Molecular cloning of a cDNA encoding the glycoprotein of hen oviduct microsomal signal peptidase

Ann L. NEWSOME,* John W. McLEAN† and Mark O. LIVELY*‡

*Department of Biochemistry, Comprehensive Cancer Center, Bowman Gray School of Medicine, Wake Forest University, Winston-Salem, NC 27103, U.S.A., and †Genentech Inc., South San Francisco, CA 94080, U.S.A.

Detergent-solubilized hen oviduct signal peptidase has been characterized previously as an apparent complex of a 19 kDa protein and a 23 kDa glycoprotein (GP23) [Baker & Lively (1987) *Biochemistry* **26**, 8561–8567]. A cDNA clone encoding GP23 from a chicken oviduct λ gt11 cDNA library has now been characterized. The cDNA encodes a protein of 180 amino acid residues with a single site for asparagine-linked glycosylation that has been directly identified by amino acid sequence analysis of a tryptic-digest peptide containing the glycosylated site. Immunoblot analysis reveals cross-reactivity with a dog pancreas protein. Comparison of the deduced amino acid sequence of GP23 with the 22/23 kDa glycoprotein of dog microsomal signal peptidase [Shelness, Kanwar & Blobel (1988) *J. Biol. Chem.* **263**, 17063–17070], one of five proteins associated with this enzyme, reveals that the amino acid sequences are 90% identical. Thus the signal peptidase glycoprotein is as highly conserved as the sequences of cytochromes *c* and *b* from these same species and is likely to be found in a similar form in many, if not all, vertebrate species. The data also show conclusively that the dog and avian signal peptidases have at least one protein subunit in common.

INTRODUCTION

Signal peptidase, the membrane-bound endoproteinase that removes signal peptides from nascent proteins as they are translocated into the lumen of the endoplasmic reticulum (Lively, 1989), is presumed to be either part of, or closely associated with, sites of translocation of proteins into the endoplasmic reticulum (Blobel & Dobberstein, 1975). Eukaryotic signal peptidases have been characterized only in dog pancreas (Evans *et al.*, 1986) and chicken oviduct (Baker & Lively, 1987), and these differ in the number of proteins associated with the detergent-solubilized purified enzyme activity. Dog signal peptidase was isolated as a complex of five proteins with estimated molecular masses of 12, 18, 21, 23 and 25 kDa (Evans *et al.*, 1986). Partially purified yeast microsomal signal peptidase also appears to be a complex of four polypeptides, one of which is a glycoprotein (YaDeau *et al.*, 1991). Hen oviduct signal peptidase (HOSP) activity is associated with an apparent complex of only two proteins, with estimated molecular masses of 19 and 23 kDa (Baker & Lively, 1987). In purified signal peptidase preparations from these two species, the 23 kDa proteins are *N*-linked glycoproteins that migrate as doublets when analysed by SDS/PAGE. Each migrates as a single protein band of approx. 20 kDa following removal of carbohydrate by treatment with *N*-Glycanase (Evans *et al.*, 1986; Baker & Lively, 1987).

It is not yet known which of the two proteins is required for catalysis of peptide bond cleavage or what other roles may be played by the other components in the process of translocation of nascent proteins into the endoplasmic reticulum. The data from purification of HOSP suggest that no more than two of the proteins isolated as part of the dog signal peptidase complex are required for proteolysis *in vitro* if the enzymes purified from these two sources are related. A comparison of molecular masses determined by SDS/PAGE analysis of the purified chicken (Baker & Lively, 1987) and dog (Evans *et al.*, 1986) enzymes

suggested that the glycoproteins may be similar because they both migrate as doublets with the same apparent molecular mass. Amino acid sequences of tryptic-digest peptides of the 23 kDa (Baker & Lively, 1987) oviduct protein were found to be similar to amino acid sequences predicted by a cDNA encoding the dog 23 kDa glycoprotein (Shelness *et al.*, 1988), suggesting strongly that the two proteins are related. In addition to the dog glycoprotein, cDNAs encoding the dog 18 kDa (Shelness & Blobel, 1990) and 21 kDa (Greenburg *et al.*, 1989) have also been cloned, revealing that the 18 kDa and 21 kDa dog proteins are closely related proteins. In the present paper we report the sequence of a cDNA encoding the hen oviduct 23 kDa glycoprotein (GP23). The results show that the signal peptidase glycoproteins are nearly identical in dogs and chickens. Their amino acid sequences have been as highly conserved as cytochromes *b* and *c* during the 300 million years since divergence of mammals from birds. These results demonstrate that, although signal peptidases from different species have been isolated with different numbers of protein constituents, the mammalian and avian signal peptidases contain at least one common protein and suggest that the common glycoprotein may play a direct role in catalysis of signal peptide bond cleavage from nascent secretory proteins during translocation into the endoplasmic reticulum.

