Topological Rules for Membrane Protein Assembly in Eukaryotic Cells*

(Received for publication, September 17, 1996, and in revised form, December 26, 1996)

Guro Gafvelin†, Masao Sakaguchi‡, Helena Andersson§, and Gunnar von Heijne¶

From the †Department of Biochemistry, Stockholm University, S-106 91 Stockholm, Sweden, the ‡Department of Molecular Biology, Graduate School of Medical Science, Kyushu University, Fukuoka 812, Japan, and the §Department of Molecular Biology, Karolinska Institute Center for Biotechnology, NOVUM, S-141 57 Huddinge, Sweden

Insertion into the endoplasmic reticulum membrane of model proteins with one, two, and four transmembrane segments and different distributions of positively charged residues in the N-terminal tail and the polar loops has been studied both in vitro and in vivo. Membrane insertion of these same constructs has previously been analyzed in Escherichia coli, thus making possible a detailed comparison between the topological rules for membrane protein assembly in prokaryotic and eukaryotic cells. In general, we find that positively charged residues have similar effects on the membrane topology in both systems when they are placed in the N-terminal tail but that the effects of charged residues in internal loops clearly differ. Our results rule out a sequential tail but suggest that the effects of charged residues in internal loops have similar effects on the membrane topology in prokaryotic and eukaryotic systems.

What features of the amino acid sequence of integral membrane proteins control their insertion and orientation in the membrane? A number of recent studies in Escherichia coli have suggested that hydrophobic segments both target bacterial inner membrane proteins to the membrane and drive their insertion into the lipid bilayer, whereas flanking positively charged amino acids determine the final orientation of the hydrophobic transmembrane segments (1).

Statistical analyses suggest that a similar “positive inside” rule is applicable also to eukaryotic membrane proteins (2–7), although the tendency for positively charged residues to be absent from extracellularly exposed parts is weaker in eukaryotic than in prokaryotic membrane proteins. Nevertheless, a handful of eukaryotic membrane proteins have been successfully expressed in E. coli with what appears to be the correct topology (8–10), and at least one example of an E. coli inner membrane protein that inserts with the correct topology into mammalian microsomes in vitro has been described (11, 12).

Mutagenesis experiments on a number of bitopic eukaryotic plasma membrane proteins with a single N-terminal transmembrane segment have revealed a clear effect of arginine and lysine residues on membrane orientation, although other factors such as the presence of negatively charged residues, the length and folding properties of the polar N-terminal tail preceding the transmembrane segment, and the length of the transmembrane segment itself have also been shown to be important (13–23). However, the effects of charged residues on the topology of polytopic eukaryotic membrane proteins have not been systematically studied so far, and it is not clear whether the topological rules relating amino acid sequence to membrane topology are the same in prokaryotic and eukaryotic cells.

Building on our earlier work on membrane protein assembly in E. coli, we now report a systematic comparative study of the insertion into the ER† membrane, both in vitro and in vivo, of a range of model proteins with one to four transmembrane segments and different distributions of positively charged residues in the flanking regions. Most of these proteins have previously been characterized in E. coli, allowing a detailed comparison between membrane protein insertion in prokaryotic and eukaryotic systems.

MATERIALS AND METHODS

Enzymes and Chemicals—Unless otherwise stated, all enzymes were from Promega. T7 DNA polymerase was from Pharmacia Biotech Inc. Endo H was from Boehringer Mannheim. (35)S-Met was from Amersham Corp. Ribonucleotides, deoxyribonucleotides, dideoxyribonucleotides, protein A-Sepharose, and the cap analog m7G(5′)ppp(5′)G were from Pharmacia. Plasmid pGEM1, transcription buffer, and rabbit reticulocyte lysate were from Promega. BHK 21 cells were grown in medium from Life Technologies, Inc. The glycosylation acceptor peptide N-benzoyl-Asn-Leu-Thr-N-methylamide and the nonacceptor peptide N-benzoyl-Asn-Leu-(allo)Thr-N-methylamide were synthesized according to Ref. 24. Oligonucleotides were from Kebco Lab (Stockholm, Sweden).

DNA Techniques—Site-specific mutagenesis was performed according to the method of Kunkel (25) as modified by Geisselsoder et al. (26). All mutants were confirmed by DNA sequencing of single-stranded M13 DNA using T7 DNA polymerase. For cloning into and expression from the pGEM1 plasmid, the 5′ end of the lep gene was modified first by the introduction of an XbaI site and second by changing the context 5′ to the initiator ATG codon to a “Kozak consensus” sequence (27). Thus, the 5′ region of the gene was modified to: ATAACCCCTCTAGACCCATG-GCGAAT(...). Mutants with four transmembrane segments were made by fusing the H1-P1-H2 region (from the upstream XbaI site to a KpnI site introduced at the codon corresponding to Asp26 in the wild type sequence) of one Lep construct to a KpnI site introduced just upstream of the start codon of a second Lep construct (see Fig. 1). The KpnI site introduced three new amino acid residues (Glu-Val-Pro) at the fusion site. The XbaI-Smal fragment carrying the constructs was cloned into pGEM1 behind the SP6 promoter. For cloning into the pSFV1 in vitro expression vector (28), constructs were polymerase chain reaction-amplified from pGEM1 using polymerase chain reaction primers introducing BonII and Smal sites flanking the relevant gene. The BanII-Smal fragment carrying the constructs was then cloned into the pSFV1 polylinker region.

