2

Decoding the Signals of Membrane Protein Sequences

GUNNAR VON HEIJNE

In order to understand membrane protein structure in a general sense, one needs an efficient conceptual framework. When discussing globular proteins, we tend to use hierarchical descriptions and distinguish what we think are semi-independent levels of increasing structural complexity (Brändén and Tooze, 1991): the amino acid sequence (primary structure); local structural elements such as helices, β-strands, and turns (secondary structure); structural units (β–α–β units, β-strand hairpins); larger folding domains (Rossman fold, TIM barrel); globular single-chain structures (tertiary structure); and oligomeric assemblies (quaternary structure). In addition to being a purely descriptive scheme, it is often assumed that the unidirectional path leading from simple, locally determined structural elements to complex, global structures also in some way approximates the real *in vivo* folding process, although this is still very much a controversial point (Dill, 1990).

A similar scheme based on a "top-down" analysis of known three-dimensional structures can of course be devised to describe integral membrane proteins (see Chapters 1 and 3), but I will argue that a more fruitful way to understand their structure is in terms of how they are synthesized, targeted to the appropriate membrane, inserted into this membrane, and finally form their full three-dimensional-structure within the confines of a lipid bilayer. Thus we need to widen our conceptual universe to include not only the static picture of the final structure but also the dynamics of the biosynthetic pathway leading to this structure. In fact, since many important aspects of membrane protein biosynthesis seem to rely on rather well-defined "signals" encoded in the protein chain (targeting signals, topological signals, packing signals), the resulting logic is a rather simple one of successive "decoding" of these signals, ultimately giving rise to the correctly folded, functional molecule in its proper location.

This chapter is organized around three main sections. First, I will give a brief introduction to the general problem of intracellular protein sorting. Then follows a discussion of how proteins insert into membranes and how their transmembrane topology is determined, and finally I try to derive some implications for membrane protein structure prediction. Throughout, the focus is on the so-called helical bundle integral membrane proteins, i.e., proteins with transmembrane segments formed by long hydrophobic α-helices. Typical examples of this structural class are bacterio-

S. H. White (ed.), *Membrane Protein Structure*
© American Physiological Society 1994

rhodopsin (Henderson et al., 1990) and the bacterial photosynthetic reaction center (Deisenhofer and Michel, 1991); in general, all bacterial inner membrane proteins, all eukaryotic membrane proteins initially inserted into the endoplasmic reticular (ER) membrane, all mitochondrial inner membrane proteins, and all chloroplast thylakoid membrane proteins most likely are of this type. The second well-characterized class of integral membrane proteins is the"β-barrel" variety, in which the transmembrane segments are amphiphilic β-strands that together form a closed β-sheet barrel with a hydrophobic outside facing the lipids and a more polar inside lining a central pore. Such structures are known to atomic resolution (e.g. Weiss et al., 1991), but relatively little information on the membrane insertion of such molecules is available (Bosch et al., 1988, 1989; Struyve et al., 1991), and they will not be discussed further here.

Targeting Signals

All cells need to be able to sort proteins between a number of subcellular compartments. In *Escherichia coli,* there are at least five well-defined compartments: the cytoplasm, the inner and outer membranes, the periplasm, and the extracellular medium. A membrane protein must thus somehow"know" that it should not remain in the cytoplasm and further whether it is supposed to go to the inner or outer membrane. The highly complex subcellular structure of eukaryotic cells makes protein sorting a very complicated business, in which proteins must not only be routed into the correct organelle but further sorted to the correct intraorganellar compartment.

In all cases known, the targeting information is encoded within the nascent polypeptide, often as an N-terminal extension that is removed by appropriately located proteases once the correct compartment has been reached (von Heijne, 1990a). The targeting signal in the nascent protein is recognized by receptors in the cytoplasm or on the surface of the organelle, and the protein is translocated across one or more membranes and finally delivered to its site of action.

Sorting Between Organelles

Major recipients of cytoplasmically synthesized proteins are mitochondria, chloroplasts, the nucleus, peroxisomes, and the organelles in the secretory pathway. Targeting signals specific for each of these organelles have been defined by, for example, gene fusion studies, and their basic designs have been elucidated by a combination of experimental and statistical techniques.

