Volume 318, number 3, 345–352 FEBS 12208 March 1993

Purification and characterization of the trefoil peptide human spasmolytic polypeptide (hSP) produced in yeast

Lars Thim, Kjeld Norris, Fanny Norris, Per F. Nielsen, Søren E. Bjørn, Mogens Christensen and Jørgen Petersen

Novo Nordisk, Novo Alle, DK-2880 Bagsvaerd, Denmark

Received 12 January 1993

Recombinant human spasmolytic polypeptide (r-hSP) has been produced in relatively large amounts in *Saccharomyces cerevisiae*. The two intronless trefoil domains of the hSP-DNA were cloned separately by PCR from human genomic DNA, and the remaining parts of the gene synthezised. Recombinant plasmids were constructed to encode a fusion protein consisting of a hybrid leader sequence and the hSP sequence. The leader sequence serves to direct the fusion protein into the secretory pathway of the cell and to expose it to the Kex 2 processing enzyme system. The secreted r-hSP was found in a glycosylated and an non-glycosylated form. The two forms of r-hSP were purified from the yeast fermentation broth by a combination of ion-exchange chromatography and preparative HPLC. The overall yield from 8 litres of fermentation broth was 160 mg r-hSP and 219 mg glycosylated r-hSP corresponding to 50% and 34%, respectively. The structure of the r-hSP and the glycosylated r-hSP was determined by amino acid analysis and carbohydrate composition analysis as well as by peptide mapping, amino acid sequencing and mass spectrometric analysis.

Pancreatic spasmolytic polypeptide; Spasmolytic polypeptide; Intestinal trefoil factor; Intestinal trefoil factor; Breast cancer associated peptide; Breast cancer associated peptide (pS2); *N*-Glycosylation; Electro-spray mass spectrometry

1. INTRODUCTION

Human spasmolytic polypeptide (hSP) belongs to a growing family of peptides containing one or more characteristic trefoil domains [1]. The trefoil domain is made up by a sequence of 38 or 39 amino acid residues in which 6 cysteine residues are linked in the configuration 1–5, 2–4 and 3–6 thus forming a characteristic three-leaved structure [1]. The trefoil family of peptides consists of rat intestinal trefoil factor, ITF [2], human breast cancer associated peptide, pS2 [3–5], spasmolytic polypeptide from pig, human and mouse (PSP, hSP, mSP) [6–8] and frog spasmolysins (xP1, xP2 and xP4) [8,10,11] all containing 1, 2 or 4 trefoil domains (Fig. 1).

The physiological function of the trefoil peptides is poorly understood, and so far only PSP has been studied in some detail. In the porcine pancreas, PSP is found in the acinar cells and is secreted in large amounts (50–100 mg/ml) into the pancreatic juice upon stimulation with pancreozymin or secretin [12–14]. PSP is resistant towards digestion by intestinal proteases within the gastrointestinal tract [12], and specific binding of PSP to rat intestinal mucosa cells and membrane preparations from these cells have been demonstrated [15,16]. In the porcine gastrointestinal tract, specific receptor-like binding to Paneth cells in the duodenum has been found [17]. These results suggest a unique intraluminal function of the peptide. A pharmacological screening has indicated that PSP has spasmolytic and gastric acid secretion inhibitory effects [18], and studies on mammalian cells have indicated a growth factor-like activity of PSP [19].

In contrast to porcine PSP, the human counterpart, hSP (Fig. 2), has been found to be expressed in the stomach, but not in the pancreas to any great extent [8]. An increased expression of hSP and pS2 has been reported in association with peptic ulcers and mucosal injury in inflammatory bowel disease [20,21] indicating a possible healing function of these peptides.

Only very limited amounts of hSP can be prepared from human tissue extraction, and in the present study, we describe a yeast expression system designed for the production of sufficient amounts of hSP for physiological and biochemical studies of the peptide.

Correspondence address: L. Thim, Department of Protein Chemistry, Pharmaceuticals Research, Novo Nordisk, Novo Alle, DK-2880 Bagsvaerd, Denmark. Fax: (45) 4444-4565.

2. MATERIALS AND METHODS

2.1. *General methods*

Standard DNA techniques were used as previously described [29]. Synthetic oligonucleotides were prepared on an automatic DNA synthesizer (380B, Applied Biosystems) using commercially available reagents. DNA sequence determinations were performed by the dideoxy chain-termination technique [30]. Polymerase chain reactions (PCR) were performed on a DNA Thermal Cycler (Perkin Elmer Cetus) using a commercial kit (GeneAmp, Perkin Elmer Cetus).

2.2. *PCR cloning of hSP*

The first trefoil domain of hSP was isolated by a PCR reaction in which 1 μg human genomic DNA (Clontech, Palo Alto, CA, USA) was used as a template. The reaction mixture contained 100 pmol each of the forward primer 1 (GGCTGAGCCCCCATAACAG) and reverse primer 2 (TGGAAACACCAGGGGAC) (Fig. 3) and was carried out in a 100 μl volume. The cycle was: 94°C for 1 min, 50°C for 1 min, and 72°C for 1 min. After 30 cycles a final cycle was performed in which the 72°C step was maintained for 10 min. The PCR product, a 115 bp fragment, was isolated by electrophoresis on a 2% agarose gel.

The 115 bp PCR fragment was digested with *Dde*I and then ligated to a 31 bp duplex formed from the oligonucleotides (GAGAAACCCTCCCCCTGCCAGTGCTCCAGGC) and (TCAGCCTGGAGCACTGGCAGGGGGAGGGTTTCTC). The ligation product was amplified by PCR using forward primer 3 (GCTGAGAGATTGGAGAAGAGAGAGAAACCCTCCCCCT) and reverse primer 2. The 3′ part of primer 3 is identical to the N-terminal encoding part of the hSP gene and the 5′ part of primer 3 is identical to the C-terminal encoding part of the hybrid leader gene (Fig. 3).