* This work was supported by grants from the Swedish Cancer Society, the Swedish Natural Sciences Research Council, and the Göran Gustafsson Foundation (to G. v. H.) and a grant from the Naito Foundation (to M. S.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence should be addressed. Fax: 46-8-15-3679; E-mail: gunnar@biokemi.su.se.

§ The abbreviations used are: ER, endoplasmic reticulum; Endo H,endo-β-N-acetylglucosaminidase; BHK, baby hamster kidney; TPCK, 1-tosylamide-2-phenylethyl chloromethyl ketone.
transformed with the plNGL1 vector (30) carrying the leader peptidase mutant WT-3K under control of the arabinose promoter was grown at 37°C in M9 minimal medium supplemented with 100 μg/ml ampicillin, 0.5% fructose, and all amino acids (50 μg/ml each) except methionine. An overnight culture was diluted 1:25 in fresh medium, shaken for 3.5 h at 37°C, induced with arabinose (0.2%) for 5 min, and labeled with [35S]methionine (150 μCi/ml). After 1 min, nonradioactive methionine was added (500 μg/ml) and incubation was stopped by chilling on ice. Cells were spun at 14,000 rpm for 2 min, resuspended in ice-cold buffer (40% w/v sucrose, 33 mM Tris pH 8.0), and incubated with lysozyme (5 μg/ml) and 1 mM EDTA for 15 min on ice. Aliquots of the cell suspension were incubated for 1 h on ice, either with 0.75 mg/ml TPCK-treated trypsin or with (0.75 mg/ml TPCK-treated trypsin inhibitor + 0.64 mg/ml phenylmethylsulfonyl fluoride). After the addition of trypsin inhibitor and phenylmethylsulfonyl fluoride, samples were acid-precipitated (trichloroacetic acid, 10% final concentration), resuspended in 10 mM Tris/2% SDS, immunoprecipitated with antisera to Lep and OmpA (an outer membrane control (31)), washed, and analyzed by SDS-polyacrylamide gel electrophoresis.

**In Vitro Transcription and Translation in Reticulocyte Lysate**—The pGEM1 plasmids carrying the relevant constructs were linearized at a SmaI site immediately downstream of the coding region prior to *in vitro* transcription. Synthesis of RNA by SP6 RNA polymerase and translation in reticulocyte lysate in the presence of dog pancreas microsomes was performed as described (32) or according to the protocol provided by Promega. Translation of polypeptides to the luminal side of the microsomes was assayed by prevention of N-linked glycosylation through competitive inhibition by the addition of a glycosylation acceptor tripeptide but not by a nonacceptor tripeptide and by proteinase K treatment of the microsomes (12). Alkaline extraction of microsomes was carried out as described (33). SDS-polyacrylamide gel electrophoresis gels were scanned in a Fuji X Cas 1000 PhosphoImaging Plate Scanner and analyzed using the MacBAS software (version 2.1).

**In Vivo Expression in BHK Cells**—The pSFV1 vector containing the relevant constructs were linearized at a unique *Sph*I site downstream of the cloned gene. The linearized plasmid was used as template for *in vitro* transcription, and mRNA was transfected into cells by electroporation as described earlier (34). 7 h post-transfection, cells were pulse-labeled with [35S]Met for 15 min, chased for 30 min, and solubilized in Nonidet P-40 (1%) lysate buffer (35). The lysate was centrifuged at 0.5% fructose, and all amino acids (50 μg/ml each) except methionine were added (500 μg/ml) and incubation was stopped by chilling on ice. Cells were spun at 14,000 rpm for 2 min, resuspended in ice-cold buffer (40% w/v sucrose, 33 mM Tris pH 8.0), and incubated with lysozyme (5 μg/ml) and 1 mM EDTA for 15 min on ice. Aliquots of the cell suspension were incubated for 1 h on ice, either with 0.75 mg/ml TPCK-treated trypsin or with (0.75 mg/ml TPCK-treated trypsin inhibitor + 0.64 mg/ml phenylmethylsulfonyl fluoride). After the addition of trypsin inhibitor and phenylmethylsulfonyl fluoride, samples were acid-precipitated (trichloroacetic acid, 10% final concentration), resuspended in 10 mM Tris/2% SDS, immunoprecipitated with antisera to Lep and OmpA (an outer membrane control (31)), washed, and analyzed by SDS-polyacrylamide gel electrophoresis.

