The NH_2 Terminus of Preproinsulin Directs the Translocation and Glycosylation of a Bacterial Cytoplasmic Protein by Mammalian Microsomal Membranes

Etta M. Eskridge and Dennis Shields

Departments of Anatomy and Structural Biology, and Developmental Biology and Cancer, Albert Einstein College of Medicine, Bronx, New York 10461

Abstract. To investigate putative sorting domains in precursors to polypeptide hormones, we have constructed fusion proteins between the amino terminus of preproinsulin (ppI) and the bacterial cytoplasmic enzyme chloramphenicol acetyltransferase (CAT). Our aim is to identify sequences in ppI, other than the signal peptide, that are necessary to mediate the intracellular sorting and secretion of the bacterial enzyme. Here we describe the in vitro translation of mRNAs encoding two chimeric molecules containing 71 and 38 residues, respectively, of the ppI NH_2 terminus fused to the complete CAT sequence. The ppI signal peptide and 14 residues of the B-chain were sufficient to direct the translocation and segregation of CAT into microsomal membrane vesicles. Furthermore, the CAT enzyme underwent N-linked glycosylation, presumably at a single cryptic site, with an efficiency that was comparable to that of native glycoproteins synthesized in vitro. Partial amino-terminal sequencing demonstrated that the downstream sequences in the fusion proteins did not alter the specificity of signal peptidase, hence cleavage of the ppI signal peptide occurred at precisely the same site as in the native precursor. This is in contrast to results found in prokaryotic systems. These data demonstrate that the first 38 residues of ppI encode all the information necessary for binding to the endoplasmic reticulum membrane, translocation, and proteolytic (signal sequence) processing.

In recent years the process of intracellular protein trafficking and secretion has been described in detail for eukaryotic cells (5) and some of the primary events in the secretory pathway have been characterized (26). However, the precise molecular mechanisms that regulate intracellular sorting of secretory proteins are still poorly understood. Our laboratory is concerned with elucidating putative sorting sequences in presecretory proteins, particularly those whose secretion is regulated in response to environmental stimuli. To this end we have been studying the biosynthesis and posttranslational processing of precursors to the pancreatic islet hormones, insulin, glucagon, and somatostatin, as models for secretory proteins. Insulin is particularly appropriate for such studies since it is one of the best characterized of all polypeptide hormones. The cDNA and gene sequence of preproinsulins (ppI)[1] have been determined from numerous species as has the x-ray crystallographic structure of the mature hormone (2, 24). In all species, ppI is synthesized in pancreatic β cells and comprises a signal peptide, the B-chain, C-peptide, and A-chain. The signal peptide is cleaved cotranslationally (4, 20) to yield nascent proinsulin (pI) in which the B- and A-chains of mature insulin are joined by the C-peptide, which is flanked by two pairs of basic amino acids. One function of the C-peptide is to facilitate the correct folding of pI so that the disulfide bridges that link the A- and B-chains in mature insulin can form efficiently (24). Proteolytic cleavage of pI to insulin occurs at the paired basic amino acids by enzyme(s) localized in the secretory granules, resulting in secretion of both insulin and the C-peptide. Although no topogenic sequences other than the signal peptide have been identified in ppI, it is possible that the correct folding of pI might be important for efficient intracellular transport of the molecule through the secretory pathway. Since ppI undergoes only two posttranslational modifications, i.e., disulfide bridge formation and proteolysis, it represents a relatively simple peptide hormone precursor in which to investigate putative topogenic sequences.

Using ppI as a model, we have previously demonstrated (4) that there is a minimum size (60–70 amino acids) for productive interaction of the nascent precursor with signal recognition particle and the endoplasmic reticulum (ER) membrane. To determine if this minimum size might also constitute the minimum amount of structural information necessary to facilitate translocation of any nascent proteins

1. *Abbreviations used in this paper*: CAT, chloramphenicol acetyltransferase; EndoH, endoglycosidase H; ER, endoplasmic reticulum; pI, proinsulin; ppI, preproinsulin.

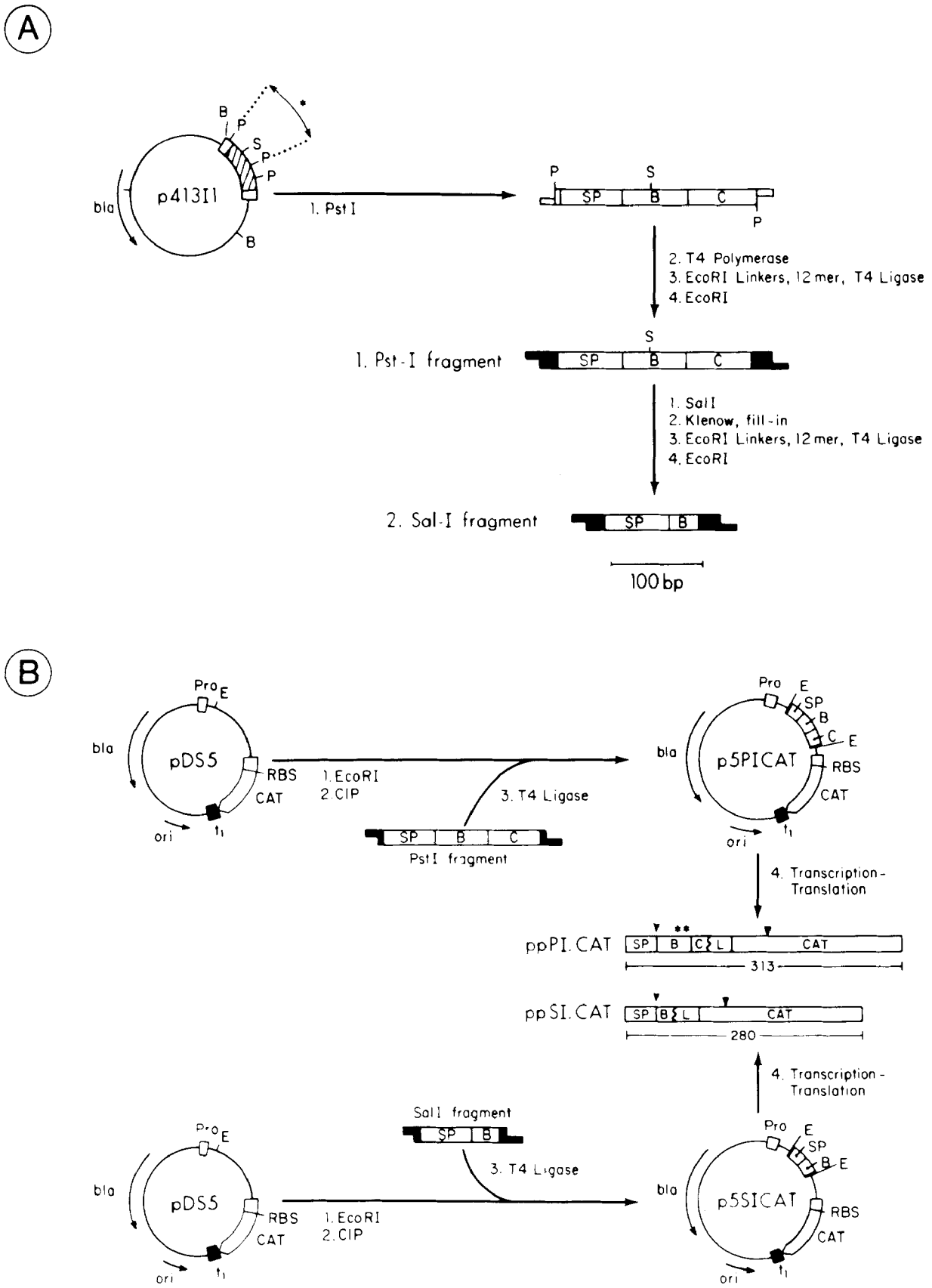

Figure 1. Construction of ppI–CAT Fusions. (*A*) Generation of ppI fragments. Clone p413-I1 was digested with PstI, and the PstI fragment of 220 bp indicated by the asterisk was gel purified. This fragment, encoding the 72 NH_2-terminal amino acids of ppI, was digested with T4 polymerase to remove single-stranded 3′ ends, then ligated to 12-bp EcoRI linkers for subcloning into the EcoRI site of pBR322. The PstI fragment (with EcoRI sticky ends) was digested with SalI, blunt-ended with the Klenow fragment of DNA polymerase I, and ligated to 12-bp EcoRI linkers as above. This fragment was then subcloned into the EcoRI site of pBR322 for amplification and purification. (*B*)