MATERIALS

Reagents and enzymes were obtained from the following sources: *Eco*RI, urea and *Escherichia coli* strains Y1090 and Y1088 were from Stratagene (La Jolla, CA, U.S.A.); nitrocellulose was from Schleicher and Schuell (Keene, NH, U.S.A.); formamide was from EM Science (Gibbstown, NH, U.S.A.); [γ - 32 P]ATP, [α - 32 P]dATP and [α - 35 S]thio[dATP] were from Amersham (Arlington Heights, IL, U.S.A.); T4 polynucleotide kinase was from Boehringer Mannheim (Indianapolis, IN, U.S.A.); DNAase I, goat anti-(rabbit IgG) antibody-horse radish

Abbreviations used: HOSP, hen oviduct signal peptidase; GP23, 23 kDa glycoprotein of hen oviduct; SPC22/23, 23 kDa glycoprotein of dog pancreas.

‡ To whom correspondence should be addressed.

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number X60795.

Table 1. Synthetic oligonucleotide probes used in cDNA screening

(a) The amino acid sequence of the signal peptidase glycoprotein tryptic-digest peptide 22T31 is given in the top row. The sequence of the synthetic oligonucleotide probe is in the middle with the corresponding sequence found in the GP23 cDNA clone below it. (b) The top row gives the amino acid sequence of signal peptidase glycoprotein tryptic-digest, peptide 24T49. The middle row shows the sequence of the synthetic oligonucleotide probe and the bottom row is the sequence found in the GP23 cDNA.

Sequences																				
(a)																				
22T31	Asn	Asn	Ala	Leu	Asn	Gln	Val	Val	Leu	Trp	Asp	Lys								
Probe 22	AA ^T _C	AA ^T _C	GC ^T _C	CTG	AA ^T _C	CA ^A _G	GTG	GTG	CTG	TGG	GA ^T _C	AAG								
cDNA	AAC	AAT	GCT	CTG	AAC	CAA	GTG	GTT	CTT	TGG	GAC	AAG								
(b)																				
24T49	Asn	Val	Thr	Leu	Thr	Leu	Ser	Trp	Asn	Val	Val	Pro	Asn	Ala	Gly	Leu	Leu	Pro	Leu	Val
Probe 24	AAT	GTG	ACC	CTG	ACC	CTG	TCC	TGG	AAT	GTG	GTG	CCC	AAT	GCT	GGC	CTG	CTG	CCC	CTG	GTG
cDNA	AAT	GTC	ACT	TTG	ACA	CTC	TCC	TGG	AAT	GTT	GTA	CCA	AAT	GCC	GGC	CTT	CTA	CCT	CTT	GTG

peroxidase conjugate was from Sigma Chemical Co. (St. Louis, MO, U.S.A.); SeaKem GTG agarose was from FMC BioProducts (Rockland, ME, U.S.A.); acrylamide was from Bio-Rad Laboratories (Richmond, CA, U.S.A.).

METHODS

Preparation of rabbit anti-GP23 antibodies

HOSP was purified from chicken oviduct as previously described (Baker & Lively, 1987). A 170 µg portion of purified HOSP was precipitated with 10% (w/v) trichloroacetic acid, washed with ice-cold 90% (v/v) acetone and subjected to SDS/PAGE in a 12.5% acrylamide gel (Laemmli, 1970). The gel was lightly stained with Coomassie Brilliant Blue R-250, and then the GP23 band was excised and the protein was electroeluted from the gel as described by Hunkapiller *et al.* (1983). Approx. 50 µg of the electroeluted GP23 in phosphate-buffered saline (140 mM-NaCl/10 mM-sodium phosphate buffer, pH 7.2) was emulsified in an equal volume of Freund's complete adjuvant for intradermal injection into a New Zealand White rabbit. The rabbit received injections of 25 µg of GP23 in Freund's incomplete adjuvant 4 and 8 weeks after the initial immunization. Blood was collected 7 days after the second booster injection and the serum was precipitated with (NH₄)₂SO₄ at 50% (w/v) saturation. The precipitated immunoglobulins were dissolved in and dialysed against phosphate-buffered saline to remove the (NH₄)₂SO₄. An equal volume of glycerol was added and the antibodies were stored at -20 °C.