Mitochondrial targeting peptides are in most cases found as N-terminal extensions that are cleaved by a mitochondrial matrix protease. They are rich in positively charged (Arg in particular) and hydroxylated (Ser, Thr) amino acids, but lack negatively charged residues (Asp, Glu). An apparently very important property is their ability to form amphiphilic α-helices with one highly charged and one hydrophobic face (von Heijne, 1986b; Gavel and von Heijne, 1990a) (Fig. 2.1).

Chloroplast transit peptides from higher plants are also N-terminal extensions and are removed in the chloroplast by a stromal processing peptidase. They are characterized by an extremely high content of hydroxylated amino acids ($\sim$30% Ser + Thr) and contain few if any negatively charged residues (von Heijne et al., 1989;

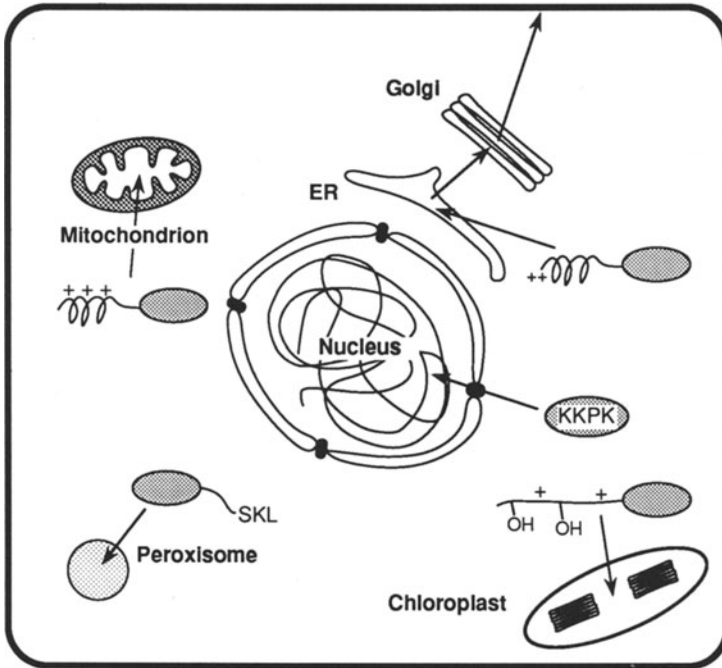

Fig. 2.1. Protein sorting pathways.

Gavel and von Heijne, 1990b). It is not clear if they are designed to have any particular conformational preferences (von Heijne and Nishikawa, 1991). Transit peptides from the green algae *Chlamydomonas reinhardtii,* in contrast, are strikingly similar to mitochondrial targeting peptides and probably form amphiphilic α-helices (Franzén et al., 1990), raising the question of how the specificity of proteins sorting between mitochondria and chloroplasts is maintained in this organism.

Nuclear location signals are not removed from the mature protein and are generally found in internal positions (Silver, 1991). Clusters of positively charged residues presumably exposed on the surface of the folded molecule signal nuclear import, and the presence of multiple copies of an import signal in a protein chain often lead to enhanced levels of import.

Peroxisomal targeting signals are less well understood, but a C-terminal tripeptide Ser-Lys-Leu (SKL) has been shown to promote peroxisomal import in a number of cases (Gould et al., 1989, 1990; Miyazawa et al., 1989). Many peroxisomal proteins lacking this signal are known, however.

Secretory signal peptides, finally, have a tripartite design, with an N-terminal positively charged region, a central hydrophobic region, and a C-terminal region that specifies the cleavage site (von Heijne, 1985, 1990b; Gierasch, 1989). This basic structure is similar throughout nature, from bacteria to man.