across the ER membrane, we have used recombinant DNA techniques to construct fusion proteins between ppI and the bacterial enzyme chloramphenicol acetyltransferase (CAT). Since CAT normally resides in the bacterial cytoplasm, the aim of these studies is to identify and characterize putative sorting domains in ppI and pI that could mediate the intracellular transport and secretion of a normally nonsecretory protein, i.e., the bacterial cytoplasmic enzyme CAT.

As a first step towards defining putative topogenic determinants in ppI, we have investigated the in vitro biosynthesis, membrane translocation, and cotranslational modifications of ppI-CAT fusions. Here we report that the ppI signal peptide and a portion of the B-chain can mediate the efficient translocation of CAT into mammalian microsomal membrane vesicles. In addition, a single cryptic N-linked glycosylation site present in the CAT molecule is apparently recognized by the microsomal membranes, such that a glycosylated form of CAT is synthesized in vitro with an efficiency that is comparable to native glycoproteins.

Materials and Methods

Materials

Plasmid pDS5 was a gift from Dr. B. Dobberstein, European Molecular Biology Laboratory, Heidelberg, Federal Republic of Germany. Rabbit anti-CAT serum was a gift from Dr. D. Wong (Albert Einstein College of Medicine), and guinea pig anti-porcine insulin was purchased from Miles Laboratories. Endoglycosidase H was a gift from Dr. P. Atkinson (Albert Einstein College of Medicine). [^{3}H]leucine and [^{35}S]cysteine were purchased from Amersham/Searle Corp., Arlington Heights, IL at the highest available specific activity. *Escherichia coli* RNA polymerase and ^{7}mGpppA were purchased from P-L Biochemicals, Inc., Milwaukee, WI. Restriction enzymes were purchased from BRL, Gaithersburg, MD or New England Biolabs, Beverly, MA, and used as recommended by the manufacturers.

Methods

Construction of an Islet cDNA Library. A pancreatic islet cDNA library was constructed from anglerfish (*Lophius americanus*) polyA-containing mRNA exactly as described by Gubler and Hoffman (8). Bacterial transformants were screened for ppI inserts by colony hybridization (6) using a nick-translated 220-bp PstI fragment of a partial ppI cDNA clone, pAFI-I1, which was isolated from a previous cDNA library (27). This clone had been shown to be specific for ppI by hybrid-select translation of anglerfish mRNA. Several positive clones were analyzed in detail by restriction mapping and by dideoxy DNA sequencing (18); one such clone, designated p413-I1, was found to contain full-length ppI cDNA.

Construction of ppI-CAT Hybrid Genes. A scheme for the construction of the hybrid cDNAs is outlined in Fig. 1. Clone p413-I1 was digested with PstI, and the 220-bp fragment, encoding the ppI signal peptide, B-chain, and a portion of the C-peptide, was purified by polyacrylamide gel electrophoresis. The Pst I fragment was blunt-ended by digestion with T4 DNA polymerase, and EcoRI linkers (12 bp) were added. 1 μg of this PstI/EcoRI fragment was digested with SalI, and the 5′ overhang was filled in with the Klenow fragment of DNA polymerase I and ligated to EcoRI linkers (12 bp) (Fig. 1 *A*). Both the PstI fragment and the SalI fragment containing EcoRI ends were ligated into pDS5 that had been digested with EcoRI and alkaline phosphatase (Fig. 1 *B*). The resulting plasmids, p5PI.CAT and p5SI.CAT, encode chimeric ppI-CAT fusions designated ppPI.CAT and ppSI.CAT, which have 313 and 280 amino acids, respectively. ppPI.CAT contains the ppI signal peptide, B-chain, and 17 residues of the C-chain, as well as 23 residues from the polylinker of pDS5 (25), and the entire CAT sequence (Fig. 1 *B*, top). ppSI.CAT possesses the ppI signal peptide, 14 amino acids of the B-chain, 23 residues of the polylinker, and the entire CAT sequence (Fig. 1 *B*, bottom). The predicted amino acid sequence encoded by the pDS5 polylinker in both plasmids is: Arg-Asn-Ser-Arg-Gly-Ser-Val-Asp-Leu-Gln-Pro-Ser-Leu-Ala-Arg-Phe-Ser-Gly-Ala-Lys-Glu-Ala-Lys.

In Vitro Transcription. DNA from the appropriate plasmids was isolated by CsCl centrifugation and was transcribed in vitro using *E. coli* RNA polymerase and the cap analogue ^{7}mGpppA exactly as described by Stueber et al. (25).

In Vitro Translation. Cell-free translation of in vitro transcribed mRNA was performed as previously described (20) using the wheat germ cell-free system containing 800 μCi/ml [^{35}S]cysteine and 1 mCi/ml [^{3}H]leucine. The isolation of anglerfish islet mRNA and canine microsomal membranes and their use in the wheat germ system was as previously described (4, 20, 21). Immunoprecipitation of the translation products was as described (13) with the following modifications: aliquots of the translation products were adjusted to 2% SDS, 4 mM L-cysteine, and incubated at 42°C for 5 min, followed by addition of 5 vol of immunoprecipitation buffer (13). The appropriate antiserum (3 μl anti-CAT, 8 μl anti-insulin) was added, and the samples incubated at 4°C overnight. Immunoprecipitates were treated with protein A Sepharose and washed four times. The final pellet was resuspended in SDS PAGE loading buffer and incubated at 60°C for 3 min, followed by alkylation and analysis by SDS PAGE on 15% polyacrylamide gels.

Assay for Translocation of CAT-related Polypeptides into Membrane Vesicles. Resistance to posttranslational proteolysis was used to assay for the segregation of nascent pI-CAT fusions into microsomal membrane vesicles. Aliquots of the translation products synthesized in the absence and presence of microsomal membranes were adjusted to 3.6 mM tetracaine and digested with 250 μg/ml each of trypsin and chymotrypsin as previously described (21). After incubation for 1 h at 0°C, the digestions were terminated by adjusting the samples to 2 mM PMSF and 800 U/ml Trasylol. Samples were then prepared for SDS PAGE.

Endoglycosidase H (EndoH). Glycosylation of pI-CAT fusions was assayed by sensitivity to digestion with EndoH. Translation products synthesized in the presence of microsomal membranes were adjusted to 1% SDS in a total volume of 18 μl and incubated at 95°C for 3 min. 6 μl of 0.1 M citrate phosphate buffer, pH 5.0, were added, followed by 1 μl of EndoH (0.04 U); 1 μl of water was added to control samples. Incubations were for 16–20 h at 37°C; the digestion was terminated by incubating the samples at 95°C for 3 min followed by precipitation in 10% cold TCA-containing 2 mM cysteine. The TCA pellets were resuspended in SDS gel loading buffer and analyzed by SDS PAGE.

