

Structural Requirements for Membrane Assembly of Proteins Spanning the Membrane Several Times

Joachim Lipp, Nicholas Flint, Marie-Theres Haeuptle, and Bernhard Dobberstein

European Molecular Biology Laboratory, D-6900 Heidelberg, Federal Republic of Germany

Abstract. We have investigated the structural requirements for the biogenesis of proteins spanning the membrane several times. Proteins containing various combinations of topological signals (signal anchor and stop transfer sequences) were synthesized in a cell-free translation system and their membrane topology was determined. Proteins spanning the membrane twice were obtained when a signal anchor sequence was followed by either a stop transfer sequence or a second signal anchor sequence. Thus, a signal anchor sequence in the second position can function as a stop transfer sequence, spanning the membrane in the op-

posite orientation to that of the first signal anchor sequence. A signal anchor sequence in the third position was able to insert amino acid sequences located COOH terminal to it. We conclude that proteins spanning the membrane several times can be generated by stringing together signal anchor and stop transfer sequences. However, not all proteins with three topological signals were found to span the membrane three times. A certain segment located between the first and second topological signal could prevent stable membrane integration of a third signal anchor segment.

INTEGRAL membrane proteins spanning the membrane several times (multiple-spanning proteins [MS proteins])¹ are characterized by the presence of hydrophobic segments that are integrated in the lipid bilayer: e.g., band 3 (Kopito and Lodish, 1985), halo-opsin (Blanck and Oesterhelt, 1987), and rhodopsin (Nathans, 1986). These segments have been proposed to function alternately as signals for membrane insertion and stop transfer (ST) (Blobel, 1980; Friedlander and Blobel, 1985). Hydrophilic segments between the membrane-spanning segments are alternately exposed on either side of the membrane. To test the requirements for membrane integration and topology of MS proteins, we strung together previously characterized hydrophobic topological sequences separated from each other by hydrophilic segments of varying size.

In proteins spanning the membrane of the ER once, two different topological signals can be found: signal sequences and ST sequences. Signal sequences mediate targeting to and insertion into the ER membrane. They can either be cleaved (in most secretory proteins) or uncleaved. Uncleaved signal sequences can be translocated (e.g., ovalbumin [Palmiter et al., 1978]) or retained in the membrane as a membrane-spanning segment. As the latter function in membrane insertion as well as in membrane anchoring, they have been termed signal anchor (SA) sequences (Lipp and Dobberstein, 1986). Proteins containing a single SA sequence span the membrane once and expose the NH₂ terminus either on

the cytoplasmic side (type II) or on the extracytoplasmic (luminal) side of the membrane (type I). SA sequences are usually found close to the NH₂ terminus of a protein. In contrast, an ST sequence is always preceded by a signal sequence and is usually found close to the COOH terminus where it spans the membrane in a way such that the COOH terminus of the protein is exposed on the cytoplasmic side.

The main characteristic feature of cleavable signal, SA, and ST sequences is a segment of hydrophobic amino acid residues. No specific amino acid sequence seems to be required for the SA or ST function (Zerial et al., 1987; Audigier et al., 1987; Mize et al., 1986). Sequences flanking the hydrophobic core region of SA and ST sequences seem to play a modulating role in membrane insertion and anchoring (Lipp and Dobberstein, 1988; Szczesna-Skorupa et al., 1988; Haeuptle et al., 1989). They determine the orientation of SA membrane proteins in the membrane and the stability with which ST sequences become embedded in the membrane (Haeuptle et al., 1989; Cutler et al., 1986).

Signal and SA sequences mediate membrane insertion by the interaction with specific receptors. The signal recognition particle binds to the signal or SA sequence and targets the nascent chain-ribosome-signal recognition particle complex to the ER membrane where insertion is mediated by docking protein (signal recognition particle receptor) and probably other components (Wiedmann et al., 1987; Toyn et al., 1988). If an ST sequence is present in the protein, complete translocation is prevented. The protein remains embedded in the membrane most likely due to the interaction of the ST sequence with the lipid bilayer of the membrane. Whether

1. *Abbreviations used in this paper:* CAT, chloramphenicol acetyltransferase; MS proteins, multiple-spanning proteins; SA, signal anchor; ST, stop transfer.

or not protein is involved in the ST function is a controversial issue (Rothman et al., 1988).

MS proteins contain several hydrophobic segments that span the membrane with alternating orientations. Membrane insertion and anchoring of MS proteins might be mediated by these hydrophobic segments. For instance, hydrophobic segments of bovine opsin when placed behind a reporter sequence can perform either membrane insertion and/or ST function (Friedlander and Blobel, 1985; Audigier et al., 1987). These experiments supported the idea that hydrophobicity might be the main feature of SA and ST sequences.

It is conceivable that in MS proteins SA and ST sequences alternate (Blobel, 1980). SA sequences then would be located at uneven positions and ST sequences would be at even positions (sequential signals). If MS proteins can be composed of sequential signals, it should be possible to assemble them by stringing together several SA and ST sequences.

We constructed hybrid proteins that contain various combinations of SA and ST sequences separated by hydrophilic segments of different sizes. Membrane insertion and translocation of the hybrid proteins was investigated using an *in vitro* translation-translocation system.

Materials and Methods

Materials

Restriction endonucleases, T4 DNA ligase, nuclease S1, alkaline phosphatase, and proteinase K were from Boehringer Mannheim GmbH (Mannheim, FRG). *Escherichia coli* RNA polymerase and ⁷methylguanosine-5'-triphosphate-adenosine-5'-monophosphate were from Pharmacia Fine Chemicals (Freiburg, FRG). L-[³⁵S]methionine was from Amersham Buchler GmbH (Braunschweig, FRG). ⁷Methylguanosine-5'-monophosphate was from Sigma Chemical GmbH (Muenchen, FRG).

Methods

The cDNA clone pgamma 2 containing the entire coding region of the human invariant chain cloned into the Pst I site of pBR322 was obtained from P. A. Peterson's laboratory (Research Institute of Scripps Clinic, La Jolla, CA) (Claesson et al., 1983). The expression plasmid pDSS has been described previously (Stueber et al., 1984).

Construction of plasmids

pInc, pInc-Inc, and pInc-Inc-Inc. pgamma 2 was digested with Sau 3A and a 642-bp fragment, encoding the 209 NH₂-terminal amino acids of the invariant chain, isolated, and ligated into the Bam HI site of pDSS. The fragment lacks the coding region for the eight COOH-terminal amino acids and the stop codon for translation. This resulted in pInc, a fusion of the invariant chain to an out-of-frame chloramphenicol acetyltransferase (CAT) sequence encoding 63 amino acid residues.

In pInc, a unique Bam HI restriction site is regenerated upon joining the 3' end of the 642-bp fragment to the CAT sequence in pDSS. This site was then used to introduce a second and third 642-bp Sau 3A fragment of invariant chain, resulting in pInc-Inc and pInc-Inc-Inc, respectively.

pInc-Inc-CAT and pInc-Inc-Inc-CAT. pInc-Inc and pInc-Inc-Inc were linearized by restriction enzyme Bam HI. Ends were overdigested with nuclease S1 (Maniatis et al., 1982), which at high concentrations can result in cleavage of double-stranded DNA. After digestion, pInc-Inc and pInc-Inc-Inc were religated. Plasmid DNA was transcribed and the resulting mRNA translated in a cell-free system. Protein products were screened for an in-frame CAT protein using anti-CAT antibodies.

pIn-Hc. pH-2^d (Bregerer et al., 1981) was digested with Pvu II and Pst I, the 313-bp fragment encoding the COOH terminus of an H-2^d antigen isolated and made blunt using T4 polymerase (Maniatis et al., 1982). pInc was digested with Pvu II and the 3.4-kb vector fragment encoding the NH₂-terminal portion (72 amino acid residues) of invariant chain was isolated, dephosphorylated by alkaline phosphatase, and ligated to the 313-bp fragment.

pIn-Hc-Ic. pInc was digested with Pvu II and the 580-bp fragment encoding the COOH-terminal portion of the invariant chain was isolated. Plasmid In-Hc was digested with Sst I and Pvu II and made blunt with T4 polymerase, and the vector fragment was ligated with the 580-bp fragment.