**RESULTS**

**Construction and Expression of Model Proteins with One, Two, and Four Transmembrane Segments**—In order to be able to make direct comparisons between the topological effects of positively charged amino acids in membrane proteins expressed in prokaryotic and eukaryotic systems, we chose the well characterized *E. coli* inner membrane protein leader peptidase (Lep) as a model protein. Lep has been used extensively in studies of membrane protein assembly in *E. coli* (1) and has also been shown to insert efficiently into ER-derived dog pancreas microsomes (11, 12).

Lep constructs with one and two transmembrane segments and various point mutations, additions, or deletions (Fig. 1) were made from the wild type gene by site-directed mutagenesis. Molecules with four transmembrane segments were constructed by fusing the H1-L1-H2 region from one construct with the entire coding region from another (36). To map the cytoplasmic or extra-cytoplasmic location of the N-terminal tail (N<sub>ext</sub> and N<sub>cyt</sub>) and the C-terminal tail (C<sub>cyt</sub> and C<sub>ext</sub>) acceptor sites for N-linked glycosylation (see [35]) that can be identified if translocated to the lumen of the ER were introduced into these regions by site-directed mutagenesis (note that all constructs with an N-terminal glycosylation acceptor site have an extra N-terminal extension of 14 residues, except in wild type-derived (WT<sup>+</sup>) construct and construct ΔH2 (Fig. 2), where the extension is 16 residues). All genes were cloned behind the SP6 promoter in the pGEM1 vector for expression *in vitro* and into the Semliki Forest virus vector pSFV1 (28) for expression *in vivo*.

**A Single N-terminal Transmembrane Segment Is Sufficient for Targeting and Insertion of a Bacterial Protein into the ER Membrane**—In *E. coli*, Lep adopts a somewhat unusual membrane topology with both the N and C termini facing the periplasm. When expressed *in vitro* in the presence of microsomes, both termini are again translocated, whereas the P1 loop remains on the cytoplasmic side of the membrane (12). To check whether both the H1 and H2 transmembrane segments are necessary for targeting and insertion into the ER membrane, transmembrane segment H2 was deleted from WT<sup>+</sup> with an N-terminal extension including a glycosylation acceptor site and a mutation (Asn<sup>214</sup>→Gln) that removes the only acceptor site in the P2 domain, and the protein was expressed in the absence and the presence of microsomes. As seen in Fig. 2, the ΔH2 mutant was glycosylated on the N-terminal acceptor site to the same extent as WT<sup>+</sup> (lanes 2 and 5). When the microsomes were treated with proteinase K, a protease-resistant fragment representing the H2-P2 domain (12) was produced from WT<sup>+</sup> (Fig. 2, lane 3), whereas mutant ΔH2 was completely degraded (Fig. 2, lane 6). After disruption of the microsomes with the detergent Triton X-100, no protease-resistant material remained for either construct (data not shown). Together, these results demonstrate that H1 has an intrinsic ability to target to the ER membrane and to insert in the same N<sub>ext</sub>C<sub>cyt</sub> orientation as adopted in *E. coli*.

**N-terminal Positively Charged Residues Control the Topology of a Protein with Two Transmembrane Segments**—Lep has a highly charged cytoplasmic loop (P1) that cannot be translocated across the inner *E. coli* membrane, presumably because the positively charged amino acids immediately downstream of...
the first hydrophobic domain (H1) prevent translocation (37, 38). Thus, even though the addition of one or more lysines to the N-tail preceding H1 can cause certain Lep constructs lacking most of the charged residues in the P1 loop to insert into the inner membrane in an inverted N_cyt-C_cyt orientation with P1 exposed to the periplasm (31, 39), no translocation of the P1 loop was observed even when as many as three lysines were added to the N-tail in wild type Lep (mutant WT-3K). This mutant was expressed in E. coli, and its topology was probed by trypsin treatment of spheroplasts followed by immunoprecipitation with a Lep polyclonal antiserum (Fig. 3 A). Very little protein remained after protease treatment (Fig. 3A, lane 2), and there was no protease-protected fragment diagnostic for translocation of the P1 domain (31). Parallel analysis of the protease-sensitivity of the mature and precursor forms of the outer membrane protein OmpA was performed to ensure that the spheroplasts were intact. The protease sensitivity of the periplasmically exposed mature OmpA demonstrates that no intact cells remained, whereas the protease-resistance of cytoplasmic, nontranslocated precursor OmpA demonstrates that the spheroplasts were intact (31). Thus, the P2 domain in WT-3K is exposed to the periplasmic space, whereas the P1 domain is cytoplasmic.