In-frame fusion of the hybrid leader gene and the first trefoil domain from hSP was obtained by overlay extension PCR [31]. The product was digested with *Eco*RI and *Ava*II and isolated as a 360 bp DNA fragment.

The second trefoil domain of hSP was PCR-cloned from human genomic DNA as described for the first domain by replacing primers 1 and 2 with forward primer 4 (TGCGTCATGGAGGTCTC) and reverse primer 5 (AGCACCATGGCACTTCAAAG) (Fig. 3). Reverse primer 5 introduces a *Nco*I site as a silent mutation. The PCR product, a 115 bp fragment, was isolated and digested with *Dde*I and *Nco*I resulting in a 91 bp fragment. To this fragment was ligated two synthetic duplexes. The first, encoding the amino acid sequence between the two trefoil domains, consisted of the oligonucleotides (GTCCCCTGGTGTTTCCACCCCCTCCCAAAGCAAGAGTCGGATCAGTGCGTCATGGAGGTC) and (TGAGACCTCCATGACGCACTGATCCGACTCTTGCTTTGGGAGGGGGTGGAAACACCAGGG). The second, a 46 bp *Nco*I-*Xba*I fragment encoding the C-terminal part of hSP, consisted of the oligonucleotides (CATGGTGCTTCTTCCCGAACTCTGTGGAAGACTGCCATTACTAAGT) and (CTAGACTTAGTAATGGCAGTCTTCCACAGAGTTCGGGAAGAAGCAC). After *Ava*II digestion a 195 bp *Ava*II-*Xba*I fragment was isolated.

A DNA construct encoding the hybrid leader fused in-frame to the entire hSP gene was obtained by ligation of the 360 bp *Eco*RI-*Ava*II fragment and the 195 bp *Ava*II-*Xba*I fragment described above to the 2.7 kb *Eco*RI-*Xba*I fragment from vector pTZ19R [32]. This construct was then transformed into *E. coli* strain MT-172 (r^-, m^+) by selection for resistance to ampicillin. DNA sequencing of the resulting plasmid, KFN-1843, showed that the correct construction had been obtained.

Trefoil family of peptides

Peptide		Number of trefoil domains	Species
Intestinal Trefoil Factor	(ITF)		Rat Human
Breast cancer associated peptide	(pS2)		Human
Spasmolytic polypeptide	(hSP) (PSP) (mSP)		Human Porcine Mouse
Spasmolysins	(xP1)		Frog
	(xP4)		Frog

Fig. 1. Trefoil family of peptides. Intestinal trefoil factor (ITF) contains one trefoil domain [2], as does the breast cancer associated pS2 peptide [3,4]. The spasmolytic polypeptides from man, pig and mouse contain two trefoil domains [1,8]. Spasmolysins from *Xenopus laevis* contain one or four trefoil domains [10]. Recently, a member of the frog trefoil family containing two domains has been described [11].

2.3. *Construction of the hSP secreting yeast strain*

Plasmid KFN-1843 described above was digested with *Eco*RI and *Xba*I. The resulting 558 bp fragment was isolated and ligated to the 9.3 kb *Nco*I-*Xba*I fragment and the 1.6 kb *Nco*I-*Eco*RI fragment both from the yeast expression vector pMT-636. Plasmid pMT-636 is derived from the *S cerevisiae–E coli* shuttle vector CPOT [25,33] by deletion of the 0.4 kb *Hpa*I-*Nru*I fragment from the Leu-2 gene. The ligation mixture was transformed into *E. coli* strain MT-172, and the hSP expression plasmid, KFN-1847, was isolated (Fig 4). Plasmid pKFN-1847 was transformed into *S cerevisiae* strain MT-663 by selection for growth on glucose as the sole carbon source. One transformant, KFN-1852, was selected for fermentation.

2.4. *Fermentation*

The transformant described above was cultivated at 30°C for 3 days in yeast peptone dextrose (YPD) medium [40] supplied with additional yeast extract (60 g/l). An OD 650 nm value of 52 was reached at the end of the fermentation.

2.5. *Purification of r-hSP*

The concentration of r-hSP in the yeast fermentation broth and fractions obtained during the purification was measured by analytical HPLC. Aliquots (usually 50–200 μl) were injected onto a Vydac 214TP54 reverse-phase C4 HPLC column (0.46 × 25 cm) equilibrated at 30°C at a flow rate of 1.5 ml/min with 0.1% (v/v) TFA in 5% (v/v) acetonitrile. The concentration of acetonitrile in the eluting solvent was raised to 65% (v/v) over 30 min. Absorbance was measured at 280 nm. The peaks eluting at 15.6 min. and 16.1 min. (Fig. 5) was found by mass spectrometry analysis to represent glycosylated r-hSP and unglycosylated r-hSP, respectively. The peptides were quantified using a calibrated PSP sample as standard as both peptides contain two Trp and two Tyr out of 106 amino acid residues.

From a 10 l fermentor, 8 l of fermentation broth was isolated by centrifugation at 3,000 rpm for 10 min. The supernatant was concentrated to 0.9 l using an Amicon ultrafiltration unit (RA 2000) equipped with an Amicon spiral ultrafiltration cartridge type S1Y3, MW cutoff 3,000 (Product No. 540620). The pH was adjusted to 1.7 and the conductivity in the resulting concentrated sample was measured to 4.7 mS.

The sample was pumped onto a Fast Flow S-Sepharose (Pharmacia) column (5 × 11 cm) with a flow rate of 40 ml/h. Previous to the application, the column was equilibrated in 50 mM formic acid buffer, pH 3.7. After application of the sample, the column was washed with 500 ml of 50 mM formic acid buffer, pH 3.7. The peptides were eluted from the column by a linear gradient between 1.5 l of 50 mM formic acid buffer, pH 3.7, and 1.5 l of 50 mM formic acid buffer, pH 3.7, containing 0.6 M NaCl. Fractions of 10 ml were collected at a flow rate of 40 ml/h and the absorbance was measured at 280 nm. Fractions were assayed for the content of r-hSP and glycosylated r-hSP in the HPLC-system as previously described. The elution profile is shown in Fig. 6. Fractions corresponding to r-hSP (fract. nos. 107–128) and glycosylated r-hSP (fract. nos. 78–95), respectively, were pooled.