pIn-Hc-Inc. pInc was cut with restriction enzymes Sma I and Xba I and the 1,500-bp fragment was isolated. pIn-Hc was linearized with Sst I and the single-stranded overhang was removed by T4 polymerase. The linearized plasmid was then cut with Xba I and the 3,000-bp fragment was isolated. Ligation of the two isolated fragments resulted in pIn-Hc-Inc.

pIn-Hc-Ic-Inc. pIn-Hc-Ic was linearized with Bam HI and the 642-bp Sau 3A fragment of pgamma 2 was ligated into the Bam HI site.

pIn-Inc and pIn-Inc-Inc. pInc was digested with Sma I and Pvu II. The 240-bp fragment encoding the NH₂-terminal portion of invariant chain was isolated and ligated into the Sma I site of pInc to form pIn-Inc or into the Sma I site of pInc-Inc to form pIn-Inc-Inc.

pIn-Inc. pInc was digested with Sma I and Xba I and the 1,540-bp fragment encoding Inc was isolated. pIn-Inc was digested with Pvu II and Xba I and the vector fragment encoding In-Inc was ligated to the 1,540-bp fragment.

pIn-Inc-Inc. pInc was linearized with Bam HI. Sticky ends were removed with nuclease S1. The plasmid was then cut with Xba I and the 3,400-bp vector fragment isolated. Plasmid In-Inc was cut with Sma I and Xba I and the fragment containing the In-Inc coding region was isolated. Fragment In-Inc was ligated into the blunt Bam HI-Xba I site. Clones were selected after *in vitro* transcription-translation for in-frame fusion proteins of the appropriate size.

In Vitro Transcription and Translation

Plasmids were transcribed *in vitro* by *E. coli* RNA polymerase in the presence of ⁷methylguanosine-5'-triphosphate-adenosine-5'-monophosphate and the resulting capped mRNA was translated in the wheat germ cell-free system as described by Stueber et al. (1984). In some of the translations, microsomal membranes derived from dog pancreas were included to test for membrane insertion (Lipp and Dobberstein, 1986).

Glycosylation of asparagine residues was blocked by the addition of the acceptor peptide benzoyl-asn-leu-thr-N-methylamide to a final concentration of 30 μM (Bause, 1983; Lau et al., 1983; Lipp and Dobberstein, 1986).

Posttranslational Assays

Proteinase K treatment of microsomal membranes was performed as previously described (Lipp and Dobberstein, 1986). Proteins were analyzed by SDS-PAGE (Laemmli, 1970) and labeled proteins were visualized by fluorography using EN³HANCE (New England Nuclear, Dreieich, FRG).

Results

We constructed fusion proteins containing SA and ST sequences to study membrane insertion and topology of MS proteins. The SA sequence was derived from the human invariant chain (Ii) and the ST sequence was from mouse H-2^d antigen (Claesson et al., 1983; Bregegere et al., 1981). Ii exposes the NH₂ terminus on the cytoplasmic side of the membrane (Lipp and Dobberstein, 1986). The ST sequence of H-2^d antigen spans the membrane in the opposite orientation (Fig. 1). Complementary DNAs were cloned into pDS vectors and transcribed *in vitro* by *E. coli* RNA polymerase, and the resulting mRNAs were translated in a wheat germ cell-free system in the absence or presence of microsomal membranes (Stueber et al., 1984). Membrane insertion and topology of the resulting proteins were assayed by using proteinase K digestion (Lipp and Dobberstein, 1986). Complementary DNAs coding for the following segments were used in the constructions (Fig. 1): (a) the 72 NH₂-terminal amino acids of invariant chain containing the SA sequence (In); (b) the 137 COOH-terminal amino acids of invariant chain containing two potential N-glycosylation sites (Ic) (Claesson et al., 1983); and (c) the 102 COOH-terminal amino acids

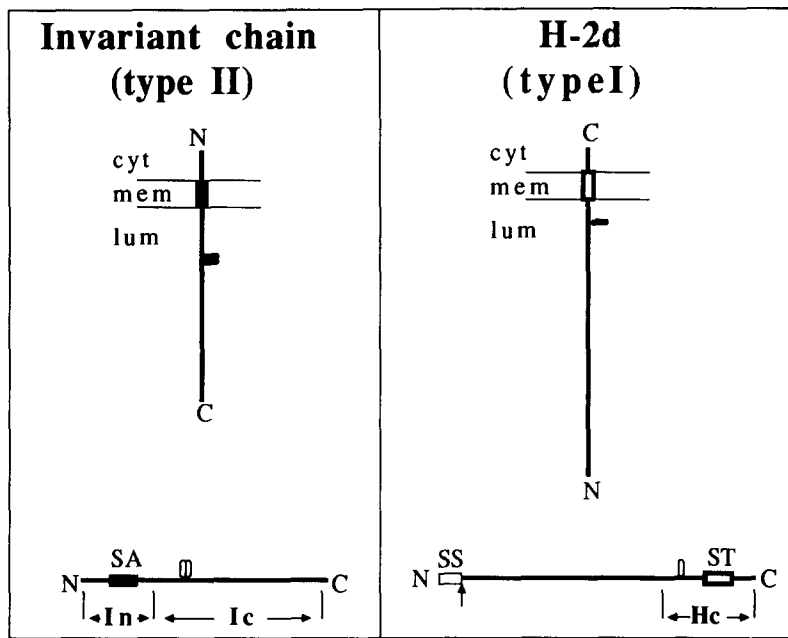


Figure 1. Membrane topology and linear structure of invariant chain and H-2^d class I antigen. Invariant chain spans the membrane once and exposes the NH₂ terminus on the cytoplasmic side (type II membrane protein). (*In*) NH₂-terminal region of invariant chain (72 amino acids) containing the SA sequence (SA). (*Ic*) COOH-terminal region of invariant chain (137 amino acids). H-2^d antigen spans the membrane once and exposes the NH₂ terminus on the luminal (extracytoplasmic) side (type I membrane protein). (*Hc*) COOH-terminal region of H-2^d containing the ST sequence (ST). (SS) cleavable signal sequence of H-2^d; (●) recognition site for N-linked glycosylation; (◐) N-glycosylated site. N, NH₂ terminus; C, COOH terminus.

of a mouse H-2^d class I antigen containing one potential N-glycosylation site (Hc) (Breggare et al., 1981).

Fusion Proteins with Two Topological Signals

We constructed fusion proteins containing either one SA and one ST sequence or two successive SA sequences. In the first construct, the segment In was fused to Hc-Ic (see Materials and Methods and Fig. 2 A). Translation of In-Hc-Ic mRNA in a wheat germ cell-free system in the absence of microsomal

membranes yielded a protein of 42 kD (Fig. 3 A, lane 1). When the *in vitro* system was supplemented with microsomal membranes a 45-kD protein appeared (Fig. 3 A, lane 2). The increase in molecular mass is consistent with the addition of one N-linked oligosaccharide side chain onto the protein backbone (Kornfeld and Kornfeld, 1985). Treatment of microsomes with proteinase K left protein fragments of 18 and 21 kD undigested. These fragments were not present when membranes were solubilized by the detergent NP-40 (Fig. 3 A, cf.