It has been shown that translocation of the P2 domain across the microsomal membrane is much less sensitive to the presence of positively charged amino acids immediately downstream of H2 than when expression is carried out in E. coli (11), suggesting that the addition of one or more lysines to the normally uncharged N-tail of Lep might result in an inverted orientation in microsomes with the highly charged P1 loop in the lumen, in contrast to what is seen in E. coli. Wild type Lep expressed in vitro in the presence of dog pancreas microsomes was efficiently glycosylated on its only potential acceptor site for N-linked glycosylation (Asn<sup>214</sup>) (Fig. 3B, top panel, lanes 1 and 2), demonstrating the luminal disposition of the P2 domain. The addition of a single N-terminal lysine between Met<sup>4</sup> and Phe<sup>5</sup> resulted in a marked reduction in glycosylation efficiency (Fig. 3B, top panel, lanes 3 and 4), and with two and four N-terminal lysines the glycosylation efficiency was only ~20%, consistent with a predominantly N_cyt-C_cyt topology. The decrease in glycosylation was paralleled by a concomitant decrease in the amount of protease-resistant H2-P2 fragment (data not shown). In all cases, both the glycosylated and non-glycosylated forms were properly integrated into the microsomal membrane as shown by their retention in the pellet after alkaline extraction (Ref. 33 and data shown for the 0K and 2K mutants in Fig. 3B, bottom panel). We conclude that positively charged residues in the N-terminal tail have a strong influence on the topology of Lep expressed both in the E. coli and microsomal systems, whereas positively charged residues in the P1 loop between the two transmembrane segments block translocation only in E. coli.

![Diagram](image)

**Fig. 2.** H1 inserts into microsomes with N_cyt-C_cyt topology. Mutants were expressed *in vitro* in rabbit reticulocyte lysate either in the absence (−) or the presence (+) of rough microsomes (RM). WT<sup>*</sup> differs from wild type Lep by an Asn<sup>214</sup> → Gln mutation in P2 and by the addition of a short N-terminal extension including a glycosylation acceptor site (compared with the sequence of NT/0K* in Fig. 1, an additional pair of residues (Val-Ala) were inserted between Pro<sup>12</sup> and Gly<sup>13</sup>). ΔH2 is the same as WT<sup>*</sup> but with transmembrane segment H2 (residues 62–76) deleted. Nonglycosylated and glycosylated products are indicated by black and white circles, respectively. The protease K (PK)-resistant fragment in WT<sup>*</sup> marked by a black square represents the luminally located H2-P2 domain.

The Length of the Loop between Two Transmembrane Segments Does Not Affect Topology—Another variable that has been shown to influence the topology of Lep-derived constructs in E. coli is the length of the P1 loop; although this 39-residue-long loop cannot be translocated (Fig. 3A), translocation is possible when its length is increased to more than ~60 residues (40). It has been speculated that this may be related to the participation of the Sec translocase in the translocation of long loops across the inner membrane; only a small number of positively charged residues would be allowed in short loops (<60 residues) that cannot use the Sec translocase for translocation, whereas longer loops that use the more permissive Sec pathway would have less stringent sequence requirements (41).

To test if the length of the P1 loop affects topology also in the ER membrane, a series of constructs with zero, one, two, and four N-terminal lysines and P1 lengths ranging from 19 to 88 residues was analyzed in the microsomal *in vitro* system. As shown in Fig. 4, the length of the P1 loop has at best a minor and not very consistent effect on the topology in the 1K series of mutants and has no effect for the 0K, 2K, and 4K series, in contrast to the situation in E. coli.

The construct with 88 residues in the P1 loop was glycosylated somewhat more efficiently than the other constructs when two and four lysines were present in the N-tail (Fig. 4B); however, in this particular case there is an additional glycosylation site in the P1 loop, ~15 residues downstream of H1. The higher glycosylation efficiency may thus result from partial glycosylation of this site rather than from translocation of the P2 domain. To ascertain that this was indeed the case, a second glycosylation site was inserted into the middle of the P1 loop, and the site in P2 was simultaneously removed. The resulting construct 2K(88*) with two N-terminal lysines and with two glycosylation sites in P1 and none in P2 was efficiently glycosylated (Fig. 4A, lanes 11 and 12); 20% of the chains were glycosylated on both sites, and 70% were glycosylated only on one, demonstrating that no more than ~10% of the molecules have the wild type orientation (or fail altogether to integrate into the membrane). N-terminal positively charged residues thus efficiently promote an inverted, N_cyt-C_cyt topology in the ER membrane irrespective of the length of the P1 loop.

