific expression of a peptide targeted for apical secretion. *J. Biol. Chem.* **268**, 6694–6702.
16. Jakowlew, S.B., Breathnach, R., Jeltsch, J.-M., Masiakowski, P. & Chambon, P. (1984). Sequence of the pS2 mRNA induced by estrogen in the human breast cancer cell line MCF-7. *Nucleic Acids Res.* **12**, 1861–1878.
17. Prud'homme, J.-F., *et al.*, & Milgrons, E. (1985). Cloning of a gene expressed in human breast cancer and regulated by estrogen in MCF-7 cells. *DNA*, **4**, 11–21.
18. Rio, M.-C., *et al.* & Chambon, P. (1988). Breast cancer-associated pS2 protein: synthesis and secretion by normal stomach mucosa. *Science*, **241**, 705–708.
19. Hauser, F. & Hoffmann, W. (1991). xP1 and xP4. P-domain peptides expressed in *Xenopus laevis* stomach mucosa. *J. Biol. Chem.* **266**, 21306–21309.
20. Hauser, F., Roeben, C. & Hoffmann, W. (1992). xP2, a new member of the P-domain peptide family of potential growth factors, is synthesized in *Xenopus laevis* skin. *J. Biol. Chem.* **267**, 14451–14451.
21. Hauser, F. & Hoffmann, W. (1992). P-domains as shuffled cysteine-rich modules in integumentary mucin C.1 (FIM-C.1) from *Xenopus laevis*. *J. Biol. Chem.* **267**, 24620–24624.
22. Bork, M. (1993). A trefoil domain in the major rabbit *zona pellucida* protein. *Protein Sci.* **2**, 669–670.
23. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.
24. Jones, T.A., Bergdoll, M. & Kjeldgaard, M. (1990). In *Crystrallography and Modelling Methods in Molecular Design*. pp. 189–199. Springer, New York.
25. Carr, M.D. (1992). ^1H NMR-based determination of the secondary structure of porcine pancreatic spasmolytic polypeptide. *Biochemistry*, **31**, 1998–2004.
26. Brunger, A.T. (1993). In *X-PLOR Version 3.1*. Yale University, New Haven, CT.
27. Gorman, M.A., De, A. & Freemont, P.S. (1992). Crystallization and preliminary X-ray diffraction studies of pancreatic spasmolytic polypeptide. *J. Mol. Biol.* **228**, 991–994.
28. Strous, G.J. & Dekker, J. (1992). Mucin-type glycoproteins. *Crit. Rev. Biochem. Mol. Biol.* **27**, 57–92.
29. Gajhede, M., Thim, L., Jørgensen, K.H. & Melberg, S.G. (1992). Pancreatic spasmolytic polypeptide: crystallization, circular dichroism analysis and preliminary X-ray diffraction studies. *Proteins*, **13**, 364–368.
30. Zhang, K.Y.J. & Main, P. (1990). Histogram matching as a new density modification technique for phase refinement and extension of protein phases. *Acta Crystallogr. A*, **46**, 41–45.
31. Jones, A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these maps. *Acta Crystallogr. A*, **47**, 110–119.
32. Evans, S.V. (1993). SETOR: hardware lighted three-dimensional solid model representation of macromolecules. *J. Mol. Graph.*, in press.
33. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: a program to check the stereochemistry of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.

Received: 7 Oct 1993; revisions requested: 18 Oct 1993; revisions received: 25 Oct 1993. Accepted: 27 Oct 1993.