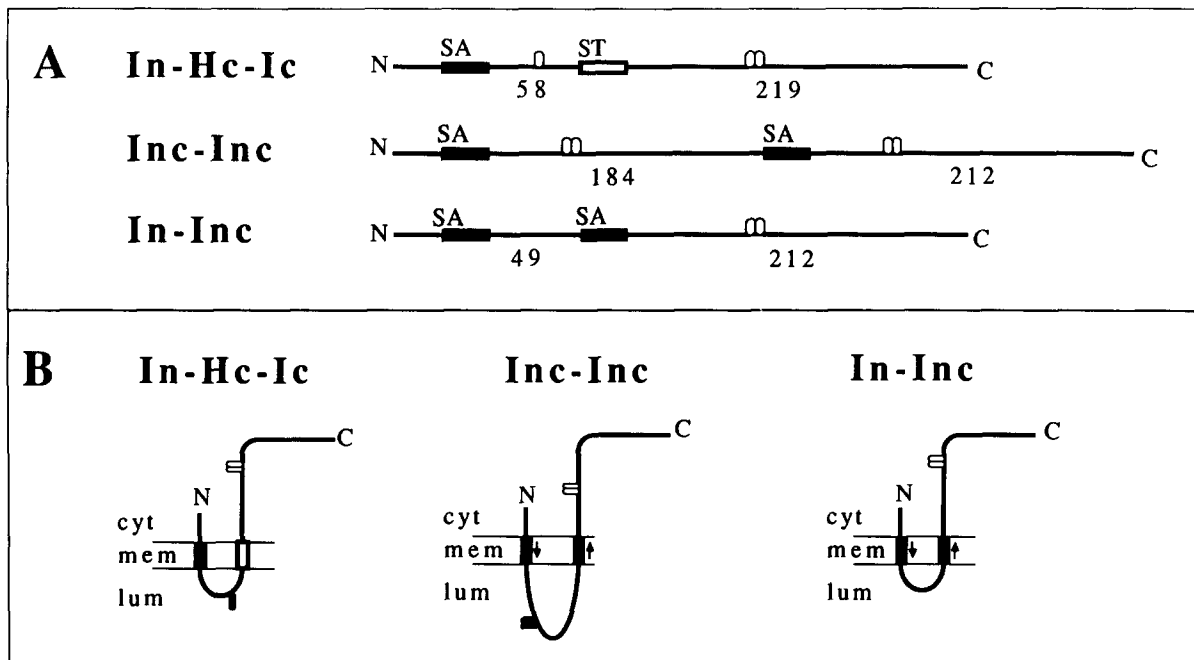


Figure 2. Linear structure and deduced membrane topology of proteins containing two topological signals. (A) Outline of the fusion proteins showing SA sequence of invariant chain (SA); ST sequence of H-2^d (ST); recognition site for N-linked glycosylation (●); N-glycosylated site (◐). Amino acid residues of hydrophilic segments are given by numbers. (B) Deduced membrane topology of fusion proteins. Cyt, cytoplasm; mem, ER membrane; lum, luminal or extracytoplasm. The arrows indicate NH₂ to COOH orientation of SA sequence.

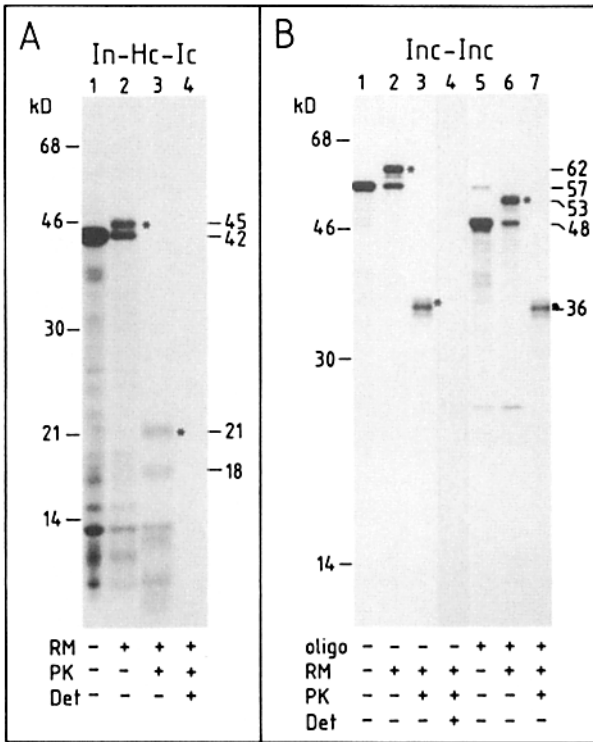


Figure 3. In vitro translation and membrane insertion of In-Hc-Ic (A) and Inc-Inc (B). (A) mRNA encoding In-Hc-Ic was translated in a cell-free system in the absence (lane 1) or presence (lane 2) of microsomes (RM) derived from dog pancreas. Aliquots of proteins translated in the presence of microsomes were assayed for protease protection by incubation with proteinase K (PK) either in the absence (lane 3) or in the presence (lane 4) of detergent NP-40 (Det). (B) Inc-Inc mRNA was translated in the absence (lanes 1-4) or in the presence (lanes 5-7) of an oligonucleotide complementary to mRNA sequences located 189 bp upstream of the stop codon of Inc-Inc. Membrane insertion of full-length and truncated Inc-Inc was assayed by posttranslational treatment of microsomes with proteinase K in the absence (lane 3 and 7) or presence of detergent NP-40 (lane 4). Glycosylated forms are indicated by an asterisk. Proteins were analyzed on a 10–15% gradient polyacrylamide gel.

lanes 3 and 4). As glycosylation is usually not complete, we suspected that the 21-kD protease-protected fragment is the glycosylated form of the 18-kD fragment. To test this, we inhibited glycosylation by adding an acceptor peptide to the translation system. The acceptor peptide benzoyl-asn-leu-thr-N-methylamide competes with newly synthesized polypeptides for N-linked glycosylation (Bause, 1983; Lau et al., 1983; Lipp and Dobberstein, 1986). When membrane-inserted but not glycosylated In-Hc-Ic was treated with proteinase K, the 21-kD band did not appear and the amount of the 18-kD fragment increased (data not shown). This indicates that the 21-kD protected fragment represents the glycosylated form of the 18-kD protein. The 18-kD segment does not contain the hydrophilic NH₂ terminus of In since it does not react with an antibody directed against this region (data not shown). This suggests that the hydrophilic segment between the SA and ST sequences is translocated across the membrane. The protein then spans the membrane twice and exposes the NH₂ and COOH termini on the cytoplasmic side (Fig. 2 B). We

conclude that the SA and ST sequences function as in the authentic proteins.

To test whether an SA sequence placed behind another SA sequence functions as an ST sequence, we cloned the SA sequence of invariant chain (Inc) in tandem (Fig. 2 A, Inc-Inc). The COOH terminus of the fusion protein consists of 63 amino acid residues encoded by an out-of-frame DNA sequence of the CAT gene (see Materials and Methods). The analysis of membrane topology of the fusion protein Inc-Inc is shown in Fig. 3 B. Translation of mRNA encoding Inc-Inc in the absence of microsomes yielded a protein of 57 kD. In the presence of microsomes a glycosylated form of 62 kD was synthesized (Fig. 3 B, lanes 1 and 2). The increase in apparent molecular mass is consistent with the addition of two oligosaccharide side chains.

Proteinase K digestion of intact microsomes revealed a protected fragment of 36-kD (Fig. 3 B, lane 3). As shown in Fig. 2 A, Inc-Inc contains four potential sites for N-linked glycosylation. The size of the protected fragment and its glycosylation is consistent with translocation of either the segment between the two SA sequences or the COOH-terminal hydrophilic segment. To determine which fragment is translocated and glycosylated, we shortened the Inc-Inc polypeptide at the COOH-terminus by using an oligonucleotide complementary to the 3' coding region (180 nucleotides upstream of the stop codon). Endogenous RNase H of the wheat germ system cleaves at the site of hybrid formation, resulting in a truncated mRNA and therefore also shortened protein product (Haeuptle et al., 1986). The shortened protein has an apparent molecular mass of 48 kD (Fig. 3 B, lane 5) and the glycosylated form has a molecular mass of 53 kD (Fig. 3 B, lane 6). Upon treatment of microsomes with proteinase K, the size of the protected protein fragment remains the same as that of the full-length Inc-Inc protein (Fig. 3 B, lanes 3 and 7). This indicates that the COOH-terminal part of Inc-Inc is located on the cytoplasmic side and that the segment located between the two SA sequences is protected and therefore translocated. The Inc-Inc protein thus spans the membrane twice and exposes the NH₂ and COOH termini on the cytoplasmic side of the membrane. Consequently, the second SA sequence must span the membrane in an orientation opposite to that of the first one (Fig. 2 B). We conclude that an SA sequence in the second position can function as an ST sequence. The same membrane topology was obtained for a protein in which the hydrophilic segment between the two SA sequences was shortened (Fig. 2, In-Inc; data not shown).