Immunoblotting

Dog pancreas rough microsomal fraction was prepared by differential centrifugation of homogenized tissue (Walter & Blobel, 1983). Triton X-100 was added to the microsomal fraction to a final concentration of 2.5% (w/v) and the solubilized proteins were separated from particulate material by centrifugation in a Beckman Airfuge at 135 000 g (*r_{av}*, 12.1 mm) for 4 min at 25 °C. A sample of this solubilized membrane preparation and a sample of partially purified HOSP were analysed by SDS/PAGE and electrophoretically transferred to a nitrocellulose membrane (Towbin *et al.*, 1979). The membrane was incubated for 2 h in Tris-buffered saline/Tween (150 mM-NaCl/10 mM-Tris/HCl buffer, pH 8.0, containing 0.05% Tween-20) containing 1% (w/v) BSA, then incubated for 3 h in rabbit anti-

GP23 immunoglobulins (1:1000) in Tris-buffered saline/Tween containing 1% (w/v) BSA. The membrane was then washed with three 10 min washes in Tris-buffered saline/Tween before incubation for 2 h with goat anti-(rabbit IgG) antibody conjugated to horseradish peroxidase (1:1000) in Tris-buffered saline/Tween containing 1% (w/v) BSA. After three 10 min washes with Tris-buffered saline/Tween, the membrane was placed in a peroxidase substrate solution (0.25 mg of *o*-dianisidine/ml and 0.005% H₂O₂ in 10 mM-Tris/HCl buffer, pH 7.5).

Amino acid sequence analysis

HOSP was partially purified from chicken oviduct as previously described (Baker & Lively, 1987). A sample of the hydroxyapatite pool was precipitated with 10% (w/v) trichloroacetic acid and subjected to SDS/PAGE. GP23 was electroeluted as described above and analysed by automated Edman degradation with an Applied Biosystems Model 475A protein sequencer. A second preparation of GP23 was digested with Tos-Phe-CH₂Cl-treated trypsin, and the tryptic-digest peptides were separated by reverse-phase h.p.l.c. (Baker & Lively, 1987). Tryptic-digest peptides were then analysed by automated Edman degradation as described above.

cDNA cloning

A chicken oviduct cDNA library in λgt11 (Conneely *et al.*, 1987) was generously provided by Dr. Mark Hughes (Baylor College of Medicine, Houston, TX, U.S.A.). The library, which contained 5.3 × 10⁶ recombinants, was prepared from total polyadenylated RNA and primed with oligo(dT). Oligonucleotide probes based on amino acid sequences of tryptic-digest peptides of GP23 (Table 1) were synthesized by using an Applied Biosystems model 380A DNA synthesizer and were end-labelled with [γ-³²P]ATP by using T4 polynucleotide kinase. Recombinant bacteriophage plaque lifts were made with nitrocellulose filters (Benton & Davis, 1977) and screened with the oligonucleotide probes. DNA inserts obtained from λgt11 clones were isolated by digestion with *Eco*RI, and subcloned into M13mp19. DNA sequencing was performed on both coding and non-coding strands by using the dideoxy chain-termination method (Sanger *et al.*, 1977) with synthetic oligonucleotide primers. Both strands were sequenced to completion by the specific primer-directed method (Strauss *et al.*, 1986).

Analysis of protein sequence

The hydropathy profile (Kyte & Doolittle, 1982) was performed on the deduced amino acid sequence of GP23 with an amino acid window of 19 by using the program PEPTIDESTRUCTURE implemented by the sequence analysis package of the Genetic Computer Group (Devereux *et al.*, 1984). Searches of GenBank, EMBL, NBRF and Swiss-Prot databases were performed with the use of FASTA (Pearson & Lipman, 1988).

RESULTS AND DISCUSSION

Immunoblot analysis of signal peptidase

As a first step towards characterization of the purified oviduct signal peptidase glycoprotein (GP23), polyclonal antibodies specific for denatured GP23 were prepared. GP23 was isolated from total oviduct rough microsomal fraction that had been stripped with ice-cold Na_2CO_3 solution as described previously (Baker & Lively, 1987). The detergent-solubilized microsomal proteins were subjected to chromatography on DEAE-cellulose, CM-cellulose and hydroxyapatite. This partially purified preparation of oviduct signal peptidase was then subjected to preparative SDS/PAGE and electroelution as described in the Methods section. The electrophoretically purified GP23, which was shown to migrate as a doublet with an apparent molecular mass of 23–24 kDa on analytical SDS/PAGE (results not shown), was used for immunization. With the use of immunoblot analysis (Towbin *et al.*, 1979), the resulting antibodies were found to be specific for the 23 kDa doublet in a partially purified preparation of HOSP (Fig. 1, lane 2). In addition, a strongly immunoreactive doublet with electrophoretic mobility identical with that of GP23 was observed in detergent-solubilized dog pancreas rough microsomal fraction (Fig. 1, lane 1). On this basis we concluded that the 23 kDa glycoproteins observed in preparations of crude solubilized dog pancreas microsomal fraction and partially purified oviduct were related proteins.