The Topology of “Nonfrustrated” Proteins Follows the Posi-
Previous topological analyses in E. coli of Lep-derived constructs with four transmembrane segments and different distributions of positively charged residues in the N-tail and the loops connecting the transmembrane segments have demonstrated that “frustrated” constructs, i.e., molecules where the distribution of positively charged residues is such that neither the N\textsubscript{ext}-C\textsubscript{ext} nor the N\textsubscript{cyt}-C\textsubscript{cyt} topology will have all the highly charged loops in the cytoplasm, often adopt a “leave one out” topology with only three of the hydrophobic segments inserted across the membrane (36).

To test if the influence of positively charged residues is equally strong in the ER, a subset of the constructs previously characterized in E. coli was expressed in the microsome system. First, two nonfrustrated constructs that adopt, respectively, the N\textsubscript{ext}-C\textsubscript{ext} (i.e. N and C terminus in the periplasm) and the N\textsubscript{cyt}-C\textsubscript{cyt} topology in E. coli were analyzed. As shown in Fig. 5A, construct 0K/3K/0K/3K with three lysines in each of the loops between the first and second and between the third and fourth transmembrane segments was efficiently glycosylated both in the C-terminal P2 domain (Fig. 5A, middle panel, lane 2) and on an Asn-Ser-Thr acceptor site introduced into the N-tail (Fig. 5A, top panel, lane 2) and thus adopts the N\textsubscript{ext}-C\textsubscript{ext} topology just as in E. coli. The nonglycosylated full-length product of construct 0K/3K/0K/3K (Fig. 5A, top panel, lane 7) has the same size as the corresponding 0K/3K/0K/3K construct expressed in E. coli (36) (Fig. 5A, top panel, lane 8).

Protease K treatment of the microsomes (Fig. 5A, top panel, lane 5) produced a protected fragment that runs at the approximate location of the weak band in lane 1 of the top panel of Fig. 5A marked by a black square in the middle and bottom panels and considerably below this band in the top panel. This suggests that the protected fragment is glycosylated when the P2 domain carries a glycosylation site (Fig. 5A, middle and bottom panels) but not when the glycosylation site is in the N-tail (Fig. 5A, top panel). Indeed, the protected fragment in the middle and bottom panels of Fig. 5A has the same size as a glycosylated, protease-protected fragment derived from microsome-integrated wild type Lep that was previously shown to correspond to the glycosylated H2-P2 domain (Ref. 12 and data not shown), and we conclude that the protected fragment in Fig. 5A results from cleavage in the L1’ loop.

With glycosylation sites simultaneously present in both the N-tail and the P2 domain, a majority of the molecules were glycosylated on both sites (Fig. 5A, bottom panel, lane 2), suggesting that all membrane-inserted molecules are glycosylated in the P2 domain and that most of these are also glycosylated on the N-tail. To rule out that the rather small number of residues between the N-terminal glycosylation site and H1 cause the somewhat inefficient glycosylation of the N-tail (12), we also made constructs where this distance was increased to 19 and 25 residues. Both were glycosylated with the same efficiency as the original construct (data not shown), suggesting that the N-terminal glycosylation site is either intrinsically less easily modified than the site in the P2 domain or that the N-tail is somewhat less efficiently translocated than P2. In any case, these data show that the majority of the molecules have the expected N\textsubscript{ext}-C\textsubscript{ext} orientation with the L1’ loop (and most likely also the L1 loop) exposed on the cytoplasmic side of the microsomal membrane.

In contrast, construct 3K/0K/3K/0K with three lysines in the N-tail and in the loop between transmembrane segments 2 and 3 was found to be essentially unglycosylated on both the N- and C-terminal acceptor sites (Fig. 5B, top panel), suggesting the same N\textsubscript{ext}-C\textsubscript{cyt} topology as observed in E. coli. The trivial explanation that this particular construct is not inserted into the membrane was ruled out by alkaline extraction of the micro-
somes (33); as seen in Fig. 5B (bottom panel), a large fraction of the molecules partitioned with the wash supernatant when translation was carried out in the absence of microsomes (or when microsomes were added post-translationally; data not shown), whereas essentially no molecules were found in this fraction when translation was carried out in the presence of microsomes, demonstrating proper membrane integration. Further, the long L1 loop in the related construct 3K/88/3K/0K (see below) is efficiently glycosylated, again consistent with proper membrane assembly and an Ncyt-Cext topology.

Both nonfrustrated constructs with a glycosylation acceptor site in the P2 domain were also expressed in vivo in BHK cells. As in vitro, construct 3K/0K/3K/0K was not glycosylated, whereas construct 0K/3K/0K/3K was fully glycosylated (Fig. 5C), demonstrating efficient targeting and insertion into the ER membrane (we cannot completely rule out that some molecules may fail to integrate and are rapidly degraded). The two lower molecular weight bands marked with black and white squares are also faintly seen in most of the in vitro experiments. These bands run at roughly the same position as non-glycosylated and glycosylated wild type Lep, respectively (data not shown), and may result either from endogenous proteolysis or from internal initiation at one or other of the four methionine codons present in the L2 loop (see Fig. 1).