Sorting Within Organelles

Sorting within the mitochondrion is believed to follow a "conservative" sorting pathway (Hartl et al., 1989), i.e., proteins destined for the inner membrane or the inter-

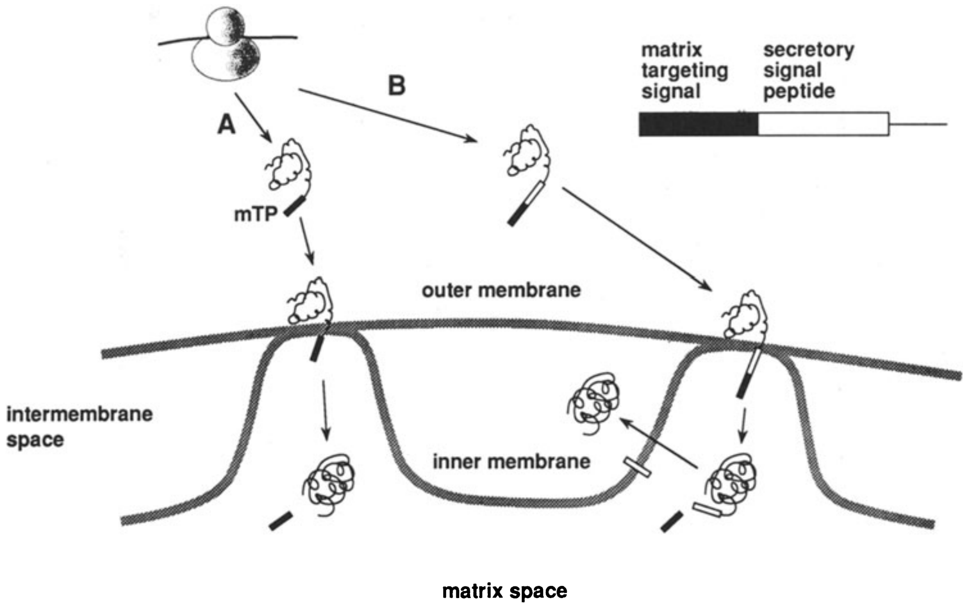

Fig. 2.2. Intramitochondrial protein sorting. Matrix-targeting signals (mTP) allow entry into the matrix space *(path A)*. Signals composed of a matrix-targeting part and a signal peptide-like part allow transport into the intermembrane space via a "conservative sorting" pathway (*B*).

membrane space are first imported all the way into the matrix and subsequently re-exported across or inserted into the inner membrane (Fig. 2.2). This two-step process is neatly mirrored in the design of the targeting peptides (von Heijne et al., 1989): an N-terminal matrix-targeting signal (positively charged amphiphilic α-helix) is immediately followed by a second cleavable targeting signal that has all the characteristics of a secretory signal peptide. It is thus thought that the mitochondrial inner membrane has a machinery for protein translocation similar to that found in bacteria.

Thylakoid lumen proteins, as well as some of the thylakoid membrane proteins, also have composite targeting peptides of this kind, but with the N-terminal part now being a chloroplast transit peptide (von Heijne et al., 1989). The analogy with secretory signal peptides goes even further in this case, since the substrate specificity of the so-called thylakoid processing peptidase responsible for removing the signal peptide-like thylakoid transfer domain is nearly identical to that of the corresponding *E. coli* enzyme (Halpin et al., 1989; Shackleton and Robinson, 1991).

Sorting in the Secretory Pathway

The secretory pathway comprises a number of subcompartments that are traversed in sequel from the ER through the Golgi stacks and the trans-Golgi network to the plasma membrane or to the lysosome (Breitfeld et al., 1989; Pugsley, 1989). Each subcompartment has its own specific complement of resident proteins that are thought to be actively retained in or recycled to that compartment in response to retention signals present in their amino acid sequence. Thus, lumenal ER proteins

have a C-terminal tetrapeptide retention signal (Lys-Asp-Glu-Leu or -KDEL) (Pelham, 1990). Resident ER membrane proteins seem to have a retention signal located in their cytoplasmically exposed parts; one or two lysines in their cytoplasmic tails have been implicated in retention, but this is still a somewhat controversial point (Gabathuler and Kvist, 1990; Jackson et al., 1990).

Transmembrane proteins can also be specifically degraded in the ER. This"quality control" feature may enable the cell to remove, for example, subunits of multichain complexes that have failed to associate properly with their partners (Klausner and Sitia, 1990). Isolated charged residues present in a transmembrane segment seem to signal ER degradation (Bonifacino et al., 1990); presumably, such degradation signals become masked when subunits associate with each other.

