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 spasmolytic 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 Å 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 Å. All data were procesed 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 Å, b = 72.6 Å and c = 181.5 Å. The standard deviations derived from counting statistics were modified according to the following scheme:

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

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 Å 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 Å, b = 72.9 Å and c = 181.7 Å, 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 Å^2. A scaled $<F_{PH}>/<F_P>$ plot against resolution indicated that the Pt^{2+} derivative was only isomorphous to 4 Å 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.

Table 2. Summary of data collection statistics.

	Native	Ag	Pt
Resolution (Å)	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 (Å)	2.5–2.63	2.80–2.95	3.00–3.16
Outermost shell completeness (%)	98.6	100	100
$I/\sigma(I) > 2$ (%)	89.3	91.0	89.9
$I/\sigma(I) > 2$ 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, $R_i = \Sigma_{hkl}\Sigma_i|I_i(hkl) - <I(hkl)>|/\Sigma_{hkl}\Sigma_i I_i(hkl)$ 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. $I/\sigma(I) > 2$ is the number of reflections with intensity greater than 2 standard deviations. These data were used in subsequent refinements.

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 accomplished 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 Å resolution was extracted from this derivative.

For the Pt^{2+} derivative a difference Patterson map was calculated to 3.0 Å resolution. This map showed only one site. A cross difference Fourier, using the phases determined from the Ag^+ derivative, using data up to 4 Å 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 x	y	z	B-factor (Å^2)	Occupancy (%)	Binding site
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. $R_{cullis} = \Sigma_{hkl} ||F_{PH} \pm F_P| - F_{H(calc)}|/\Sigma_{hkl} |F_{PH} - F_H|$. The phasing power is defined as: $\Sigma_{hkl} F_H/\Sigma_{hkl} |F_{PH(obs)} - F_{PH(calc)}|$. 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.