We conclude that nonfrustrated constructs with four transmembrane segments based on a bacterial inner membrane protein can be targeted and inserted into the ER membrane both in vitro and in vivo and that they adopt the same topology as in E. coli.

Membrane Insertion of Frustrated Proteins—Two frustrated constructs, previously shown to adopt leave one out topologies in E. coli, were also tested in the ER system. Construct 0K/3K/3K/0K was glycosylated on the N terminus and in the P2 domain both in vitro and in vivo and was doubly glycosylated when acceptor sites were simultaneously present in these two regions (Fig. 6A, bottom), suggesting that the topology is predominantly Ncyt-Cext with four transmembrane segments (Fig. 6B, bottom panels). This is in contrast to E. coli, where this construct adopts a leave one out topology with only three transmembrane segments (36).

On the other hand, the N-tail in construct 3K/0K/0K/3K was not glycosylated, whereas the P2 domain was efficiently glycosylated (Fig. 6A, left top panels). When expressed in vivo, this construct was again fully glycosylated in the P2 domain (Fig. 6A, right top panel). As in E. coli, the dominating topology is thus Ncyt-Cext, with only three transmembrane segments (Fig. 6B). It is formally possible that the second hydrophobic segment is translocated into the lumen with the first, third, and fourth segments spanning the membrane; however, given that there is no precedence for the translocation of hydrophobic segments across the ER membrane, we consider this unlikely. Because this construct is slightly less efficiently glycosylated in the P2 domain in vitro than the nonfrustrated construct 0K/3K/0K/3K (65% versus 75%) and because most of the nonglycosylated molecules are found in the membrane fraction after alkaline extraction in both cases (data not shown), a minor proportion of the molecules (~10%) may have the P2 domain in the cytoplasm. This is not the case in vivo, however. Which of the two uncharged loops that is exposed to the luminal side cannot be determined with the glycosylation assay. Protease treatment of the microsomes did not give interpretable results for this mutant, because no unique protease-resistant fragment was produced.

In order to address this point further, two constructs where the L1 loop between the first and second transmembrane segments is 88 rather than 25 residues long were also tested. A
glycosylation acceptor site placed in this loop is far enough from transmembrane segment H1 (15 residues) to be accessible to the lumenally disposed glycosyltransferase, and the location of the loop can thus be mapped. In the nonfrustrated construct 3K/88/3K/0K, the long loop was quite efficiently glycosylated (Fig. 7A, lane 3), and no doubly glycosylated molecules were apparent when the acceptor site in the P2 domain was also present (Fig. 7A, lane 2), demonstrating that the dominating topology is N\textsubscript{ac}-C\textsubscript{cyt}, as expected. The slight increase in glycosylation efficiency (from 45 to 55%; compare lanes 2 and 3, Fig. 7A) when the acceptor site in the P2 domain was added suggests that a small fraction of the molecules may have the long loop in the cytoplasm and the P2 domain in the lumen.

In the frustrated construct 3K/88/0K/3K, both the long loop and the P2 domain could be glycosylated. With one acceptor site in the long loop and one in the P2 domain, molecules with
both one and two glycosyl moieties were apparent (Fig. 7A, lane 4); with an additional acceptor site present in the P2 domain, the molecules were glycosylated either once or on all three sites (Fig. 7A, lane 5). The absence of twice glycosylated molecules modified only in the P2 domain in the latter construct demonstrates that no molecules have a topology with the P2 domain and the short L2 loop between the second and third transmembrane segments in the lumen. Thus, the majority of the molecules have both the long loop and the P2 domain in the lumen (N\textsubscript{cyt}-C\textsubscript{ext}, leave one out topology) (Fig. 7A), although the singly glycosylated molecules in lane 5 of Fig. 7A suggest the additional presence of molecules with four transmembrane segments and N\textsubscript{cyt}-C\textsubscript{cyt} orientation. Similar proportions of nonglycosylated and glycosylated molecules were found in the membrane pellet after alkaline extraction (data not shown).

**DISCUSSION**

We have constructed a number of model proteins with either one, two, or four hydrophobic transmembrane segments and have analyzed their membrane assembly both in vitro in dog pancreas microsomes and in vivo in BHK cells. Most of these constructs have previously been analyzed in E. coli (36, 39, 40), and we can thus make a detailed comparison between the sequence determinants of membrane protein topology in prokaryotic and eukaryotic cells.

In general, we find that positively charged residues placed in the N-tail have similar effects on the topology of model proteins in both the prokaryotic and eukaryotic systems, i.e. that such residues prevent translocation of the N-tail across the membrane. However, there are also some obvious differences.