Partial NH_2-terminal Sequencing of the Fusion Proteins. Appropriate bands were located by autoradiography, excised from the dried gel, and subjected to up to 40 cycles of automated Edman degradation using a spinning cup sequencer (model 890C; Beckman Instruments Inc., Fullerton, CA), as previously described (20, 22), with the following modification. Radiolabeled polypeptides were electrophoretically eluted as follows: the gel slices were rehydrated in electrophoresis tank buffer (0.05 M Tris, 0.38 M glycine) containing 2% SDS; the rehydrated gel pieces were placed in an electrophoretic concentrator elution chamber (model 1750; Isco, Inc., Lincoln, NE) containing 0.01 M Tris, 0.077 M glycine, and 0.1% SDS; the apparatus was filled with electrophoresis tank buffer containing 0.1% SDS. Electroelution was performed at 1 W (constant power) overnight. The eluate (200 μl) was diluted with an equal volume of sterile water, dialyzed briefly against water, and loaded into the spinning cup of the sequencer containing 4 mg polybrene and 1 mg of myoglobin (27).

Generation of p5PI.CAT and p5SI.CAT. pDS5 was digested with EcoRI and dephosphorylated with calf intestinal phosphatase (CIP). The vector was then ligated with either the PstI fragment or the SalI fragment, both containing EcoRI cohesive ends, and the ligation reaction used to transform *E. coli* strain MC1000. Positive clones were identified by colony hybridization using the nick-translated PstI fragment of p413-I1 as a probe. Those clones containing either the PstI or SalI fragment were amplified and plasmid DNA was prepared. The plasmids p5PI.CAT and p5SI.CAT were subjected to coupled transcription and translation, yielding the two fusion proteins ppPI.CAT and ppSI.CAT. bla, β-lactamase gene; SP, signal peptide; B, B-chain; C, C-peptide; P, PstI; S, SalI. Hatched bar in p413-I1 represents ppI coding region; open bar represents ppI coding region in the fragments; filled-in region represents EcoRI linkers. L, polylinker region of pDS5; Pro, coliphage promoter; RBS, prokaryotic ribosome binding site; t_i, transcription terminator of the rnnB operon in *E. coli*; and E, EcoRI. Arrowhead indicates the signal cleavage site; triangle indicates the cryptic site for N-linked glycosylation in the CAT protein; double asterisks indicate the paired basic cleavage site Lys-Arg between the B-chain and the C-peptide of ppPI.CAT.

Results

Construction of ppI-CAT Fusions

The construction of the two ppI-CAT fusion proteins is diagrammed in Fig. 1. Plasmid p5PI.CAT encodes a hybrid protein of 313 amino acids comprising the anglerfish ppI signal peptide (24 residues), the complete B-chain (30 residues), the first 17 residues of the connecting peptide (C-peptide), a 23 amino acid peptide encoded by the linker region of pDS5 (see Materials and Methods), and the complete CAT sequence (219 amino acids). Plasmid p5SI.CAT encodes a hybrid protein of 280 amino acids possessing the complete ppI signal peptide, the first 14 residues of the insulin B-chain, the same 23 residue polypeptide encoded by the linker region of pDS5, and the complete CAT sequence. In addition, both plasmids encode a cryptic site for N-linked glycosylation (Asn-Gln-Thr) present at residues 34 through 36 of the native CAT molecule; this sequence starts at position 128 in ppPI.CAT and at residue 95 in ppSI.CAT. The structure of these plasmids was confirmed by detailed restriction digests, DNA sequencing, and partial amino acid sequence analysis (Fig. 5) of the translation products.

In Vitro Biosynthesis of Fusion Proteins

Initially we determined if the NH_2-terminus of ppI could direct the translocation of CAT into mammalian microsomal membranes. To this end, RNA transcribed in vitro from plasmids p5PI.CAT and p5SI.CAT was translated in the wheat germ cell-free system in the absence and presence of microsomal membranes (Fig. 2). In the absence of microsomes the translation products encoded by both plasmids p5PI.CAT and p5SI.CAT were significantly larger than anglerfish ppI or native CAT (Fig. 2 *A*, lanes *1*, *3*, *5*, and *7*) and were of the expected size for the two predicted fusion proteins, i.e., M_r 31,000 for ppPI.CAT and M_r 28,000 for ppSI.CAT. To confirm that these products were fusion proteins between ppI and CAT, the translation products were treated with antibodies directed against porcine insulin or CAT (Fig. 2, *B* and *C*, lanes *1* and *3*, respectively). Both ppPI.CAT and ppSI.CAT were immunoprecipitated with anti-insulin and anti-CAT antibodies, indicating they had the predicted antigenic determinants. Some cross-reactivity between the anti-insulin serum and the CAT translation products generated from transcription of the parent vector pDS5 was noted (Fig. 2 *B*, lanes *5* and *6*). This cross-reactivity, which was variable with different batches of antisera, is most likely due to the presence of trace levels of endogenous CAT antigen, synthesized by *B. subtilis*, a component of Freund's complete adjuvant used for the initial immunization of the animals. Consequently, the final serum may contain a low level of CAT antibodies in addition to anti-insulin antibodies. Both ppI and pI (Fig. 2 *B*, lanes *7* and *8*) were efficiently recognized by the anti-insulin antibody, while, as expected, the anti-CAT antibodies showed no cross-reactivity with ppI or pI (Fig. 2 *C*, lanes *7* and *8*).

In the presence of microsomal membranes, nascent ppI was cotranslationally cleaved to pI (Fig. 2 *B*, lanes *7* and *8*; reference 20). Surprisingly, the translation products from ppPI.CAT and ppSI.CAT mRNA synthesized in the presence of membranes were processed to two forms. One form was of slightly faster mobility (Fig. 2, *A*, *B*, and *C*; lanes *2* and