Glycosylated r-hSP and r-hSP were further purified by preparative HPLC chromatography. Pooled fractions (approx. 200 ml) were pumped onto a Vydac 214TP1022 C4 column (2.2 × 25 cm) equilibrated in 0.1% (v/v) TFA. The column was washed with 100 ml of 0.1% (v/v) TFA in 10% (v/v) MeCN. The peptides were eluted at 25°C and at a flow rate of 5 ml/min with a linear gradient (650 ml) formed from MeCN/H_2O/TFA (10.0 : 89.9 : 0.1 v/v/v) and MeCN/H_2O/TFA (60.0 : 39.9 : 0.1 v/v/v). UV-absorption was monitored at 280 nm, and fractions corresponding to 10 ml were collected and analysed for the content of r-hSP or glycosylated r-hSP. Fig. 7 shows the preparative HPLC purification of r-hSP (Fig. 7A) and glycosylated r-hSP (Fig. 7B). Fractions corresponding to the bars were pooled, and the volume

reduced to 30% by vacuum centrifugation. From the two resulting pools, r-hSP and glycosylated r-hSP were isolated by lyophilization.

2.6. *Characterization of r-hSP and glycosylated r-hSP*

Amino acid composition analysis were carried out by hydrolysis of 50 μg peptide with 6 M HCl for 24 h at 110°C as previously described [6]; no correction for loss during hydrolysis was carried out. Amino acid sequence analysis was determined by automated Edman degradation using an Applied Biosystems Model 470A gas-phase sequencer [22]. Carbohydrate composition analysis was carried out by hydrolysis of 50 μg peptide with 2 M HCl for 1 h, 2 h and 4 h at 100°C and monosaccharides were separated on a CarboPac PAI (Dionex, Sunnyvale, CA) column (4 × 250 mm) eluted with 14 mM NaOH. The monosaccharides were detected by pulsed amperometric detection (Dionex PAD-detector). The amount of monosaccharides was corrected to zero time of hydrolysis and calculated as nmol of monosaccharide per nmol of peptide.

Mass spectrometry analysis was performed using an API III LC/MS/MS system (Sciex, Thornhill, Ont., Canada). The triple quadrupole instrument has a mass-to-charge (m/z) range of 2400 and is fitted with a pneumatically assisted electrospray (also referred to as ion-spray) interface [23,24]. Sample introduction was done by a syringe infusion pump (Sage Instruments, Cambridge, MA) through a fused capillary (75 μm i.d.) with a liquid flow-rate set at 0.5–1 μl/min. The instrument m/z scale was calibrated with the singly-charged ammonium adduct ions of poly(propylene glycols) (PPGs) under unit resolution.

The accuracy of mass measurements is generally better than 0.02%.