In constructs with two transmembrane segments, the length of the connecting loop does not seem to be important for its ability to be translocated across the ER membrane (Fig. 4), whereas in E. coli short and long loops apparently use different mechanisms for translocation and react differently to the presence of positively charged residues (40–43). Thus, when two lysines are added to the N terminus of wild type Lep, most of the molecules insert into the ER membrane in an inverted N\textsubscript{cyt}-C\textsubscript{ext} orientation (Fig. 3), whereas addition of three lysines to the N-tail has no effect on the orientation when the protein is expressed in E. coli. Our data thus suggest that positively charged residues placed in the N-tail are critical topological determinants in both E. coli and mammalian ER, whereas charged residues in the following P1 loop have a major effect on the topology only in E. coli.

Three of the four constructs with four transmembrane segments also behave similarly in the two systems. The nonfrustrated constructs 3K/0K/3K/0K and 0K/3K/0K/3K insert according to the positive inside rule both in E. coli (36) and in the
which highly charged loops are translocated in the eukaryotic system may be related to the absence of a membrane potential in the ER (cf. Ref. 45) and may explain the earlier observation that extra- cytoplasmic loops in eukaryotic plasma membrane proteins have on average a higher content of positively charged residues than found in bacterial inner membrane proteins (2–4).

We have previously suggested (36, 46) that the so-called Sec-independent mechanism of membrane protein assembly in E. coli is based on helical hairpins composed of two neighboring hydrophobic segments connected by a short loop lacking in positively charged residues that insert en bloc. A purely sequential insertion process, on the other hand, where the orientation of all downstream transmembrane segments is dictated by the orientation of the most N-terminal one (47), is inconsistent with the leave one out topologies found in E. coli and with the N<sub>ext</sub>-C<sub>ext</sub> topology found for constructs 3K/OK/OK/3K an 3K/88/OK/3K in the ER, as well as with the results of a recent study on the topology of P-glycoprotein mutants in the ER (48). Possibly, proteins with widely spaced hydrophobic segments insert in a sequential fashion (49, 50), whereas regions with closely spaced hydrophobic segments insert according to a helical hairpin mechanism.

Finally, we note that the observation that the topological information present in certain of our constructs is interpreted differently in prokaryotic and eukaryotic cells suggests that problems encountered when trying to overexpress eukaryotic membrane proteins in prokaryotic hosts may in some cases be related to incorrect topologies. If so, redesigning the protein to conform better to the stricter prokaryotic version of the positive inside rule may be one way to obtain higher expression levels (51).

Acknowledgments—Glycosylation acceptor and nonacceptor tripeptides were kindly provided by Dr. Henrik Garoff (Karolinska Institute Center for Biotechnology).

REFERENCES

1. von Heijne, G. (1994) Annu. Rev. Biophys. Biomol. Struct. 23, 167–192
2. von Heijne, G., and Gavel, Y. (1988) Eur. J. Biochem. 174, 671–678
3. Sipos, L., and von Heijne, G. (1993) Eur. J. Biochem. 213, 1333–1340
4. Wallin, E., and von Heijne, G. (1993) Protein Eng. 6, 683–688
5. Jones, D. T., Taylor, W. R., and Thornton, J. M. (1994) Biochemistry 33, 3038–3049
6. Hartmann, E., Rapoport, T. A., and Lodish, H. F. (1989) Proc. Natl. Acad. Sci. U.S.A. 86, 5786–5790
7. Landolt-Marticorena, C., Williams, K. A., Deber, C. M., and Reithmeier, R. A. (1993) J. Mol. Biol. 229, 602–608
8. Hennessey, E. S., Hashemzadeh-Bonehi, L., Hunt, L. A., and Broome-Smith, J. (1995) FEBS Lett. 361, 159–161
9. Grishhammer, R., and Tate, C. G. (1995) Q. Rev. Biochem. 28, 315–422
10. Geller, D., Taglicht, D., Edgar, R., Tam, A., Pines, O., Michaelis, S., and Bibi, E. (1996) J. Biol. Chem. 271, 13746–13753
11. Johansson, M., Nilsson, I., and von Heijne, G. (1993) Mol. Gen. Genet. 239, 251–256
12. Nilsson, I., and von Heijne, G. (1993) J. Biol. Chem. 268, 5798–5801
13. Denzer, A. J., Nahholz, C. E., and Spiess, M. (1995) EMBO J. 14, 6311–6317
14. Szczesna-Skorupa, E., Brown, N., Mead, D., and Kemper, B. (1998) Proc. Natl. Acad. Sci. U.S.A. 85, 738–742
15. Szczesna-Skorupa, E., and Kemper, B. (1989) J. Cell Biol. 108, 1237–1243
16. Parks, G. D., and Lamb, R. A. (1991) Cell 64, 777–787
17. Parks, G. D., and Lamb, R. A. (1993) J. Biol. Chem. 268, 19101–19109
18. Sato, T., Sakaguchi, M., Mihara, K., and Omura, T. (1990) EMBO J. 9, 2391–2397
19. Beltzer, J. P., Fiedler, K., Fuhrer, C., Geffen, I., Handschin, C., Wessels, H. P., and Spiess, M. (1991) J. Biol. Chem. 266, 973–978
20. Andrews, D. W., Young, J. C., Mirsky, L. F., and Czarnota, G. J. (1992) J. Biol. Chem. 267, 7761–7769
21. Sakaguchi, M., Tomiyoshi, R., Kuriwa, T., Mihara, K., and Omura, T. (1992) Proc. Natl. Acad. Sci. U.S.A. 89, 16–19
22. Spiess, M. (1995) FEBS Lett. 369, 76–79
23. Harley, C. A., and Tipper, D. J. (1996) J. Biol. Chem. 271, 24625–24633
24. Erickson, B. W., and Merrifield, R. B. (1970) in The Proteins (Hill, R. L., and Neurath, H., eds) Vol. 2, pp. 555–527, Academic Press, London
25. Kunkel, T. A. (1985) Proc. Natl. Acad. Sci. U.S.A. 82, 488–492
26. Geisselsoneder, J., Witney, F., and Yuckenberg, P. (1987) BioTechniques 5, 786–781
27. Kozak, M. (1989) Mol. Cell. Biol. 9, 5073–5080
28. Liljestro¨m, P., and Garoff, H. (1991) Bio/Technology 9, 1356–1361
29. Dalbey, R. E., and Wickner, W. (1986) J. Biol. Chem. 261, 13844–13849