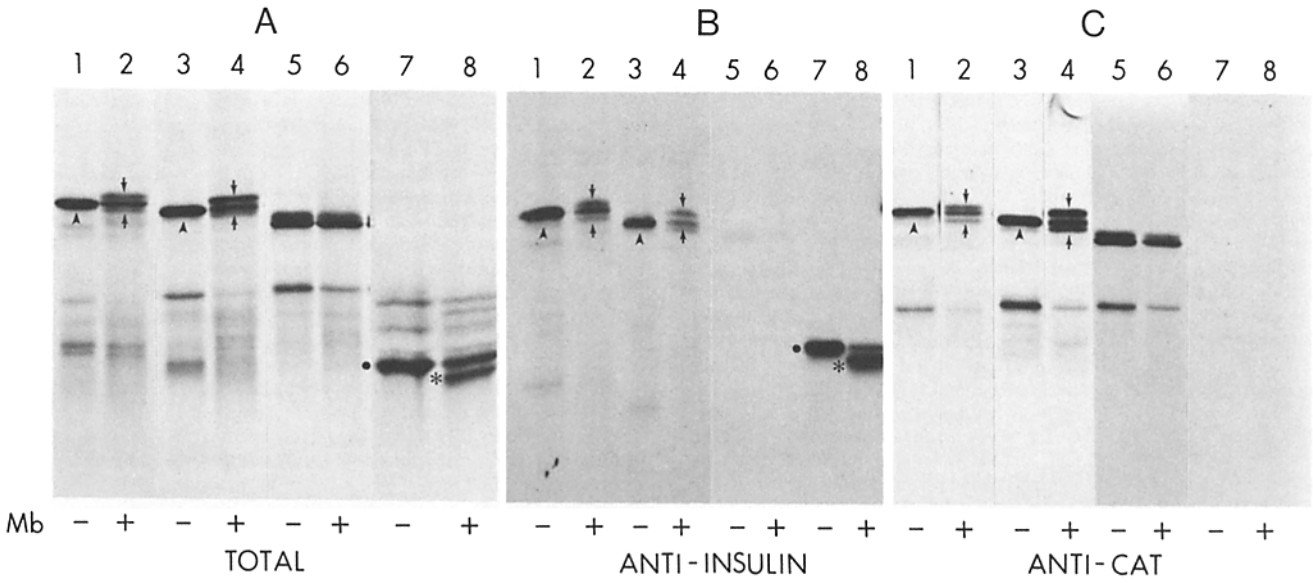

Figure 2. Characterization of ppI-CAT fusions. Plasmids p5PI.CAT, p5SI.CAT, and pDS5 were transcribed in vitro (25) and the RNA translated in the wheat germ system in the absence (−) or presence (+) of 2.5 A_{260}/ml microsomal membranes (Mb). Total islet mRNA (lanes *7* and *8*) was translated under identical conditions (see Materials and Methods). After translation, aliquots were either analyzed directly by electrophoresis on 15% acrylamide SDS gels (*A*) or treated with anti-insulin antibodies (*B*) or anti-CAT antiserum (*C*) followed by SDS PAGE. The samples in lanes *1–8* are identical in each panel. Translation of products of p5PI.CAT, lanes *1* and *2*; lane *1*, ppPI.CAT, arrowhead; lane *2*, pPI.CAT, upward pointing arrowhead indicates fusion protein lacking its signal peptide, downward pointing arrow indicates putative glycosylated form of fusion protein. Translation of products of p5SI.CAT, lanes *3* and *4*; lane *3*, ppSI.CAT, arrowhead; lane *4*, pSI.CAT, upward arrow indicates fusion protein lacking its signal peptide, downward pointing arrow indicates glycosylated form of fusion protein. Translation products of pDS5-CAT, lanes *5* and *6*. Translation products of anglerfish islet mRNA, lanes *7* and *8*; dot (lane *7*) ppI; asterisk (lane *8*), pI (20).

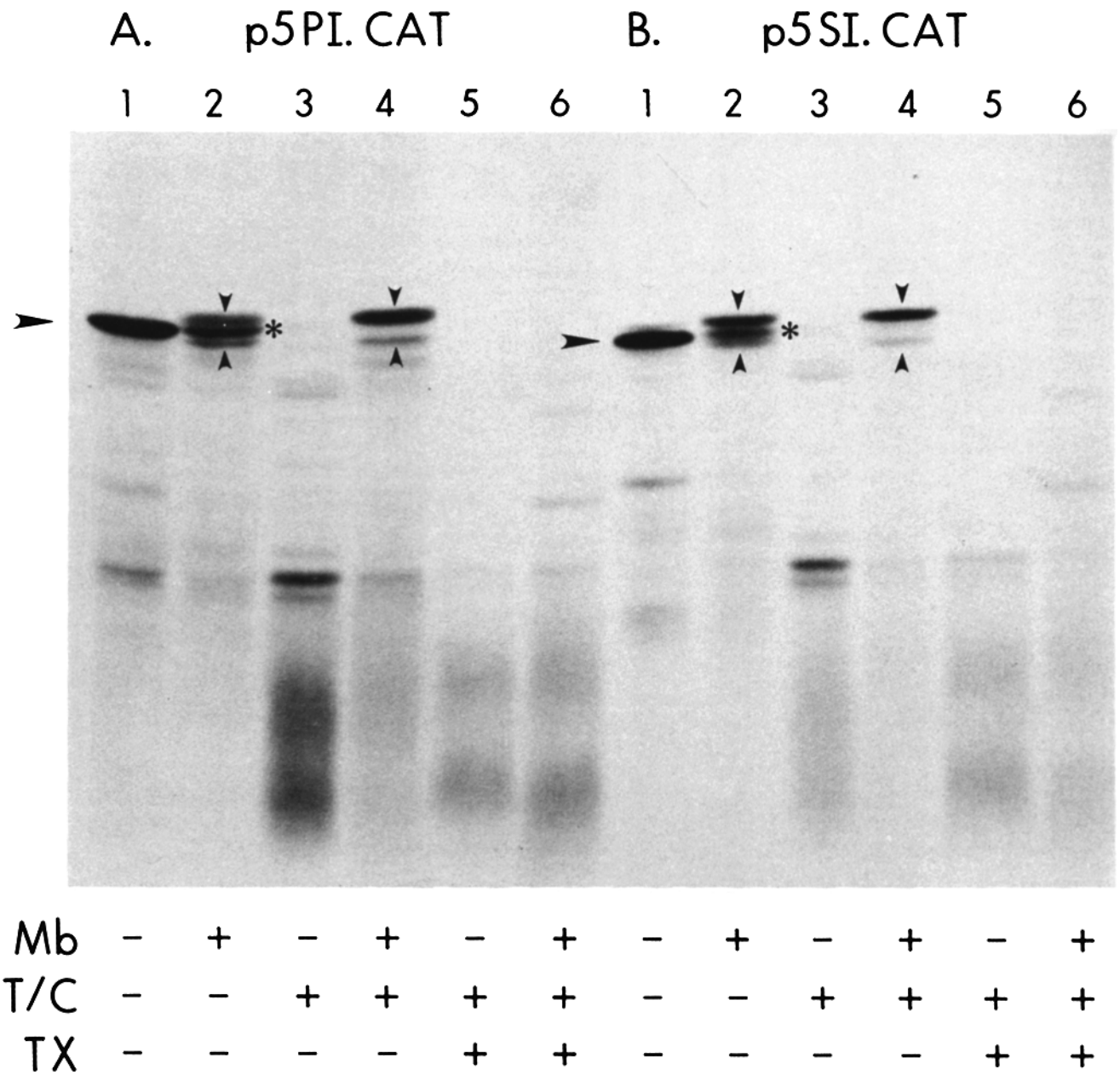

Figure 3. Segregation of ppI-CAT fusions. Plasmids p5PI.CAT and p5SI.CAT were transcribed in vitro and the RNA translated in the absence (−) and presence (+) of 3 A_{260}/ml microsomal membranes. After translation, aliquots were adjusted to 10 mM $CaCl_2$, 3 mM tetracaine and treated with either protease (T/C, 250 μg/ml each of trypsin and chymotrypsin) or with protease in the presence of 1% Triton X-100 (TX). Samples were incubated for 1 h at 4°C and then treated with 2 mM PMSF and prepared for SDS PAGE. (*A*) Products of p5PI.CAT. (*B*) Products of p5SI.CAT. Lane *1*, products synthesized in the absence of membranes; lane *2*, products synthesized in the presence of membranes; lanes *3* and *4*, as lanes *1* and *2* but treated with trypsin and chymotrypsin; lanes *5* and *6*, as *3* and *4* except that proteolysis was performed in the presence of 1% Triton X-100. Large arrowhead (left of lanes *A 1* and *B 1*), ppPI.CAT and ppSI.CAT, respectively. Lanes *2* and *4*, downward and upward pointing arrowheads, protease resistant forms of fusion protein; lane *2*, asterisks, residual ppPI.CAT and ppSI.CAT, respectively.