An SA Sequence in the Third Position Initiates a Second Round of Membrane Insertion

In proteins spanning the membrane twice the amino acid residues after the ST sequence remain on the cytoplasmic side. To obtain a protein spanning the membrane more than twice a new round of membrane insertion is required. Therefore we engineered an SA sequence into the third position after either an ST sequence (In-Hc-Ic or In-Hc-Ic-Inc) or an SA sequence (In-Inc or In-Inc-Inc). A schematic outline of these constructs is given in Fig. 4.

When mRNA coding for In-Hc-Inc was translated in a cell-free system, the major translation product was a protein of 50 kD (Fig. 5 A, lane 1). Upon addition of microsomes, a major new form of ~60 kD was synthesized (Fig. 5 A, lane 2). The increase in molecular mass correlates well with the

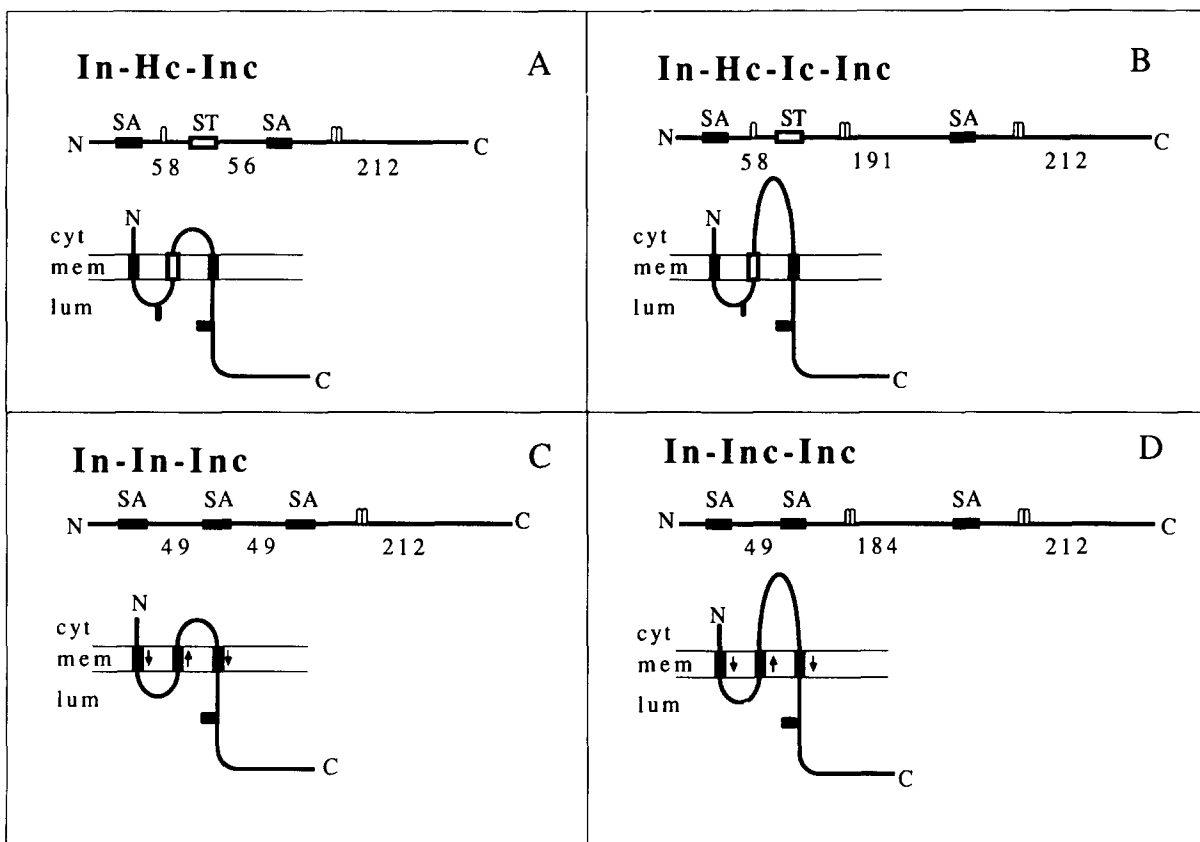


Figure 4. Linear structure and deduced membrane topology of proteins containing three topological signals. (SA) SA sequence of invariant chain; (ST) ST sequence of H-2^d; (○) recognition site for N-linked glycosylation; (■) N-glycosylated site. Amino acid residues of hydrophilic segments are given by numbers. The arrows indicate NH₂ to COOH orientation of the SA sequences.

addition of three oligosaccharide side chains. Treatment of microsomes with proteinase K revealed two protected protein doublets of 31/36 kD and 18/21 kD (Fig. 5 A, lane 3). These fragments were not protected when microsomes were solubilized by detergent before digestion with proteinase K (Fig. 5 A, lane 4). Because membrane-inserted In-Hc-Inc was found to be glycosylated, we also determined glycosylation of the protected fragments. Translation was carried out in the presence of acceptor peptide to inhibit N-glycosylation. Only the lower size polypeptides of the two doublets, 31 and 18 kD in size, were found to be protected (Fig. 5 A, lanes 5–7). Thus, the 36- and the 21-kD fragments represent glycosylated forms. The size difference, when compared with the unglycosylated species, suggests that two and one N-linked carbohydrate side chains are bound to the 36- and 21-kD fragments, respectively.

The sizes of the protected fragments and their glycosylation suggests that In-Hc-Inc spans the membrane three times. The hydrophilic segment between the first SA and the ST sequence (21 kD) forms a loop on the luminal side and is N-glycosylated at one site. The second SA sequence also spans the membrane, and the COOH-terminal hydrophilic segment with two oligosaccharide side chains (36 kD) is exposed on the luminal side (Fig. 4 A). These results demonstrate that an SA sequence in the third position can mediate translocation of COOH-terminally located amino acid sequences across the ER membrane.

An SA Sequence in the Second Position Does Not Interfere with Reinsertion of a Subsequent SA Sequence

We have shown above that two SA sequences in tandem result in a membrane protein that spans the membrane twice. The second SA sequence spans the membrane in the opposite orientation to the first one and functions as an ST sequence. It is thus conceivable that its insertion is not in all aspects identical to the insertion and anchoring of an authentic ST sequence. We therefore tested whether or not an SA sequence in the second position could have an effect on the reinsertion function of an SA sequence in the third position.

Protein In-In-Inc, which contains three successive SA sequences, was constructed. It contains two potential N-glycosylation sites in the COOH-terminal hydrophilic segment, which is identical to the COOH-terminal segment of In-Hc-Inc. In-In-Inc, synthesized in the absence of membranes, has a molecular mass of 49 kD (Fig. 5 B, lane 1). When synthesized in the presence of membranes it has a molecular mass of 55 kD, consistent with the addition of two N-linked oligosaccharide side chains (Fig. 5 B, lane 2). When treated with proteinase K, we obtained protected fragments of 36 and 14 kD, respectively (Fig. 5 B, lane 3). The protected 36-kD protein is common to In-Hc-Inc and In-In-Inc and therefore corresponds in size to a fragment derived from the COOH terminus (cf. Fig. 4, A and C, and Fig. 5). The segment between