cDNA cloning and characterization

To characterize GP23 further, a cDNA clone for the protein was isolated and characterized. GP23 was isolated by electrophoresis and electroelution, then subjected to *N*-terminal amino

acid sequence analysis. A single *N*-terminal sequence was observed with the following sequence: Met-Asn-Thr-Val-Leu-Ser-Arg-Ala-Asn-Ser-Leu-Phe-Ala-Phe-Ser-. Additional amino acid sequences of tryptic-digest peptides purified by h.p.l.c. were obtained as described previously (Baker & Lively, 1987), and these sequences were used to design synthetic oligonucleotide probes for cDNA clone selection (Table 1). Codon usage tables (Aota *et al.*, 1988) and rules for codon selection (Lathe, 1985) were applied in the design of the oligonucleotide probes. Probe 22 was a mixture of 64 oligonucleotide sequences and probe 24 was a unique 60-nucleotide-residue sequence. Each synthetic oligonucleotide probe was end-labelled with ^{32}P and used to screen approx. 7×10^5 bacteriophages for an oligo(dT)-primed chicken oviduct cdNA library cloned in $\lambda\text{gt}11$ (Conneely *et al.*, 1987). Duplicate sets of nitrocellulose filters were screened with the ^{32}P -labelled oligonucleotide probes and each of the probes was hybridized to DNA from eight plaques. The bacteriophages from these plaques were purified and DNA was isolated.

cDNA inserts were examined by electrophoresis in 1% agarose gels (Maniatis *et al.*, 1982). Four of the clones were found to have inserts of approx. 1000 bp, one of which was subcloned into the *EcoRI* site of M13mp19 (Messing *et al.*, 1981). Template DNA from that subclone was subjected to dideoxy chain-termination sequencing (Sanger *et al.*, 1977).

The complete sequence of the 799 bp cDNA clone and the deduced amino acid sequence are given in Fig. 2. The cloned DNA has a 5' untranslated region of 68 bp, an open reading frame of 540 bp and a 3' untranslated region of 188 bp that contains the polyadenylation signal, AATAAA, 35 nucleotide residues upstream of the poly(A)-addition site (Proudfoot & Brownlee, 1976). The initiator methionine residue was identified by comparison of the predicted amino acid sequence with the *N*-terminal sequence of the intact protein, which begins with that methionine residue. Additionally, the nucleotide sequence upstream from the AUG is consistent with the requirements for eukaryotic initiation sites (Kozak, 1984).

Although it is possible the GP23 could be synthesized as a precursor that undergoes limited proteolysis before insertion into the lipid bilayer of the endoplasmic reticulum, this seems unlikely. Translation of the open reading frame upstream from the observed *N*-terminal methionine residue of GP23 predicts the following sequence: Leu-Pro-Tyr-Pro-Ala-Pro-Ser-Ser-Gly-Cys-Ser-Pro-Gly-Ala-Arg-Thr-Ala-Arg-Ala-Gly-Gly-Glu. The sequence is inconsistent with common cleavage sites for precursor processing of prohormones and related proteins that typically end with one or two lysine or arginine residues (Devi, 1991). This sequence also does not contain the structural features of a typical signal peptide. Glutamic acid is rarely, if ever, present as the C-terminal residue of a signal peptide (von Heijne, 1986). Furthermore, the *N*-terminus of this sequence lacks the predominant hydrophobic domain characteristic of signal peptides that target proteins to the endoplasmic reticulum. Therefore it appears unlikely that GP23 is synthesized as a precursor with an *N*-terminal cleaved signal peptide. This conclusion is consistent with the structure of the dog microsomal signal peptidase glycoprotein, which also appears to be synthesized without a precursor form (Shelness *et al.*, 1988). Furthermore, although GP23 is clearly not structurally related to leader peptidase (signal peptidase I) from *E. coli*, that enzyme is a membrane-bound protein that is also synthesized without an *N*-terminal cleaved signal peptide (Wolfe *et al.*, 1983).

The deduced amino acid sequence of GP23 has 180 amino acid residues with a single site predicted for *N*-linked glycosylation at Asn-141 and a calculated molecular mass of 20200 Da. This molecular mass is consistent with the electrophoretic mobility of the deglycosylated protein produced by treatment with

Fig. 1. Glycoprotein components of hen and dog signal peptidases are immunologically cross-reactive

Preparations of partially purified HOSP and solubilized dog pancreas microsomal fraction were separated by SDS/PAGE and electrophoretically transferred to nitrocellulose. The nitrocellulose blot was incubated with rabbit anti-GP23 antibodies followed by goat anti-(rabbit IgG) antibody conjugated to horseradish peroxidase. Bound antibodies were detected by assaying for peroxidase activity as described in the Methods section. Lane 1, dog pancreas microsomal proteins solubilized in 2.5% Triton X-100; lane 2, partially purified HOSP after DEAE-cellulose chromatography.