4, lower arrow) and was presumably the fusion protein minus its signal peptide. The second form of processed fusion protein migrated more slowly than the precursor (Fig. 2, *A*, *B*, and *C*; lanes *2* and *4*, downward pointing arrows). No such processed forms of CAT were seen when native CAT, encoded by pDS5, was synthesized in the absence or presence of microsomal membranes (Fig. 2, *A* and *C*, lanes *5* and *6*). The translation products from pDS5 appeared to migrate as a doublet on SDS gels; the reason for this is unclear. However, the appearance of this doublet was unaffected by the presence or absence of microsomal membranes, indicating that it is not due to the incorporation of CAT into microsomes.

Segregation of the Fusion Proteins

To further analyze the nature of the fusion proteins synthesized in the presence of microsomal membranes, the translation products were assayed for translocation into the microsomal vesicles by determination of their sensitivity to protease digestion (Fig. 3). Aliquots of the translation prod-

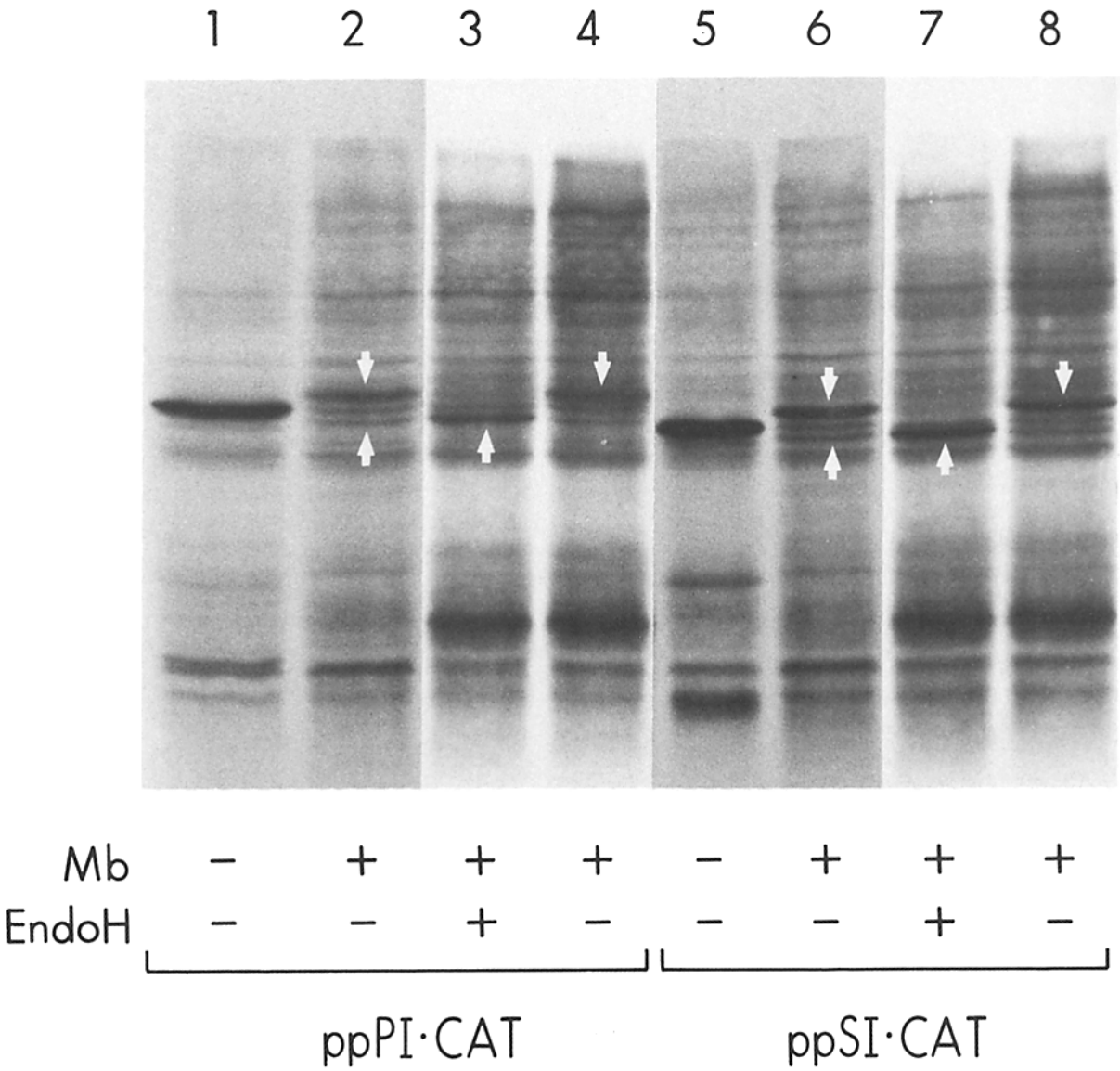

Figure 4. N-linked glycosylation of ppI–CAT fusions. Plasmids p5PI.CAT and p5SI.CAT were transcribed and translated as outlined in Figs. 2 and 3. At the end of the incubation, samples were made 1% SDS and incubated for 3 min at 95°C. Samples were then adjusted to 25 mM citrate, pH 5.0, and digested with 0.04 U EndoH for 16 h at 37°C. Lane *1*, ppPI.CAT synthesized in the absence of membranes; lane *2*, pPI.CAT synthesized in the presence of microsomal membranes; lane *3*, as lane *2* but treated with EndoH; lane *4*, as lane *3*, not treated with EndoH; lane *5*, ppSI.CAT synthesized in the absence of membranes; lane *6*, pSI.CAT synthesized in presence of membranes; lane *7*, as lane *6* but treated with EndoH; lane *8*, as lane *7* but not treated with EndoH. (*Downward pointing arrows*) glycosylated form of fusion protein. (*Upward pointing arrows*) nonglycosylated forms of fusion protein lacking its signal peptide.

ucts were treated with a mixture of trypsin and chymotrypsin in the absence and presence of the detergent Triton X-100. Fusion proteins ppPI.CAT and ppSI.CAT synthesized in the absence of membranes were completely sensitive to protease treatment (Fig. 3, *A* and *B*, lane *3*) indicating that these proteins were not intrinsically protease-resistant. In contrast, the two putative processed forms of each fusion protein, pPI.CAT and pSI.CAT, synthesized in the presence of membranes, were protease-resistant (Fig. 3, *A* and *B*, lane *4*), indicating that they were shielded by the membrane bilayer. This was confirmed when proteolysis was performed in the presence of Triton X-100. In this case, these products were completely digested (lane *6*). These data indicate that the two processed forms of each fusion protein were completely segregated into the cisternae of the microsomal vesicles.

Glycosylation of ppI-CAT Fusions

The appearance of a slower migrating form of processed fusion proteins, resulting from synthesis in the presence of microsomal membranes, was similar to that seen for numerous glycoproteins synthesized in vitro (e.g., VSV-G protein (13)). We therefore hypothesized that the molecules of slower electrophoretic mobility were glycosylated forms of the fusion proteins, whereas those migrating faster than ppPI.CAT or ppSI.CAT corresponded to processed molecules in which

the insulin signal peptide was cleaved but which were not glycosylated. Although neither ppI nor CAT are normally glycosylated in vivo, CAT does encode a cryptic recognition site for N-linked glycosylation: Asn-Gly-Thr, at residues 34–36 of the native molecule. This site appears at residues 128–130 in ppPI.CAT and residue 95–97 in ppSI.CAT. Since ppI sequences in both fusions mediated translocation of CAT into microsomal vesicles, it is possible that this site was recognized by the glycosylation enzymes in the microsomal membranes. The prediction that the higher molecular weight forms of the fusion proteins (Fig. 4, lanes *2* and *6*) were glycoproteins was tested by subjecting the translation products to treatment with EndoH. Since high mannose core oligosaccharides added to nascent glycoproteins in the ER can be cleaved by EndoH digestion, the glycosylated fusion proteins should be sensitive to this enzyme. EndoH digestion was performed on translation products generated from p5PI.CAT and p5SI.CAT synthesized in the presence of microsomal membranes, and the samples were analyzed by SDS PAGE (Fig. 4). In the presence of EndoH (lanes *3* and *7*), the slower migrating species of both fusions (upper arrows) was quantitatively converted to a form which then co-migrated with the faster migrating species observed in the presence of membranes (lanes *2* and *6*, lower arrows); i.e., after removal of the single N-linked carbohydrate chain, the resulting product co-migrated with a protein that represented the fusion protein minus its putative signal peptide.