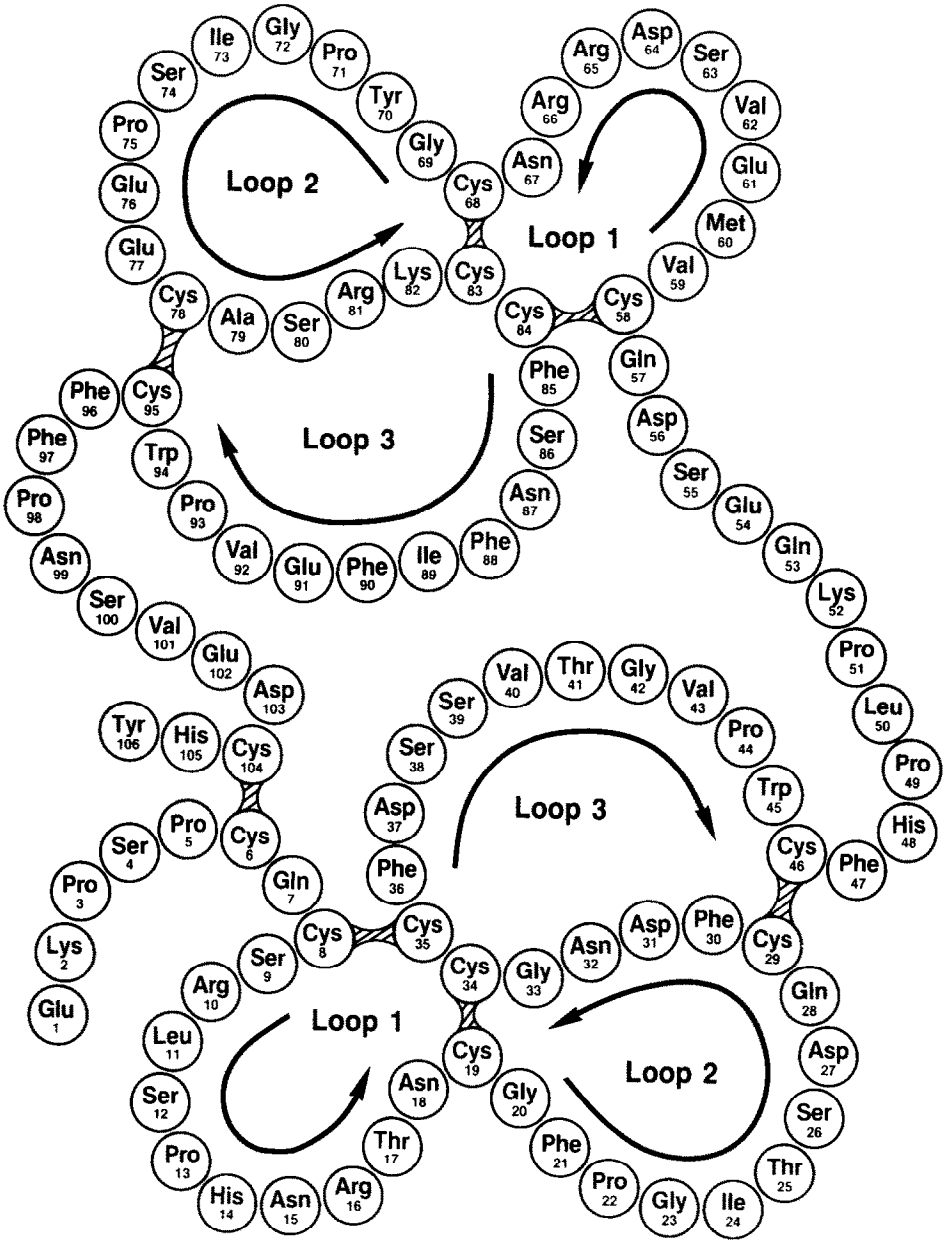

Fig. 2. Proposed structure of human spasmolytic polypeptide, hSP. The primary amino acid sequence is taken from Tomasetto et al. [8], and the disulphide bonds are placed in homology to PSP [1].

```
EcoRI
GAATTCCATTCAAGAATAGTTCAAACAAGAAGATTACAAACTATCAATTTCATACACAAT

ATAAACGACCAAAAGAATGAAGGCTGTTTTCTTGGTTTTGTCCTTGATCGGATTCTGCTG
                  M  K  A  V  F  L  V  L  S  L  I  G  F  C  W

GGCCCAACCAGTCACTGGCGATGAATCATCTGTTGAGATTCCGGAAGAGTCTCTGATCAT
 A  Q  P  V  T  G  D  E  S  S  V  E  I  P  E  E  S  L  I  I

CGCTGAAAACACCACTTTGGCTAACGTCGCCATGGCTGAGAGATTGGAGAAGAGAgagaa
 A  E  N  T  T  L  A  N  V  A  M  A  E  R  L  E  K  R ↑E  K
                      Dde I   1 →
accctccccctgccagtgctccaggcTGAGCCCCCATAACAGGACGAACTGCGGCTTCCC
 P  S  P  C  Q  C  S  R  L  S  P  H  N  R  T  N  C  G  F  P

TGGAATCACCAGTGACCAGTGTTTTGACAATGGATGCTGTTTCGACTCCAGTGTCACTGG
 G  I  T  S  D  Q  C  F  D  N  G  C  C  F  D  S  S  V  T  G
AvaII  ← 2                                              4 →
Ggtcccctggtgtttccaccccctcccaaagcaagagtcggatcagtgcgtcatggaggt
 V  P  W  C  F  H  P  L  P  K  Q  E  S  D  Q  C  V  M  E  V
Dde I
cTCAGACAGAAGAAACTGTGGCTACCCGGGCATCAGCCCCGAGGAATGCGCCTCTCGGAA
 S  D  R  R  N  C  G  Y  P  G  I  S  P  E  E  C  A  S  R  K
                        ← 5    Nco I
GTGCTGCTTCTCCAACTTCATCTTTGAAGTGCCatggtgcttcttcccgaactctgtgga
 C  C  F  S  N  F  I  F  E  V  P  W  C  F  F  P  N  S  V  E
                 Xba I
agactgccattactaagtCTAGA
 D  C  H  Y
```

Fig. 3. Nucleotide sequence and corresponding amino acid sequence of the 563 bp *Eco*RI-*Xba*I fragment encoding the leader – hSP fusion protein. The Kex 2 processing site is indicated by a vertical arrow. The leader and the PCR cloned parts of the hSP gene are shown in capital letters, while the synthetic parts are shown in small letters. The underlined sequences correspond to the PCR primers with horizontal arrows indicating the direction. Restriction sites relevant for the construction are shown.

3. RESULTS

3.1. *Expression and purification*

DNA fragments encoding the two trefoil domains of hSP were isolated by PCR from human genomic DNA using primers based on the published cDNA sequence [8]. The full-length hSP gene was obtained from the PCR cloned fragments by addition of synthetic DNA fragments. The hSP gene was fused in-frame to a hybrid yeast leader sequence by overlay extension PCR [31] (Fig. 3). The hybrid leader is based on the mouse salivary amylase signal peptide [34] and the *S. kluyveri* α mating factor leader sequence [35] and is further modified near the Kex 2 cleavage site for efficient processing [36,41].

hSP expression vector

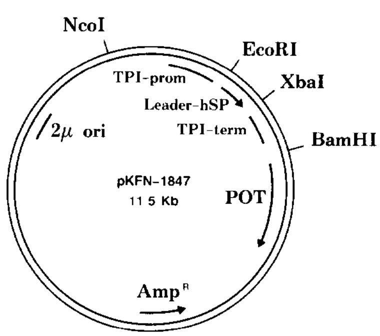

Fig. 4. *S cerevisiae* plasmid for the expression and secretion of hSP. TPI-prom. and TPI-term. are *S. cerevisiae* triosephosphate isomerase transcription promoter and terminator sequences, respectively. POT is a selective marker, the *Schizosaccharomyces pombe* triosephosphate isomerase gene. Only restriction sites relevant for the construction of the plasmid have been indicated.

The yeast expression plasmid pKFN-1847 contains the leader-hSP gene inserted between the *S. cerevisiae* triosephosphate isomerase promoter and terminator [37]. The expression vector (Fig. 4) also contains the *Schizosaccharomyces pombe* TPI gene (POT) [38].

The plasmid was transformed into the yeast strain MT-663, carrying a deletion in the TPI gene, by selecting for growth on glucose.