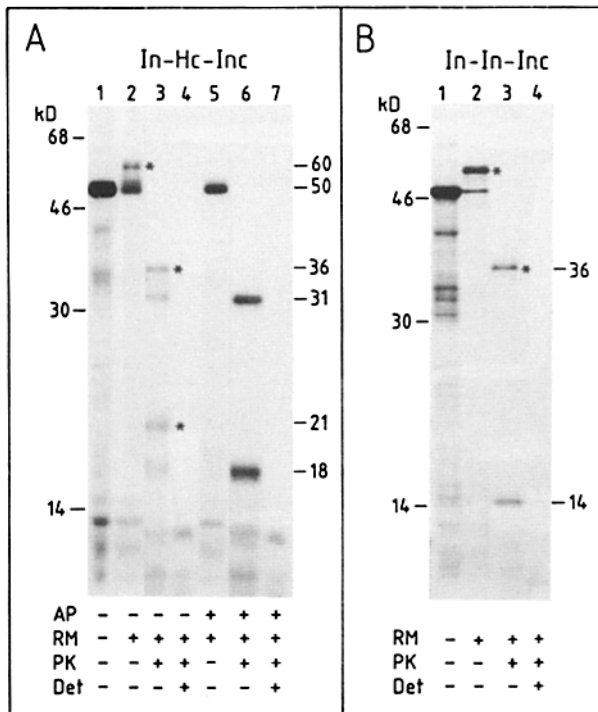


Figure 5. In vitro translation and membrane insertion of In-Hc-Inc (A) and In-In-Inc (B). mRNAs encoding In-Hc-Inc (A, lanes 1-7) or In-In-Inc (B, lanes 1-4) were translated in a wheat germ cell-free system in the absence (lanes 1) or presence (lanes 2-7) of microsomes (RM). Translations as shown in A, lanes 5-7, were done in the presence of acceptor peptide. Aliquots of proteins translated in the presence of microsomes were assayed for protease protection by incubation with proteinase K (PK) in the absence (lanes 3 and 6) or in the presence (lanes 4 and 7) of detergent NP-40 (Det). Glycosylated forms are indicated by an asterisk.

the first and the second hydrophobic sequence of In-In-Inc contains no glycosylation site (Fig. 4 C). The small protected segment is of similar size to the unglycosylated small protected fragment of In-Hc-Inc (cf. Fig. 5, A, lane 6, and B, lane 3, and Fig. 4, A and C). This indicates an identical membrane topology for In-In-Inc and In-Hc-Inc; i.e., both proteins spanning the membrane three times.

Reinsertion Is Independent of the Size of the Segment between the Second and the Third Topological Signal

Some hydrophilic segments preceding a signal sequence have been shown to interfere with membrane insertion (Yost et al., 1983). We therefore investigated whether the size of the segment between the second and third topological signal affects insertion of COOH-terminal protein segments. Fusion proteins In-Hc-Ic-Inc and In-Inc-Inc (Fig. 4, B and D) were constructed which differ from In-Hc-Inc and In-In-Inc (Fig. 4, A and C) only in the length of the segment between the second and the third topological signal. When proteins In-Hc-Ic-Inc and In-Inc-Inc were analyzed for their membrane topology, protease-protected fragments were found to be of identical molecular mass as those generated from In-Hc-Inc and In-In-Inc, respectively (data not shown). This indicated that In-Hc-Ic-Inc and In-Inc-Inc display the same membrane topology as In-Hc-Inc and In-In-Inc (Fig. 4, B and D).

A Luminal Domain Can Prevent Stable Translocation of a COOH-terminal Protein Segment

Some of the fusion proteins with three topological signals displayed an unexpected membrane topology. In these proteins the COOH-terminal hydrophilic segment was found to be glycosylated but not translocated. In these proteins the segment between the first two SA sequences had been extended to 180 amino acids (Fig. 6). Inc-Inc-Inc contains three segments of nearly full-length invariant chain. Each of these segments contains two potential sites for N-linked glycosylation. The analysis of the membrane topology of Inc-Inc-Inc is shown in Fig. 7.

Translation of mRNA coding for Inc-Inc-Inc in the absence of microsomes results in a protein of 73 kD. Translation in the presence of microsomes yields additional proteins of ~80 and 85 kD (Fig. 7, lanes 1 and 2). This increase in molecular mass is consistent with the addition of two and four N-linked oligosaccharide side chains, respectively. To determine which parts of the protein have been translocated, we treated microsomes with proteinase K. One protein fragment of 36 kD was protected by the membrane (Fig. 7, lane 3). Two protected fragments would have been expected for a protein spanning the membrane three times. Since it was possible that the two protected fragments were of identical size, we truncated the COOH terminus of Inc-Inc-Inc by 63 amino acids using a complementary oligonucleotide (Haeupfle et al., 1986). In the absence of microsomes a shortened polypeptide of 66 kD was synthesized. In the presence of microsomes three proteins were synthesized: a nonglycosylated form of 66 kD and glycosylated forms of 72 and 78 kD (Fig. 7, lanes 5 and 6). After protease digestion, one protein fragment of 36 kD was obtained (Fig. 7, lane 7), indicating that the shortened COOH-terminal portion is not protected from protease digestion and thus not translocated across the membrane.

The First and the Third Inc Segment in Inc-Inc-Inc Become Glycosylated

Four out of six potential sites for N-linked glycosylation in Inc-Inc-Inc become glycosylated, but only one fragment was protected from protease digestion (Fig. 7, lanes 2 and 3). So it remained unclear which of the three Inc segments is glycosylated and translocated. To determine the glycosylated sites in Inc-Inc-Inc, we interrupted glycosylation at various times during membrane insertion. Nonionic detergents like NP-40 disrupt membranes and thereby prevent N-linked glycosylation. They do not interfere with protein elongation. As a point of reference for glycosylation we used protein Inc-Inc, which is identical to the NH₂-terminal two thirds of Inc-Inc-Inc. Inc-Inc spans the membrane twice and exposes the COOH terminus on the cytoplasmic side. The segment between the two SA sequences is glycosylated (see Fig. 2 B and Fig. 3 B).

Inc-Inc and Inc-Inc-Inc were translated in the absence (Fig. 8 A, lane -) and presence (Fig. 8 A, lanes 1-14) of microsomes. Translation was synchronized after 9 min by ⁷methylguanosine-5'-monophosphate addition. At the times indicated in Fig. 8 B, microsomes were solubilized by the detergent NP-40 to prevent further glycosylation. Incubation was continued to allow synthesis of full-length proteins. Glycosylation is detected by a shift to a higher molecular mass, reflecting addition of a pair of oligosaccharide side chains.

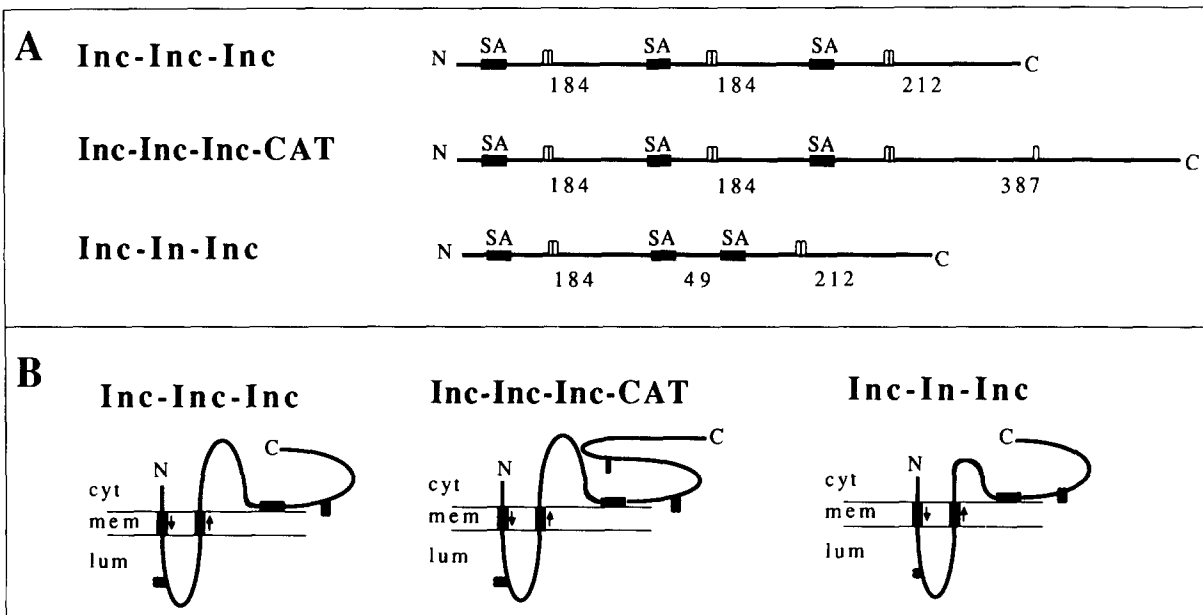


Figure 6. Linear structure (A) and suggested membrane topology (B) of proteins with three SA sequences. (SA) SA sequence of invariant chain; (ST) ST sequence of H-2^d antigen; (CAT) chloramphenicol-acetyltransferase; (O) recognition site for N-linked glycosylation; (●) N-glycosylated site. Amino acid residues of hydrophilic segments are given by numbers. The arrows indicate NH₂ to COOH orientation of SA sequences.