6	CCCTACCCCGCCCCGAGCAGCGGCTGCAGCCCGGGAGCCCGCACGGCGAGAGCGGGGGGGAG	CGCTC 5 68
69	ATG AAC ACG GTG CTG TCC CGG GCC AAC TCG CTC TTC GCC TTC TCG CTG <u>M N T V L S R A N S L F A F S L</u>	116
117	AGC GTG ATG GCG GCG CTC ACC TTC GGC TGC TTC ATC ACC ACC GCC TTC <u>S V M A A L T F G C F I T T A F</u>	164
165	AAG GAG CGC AGC GTG CCC GTC AGC ATC GCC GTG TCC CGG GTC ACG CTA <u>K E R S V P V S I A V S B V T L</u>	212
213	AGA AAT GTA GAA GAC TTC ACT GGA CCT AGA GAA AGA AGT GAT CTG GCG <u>R N V E D F T G P R E R S D L A</u>	260
261	TTC GTC ACG TTT GAC ATT ACT GCA GAT TTG CAG AGT ATA TTT GAC TGG <u>F V T F D I T A D L Q S I F D W</u>	308
309	AAT GTT AAA CAA TTG TTT CTA TAT TTG TCT GCA GAA TAT TCA ACA AAA <u>N V K Q L F L Y L S A E Y S T K</u>	356
357	AAC AAT GCT CTG AAC CAA GTG GTT CTT TGG GAC AAG ATC ATT TTG AGA <u>N N A L N O V V L W D K I I L R</u>	404
405	GGA GAT AAT CCA AGG CTG TTC TTA AAA GAC ATG AAG TCA AAG TAC TTT <u>G D N P R L F L K D M K S K Y F</u>	452
453	TTC TTT GAT GAT GGA AAT GGT CTC AAG GGA AAC AGG AAT GTC ACT TTG <u>F F D D G N G L K G N R N V T L</u>	500
	●	
501	ACA CTC TCC TGG AAT GTT GTA CCA AAT GCC GGC CTT CTA CCT CTT GTG <u>T L S W N V V P N A G L L P L V</u>	548
549	ACA GGA TCA GGA CAT ATG TCT GTA CCT TTC CCA GAT ACC TAT GAA ACA <u>T G S G H M S V P F P D T Y E T</u>	596
597	ACA AAA AGT TAT TAAATTATAATACCGAAGCAACATATTTTATACCTGTATATTGTGA <u>T K S Y</u>	655
656	ATAAATTTTATTGCGGTTTCTTCACACATCCACACAACATTCTCGTTGAAAGGTACTTAATT	718
719	TCCTCAGGTTCAACATCATGAAAAAAGAAAAATTGAATTTTTTAAATAAACCTTGAAACT	781
782	GAAACCTAAAGGAAAGGTAAAAA.....	

Fig. 2. Sequence of a cDNA clone of HOSP glycoprotein

The nucleotide and amino acid sequences for GP23 are given. All sequences determined by protein sequence analysis of either intact or trypsinized GP23 are underlined. The *N*-linked glycosylation site is indicated by the solid dot (●) below Asn-141 and the polyadenylation signal is underlined with a dotted line.

N-Glycanase (Baker & Lively, 1987), an enzyme that is specific for hydrolysis of asparagine-linked glycoproteins (Tarentino *et al.*, 1985). Amino acid sequence analysis of five tryptic-digest peptides and of the intact protein identified 74 out of 180 residues of the GP23 protein (see underlined sequences in Fig. 2) without any discrepancies with the sequence deduced from the cDNA. A single tryptic-digest glycopeptide of 40 residues was isolated as the last major peak eluted upon reverse-phase h.p.l.c. of the tryptic digest of GP23 (Baker & Lively, 1987). In two independent analyses this peptide gave a single amino acid sequence except for the first cycle of Edman degradation, which contained only background amino acids. The initial sequencing yield of one sequencer run was 92 pmol of peptide. No amino acid peaks greater than 15 pmol were observed in the first cycle of automated Edman degradation compared with 80 pmol of valine in the second cycle. This result is consistent with the presence of carbohydrate on Asn-141, which would result in failure to identify any residue in that cycle. Although there could be other explanations for the failure to identify an amino acid derivative at the first position of this peptide, the weight of evidence is consistent with the interpretation that Asn-141 is the site of glycosylation of GP23.

The predicted pI of GP23 in the absence of its carbohydrate is 9.19. This calculated value is consistent with its chromatographic properties during purification, because the protein does not bind to DEAE-cellulose at pH 8.2 but does bind to CM-cellulose at pH 5.8 (Baker & Lively, 1987).