FIG. 7. Topology of constructs with one long internal loop. A. Constructs 3K/88/3K/0K (lanes 1–3) and 3K/88/0K/3K (lanes 4 and 5) were expressed either in the absence (−) or the presence (+) of rough microsomes. Glycosylation acceptor sites in the long loop and P2 are indicated by an asterisk. Non-glycosylated products are indicated by a black circle, singly glycosylated products are indicated by a white circle, doubly glycosylated products are indicated by two white circles, and triply glycosylated products are indicated by three white circles. B. Dominating topologies suggested by the glycosylation data in panel A. Nonmodified glycosylation sites are circled.
Membrane Protein Topology

30. Johnston, S., Lee, J. H., and Ray, D. S. (1985) Gene (Amst.) 34, 137–145
31. von Heijne, G. (1989) Nature 341, 456–458
32. Liljestrom, P., and Garoff, H. (1991) J. Virol. 65, 147–154
33. Sakaguchi, M., Mihara, K., and Sato, R. (1987) EMBO J. 6, 2425–2431
34. Liljestrom, P., Lusa, S., Huylenebroek, D., and Garoff, H. (1991) J. Virol. 65, 4107–4113
35. Wahlberg, J. M., Boere, W. A., and Garoff, H. (1989) J. Virol. 63, 4991–4997
36. Gafvelin, G., and von Heijne, G. (1994) Cell 77, 401–412
37. von Heijne, G., Wickner, W., and Dalbey, R. E. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 3563–3566
38. Andersson, H., and von Heijne, G. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 9751–9754
39. Nilsson, I. M., and von Heijne, G. (1990) Cell 62, 1135–1141
40. Andersson, H., and von Heijne, G. (1993) EMBO J. 12, 685–691
41. von Heijne, G. (1994) FEBS Lett. 346, 69–72
42. Andersson, H., and von Heijne, G. (1994) EMBO J. 13, 2267–2272
43. Saäf, A., Andersson, H., Gafvelin, G., and von Heijne, G. (1995) Mol. Membr. Biol. 12, 209–215
44. von Heijne, G. (1986) J. Mol. Biol. 189, 239–242
45. Andersson, H., and von Heijne, G. (1994) EMBO J. 13, 2267–2272
46. Cao, G., Cheng, S., Whitley, P., von Heijne, G., Kuhn, A., and Dalbey, R. E. (1988) J. Biol. Chem. 263, 26888–26893
47. Blobel, G. (1980) Proc. Natl. Acad. Sci. U. S. A. 77, 1496–1500
48. Zhang, J.-T., Lee, C. H., Duthie, M., and Ling, V. (1995) J. Biol. Chem. 270, 1742–1746
49. Wessels, H. P., and Spiess, M. (1988) Cell 55, 61–70
50. Lipp, J., Flint, N., Haeuptle, M.-T., and Dobberstein, B. (1989) J. Cell Biol. 109, 2013–2022
51. Kiefer, H., Krieger, J., Olszewski, J. D., von Heijne, G., Prestwich, G. D., and Breer, H. (1996) Biochemistry 35, 16077–16084