Glycosylation of the first sites in Inc-Inc as well as of Inc-Inc-Inc occurred after ~20 min (Fig. 8 A, lanes 2 and 3). Inc-Inc as expected was not further glycosylated. Inc-Inc-Inc was further glycosylated after ~70 min of translation (Fig. 8 A, lane 11). To localize the protein portions that have been glycosylated, we plotted the appearance of glycosylated forms against the time of synthesis (Fig. 8 B). The first glycosylation appears after 20–25 min of translation. Assuming linear elongation, glycosylation of the second pair of glycosylation sites should occur after 40–50 min and glycosylation of the third pair after ~70 min. The appearance of glycosylated forms after 20 and 70 min strongly suggests that in Inc-Inc-Inc the first and the third Ic portion have been glycosylated.

As shown by truncation of the COOH terminus, the COOH-terminal portion is not found protected against protease digestion. This result was unexpected since glycosylation is known to occur on the luminal side of the membrane. Possibly the size of the hydrophilic COOH-terminal segment could affect stable translocation across the membrane. To test that possibility, we elongated Inc-Inc-Inc by fusing the cytoplasmic protein CAT to the COOH terminus (Inc-Inc-Inc-CAT). As control for the absence of a topological signal in CAT, we fused CAT to Inc-Inc. Inc-Inc has been shown to span the membrane twice and to expose the COOH terminus on the cytoplasmic side. Translation of mRNA encoding Inc-Inc-CAT in the absence and presence of microsomes yields proteins of 75 and ~80 kD, respectively (Fig. 9, lanes 1 and 2). Treatment of microsomes with proteinase K results in a protected fragment of 36 kD (Fig. 9, lane 3). CAT thus has no influence on membrane insertion and membrane topology of Inc-Inc.

Translation of Inc-Inc-Inc-CAT encoding mRNA in the absence of microsomes results in a protein of 96 kD (Fig. 9, lane 5); in the presence of microsomes, several glycosylated

forms of different sizes (96, 102, 108, and 111 kD) were obtained (Fig. 9, lane 6). The increase in size is consistent with two, four, and five oligosaccharide side chains. In invariant chains even numbers of oligosaccharide side chains are always attached. The CAT portion provides one additional site for N-linked glycosylation which is used upon translocation across ER membranes (Lipp and Dobberstein, 1986). As five sites were glycosylated, we can conclude that the CAT portion of Inc-Inc-Inc-CAT is glycosylated and inserted into the membrane. When protein fragments protected from protease digestion were analyzed by SDS-PAGE, we still found only one fragment of 36 kD protected (Fig. 9, lane 7), indicating that the COOH terminus, although glycosylated, is sensitive to exogenously added protease.

We next asked whether the size of the hydrophilic segment between the second and the third topological signal could affect stable translocation of the COOH terminus. Protein Inc-In-Inc, in which the third hydrophilic segment was shortened, was constructed (Fig. 6). When the topology of this protein was determined in the cell-free system it was found that the COOH terminus becomes glycosylated but not stably translocated (data not shown).

A common feature among the proteins that can not stably translocate the COOH terminus is the first Ic domain. In contrast, proteins that have a different segment at this position (In-In-Inc and In-Inc-Inc) can stably translocate the COOH terminus (cf. Fig. 4, C and D). We conclude that the first luminal domain can affect stable translocation of COOH-terminally located amino acid sequences without interfering with their glycosylation.

Discussion

To characterize signals mediating membrane insertion and

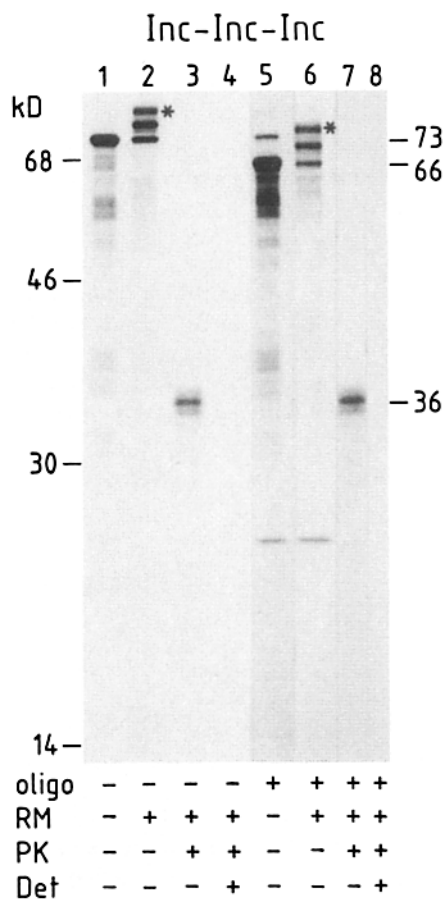


Figure 7. In vitro translation and membrane insertion of Inc-Inc-Inc. Inc-Inc-Inc mRNA was translated in the absence (lanes 1-4) or in the presence (lanes 5-8) of an oligonucleotide complementary to mRNA sequences located 189 bp upstream of the stop codon of Inc-Inc-Inc. The truncated mRNA primes for a polypeptide species shortened at its COOH terminus by 63 amino acid residues. Membrane insertion of full-length and truncated Inc-Inc-Inc was assayed by posttranslational treatment of microsomes with proteinase K in the absence (lanes 3 and 7) and presence of detergent NP-40 (lanes 4 and 8). Addition of oligonucleotide (*oligo*), rough microsomes (*RM*), proteinase K (*PK*), and detergent NP-40 (*Det*) is indicated at the bottom of each lane. Fully glycosylated forms are indicated by an asterisk.

anchoring of MS proteins, we constructed fusion proteins containing two or more topological signals. Membrane topology of these proteins was determined by (a) characterizing membrane-integrated and -translocated protein fragments protected from protease digestion; (b) analyzing N-glycosylated sites; (c) inhibiting N-linked glycosylation (Lau et al., 1983; Bause, 1983; Lipp and Dobberstein, 1986); and (d) analyzing membrane-integrated polypeptide chains shortened at the COOH terminus. (Haeuptle et al., 1986).

A Protein Containing an SA Sequence and an ST Sequence Spans the Membrane Twice

We have shown that the protein In-Hc-Ic containing two topological signals, an SA sequence in the first position followed by an ST sequence in the second position, spans the membrane twice and exposes the NH₂ and COOH termini

on the cytoplasmic side of the membrane. The hydrophilic segment between the two topological signals is located on the extracytoplasmic (luminal) side of the membrane. From the molecular mass of the membrane-protected fragment and its glycosylation, any other membrane topology of In-Hc-Ic could be ruled out. The topological signal in the first position functions as an SA sequence as in the authentic type II protein invariant chain (Ii). The second topological signal functions as an ST sequence and spans the membrane identically to the authentic protein H-2^d. H-2^d is a type I membrane protein spanning the membrane once and exposing the COOH terminus on the cytoplasmic side. Type I membrane proteins, like G protein of vesicular stomatitis virus (Rose et al., 1980) or class I and II histocompatibility antigens (Dobberstein et al., 1982; Kaufman and Strominger, 1982), contain a cleavable signal sequence in the first position and an ST sequence close to the COOH terminus. Upon membrane insertion, the signal sequence is cleaved by signal peptidase and translocation of the COOH terminus is blocked by the ST sequence. It is very likely that membrane insertion of In-Hc-Ic follows