The combined data from sequence analysis and molecular cloning revealed that the tryptic-digest maps reported in our characterization of purified HOSP had a contaminating protein (Baker & Lively, 1987). One peptide previously reported with the tryptic-digest sequences of purified GP23 (Baker & Lively, 1987) was not found within the coding sequence of the complete GP23 cDNA sequence. This peptide, Tyr-Val-Glu-Asn-Phe-Arg, originated from a contaminating protein with an electrophoretic mobility slightly faster than GP23 that was present in the hydroxyapatite chromatography pool used for purification of GP23 by SDS/PAGE. Because the *N*-terminal sequence of another purified batch of GP23 yielded only one amino acid sequence that was consistent with that of the cloned sequence and because of the correspondence of the predicted amino acid sequence from our clone with that of the canine glycoprotein (Shelness *et al.*, 1988), we conclude that the unknown peptide is unrelated to GP23.

Analysis of deduced amino acid sequence

Analysis for membrane-spanning regions (Klein *et al.*, 1985) suggests a single transmembrane segment from residues 12 to 28, which may anchor the protein in the lipid bilayer near its *N*-terminus. This region coincides with the observed hydrophobic peak at the *N*-terminus observed in the hydropathy plot (Fig. 3) (Kyte & Doolittle, 1982). A second hydrophobic region near the *C*-terminus between residues 144 and 162 closely follows the site of *N*-linked glycosylation and is potentially a second membrane-

Fig. 3. Hydropathy profile of GP23

The hydropathy index was calculated by using the method of Kyte & Doolittle (1982) with a window size of 19 amino acid residues. The positions of Cys-26, Asn-141 (the glycosylation site) and His-165 are indicated.

spanning domain or a buried hydrophobic region in a globular domain of the protein. These analyses predict a model of a transmembrane protein anchored by the *N*-terminus in the cytoplasm and with a larger globular domain in the lumen of the endoplasmic reticulum. Or, if the *C*-terminal hydrophobic domain spans the membrane, GP23 could be a bitopic membrane protein with both termini anchored in the bilayer. Either model is consistent with previously reported data that showed that the signal peptidase active site in intact microsomal vesicles is protected from exogenously added proteolytic enzymes and must therefore be located on the luminal side of the endoplasmic reticulum (Walter *et al.*, 1979).

The deduced amino acid sequence was compared with current releases of DNA (GenBank release 67.0) and protein (National Biomedical Research Foundation release 27.0; Swiss-Prot Protein Sequence release 17) databases with the use of FASTA (Pearson & Lipman, 1988; Devereaux *et al.*, 1984). The only significant match found was the dog SPC 22/23 cDNA sequence (Shelness *et al.*, 1988). No other proteins in the database at the time of this comparison have any significant similarity to the amino acid sequence of the signal peptidase glycoproteins.

Importantly, the sequence is also not related to *E. coli* leader peptidase (Zwizinski & Wickner, 1980; Müller *et al.*, 1982; Wolfe *et al.*, 1983), signal peptidase II of *E. coli* (Innis *et al.*, 1984) or mitochondrial processing peptidase (Hawltischek *et al.*, 1988; Pollock *et al.*, 1988). It is somewhat surprising that GP23 is so different from *E. coli* leader peptidase because the two enzymes have similar substrate-specificities (Rogenkamp *et al.*, 1981; Müller *et al.*, 1982; Watts *et al.*, 1983).

The chicken and dog glycoproteins are each composed of exactly 180 amino acid residues, of which 90% are identical (Fig. 4). Allowing for conservative substitutions following Dayhoff's log-odds scoring system (Dayhoff, 1978; Feng *et al.*, 1985) or the structure-genetic matrix scoring system (Doolittle, 1979), the degree of similarity of these two amino acid sequences is 97%. The rate of evolution of the signal peptidase glycoprotein has been quite low since the time of divergence of mammals and birds some 300 million years ago. This rate of divergence is similar to that of cytochrome *c* in chickens (Limbach & Wu, 1983) and dogs (McDowall & Smith, 1965), whose 104-residue sequences are 90% identical. The evolutionary rate of change of the signal peptidase glycoprotein is also comparable with that of cytochrome *b* (McDowall & Smith, 1965; Limbach & Wu, 1983) but lower than that of β -globin (Brimhall *et al.*, 1977; Knochel *et al.*, 1982). This high degree of conservation suggests that signal peptidase may be found in essentially the same form in other vertebrate species.