Some resident Golgi membrane proteins apparently have a retention signal located in their transmembrane segment(s) (Machamer and Rose, 1987; Colley et al., 1989); other Golgi-specific retention signals have not yet been defined. Lumenal lysosomal proteins carry a mannose-6-phosphate modification that serves as the targeting signal and ensures routing to the lysosome from the trans-Golgi network (Kornfeld and Mellman, 1989; Baranski et al., 1990); lysosomal membrane proteins, on the other hand, carry no such modification, and targeting information is believed to be present in their cytoplasmic domains (Peters et al., 1990).

A final aspect of membrane protein sorting is that of endocytosis. Many plasma membrane receptors continually cycle back and forth between the plasma membrane and the trans-Golgi network via endocytosis through coated pits. A critical tyrosine placed in the cytoplasmic tail near the membrane has been implicated as being part of the endocytosis signal (Jing et al., 1990; Ktistakis et al., 1990; Bansal and Gierasch, 1991; Eberle et al., 1991).

Topological Signals

Once a protein has reached its target membrane, it needs to insert into that membrane in its correct orientation. Once inserted, changes in the orientation of the whole protein or of individual transmembrane segments (flip-flop) would seem to be impossible on energetic grounds and has never been observed experimentally. It is during the insertion process that the number of transmembrane segments in the final structure as well as their orientation are decided. For a multispanning (polytopic) protein with most of its chain embedded in the membrane, the membrane insertion event is thus the most important step on the folding pathway; what remains after this step has been completed is only the final packing of the transmembrane segments against each other.

Topological Signals: An Overview

Most membrane proteins use one of the machineries normally used to translocate proteins across membranes for their insertion. From this point of view, a membrane protein can be regarded as a partially translocated protein—a molecule that, in addition to the normal targeting signal(s), contains additional signals that cannot be

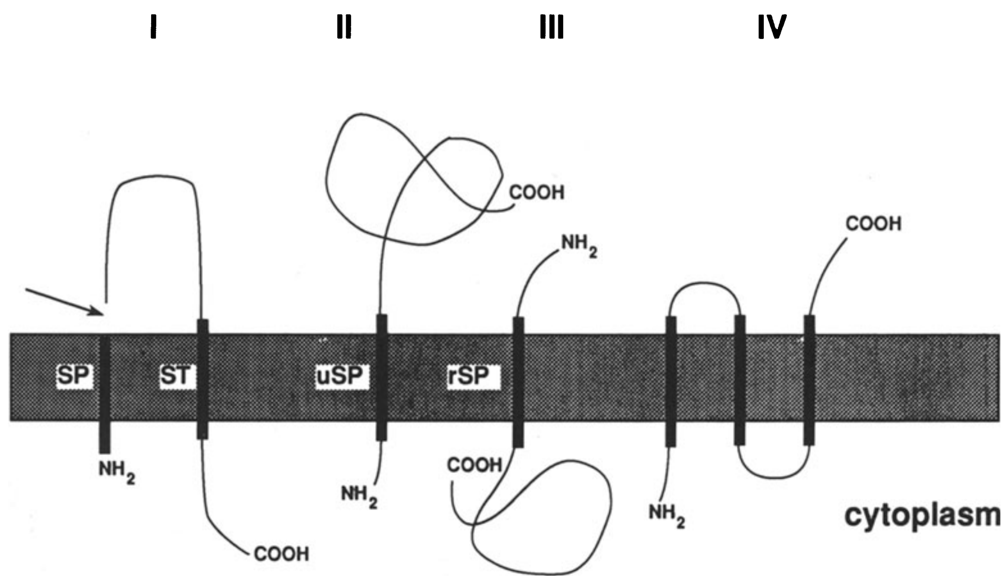

Fig. 2.3. Classification of integral membrane proteins. *SP,* signal peptide; *ST,* stop-transfer sequence; *uSP,* uncleaved signal peptide; *rSP,* reverse signal peptide.

translocated across the membrane but rather get stuck and provide a permanent transmembrane anchoring.