The expression level of r-hSP in the present yeast system is approx. 120 mg/l. This is 10 times higher than what we have previously obtained with different insulin precursors in a similar system [25]. However, the present yeast cells were grown on a complex medium in a fermentor as compared to the insulin precursor producing yeast cells which were grown on a minimal medium in shake flasks. As can be seen from Fig. 5, the yeast supernatant contains two forms of r-hSP; one eluting at R_t = 15.6 min. and one eluting at R_t = 16.1 min. These two forms were purified separately, and by using the analytical HPLC-system (Fig. 5), these two forms can be quantified individually during the different steps of the purification.

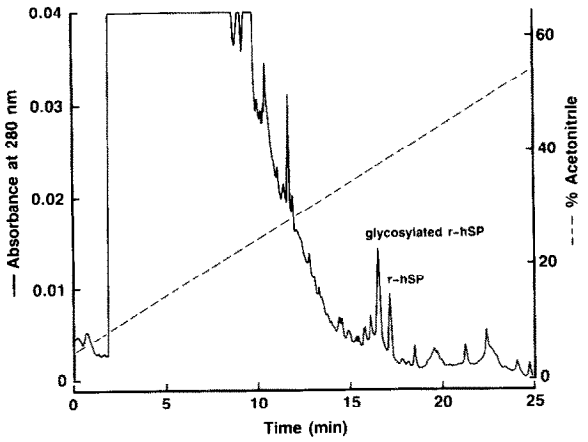

Fig. 5. Reversed-phase HPLC on a Vydac 214TP54 column of yeast fermentation broth. The two peaks corresponding to r-hSP and glycosylated r-hSP are indicated. The dashed line shows the concentration of acetonitrile in the eluting solvent.

After the initial concentration of the yeast supernatant by ultrafiltration, the first purification step was cationic exchange chromatography on a Fast Flow S column. Fig. 6 shows the elution profile from the column including the amount of r-hSP and glycosylated r-hSP determined in the fractions. A complete separation of the two forms of r-hSP was obtained in this step.

The fractions from the Fast Flows S column were pooled as indicated in Fig. 6, and the two peptides were further purified by preparative HPLC (Fig. 7). The r-hSP and glycosylated r-hSP were recovered from the fractions indicated in Fig. 7A and B by vacuum centrifugation and lyophilization. The purification is summarized in Table I. The overall yield of r-hSP and glycosylated r-hSP from 8 litres of fermentation broth was 160 mg and 219 mg corresponding to 50% and 34%, respectively.

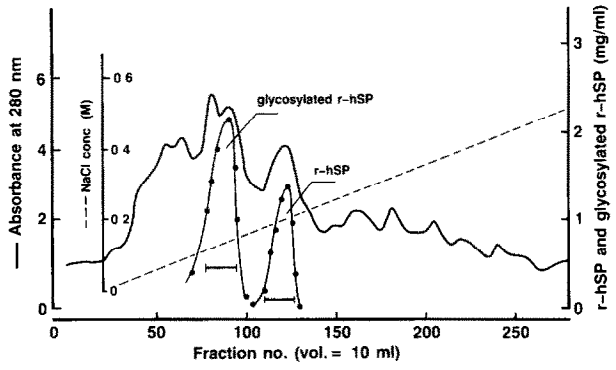

Fig. 6. Ion exchance chromatography on a Fast Flow S column of concentrated yeast supernatant. The amount of r-hSP and glycosylated r-hSP were determined by the use of the HPLC system shown in Fig. 5. The bars indicate the fractions pooled for further purification of r-hSP and glycosylated r-hSP. The dashed line shows the concentration of NaCl in the eluting solvent. For details, see Materials and Methods.

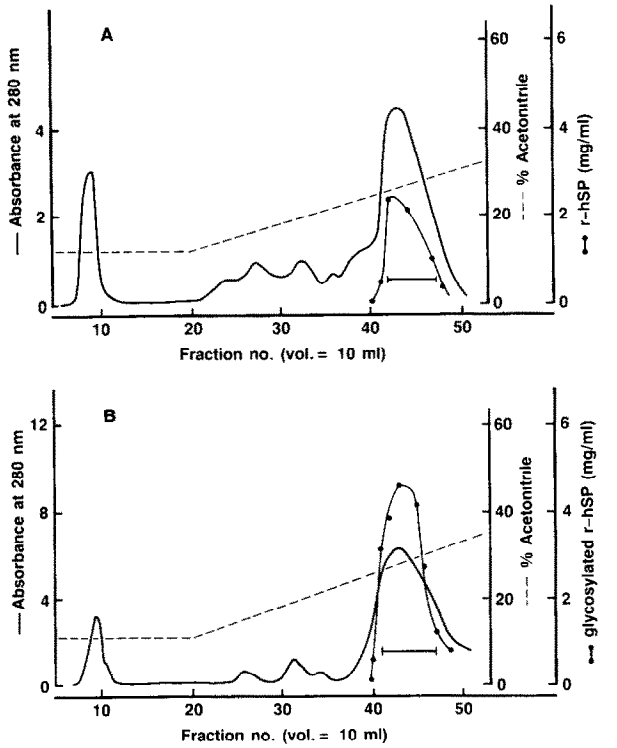

Fig. 7. Final purification of r-hSP (A) and glycosylated r-hSP (B) on a preparative reversed-phase HPLC Vydac 214TP1022 column. The bars indicate the fractions pooled for lyophilization. The dashed lines show the concentration of acetonitrile in the eluting solvent. For details, see Materials and Methods.

3.2. *Characterization of r-hSP and glycosylated r-hSP*

Fig. 8 shows the purity of r-hSP and glycosylated r-hSP as analysed by analytical HPLC. From these results none of the peptides looks completely pure. However, upon rechromatography of material eluting in the minor as well as the major peak, similar chromatograms were obtained for both peptides (results not shown). This seems to indicate that the double peak observed for both r-hSP and glycosylated r-hSP reflects an atypical behaviour of these peptides on reverse phase columns rather than impurities in the preparations. We have previously observed a similar behaviour of highly purified porcine PSP on reverse phase columns (L. Thim and K.H. Jørgensen, unpublished observations).