During the course of these experiments, we noted that both proteolytic processing of the signal peptide and glycosylation of pPI.CAT and pSI.CAT were particularly efficient, as judged by autoradiographic intensity (Figs. 3 and 4). To quantitate the relative efficiencies of proteolytic processing and glycosylation, gel bands corresponding to both the glycosylated and unglycosylated forms of each fusion protein were excised from the dried gels, solubilized, and their radioactivity determined directly (Table I). In several experiments in which 43% of nascent ppI was processed to pI, the fusion proteins were processed with an efficiency of 65–70%. Similarly, glycosylation of the fusion proteins was particularly efficient, ~45% of pPI.CAT and 68% of pSI.CAT were glycosylated. These values are comparable to the efficiency of glycosylation of native glycoproteins synthesized under these conditions (data not shown).

Table I. Efficiency of Processing and Glycosylation of ppI–CAT Fusions

Protein	Processing*	Glycosylation‡
	%	%
ppI	43.4 (5)	NA
ppPI.CAT	66.5 (3)	45.6 (3)
ppSI.CAT	69.6 (3)	68.0 (3)

The appropriate polypeptides were excised from the dried gel (Fig. 4) solubilized in 30% H_2O_2 (4), and the radioactivity determined by liquid scintillation counting. Numbers in parenthesis represent total number of experiments. NA, not applicable.

$$^{*}\text{Processing} \frac{\text{cpm in processed forms (upper and lower)}}{\text{(cpm in precursor + cpm in processed forms)}} \times 100$$

$$^{\ddagger}\text{Glycosylation} \frac{\text{cpm in upper processed form}}{\text{(cpm in lower + cpm in upper forms)}} \times 100$$

Partial NH_2-terminal Sequencing of the Fusion Proteins

The antibody precipitation data (Fig. 2) indicated that both ppPI.CAT and ppSI.CAT were fusion proteins between ppI and CAT. However, since the anti-insulin antibodies had some cross-reactivity with authentic CAT, it was necessary to unequivocally demonstrate that the fusion proteins were indeed those predicted. We therefore subjected ppPI.CAT and ppSI.CAT to partial NH_2-terminal sequencing. Previous studies (9, 22) had shown that leucine residues were present at positions 3, 5, 10, 12, 13, and 14 of the signal peptide, and leucine and cysteine were present at positions 30 and 31, respectively, of ppI (corresponding to residues 7 and 8 of the insulin B-chain). Consequently, ppPI.CAT and ppSI.CAT were synthesized in the presence of [^{3}H]leucine and [^{35}S]cysteine and electrophoresed on 15% polyacrylamide gels. The appropriate bands were localized by autoradiography, the proteins eluted and subjected to microsequencing (Fig. 5, *A* and *B*). Leucine and cysteine residues were found at the expected positions; these results not only confirm the accuracy of the antibody data but conclusively demonstrate that the fusion proteins contained the ppI signal peptide.

It is possible that foreign sequences downstream from the signal peptide might influence the site of signal peptidase cleavage (1). Therefore it was of interest to determine if both of the precursor fusion proteins were correctly cleaved by signal peptidase, particularly since ppSI.CAT contained only fourteen residues of the insulin B-chain. The polypeptides corresponding to the glycosylated forms of pPI.CAT and pSI.CAT (Fig. 4, upper band) synthesized in the presence of [^{3}H]leucine, [^{35}S]cysteine and microsomal membranes, were eluted from gels and also subjected to microsequencing, (Fig. 5, *C* and *D*). Previously, it had been shown that cleavage of the signal peptide occurred between residues 24 and 25 of anglerfish ppI (20). Consequently, if ppPI.CAT were accurately cleaved by signal peptidase, leucine residues would be present at positions 7, 12, 16, and 18, and cysteine at residues 8 and 20. The data (Fig. 5 *C*) show that this was the case and is consistent with correct cleavage of the signal peptide. The sequence data from the glycosylated form of pSI.CAT (Fig. 5 *D*) also demonstrated leucine residues at positions 7 and 12 and a cysteine residue at position 8. In addition, novel leucine residues were found at positions 23 and 27. These leucines correspond to those predicted from DNA sequencing of the linker region from pDS5. Most importantly, the data demonstrate that the signal peptide of ppSI.CAT was also cleaved at the same position as in the native precursor, even though the fusion protein contains only fourteen residues of the insulin B-chain. Thus, our results demonstrate that the downstream sequences do not influence the site of cleavage of the ppI signal peptide.

Discussion

Most small polypeptide hormones (less than ~50 amino acids) are synthesized as part of a larger precursor molecule; in some cases the precursor may also be a polyprotein containing repeating units of the same peptide or several different hormones (3). We are attempting to decode putative sorting information that may be present in the proregions of a variety of diverse peptide hormone precursors. To this end, we have synthesized chimeric genes encoding variable amounts of the NH_2-terminus of ppI fused to CAT, a bac-