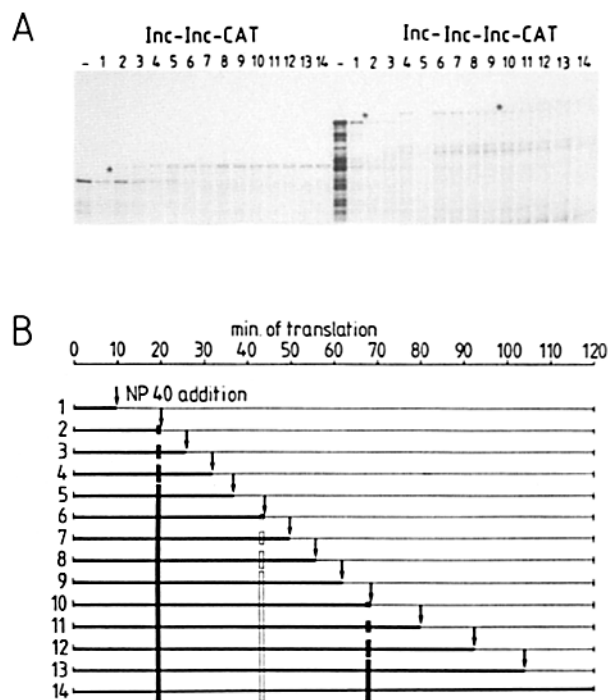


Figure 8. Time course of glycosylation of Inc-Inc-CAT and Inc-Inc-Inc-CAT. (A) mRNA encoding Inc-Inc-CAT and Inc-Inc-Inc-CAT was translated in the absence (lanes -) or in the presence (lanes 1-14) of rough microsomes. Translation was synchronized after 9 min by the addition of 2 mM methylguanosine-5'-monophosphate. Translation was allowed to continue for 120 min. At time points indicated by arrows in B (bold lines 1-14 in B correspond to lanes 1-14 in A), further glycosylation was abolished by solubilizing rough microsomes with detergent NP-40 (0.3% final concentration). Appearance of glycosylation onto different sites is indicated by asterisks. Proteins were analyzed on a 10% polyacrylamide gel. (B) Graphic representation of glycosylation occurring during synthesis of Inc-Inc-CAT and Inc-Inc-Inc-CAT. In samples 1-14, time of translation in the absence of detergent is indicated by a bold line. Addition of detergent NP-40 is indicated by arrows. Steps of glycosylation are indicated by vertical solid bars. Open vertical bars indicate potential glycosylation sites that are not used.

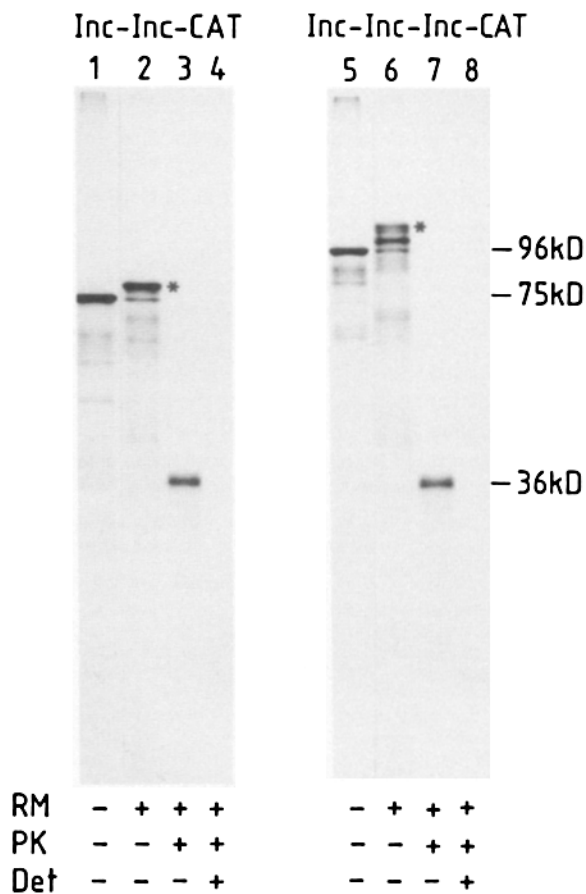


Figure 9 In vitro translation and membrane insertion of Inc-Inc-CAT and Inc-Inc-Inc-CAT. mRNA encoding Inc-Inc-CAT (lanes 1–4) or Inc-Inc-Inc-CAT (lanes 5–8) was translated either in the absence (lanes 1 and 5) or presence (lanes 2 and 6) of microsomes (RM). Aliquots of proteins translated in the presence of microsomes were assayed for protease protection by incubation with proteinase K (PK) either in the absence (lanes 3 and 7) or in the presence (lanes 4 and 8) of detergent NP-40 (Det). Fully glycosylated forms are indicated by an asterisk.

the same mechanism except that the SA sequence in the first position is not cleaved. As a consequence, the protein is anchored by the first topological signal and spans the membrane twice.

An SA Sequence in the Second Position Can Function as an ST Sequence

The protein Inc-Inc containing two identical SA sequences in tandem spans the membrane twice and exposes the NH₂ and COOH termini on the cytoplasmic side. Thus, the SA sequence in the first position mediates membrane insertion and the identical SA sequence in the second position functions as an ST sequence. Recently, Wessels and Spiess (1988) have similarly shown that the SA sequence of asialoglycoprotein receptor functions as an ST sequence when placed behind an identical SA sequence. In addition, some signal sequences that are usually cleaved by signal peptidase but contain a hydrophobic segment sufficient to span the membrane can function as an SA (Hauptle et al., 1989) or ST sequence (Coleman et al., 1985). Likewise, ST sequences have been shown to function in ER membrane targeting and

insertion of proteins, for example, if placed in the position of a cleavable signal sequence (Mize et al., 1986) or if replaced for the hydrophobic core segment of an SA sequence (Zerial et al., 1987). For the MS protein opsin Audigier et al. (1987) have shown that some of the membrane-spanning segments can also function as SA sequences when placed in the first position. However, not all ST sequences in an NH₂-terminal position function as an SA sequence (Yost et al., 1983).

From all these observations, we conclude that it is the position within the polypeptide chain and the relative position to other topological signals that decide on the SA or ST function. An ST sequence does not seem to have unique features that can not be met by an SA sequence. In this view, the essential element for both membrane insertion and ST function is the hydrophobicity of a protein segment.

Membrane Insertion of Proteins Spanning the Membrane Three Times

Several proteins with three SA or ST sequences in different combinations could be shown to span the membrane three times. For the native MS proteins bovine opsin and bacterial MalF, it has been shown that internally located membrane-spanning segments can perform membrane insertion when fused to a reporter molecule (Friedlander and Blobel, 1985; Audigier et al., 1987; Boyd et al., 1987). This supports the idea that insertion of MS proteins into the membrane can occur by consecutive membrane insertion and ST events. The stepwise N-linked glycosylation observed in the time course experiment is also consistent with a sequential membrane insertion of SA and ST sequences. N-linked glycosylation is carried out by the enzyme oligosaccharyl transferase which is located in the ER membrane (Kornfeld and Kornfeld, 1985). The timed order of appearance of N-linked glycosylation in Inc-Inc-Inc (see Fig. 8 A) suggests that the first segment has been translocated and glycosylated before the COOH-terminal segment. Similar results were recently described by Wessels and Spiess (1988). They found that in an MS protein, the first luminal segment was already protected from protease after half of the time necessary to synthesize detectable amounts of the full-length protein.

Most unexpected in this study was the finding that a protein portion of an MS protein had been translocated across the membrane but apparently did not remain on the luminal side. The protein portion comprised the COOH-terminal hydrophilic segment that followed a second functional SA sequence. Translocation of this segment across the membrane was deduced from its glycosylation; its presence on the cytoplasmic side was deduced from sensitivity to exogenously added protease. Translocation of this COOH-terminal segment was not dependent on its size since an increase in size did not yield a protected fragment. Surprisingly, we found that it is the hydrophilic segment between the first SA sequence and the following ST sequence that influences stable translocation of the COOH-terminal segment.

How could we possibly explain that the COOH-terminal segment in some of the proteins is not stably translocated across the membrane? First, it is conceivable that an interaction between the translocated segments prevents stable translocation; second, binding to specific ER proteins could also be required to retain the COOH-terminal protein portion on the luminal side. A potential candidate for binding to trans-

located but not completely folded proteins is the glucose-regulated protein (GRP 78) (Pelham, 1986). And, third, translocation of proteins across the ER membrane might require a transient interaction with components of a translocation complex in the lipid bilayer.