Table II shows the amino acid sequencing results obtained on r-hSP and glycosylated r-hSP. The average repetitive yield was 94.4% (r-hSP) and 94.6% (glycosylated r-hSP), respectively. In both cases the first 40 residues of the two peptides were confirmed by the sequence analysis. In the glycosylated hSP, no PTH-amino acids was found in Edman degradation cycle No. 15. The hSP sequence from residue 15–17 (Asn-Arg-Thr) corresponds to a classical consensus sequence for *N*-glycosylation of Asn-15.

The carbohydrate composition analysis of glysocylated r-hSP showed the presence of 12.8 nmol mannose

Table I

Purification of r-hSP and glycosylated r-hSP from yeast supernatant

Step	Volume (ml)	Amount (mg)		Yield (%)	
		r-hSP	Glycosylated r-hSP	r-hSP	Glycosylated r-hSP
Yeast supernatant	8000	320	640	100	100
Ultrafiltration	900	207	405	65	63
Ion exchange chromatography					
Pool 1	160		275		43
Pool 2	220	182		57	
Prep HPLC					
Pool 1	54		219		34
Pool 2	80	160		50	

(Man) and 1.6 nmol of *N*-acetyl glucoseamine (GlcNAc) per nmol of r-hSP. By peptide mapping of r-hSP and glycosylated r-hSP in combination with mass spectrometry and sequencing analysis (results not shown), no other residue besides Asn-15 of the glycosylated r-hSP was found to be modified, i.e. no *O*-glycosylation was found.

In Fig. 9, the electro-spray mass spectrometry (ESMS) analysis is shown for r-hSP and glycosylated r-hSP. Fig. 9A and C are original mass spectra displaying characteristics series of multiply charged protonated ions always observed in ESMS spectra of proteins. Fig. 9B and D are the corresponding computer reconstructed mass spectra from which the molecular weight of individual components may be read directly. As can be seen from Fig. 9B, the MW found for r-hSP is 11 961.5 ± 2 which is in very good agreement with a calculated mass of 11 961.3. Fig. 9D shows the reconstructed ion spray mass spectrum of the glycosylated r-hSP. From the sequence analysis and the carbohydrate composition analysis, it is known that only Asn-

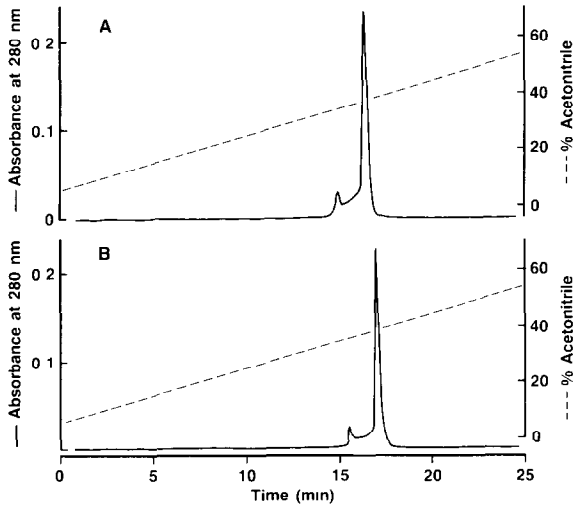

Fig. 8. Reversed-phase HPLC on a Vydac 214TP54 column of purified, glycosylated r-hSP (A) and r-hSP (B). The dashed lines show the concentration of acetonitrile in the eluting solvent.

Table II

Amino acid sequence analysis of r-hSP and glycosylated r-hSP

Cycle no.	PTH-amino acid	Yield (pmol)	
		r-hSP	Glycosylated r-hSP
1	Glu	4304	8853
2	Lys	6925	8292
3	Pro	6027	12837
4	Ser	2890	5602
5	Pro	4336	8802
6	(Cys)	ND	ND
7	Gln	3388	5689
8	(Cys)	ND	ND
9	Ser	1279	2417
10	Arg	1876	2523
11	Leu	2277	4290
12	Ser	877	1790
13	Pro	1545	2963
14	His	517	574
15	Asn	1202	0*
16	Arg	959	1471
17	Thr	978	2172
18	Asn	1066	1509
19	(Cys)	ND	ND
20	Gly	836	1857
21	Phe	993	1958
22	Pro	843	1839
23	Gly	785	2049
24	Ile	640	1400
25	Thr	589	1454
26	Ser	274	621
27	Asp	581	1391
28	Gln	445	952
29	(Cys)	ND	ND
30	Phe	623	1562
31	Asp	483	1210
32	Asn	369	823
33	Gly	359	885
34	(Cys)	ND	ND
35	(Cys)	ND	ND
36	Phe	422	1094
37	Asp	268	783
38	Ser	127	324
39	Ser	145	394
40	Val	298	827

ND = not determined. * = no trace of PTH-Asn or PTH-Asp was seen in cycle No. 15 of glycosylated r-hSP.