One useful classification based on the kinds of signals present in a protein is shown in Figure 2.3 (von Heijne, 1988). Class I proteins have a cleavable signal peptide and a second hydrophobic stop-transfer sequence; the orientation of the mature protein is N_{out}–C_{in}. Class II proteins have an uncleavable signal peptide, i.e., a signal peptide lacking the C-terminal cleavage domain, and hence get anchored to the membrane in the N_{in}–C_{out} orientation. Class III proteins also lack a cleavable signal peptide and have a hydrophobic transmembrane segment close to the N terminus (a "reverse signal peptide" [Dalbey, 1990]), but this segment is now oriented N_{out}–C_{in}, i.e., opposite to the class II proteins. Class IV proteins have multiple transmembrane segments, and examples are known with all possible combinations of N_{in}, N_{out}, C_{in}, and C_{out} orientations.

Statistical Studies: The "Positive Inside" Rule

Proteins with a cleavable signal peptide (class I) are always oriented N_{out}, but it is not immediately clear what feature(s) of the nascent chain determine the orientation of the other classes. Early statistical studies of both single-spanning (classes II and III) and multispanning proteins gave the first hint that topological determinants should be sought in the polar tail and loop regions flanking the transmembrane segments and that the distributions of positively charged residues (Arg, Lys) seemed to be particularly well correlated with the topology (von Heijne, 1986a,d). In a sample of bacterial inner membrane proteins, a fourfold higher frequency of Arg + Lys was found in the cytoplasmic than in the periplasmic flanking regions, and similar,

though less extreme, biases have since been found in proteins from other membrane systems as well (see below). Thus a "positive inside" rule (von Heijne and Gavel, 1988) was proposed to account for the topology of most integral membrane proteins.

Membrane Protein Topogenesis in Escherichia coli

Translocation of proteins across the inner membrane of *E. coli* normally requires the participation of the *sec* machinery (Schatz and Beckwith, 1990; Wickner et al., 1991). The nascent precursor protein first binds to a cytoplasmic chaperone such as SecB, is then transferred to SecA (a large peripheral membrane protein with ATPase activity), is translocated across the membrane in a process involving the integral membrane components SecY and SecE (and possibly the late-acting factors SecD and SecF), and is finally released from the membrane after the signal peptide has been cleaved by the Lep protease.

Many inner membrane proteins also need the *sec* machinery for translocation of their periplasmic domain(s), but some do not (notably those with only short periplasmic domains; see below). Obviously, the topology may be determined in different ways during *sec*-dependent and *sec*-independent membrane insertion, although it appears from the data at hand that positively charged amino acids are important in both contexts.

Most of what is known about membrane protein insertion in *E. coli* comes from studies on three proteins: phage M13 coat protein, leader peptidase (Lep), and MalF (Fig. 2.4). M13 coat is made with a cleavable signal peptide, spans the membrane once, and inserts independently of the *sec* machinery (Wickner, 1988). Lep has no cleavable signal peptide and two transmembrane segments, and translocation of the C-terminal but not the N-terminal periplasmic domain is *sec* dependent (Wolfe et al., 1985; R. Dalbey, personal communication). MalF, finally, has eight transmembrane segments, and it is not clear which periplasmic parts, if any, are *sec* dependent (McGovern and Beckwith, 1991).

The importance of cytoplasmically located positively charged residues for proper insertion of the M13 coat protein was established quite early (Kuhn et al., 1986a,b), although it could not be determined whether the molecules that failed to insert in