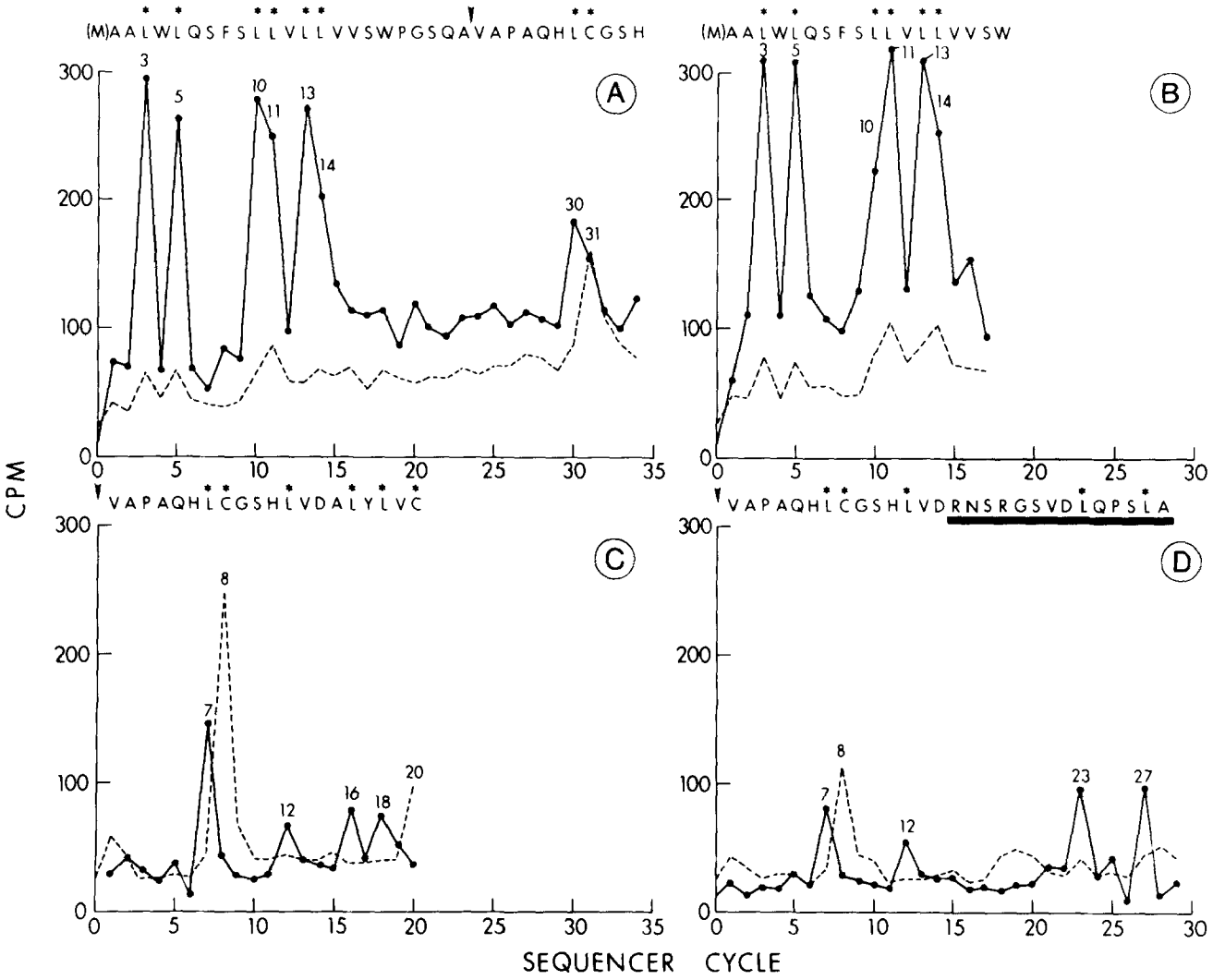

Figure 5. Partial NH_2-terminal sequence of ppI–CAT and pI–CAT fusions. Plasmids p5PI.CAT and p5SI.CAT were transcribed and translated in the wheat germ system containing both [^{35}S]cysteine and [^{3}H]leucine (1 mCi/ml of each) in the absence and presence of microsomal membranes and the translation products resolved by SDS PAGE. After autoradiography, bands corresponding to ppPI.CAT (*A*), ppSI.CAT (*B*), the glycosylated forms of pPI.CAT (*C*), and pSI.CAT (*D*) were excised and the polypeptides electrophoretically eluted. The eluted samples were applied to a sequencer (model 890C; Beckman Instruments, Inc.) and subjected to automated Edman degradation. (*Solid line*) [^{3}H]leucine; (*dashed line*) [^{35}S]cysteine; (*asterisks*) known or predicted leucine and cysteine residues that were confirmed in this analysis; (*arrowhead*; *A*, *C*, and *D*) site of signal peptide cleavage (20). (*Heavy black line*; *D*) indicates novel amino acids encoded by the linker region of pDS5 (see Materials and Methods). Single letter code: A, Ala; C, Cys; D, Asp; F, Phe; G, Gly; H, His; L, Leu; M, Met; N, Asn; P, Pro; Q, Glu; R, Arg; S, Ser; V, Val; W, Trp; Y, Tyr.

terial cytoplasmic enzyme. The rationale for these experiments is to identify sequence information within ppI that could mediate the sequestration of CAT molecules into the secretory pathway, perhaps ultimately leading to its secretion. Since we had shown (4) that ppI interacts with the ER when about half the molecule has been synthesized, we postulated that sorting information in the first 60 residues of ppI, which includes the NH_2-terminal signal peptide, should be sufficient to effect translocation of any protein across the ER membrane.

To test this hypothesis directly, we have constructed two fusion proteins. One, ppPI.CAT, contained the first 71 residues of ppI, including the complete signal peptide and B-chain and part of the C-peptide fused to CAT. The other fusion protein, ppSI.CAT, possessed only 38 amino acids of ppI (the signal peptide plus 14 residues of the B-chain) fused to CAT. It is noteworthy that the polylinker sequence, present in both fusion proteins, contained seven charged residues; consequently, it might be expected that these could interfere with membrane translocation. However, both fusions, ppPI.CAT and ppSI.CAT, were capable of targeting to microsomal membranes, as well as translocating through the lipid bilayer, suggesting that local charge effects per se may not necessarily inhibit translocation. Since the relative efficiencies of glycosylation and signal peptide cleavage were virtually identical in the two constructions (Table I), it is possible that ppSI.CAT, as well as ppPI.CAT, contains all the necessary structural domains to effect efficient translocation of foreign proteins into the lumen of the ER. This result was somewhat surprising, since a construction of 45 residues, which encodes only the first 38 amino acids of ppI and no CAT sequences, bound poorly to microsomal membranes and was

incompetent for translocation (Eskridge, E., and D. Shields, manuscript submitted for publication). These results suggest that, providing the 38 NH_2-terminal residues of ppI can assume an appropriate conformation, by virtue of being fused to foreign sequences, in this case CAT, targeting and translocation domains in the nascent precursor can be recognized by the ER translocation machinery.

Our data contrast with results seen from an analogous construction in prokaryotes. When the signal peptide and the first 15 amino acids of *LamB* were fused to the *LacZ* gene and expressed in *E. coli*, β-galactosidase remained in the cytoplasm, suggesting that additional *LamB* sequences are necessary to direct insertion of the hybrid protein into the inner membrane (15). Similarly, when the β-lactamase signal peptide was fused to chicken triosephosphate isomerase (11) only 30% of the fusion protein was targeted to the membrane in vivo; no signal peptide cleavage or translocation across the membrane was observed. At present we do not know if the insulin signal sequence alone is sufficient to mediate translocation of CAT into microsomal vesicles of if the fourteen residues of the B-chain are also required; these experiments are currently in progress.

It is noteworthy that the downstream sequences in both fusion proteins had no influence on the specificity of cleavage by signal peptidase. This was particularly striking in the case of ppSI.CAT, which contained only 14 amino acids of the B-chain fused to 242 foreign residues. Nevertheless, cleavage of the ppI signal peptide occurred at precisely the same site as in the native precursor, i.e., between Ala_{24} and Val_{25}; this was also the case for the ppPI.CAT fusion. These results contrast with recent data on the signal peptide cleavage of *Staphylococcus aureus* protein A (1). In this case, when an internal IgG-binding fragment of protein A was inserted immediately adjacent to the signal sequence (replacing the normal sequence), incorrect cleavage of the signal peptide was observed and transport into the periplasm was significantly less efficient than for wild-type protein A (1). These data suggest that in prokaryotic cells, at least gram-positive bacteria, the structure of the polypeptide chain distal to the site of signal cleavage may affect proteolytic processing by signal peptidase.

Several experiments have demonstrated that prokaryotic signal sequences, e.g., *E. coli* β-lactamase, can be efficiently recognized by the eukaryotic translocation apparatus, such that proteolytically processed β-lactamase was sequestered into mammalian microsomal membranes in vitro (12, 16). Lingappa et al. (14) also showed that the signal peptide and the first five amino acids of β-lactamase were sufficient to effect translocation of normally cytoplasmic globin chains into the lumen of microsomal membranes. In contrast to studies on prokaryotic cells (15), these results indicated that relatively little structural information other than the signal sequence may be needed to effect translocation of a protein across the ER membrane. Indeed, very recent experiments (17) suggest that a precise fusion of the β-lactamase signal peptide to appropriately engineered α-globin chains is sufficient to mediate translocation into the ER vesicles. Our results demonstrate that a eukaryotic signal sequence mediates translocation of a prokaryotic cytoplasmic protein across mammalian membranes. As such, this should enable us to distinguish between topogenic sequences sufficient for translocation into the ER lumen from those putative domains needed to mediate distal sorting events in the eukaryotic secretory pathway.