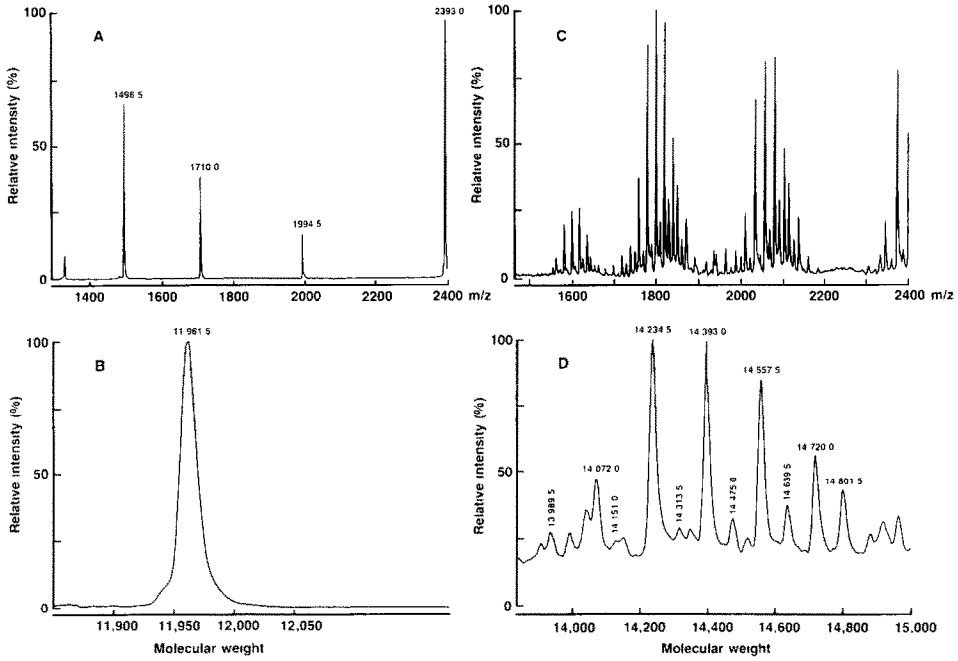

Fig. 9. Mass spectra of purified r-hSP (A and B) and glycosylated r-hSP (C and D). A and C show the original mass spectrum of r-hSP and glycosylated r-hSP, respectively. B and D show the reconstructed mass spectrum for r-hSP and glycosylated r-hSP on the basis of A and C.

15 is glycosylated and that only two monosaccharide residues, mannose and *N*-acetyl glucoseamine, occur in the glycosylated form of r-hSP. From these results in combination with the mass spectrometry data, it is possible to deduce the different glycosylated forms of r-hSP (Table III).

Molecular weights corresponding to two series of carbohydrate side chains can be deduced from the combination of carbohydrate composition data and ISMS-data, namely $(GlcNAc)_2(Hex)_{10-15}$ and $(Hex)_{13-17}$ (Table III). As mannose is the only hexose in the glycosylated r-hSP, and as Asn-15 is the only glycosylated residue, it seems reasonable to conclude that the structure of the glycosylation site is Asn-$(GlcNAc)_2$-$(Man)_{10-15}$. The observed Asn-$(Hex)_{13-17}$ forms are thus most likely to arise from fragmentation in the mass spectrometre, by which the two GlcNAc residues lose an acetyl group and are converted into two hexoses.

The structure of Asn-$(GlcNAc)_2$-$(Man)_{10-15}$ has previously been reported as high mannose type of *N*-glycosylation for other peptides and proteins expressed in yeast [26].

4. DISCUSSION

In the present study, we describe the application of a yeast expression system for the production of mg to g amount of r-hSP. The hSP gene was constructed by a combination of PCR cloning of each of the two intronless trefoil domains and synthetic DNA chemistry. Initial attempts to clone the entire hSP gene by PCR from human genomic DNA were not successful. One explanation for this could be the presence of introns of unknown size flanking each trefoil domain. In the human oestrogen-responsive gene pS2 the single trefoil domain is flanked by 3 kb and 0.8 kb introns [39].

Due to the high level of expression, the purification of the yeast produced hSP is relatively simple resulting in an overall purification yield of 50% and 34% for r-hSP and glycosylated r-hSP, respectively. The primary amino acid sequence of the r-hSP was shown by automated Edman degradation, peptide mapping and mass

Table III

Mass analysis of glycosylated r-hSP

Structure	Calculated M.W.	M.W. found by ESMS (Fig. 9D)
hSP + 2 GlcNAc + 10 Man	13,989.1	13,989.5
hSP + 2 GlcNAc + 11 Man	14,151.2	14,151.0
hSP + 2 GlcNAc + 12 Man	14,313.4	14,313.5
hSP + 2 GlcNAc + 13 Man	14,475.5	14,475.0
hSP + 2 GlcNAc + 14 Man	14,639.7	14,639.5
hSP + 2 GlcNAc + 15 Man	14,799.8	14,801.5
hSP + 13 Man	14,069.1	14,072.0
hSP + 14 Man	14,231.3	14,232.5
hSP + 15 Man	14,393.4	14,393.0
hSP + 16 Man	14,555.5	14,557.5
hSP + 17 Man	14,717.7	14,720.0

spectrometry analysis to be identical to the sequence of hSP derived from cDNA sequencing previously described [8]. In the yeast-expressed hSP, the N-terminal amino acid residue is glutamic acid in contrast to the natural occurring porcine PSP in which the N-terminal glutamine has cyclized to pyrrolidone carboxylic acid [17]. It is not known if the N-terminal residue of naturally occurring hSP is glutamic acid or pyrrolidone carboxylic acid.

No free SH-groups were found in r-hSP indicating that all 14.5 Cys-residues form disulphide bonds. The secondary structure of hSP including the seven disulphide bonds (Fig. 2) has been placed by homology to porcine PSP [1]. Final proof of the disulphide bond configuration in hSP as well as PSP may come from 3-dimensional structure analysis by X-ray [27] or NMR [28].

The yeast-expressed r-hSP is secreted as two different forms into the culture media; one form is *N*-glycosylated at Asn-15, the structure being a high mannose form of Asn-$(GlcNAc)_2$-$(Man)_{10-15}$; the other form of r-hSP is non-glycosylated. The two forms are secreted from the yeast cell in the ratio of 2:1. In man, hSP is expressed in normal stomach epithelium [8] and probably secreted into the gastric juice. However, as the peptide has not been purified and characterized on protein level, it is not known whether or not the naturally occurring hSP is glycosylated. The biological function of hSP as well as other members of the trefoil family (PSP, ITF, pS2, frog skin peptides) are poorly understood so far. As hSP is difficult to purify in any significant amount from tissue extraction and extremely difficult to obtain by peptide synthesis, the yeast expression system described in the present study may become a valuable tool in obtaining pure peptide for physiological and pathophysiological studies.