We do not yet know which of the subunits purified with signal peptidase activity is, or are, sufficient for proteolysis *in vitro*, but our data clearly indicate that GP23 is one of two possible candidates for the catalytic subunit of HOSP. Our previous data narrowed the possibilities to GP23 and a 19 kDa protein (Baker & Lively, 1987), and the current study proves that the glycoproteins are essentially identical in chickens and dogs. Furthermore, we also know that the 19 kDa oviduct protein is closely related to the dog 18 kDa and 21 kDa proteins (Shelness & Blobel, 1990; Greenburg *et al.*, 1989) since we have amino acid sequences of three tryptic-digest peptides of the oviduct 19 kDa protein. One of these peptides (Baker & Lively, 1987) precisely matches residues 140–147 of SPC21 (Greenburg *et al.*, 1989), and the other two sequences (A. L. Newsome & M. O. Lively, unpublished work) are very similar to sequences found in the dog SPC21 (Greenburg *et al.*, 1989) and SPC18 (Shelness & Blobel, 1990) proteins. The dog 18 kDa and 21 kDa protein sequences

Hen	M N T V L S R A N S L F A F S L S V M A A L T F G C F I T T A F	32
Dog	M N T V L S R A N S L F A F S L S V M A A L T F G C F I T T A F	
Hen	K E R S V P V S I A V S R V T L R N V E D F T G P R E R S D L A	64
Dog	K D R S V P V R L H V S R I M L K N V E D F T G P R E R S D L G	
Hen	F V T F D I T A D L Q S I F D W N V K Q L F L Y L S A E Y S T K	96
Dog	F I T F D I T A D L E N I F D W N V K Q L F L Y L S A E Y S T K	
Hen	N N A L N Q V V L W D K I I L R G D N P R L F L K D M K S K Y F	128
Dog	N N A L N Q V V L W D K I V L R G D N P K L L L K D M K T K Y F	
Hen	F F D D G N G L K G N R N V T L T L S W N V V P N A G L L P L V	160
Dog	F F D D G N G L K G N R N V T L T L S W N V V P N A G I L P L V	
Hen	T G S G H M S V P F P D T Y E T T K S Y	180
Dog	T G S G H V S V P F P D T Y E I T K S Y	

Fig. 4. Comparison of amino acid sequences of chicken and dog signal peptidase glycoproteins

The sequence of oviduct GP23 is given on the upper lines and the amino acid sequence of dog SPC 22/23 is given on the lower lines. Unbroken vertical lines indicate identity and vertical dotted lines indicate conservative amino acid substitutions.

are similar to that of the yeast SEC11 gene product, which has been shown by genetic analysis to be required for cell growth. SEC11 mutants are defective in processing of signal peptides (Böhni *et al.*, 1988). Therefore both proteins purified with oviduct signal peptidase are also present in purified dog signal peptidase, and these two proteins may be an enzymatically active subset of the signal peptidase complex *in vitro* that is found in the endoplasmic reticulum *in vivo*.

None of the signal peptidase proteins identified thus far appear to be related to any of the known proteinase families. It may be important to note that the sequence of GP23 contains a single cysteine residue, Cys-26, and a single histidine residue, His-165. If GP23 is the catalytic subunit, signal peptidase could be a thiol-dependent proteinase, since enzymes of this class require both a cysteine residue and a histidine residue in the active centre (Polgár & Halász, 1982). Interestingly, dog SPC 22/23 contains two histidine residues, but only His-165 is conserved in both species, suggesting that it may have a functional role. Cys-26, which is present in both proteins, is located in the C-terminal end of the putative membrane-spanning domain, which would place it near the luminal face of the endoplasmic reticulum bilayer. If the predicted orientation in the membrane is correct, Cys-26 may be in a hydrophobic environment (Fig. 3) near the luminal surface of the endoplasmic reticulum, where signal peptidase is believed to act on nascent proteins (Walter *et al.*, 1979).

We thank Mark Morris of the Protein Analysis Core Laboratory of the Comprehensive Cancer Center of Wake Forest University for his technical assistance and expertise in protein sequence analysis, Dr. Mark Hughes of Baylor College of Medicine for his gift of the cDNA library used in cloning, Dr. Keith Baker for preparing tryptic-digest peptides of GP23, and the Genentech oligonucleotide synthesis group. This work was supported by National Institutes of Health Grant GM 32861 (M.O.L.) and by Genentech Inc. The Protein Analysis Core Laboratory is supported in part by National Institutes of Health Grants CA 12197 and RR 04869 as well as a grant from the North Carolina Biotechnology Center.