Two recent reports have also demonstrated cryptic glycosylation of normally nonglycosylated proteins (19, 23). Spiess and Lodish (23) showed that fusion of the membrane-anchor domain of the asialoglycoprotein receptor to rat α-tubulin was sufficient to mediate translocation and glycosylation of normally cytoplasmic tubulin by microsomal membranes in vitro. Sharma et al. (19) constructed a chimeric gene comprising the influenza virus hemagglutinin signal peptide and SV40 large T antigen. Expression of this gene in 3T3 cells resulted in T antigen being exclusively localized to the ER, where both signal peptide cleavage and core oligosaccharide addition occurred. However, T antigen remained in the ER and was not transported through the secretory pathway. In this context, the ppI-CAT fusions described here, as well as several under construction, offer an excellent experimental system in which to investigate the function of the pI B-chain and C-peptides, as well as N-linked glycosylation, on the secretion of the CAT enzyme from several different cell types. In particular it will be of interest to determine if glycosylation plays a role in either the translocation of CAT molecules from the ER to the Golgi, c.f. yeast proalpha factor (10), or in post-Golgi sorting events (7). These experiments are currently in progress.

We thank Mr. M. Brenner for help with the sequencing experiments, Dr. Reza Green for helpful suggestions, and Ms. R. Romita for typing the manuscript.

This work was supported by National Institutes of Health grant AM-21860. Dr. Shields was supported by Career Development Award AM-01208 and a Hirschl Scientist Award. Dr. Eskridge was a predoctoral fellow of the National Science Foundation and is supported by training grant T32AM-07513.

Received for publication 20 June 1986, and in revised form 6 September 1986.

References

1. Abrahmsen, L., T. Moks, B. Nilsson, U. Hellman, and M. Uhlen. 1985. Analysis of signals for secretion in the staphylococcal protein A gene. *EMBO (Eur. Mol. Biol. Organ.) J.* 4:3901–3906.

2. Blundell, T. L., J. E. Pitts, and S. P. Wood. 1982. The conformation and molecular biology of pancreatic hormones and homologous growth factors. *CRC Crit. Rev. Biochem.* 13:141–213.

3. Douglass, J., O. Civelli, and E. Herbert. 1984. Polyprotein gene expression: generation of diversity of neuroendocrine peptides. *Annu. Rev. Biochem.* 53:665–715.

4. Eskridge, E., and D. Shields. 1983. Cell-free processing and segregation of insulin precursors. 1983. *J. Biol. Chem.* 258:11487–11491.

5. Garoff, H. 1985. Using recombinant DNA techniques to study protein targeting in the eucaryotic cell. *Annu. Rev. Cell Biol.* 1:403–445.

6. Gergen, J. P., R. H. Stern, and P. C. Wensink. 1979. Filter replicas and permanent collections of recombinant DNA plasmids. *Nucleic Acids Res.* 7:2115–2136.

7. Guan, J.-L., C. E. Machamer, and J. K. Rose. 1985. Glycosylation allows cell-surface transport of an anchored secretory protein. *Cell.* 42:489–496.

8. Gubler, U., and B. J. Hoffman. 1983. A simple and very efficient method for generating cDNA libraries. *Gene. (Amst.).* 25:263–269.

9. Hobart, P. M., L.-P. Shen, R. Crawford, R. Pictet, and W. J. Rutter. 1980. Comparison of the nucleic acid sequence of anglerfish and mammalian insulin mRNA's from cloned cDNA's. *Science (Wash. DC).* 210:1360–1363.

10. Julius, D., R. Schekman, and J. Thorner. 1984. Glycosylation and processing of prepro-α-factor through the yeast secretory pathway. *Cell.* 36:309–318.

11. Kadonaga, J. T., A. E. Gautier, D. R. Straus, A. D. Charles, M. D. Edge, and J. R. Knowles. 1984. The role of the β-lactamase signal sequence in the secretion of proteins by *Escherichia coli*. *J. Biol. Chem.* 259:2149–2154.

12. Kronenberg, H. M., B. J. Fennick, and T. J. Vasicek. 1983. Transport and cleavage of bacterial pre-β-lactamase by mammalian microsomes. *J. Cell Biol.* 96:1117–1119.

13. Lingappa, V. R., F. N. Katz, H. F. Lodish, and G. Blobel. 1978. A sig-

nal sequence for the insertion of a transmembrane glycoprotein. *J. Biol. Chem.* 253:8667–8670.

14. Lingappa, V. R., J. Chaidez, C. S. Yost, and J. Hedgpeth. 1984. Determinants for protein localization: β-lactamase signal sequence directs globin across microsomal membranes. *Proc. Natl. Acad. Sci. USA.* 81:456–460.

15. Moreno, F., A. V. Fowler, M. Hall, T. J. Silhavy, I. Zabin, and M. Schwartz. 1980. A signal sequence is not sufficient to lead β-galactosidase out of the cytoplasm. *Nature (Lond.).* 286:356–359.

16. Muller, M., I. Ibrahimi, C. N. Chang, P. Walter, and G. Blobel. 1982. A bacterial secretory protein requires signal recognition particle for translocation across mammalian endoplasmic reticulum. *J. Biol. Chem.* 257:11860–11863.

17. Perera, E., R. E. Rothman, V. R. Lingappa. 1986. Uncoupling translocation from translation: implications for transport of proteins across membranes. *Science (Wash. DC).* 232:348–352.

18. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463–5467.

19. Sharma, S., L. Rodgers, J. Brandsma, M.-J. Gething, and J. Sambrook. 1985. SV40 T antigen and the exocytotic pathway. *EMBO (Eur. Mol. Biol. Organ.) J.* 4:1479–1489.

20. Shields, D., and G. Blobel. 1977. Cell-free synthesis of fish preproinsulin, and processing by heterologous mammalian microsomal membranes. *Proc. Natl. Acad. Sci. USA.* 74:2059–2063.

21. Shields, D., and G. Blobel. 1978. Efficient cleavage and segregation of nascent presecretory proteins in a reticulocyte lysate supplemented with microsomal membranes. *J. Biol. Chem.* 253:3753–3756.

22. Shields, D. 1981. Cell-free synthesis of Angler fish preproinsulin: complete amino acid sequence of the signal peptide. *Biochem. Biophys. Res. Commun.* 98:242–249.

23. Spiess, M., and H. F. Lodish. 1986. An internal signal sequence: the asialoglycoprotein receptor membrane anchor. *Cell.* 44:177–185.

24. Steiner, D. F., S. J. Chan, J. M. Welsh, and S. C. M. Kwok. 1985. Structure and evolution of the insulin gene. *Annu. Rev. Genet.* 19:463–484.

25. Stueber, D., I. Ibrahimi, D. Cutler, B. Dobberstein, and H. Bujard. 1984. A novel *in vitro* transcription-translation system: accurate and efficient synthesis of single proteins from cloned DNA sequences. *EMBO (Eur. Mol. Biol. Organ.) J.* 3:3143–3148.

26. Walter, P., R. Gilmore, and G. Blobel. 1984. Protein translocation across the endoplasmic reticulum. *Cell.* 38:5–8.

27. Warren, T. G., and D. Shields. 1984. Cell-free biosynthesis of multiple prepromatostatins: characterization by hybrid selection and amino-terminal sequencing. *Biochemistry.* 23:2684–2690.