REFERENCES

[1] Thim, L. (1989) FEBS Lett. 250, 85–90.
[2] Suemori, S., Lynch-Devaney, K. and Podolsky, D.K (1991) Proc. Natl. Acad. Sci. USA 88, 11017–11021.
[3] Jakowlew, S.B., Breathnach, R., Jeltsch, J.M., Masiakowski, P. and Chambon, P. (1984) Nucleic Acids Res. 12, 1861–1878.
[4] Prud'homme, J.-F., Fridlansky, F., Le Cunff, M., Atger, M., Mercier-Bodart, C., Pichon, M.-F. and Milgrom, E. (1985) DNA 4, 11–21.
[5] Rio, M.C., Bellocq, J.P., Daniel, J.Y., Tomasetto, C., Lathe, R., Chenard, M.P., Batzenschlager, A. and Chambon, P. (1988) Science 241, 705–708.
[6] Thim, L., Thomsen, J., Christensen, M. and Jørgensen, K.H. (1985) Biochim. Biophys. Acta 827, 410–418.
[7] Rose, K., Savoy, L.-A., Thim, L., Christensen, M. and Jørgensen, K.H. (1989) Biochim. Biophys. Acta 998, 297–300.
[8] Tomasetto, C., Rio, M.-C., Gautier, C., Wolf, C., Hareuveni, M., Chambon, P. and Lathe, R. (1990) EMBO J. 9, 407–414.
[9] Hoffmann, W. (1988) J. Biol. Chem. 263, 7686–7690.
[10] Hauser, F. and Hoffmann, W. (1991) J. Biol. Chem. 266, 21306–21309.
[11] Hauser, F., Roeben, C. and Hoffmann, W. (1992) J. Biol. Chem. 267, 14451–14455.
[12] Jørgensen, K.H., Thim, L. and Jacobsen, H.E. (1982) Regul. Peptides 3, 207–219.
[13] Thim, L., Jørgensen, K.H. and Jørgensen, K.D. (1982) Regul. Peptides 3, 221–230.
[14] Rasmussen, T.N., Raabjerg, L., Poulsen, S.S., Thim, L. and Holst, J.J. (1992) Histochemistry 98, 113–119.
[15] Frandsen, E.K., Jørgensen, K.H. and Thim, L. (1986) Regul. Peptides 16, 291–297.
[16] Frandsen, E.K. (1988) Regul. Peptides 20, 45–52.
[17] Rasmussen, T.N., Raabjerg, L., Poulsen, S.S., Thim, L. and Holst, J.J. (1992) Am. J. Physiol. (in press)
[18] Jørgensen, K.D., Diamant, B., Jørgensen, K.H. and Thim, L. (1982) Regul. Peptides 3, 231–243.
[19] Hoosein, N.M., Thim, L., Jørgensen, K.H. and Brattain, M.G. (1989) FEBS Lett. 247, 303–306.
[20] Wright, N.A., Poulsom, R., Stamp, G.W., Hall, P.A., Jeffery, R.E., Longcroft, J.M., Rio, M.-C., Tomasetto, C. and Chambon, P (1990) J. Pathol. 162, 279–284.
[21] Rio, M.-C., Chenard, M.-P., Wolf, C., Marcellin, L., Tomasetto, C., Lathe, R., Bellocq, J.P. and Chambon, P. (1991) Gastroenterology 100, 375–379.
[22] Thim, L., Hansen, M.T. and Soerensen, A.R. (1987) FEBS Lett. 212, 307–312.
[23] Bruins, A.P., Covey, T.R. and Henion, J.D. (1987) Anal. Chem. 59, 2642–2646.
[24] Covey, T.R., Bonner, R.F., Shushan, B.I. and Henion, J D. (1988) Rapid Commun. Mass Spectrom. 2, 249–256.
[25] Thim, L., Hansen, M.T., Norris, K., Hoegh, I., Boel, E., Forstrom, J., Ammerer, G. and Fiil, N.P. (1986) Proc. Natl. Acad. Sci. USA 83, 6766–6770.
[26] Poulter, L. and Burlingame, A.L. (1990) in: Methods in Enzymology (McCloskey, J.A. ed.) vol. 193, pp. 661–689, Academic Press, San Diego, CA.
[27] Gajhede, M., Thim, L., Jørgensen, K.H. and Melberg, S.G. (1992) Proteins: Structure, Function, and Genetics 13, 364–368.
[28] Carr, M.D. (1992) Biochemistry 31, 1998–2004.
[29] Sambrook, J., Fritch, E.F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[30] Sanger, F., Micklen, S., and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[31] Horton, R.M., Hunt, H.D., Ho, S.N., Pullen, J.K., and Pease, L.R. (1989) Gene 77, 61–68.
[32] Mead, D.A., Szczesna-Skorupa, E. and Kemper, B. (1986) Prot. Engin. 1, 67–74.
[33] Kawasaki, G. (1984) 12th International Conference on Yeast Genetics and Molecular Biology, Sept. 17–24, 1984, Edinburgh, Scotland, Abstr. P15.
[34] Hagenbüchle, O., Tosi, M., Schibler, U., Bovey, R., Wellauer, P.K., and Young, R.A. (1981) Nature 289, 643–646.
[35] Egel-Mitani, M. and Hansen, M.T. (1987) Nucleic Acids Res. 15, 6303–6304.
[36] Christiansen, L. and Norris, K., personal communication.
[37] Alber, T. and Kawasaki, G. (1982) J. Mol. Appl. Genet. 1, 419–434.
[38] Russell, P.R. (1985) Gene 40, 125–130.
[39] Jeltsch, J.M., Roberts, M., Schatz, C., Garnier, J.M., Brown, A.M.C., and Chambon, P. (1987) Nucleic Acids. Res. 15, 1401–1414.
[40] Sherman, F., Fink, G.R. and Hicks, J.B. (1981) Methods in Yeast Genetics, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[41] Norris, K., Norris, F. and Bjoern, S.E. (1990) International Patent Application WO 90/10075.