Relocalization of a membrane-inserted protein onto the cytoplasmic side of the membrane has recently been reported by two groups. Trun and Silhavy (1989) observed that a mutation in the *prlC* gene of *E. coli* caused the accumulation of processed secretory proteins in the cytoplasm. Similarly Garcia et al. (1988) found that the major hepatitis B virus precore protein is processed by signal peptidase but not sequestered into the lumen of microsomes. An understanding of these phenomena will require the characterization of the components involved in protein translocation and the analysis of their interaction with the translocated polypeptides.

We thank Klaus Lingelbach, Stephen High, and Karin Roemisch for critical reading and helpful comments.

This work was supported by grant Do 199/4-4 from the Deutsche Forschungsgemeinschaft.

Received for publication 24 April 1989 and in revised form 11 July 1989.

References

- Audigier, Y., M. Friedlander, and G. Blobel. 1987. Multiple topogenic sequences in bovine opsin. *Proc. Natl. Acad. Sci. USA* 84:5783-5787.
- Bause, E. 1983. Structural requirements of N-glycosylation of proteins. *Biochem. J.* 209:331-336.
- Blanck, A., and D. Oesterhelt. 1987. The halo-opsin gene. II. Sequence, primary structure of halorhodopsin and comparison with bacteriorhodopsin. *EMBO (Eur. Mol. Biol. Organ.) J.* 6:265-273.
- Blobel, G. 1980. Intracellular protein topogenesis. *Proc. Natl. Acad. Sci. USA* 77:1496-1500.
- Boyd, D., C. Manoil, and J. Beckwith. 1987. Determinants of membrane topology. *Proc. Natl. Acad. Sci. USA* 84:8525-8529.
- Bregegere, F., J. P. Abastado, S. Kvist, L. Rask, J. L. Lalanne, H. Garoff, B. Cami, K. Wiman, D. Larhammar, P. A. Peterson, G. Gachelin, P. Kourilsky, and B. Dobberstein. 1981. Structure of C-terminal half of two H-2 antigens from cloned mRNA. *Nature (Lond.)* 292:78-81.
- Claesson, L., D. Larhammar, L. Rask, and P. A. Peterson. 1983. cDNA clone for the human invariant chain of class II histocompatibility antigens and its implications for the protein structure. *Proc. Natl. Acad. Sci. USA* 80:7395-7399.
- Coleman, J., M. Inukai, and M. Inouye. 1985. Dual functions of the signal peptide in protein transfer across the membrane. *Cell* 43:351-360.
- Cutler, D. F., P. Melancon, and H. Garoff. 1986. Mutants of the membrane-binding region of semliki forest virus E2 protein. II. Topology and membrane binding. *J. Cell Biol.* 102:902-910.
- Dobberstein, B., S. Kvist, and L. Roberts. 1982. Structure and biosynthesis of histocompatibility antigens (H-2, HLA). *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 300:161-172.
- Friedlander, M., and G. Blobel. 1985. Bovine opsin has more than one signal sequence. *Nature (Lond.)* 318:338-343.
- Garcia, P. D., J.-H. Ou, W. J. Rutter, and P. Walter. 1988. Targeting of the Hepatitis B virus precore protein to the endoplasmic reticulum membrane: after signal peptide cleavage translocation can be aborted and the product released into the cytoplasm. *J. Cell Biol.* 106:1093-1104.
- Hauptle, M. T., R. Frank, and B. Dobberstein. 1986. Translation arrest by oligodeoxynucleotides complementary to mRNA coding sequences yields polypeptides of predetermined length. *Nucleic Acids Res.* 14:1427-1448.
- Hauptle, M.-T., N. Flint, N. M. Gough, and B. Dobberstein. 1989. A tripartite structure of the signals that determine protein insertion into the ER membrane. *J. Cell Biol.* 108:1227-1236.
- Kaufman, J. F., and J. L. Strominger. 1982. HLA-DR light chain has a polymorphic N-terminal region and a conserved immunoglobulin-like C-terminal region. *Nature (Lond.)* 297:694-697.
- Kopito, R. R., and H. F. Lodish. 1985. Primary structure and transmembrane orientation of the murine anion exchange protein. *Nature (Lond.)* 316:234-238.
- Kornfeld, R., and S. Kornfeld. 1985. Assembly of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.* 54:631-664.
- Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (Lond.)* 227:680-685.
- Lau, J. T. Y., J. K. Wely, P. Shenbagamurthi, F. Naider, and W. J. Lennarz. 1983. Substrate recognition by oligosaccharyl transferase: inhibition of translational glycosylation by acceptor peptides. *J. Biol. Chem.* 258:15255-15260.
- Lipp, J., and B. Dobberstein. 1986. The membrane spanning segment of invariant chain (I gamma) contains a potentially cleavable signal sequence. *Cell* 46:1103-1112.
- Lipp, J., and B. Dobberstein. 1988. Signal and membrane anchor function overlap in the type II membrane protein IgammaCAT. *J. Cell Biol.* 106:1813-1820.
- Maniatis, T., E. G. Fritsch, and J. Sambrook. 1982. Molecular cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 545 pp.
- Mize, N. K., D. W. Andrews, and V. R. Lingappa. 1986. A stop transfer sequence recognizes receptors for nascent chain translocation across the endoplasmic reticulum membrane. *Cell* 47:711-719.
- Nathans, J., D. Thomas, and D. S. Hogness. 1986. Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. *Science (Wash. DC)* 232:193-202.
- Palmiter, R. D., J. Gagnon, and K. A. Walsh. 1978. Ovalbumin: a secreted protein without a transient hydrophobic leader sequence. *Proc. Natl. Acad. Sci. USA* 75:94-98.
- Pelham, H. R. B. 1986. Speculations on the functions of the major heat shock and glucose-regulated proteins. *Cell* 46:959-961.
- Rose, J. K., W. J. Welch, B. M. Sefton, F. S. Esch, and N. C. Ling. 1980. Vesicular stomatitis virus glycoprotein is anchored in the viral membrane by a hydrophobic domain near the COOH terminus. *Proc. Natl. Acad. Sci. USA* 77:3884-3888.
- Rothman, R. E., D. W. Andrews, M. C. Calayag, and V. R. Lingappa. 1988. Construction of defined polytopic integral transmembrane proteins: the role of signal and stop transfer sequence permutations. *J. Biol. Chem.* 263:10470-10480.
- Stueber, D., I. Ibrahim, 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.
- Szczesna-Skorupa, E., N. Browne, D. Mead, and B. Kemper. 1988. Positive charges at the NH₂ terminus convert the membrane-anchor signal peptide of cytochrome P-450 to a secretory signal peptide. *Proc. Natl. Acad. Sci. USA* 85:738-742.
- Trun, N. J., and T. J. Silhavy. 1989. PrlC, a suppressor of signal sequence mutations in *Escherichia coli*, can direct the insertion of the signal sequence into the membrane. *J. Mol. Biol.* 205:665-676.
- Toyn, J., A. R. Hibbs, P. Sanz, J. Crowe, and D. I. Meyer. 1988. In vivo and in vitro analysis of ptl1, a yeast ts mutant with a membrane-associated defect in protein translocation. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:4347-4353.
- Wessels, H. P., and M. Spiess. 1988. Insertion of multispinning membrane protein occurs sequentially and requires only one signal sequence. *Cell* 55:61-70.
- Wiedmann, M., T. V. Kurzchalia, E. Hartmann, and T. A. Rapoport. 1987. A signal sequence receptor in the endoplasmic reticulum membrane. *Nature (Lond.)* 328:830-833.
- Yost, C. S., J. Hedgpeth, and V. R. Lingappa. 1983. A stop transfer sequence confers predictable transmembrane orientation to a previously secreted protein in cell-free systems. *Cell* 34:759-766.
- Zerial, M., D. Huylebroeck, and H. Garoff. 1987. Foreign transmembrane peptides replacing the internal signal sequence of transferrin receptor allow its translocation and membrane binding. *Cell* 48:147-155.