REFERENCES

- Aota, S., Gojobori, T., Ishibashi, F., Maruyama, T. & Ikemura, T. (1988) *Nucleic Acids Res. Suppl.* **16**, 315–342
- Baker, R. K. & Lively M. O. (1987) *Biochemistry* **26**, 8561–8567
- Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182
- Blobel, G. & Dobberstein, B. (1975) *J. Cell Biol.* **67**, 835–851
- Böhni, P. C., Deshaies, R. J. & Schekman, R. W. (1988) *J. Cell Biol.* **106**, 1035–1042
- Brimhall, B., Duerst, M. & Jones, R. T. (1977) *J. Mol. Evol.* **9**, 231–235
- Conneely, O. M., Dobson, A. D., Tsai, M. J., Beattie, W. G., Toft, D. O., Huckaby, C. S., Zarucki, T., Schrader, W. T. & O'Malley, B. W. (1987) *Mol. Endocrinol.* **1**, 517–525
- Dayhoff, M. O. (1978) *Atlas of Protein Sequence and Structure*, vol. 5, suppl. 3, pp. 345–358, National Biomedical Research Foundation, Washington
- Devereaux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395
- Devi, L. (1991) *FEBS Lett.* **280**, 189–194
- Doolittle, R. F. (1979) in *The Proteins* (Neurath, H. & Hill, R. L., eds.), vol. 4, pp. 1–118, Academic Press, New York
- Evans, E. A., Gilmore, R. & Blobel, G. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 581–585
- Feng, D. F., Johnson, M. S. & Doolittle, R. F. (1985) *J. Mol. Evol.* **21**, 112–125
- Greenburg, G., Shelness, G. S. & Blobel, G. (1989) *J. Biol. Chem.* **264**, 15762–15765
- Hawltischek, G., Schneider, H., Schmidt, B., Tropschug, M., Hartl, F.-U. & Neupert, W. (1988) *Cell* **53**, 795–806
- Hunkapiller, M. W., Lujan, E., Ostrander, F. & Hood, L. E. (1983) *Methods Enzymol.* **91**, 227–236
- Innis, M. A., Tokunaga, M., Williams, M. E., Loranger, J. M., Chang, S.-Y. & Wu, H. C. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3708–3712
- Klein, P., Kanehisa, M. & DeLisi, C. (1985) *Biochim. Biophys. Acta* **815**, 468–476
- Knochel, W., Wittig, B., Wittig, S., John, M. E., Grundmann, U., Oberthur, W., Godovac, J. & Braunitzer, G. (1982) *Nature (London)* **295**, 710–712
- Kozak, M. (1984) *Nucleic Acids Res.* **12**, 857–872
- Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **175**, 105–132
- Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685
- Lathe, R. (1985) *J. Mol. Biol.* **183**, 1–12
- Limbach, K. J. & Wu, R. (1983) *Nucleic Acids Res.* **11**, 8931–8950
- Lively, M. O. (1989) *Curr. Opin. Cell Biol.* **1**, 1188–1193
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor
- McDowall, M. A. & Smith, E. L. (1965) *J. Biol. Chem.* **240**, 4635–4647
- Messing, J., Crea, R. & Seeburg, P. H. (1981) *Nucleic Acids Res.* **9**, 309–321
- Müller, M., Ibrahim, I., Chang, C. N., Walter, P. & Blobel, G. (1982) *J. Biol. Chem.* **257**, 11860–11863
- Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2444–2448
- Polgár, L. & Halász, P. (1982) *Biochem. J.* **207**, 1–10
- Pollock, R. A., Hartl, F.-U., Cheng, M. Y., Ostermann, J., Horwich, A. & Neupert, W. (1988) *EMBO J.* **7**, 3493–3500
- Proudfoot, N. J. & Brownlee, G. G. (1976) *Nature (London)* **263**, 211–214
- Rogenkamp, R., Kusterman-Kuhn, B. & Hohenberg, C. P. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 4466–4470
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- Shelness, G. S. & Blobel, G. (1990) *J. Biol. Chem.* **265**, 9512–9519
- Shelness, G. S., Kanwar, Y. S. & Blobel, G. (1988) *J. Biol. Chem.* **263**, 17063–17070
- Strauss, E. C., Kober, J. A., Siu, G. & Hood, L. E. (1986) *Anal. Biochem.* **154**, 353–360
- Tarentino, A. L., Gomez, C. M. & Plummer, T. H., Jr. (1985) *Biochemistry* **24**, 4665–4671
- Towbin, H., Staehelin, T. & Gordon, J. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4350–4354
- von Heijne, G. (1986) *Nucleic Acids Res.* **14**, 4683–4690
- Walter, P. & Blobel, G. (1983) *Methods Enzymol.* **96**, 84–93
- Walter, P., Jackson, R. C., Marcus, M. M., Lingappa, V. R. & Blobel, G. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1795–1799
- Watts, S., Wickner, W. & Zimmerman, R. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2809–2813
- Wolfe, P. B., Wickner, W. & Goodman, J. M. (1983) *J. Biol. Chem.* **260**, 12073–12080
- YaDeau, J. T., Klein, C. & Blobel, G. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 517–521
- Zwizinski, C. & Wickner, W. (1980) *J. Biol. Chem.* **255**, 7973–7977