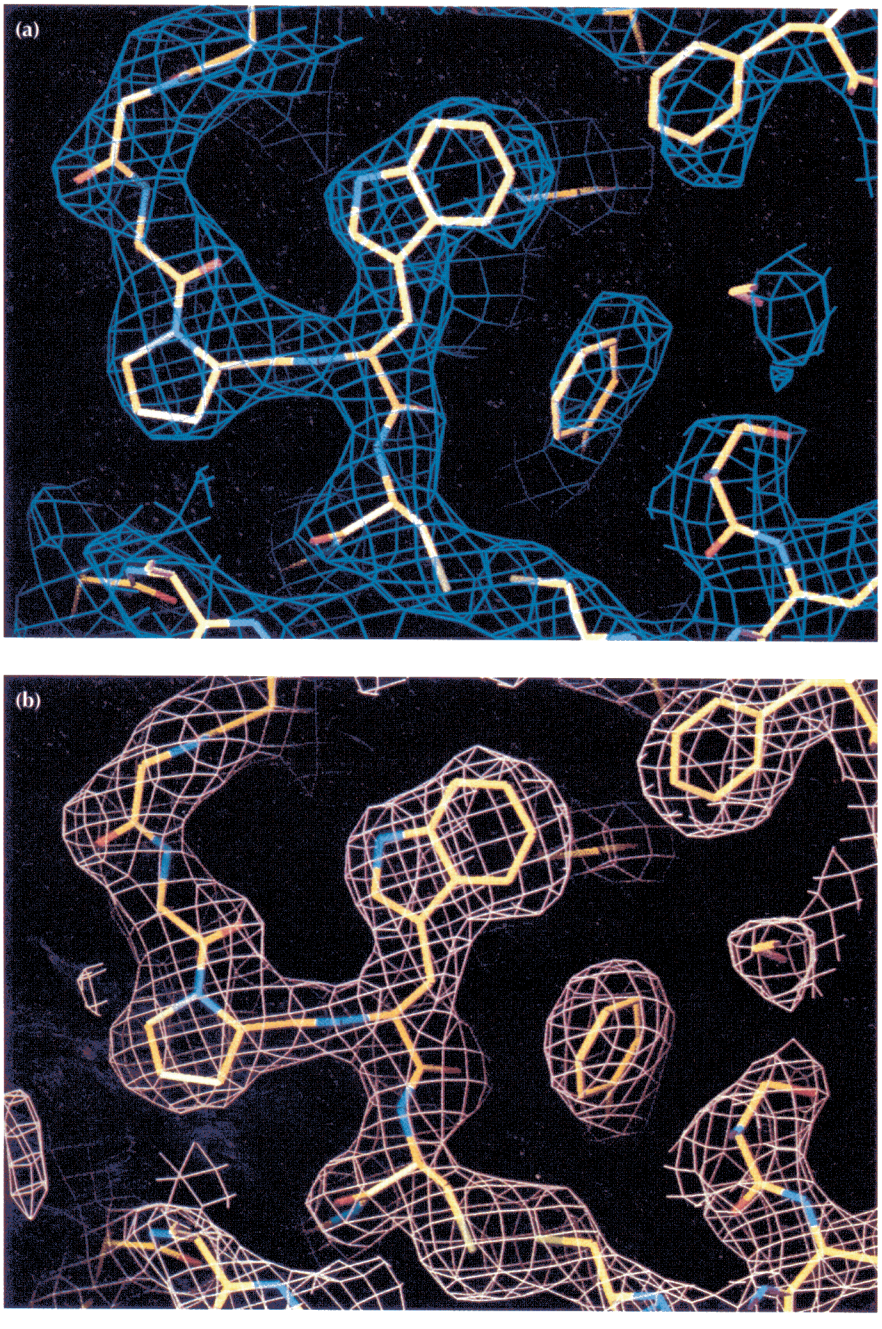

Fig. 9. Views of the electron density of the calculated maps in the region around Trp45. **(a)** The SQUASH-treated MIR map at 3 Å resolution contoured at the 1.0 rmsd level. **(b)** $2F_o$–F_c map contoured at 1.3 rmsd level.

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 Å^2, 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. $R = \Sigma_{hkl}|F_O - F_C|/\Sigma F_O$. 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 Å^2 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

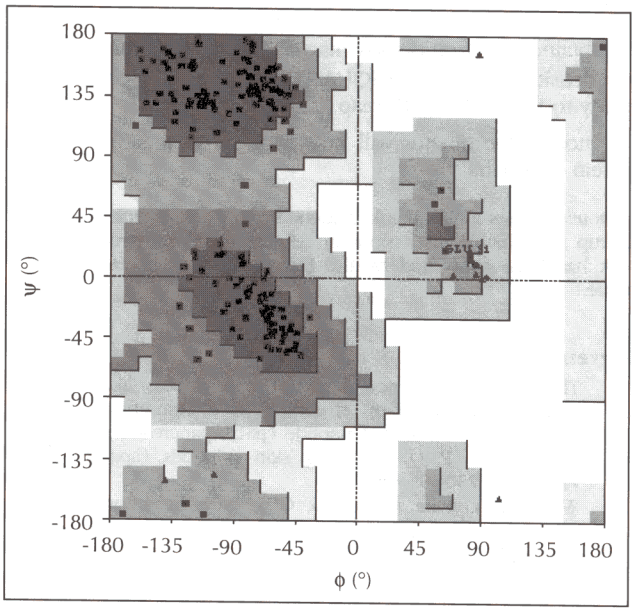

Fig. 10. Ramachandran plot including both molecules in the asymmetric unit. The plot is produced with program PROCHECK [33]. Levels, indicated by black, dark grey, medium grey and light grey shading, represent regions which are most favoured, additional allowed, generously allowed and disallowed, respectively.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

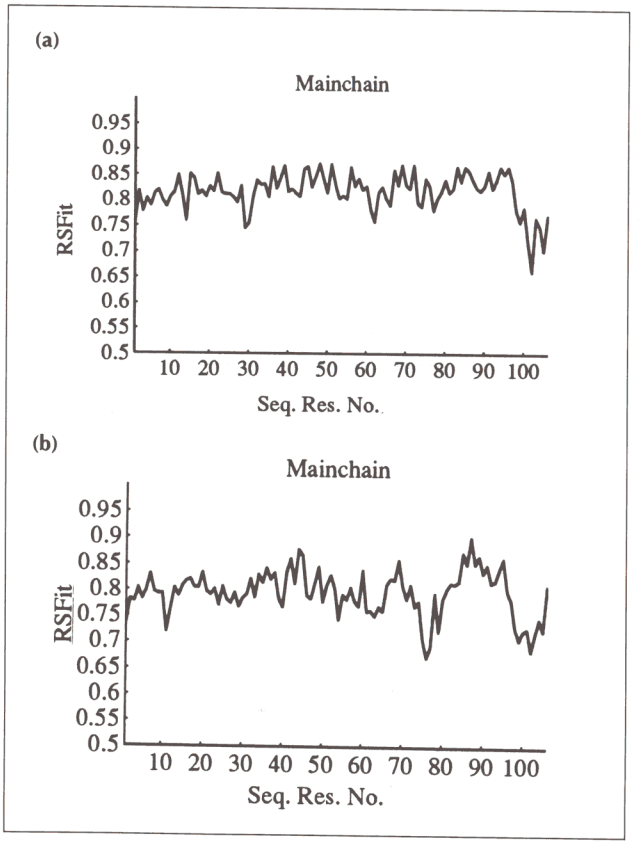

Fig. 11. Plot of the real space R-value *versus* residue number **(a)** for molecule A and **(b)** for molecule B in the asymmetric unit [31].

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., Weström, 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.* & Chamson, 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 *Crystallography and Modelling Methods in Molecular Design*. pp. 189-199. Springer, New York.
25. Carr, M.D. (1992). ^{1}H 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**.