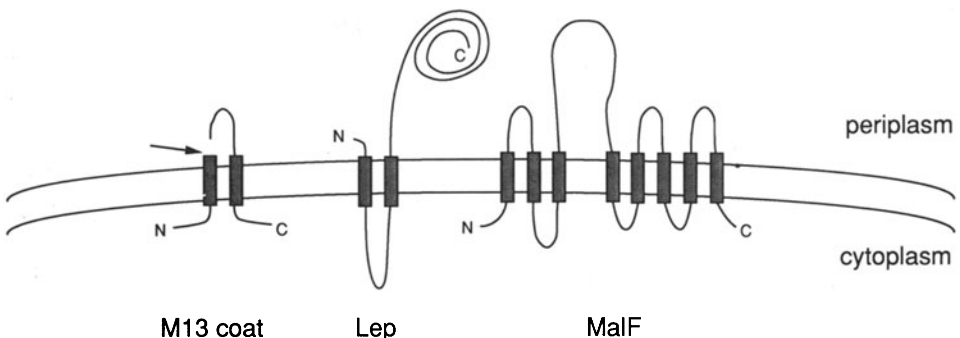

Fig. 2.4. Topology of the phage M13 coat, leader peptidase *(Lep)*, and the MalF proteins in the inner membrane of *E. coli*. The *arrow* marks the cleavage-site in the M13 coat protein signal peptide.

their normal orientation simply did not insert at all or inserted in some other orientation.

Studies of various MalF–PhoA fusion proteins have also shown that positively charged residues in cytoplasmic loops are important for preserving the transmembrane orientation (Boyd and Beckwith, 1989; Ehrmann and Beckwith, 1991; McGovern et al., 1991), but the most direct evidence for the role of positively charged residues as topological determinants has been obtained from studies of Lep-derived constructs in which it has been possible to "invert" the orientation of the molecule by redistribution of lysines and arginines. Both single-spanning and double-spanning Lep derivatives can be reoriented by the simultaneous removal of positively charged residues from their cytoplasmic regions and addition of such residues to the periplasmic regions (von Heijne et al., 1988; von Heijne, 1989; Andersson et al., 1992). Negatively charged residues have little effect on the topology by themselves (Nilsson and von Heijne, 1990), but apparently can, when suitably positioned, "neutralize" the effect of a nearby positively charged residue (Andersson and von Heijne, 1993a). Furthermore, insertion of the "inverted" double-spanning Lep constructs is *sec*-independent, in contrast to insertion of the wild-type molecule (von Heijne, 1989).

It is thus clear that positively charged amino acids are the most important topological determinants in *E. coli* (Boyd and Beckwith, 1990; Dalbey, 1990; von Heijne and Manoil, 1990; Yamane et al., 1990). However, one could still argue that they would act only to determine the orientation of the first one or two transmembrane segments in a multispanning protein such as MalF and that the more downstream segments would simply follow the dictates of their upstream neighbors, i.e., that the insertion process would progress sequentially from the N terminus toward the C terminus. The other extreme would be a locally determined insertion process, where individual "helical hairpins" (DiRienzo et al., 1978; von Heijne and Blomberg, 1979; Engelman and Steitz, 1981) composed of two neighboring transmembrane segments and the periplasmic loop in between insert independently of the rest of the chain. The statistical studies discussed above would favor the latter model, since the charge bias between periplasmic and cytoplasmic loops is equally strong in the early and late parts of multispanning proteins (von Heijne, 1992). Indeed, it has recently been reported both for MalF and LacY (which has 12 transmembrane segments) that deletions of single transmembrane segments at different positions often have no effects on the orientation of the downstream parts of the molecule (Bibi et al., 1991; Ehrmann and Beckwith, 1991; McGovern et al., 1991). This result cannot be explained on the basis of a sequential model (according to which the orientation of the downstream parts would be inverted) and strongly suggests that the topology is a local property of each helical hairpin.

Finally, why do some proteins apparently insert independently of the *sec* machinery, whereas others absolutely depend on it? First, the low frequency of positively charged residues in the periplasmic parts is only observed for relatively short loops, whereas loops longer than 60–80 residues have frequencies similar to what is found in soluble periplasmic proteins (von Heijne and Gavel, 1988). This might suggest that short loops for some reason cannot engage the *sec* machinery and hence have greater restrictions on the kinds of sequences that can be efficiently translocated. In agreement with this idea, it has been found that the insertion of the normally *sec*-independent M13 coat protein is *sec*-dependent in a construct in which the periplas-

mic domain has been increased in length from some 20 to about 100 residues (Kuhn, 1988). Also, translocation of the outer membrane protein OmpA through the inner membrane (a *sec*-dependent process) requires a chain length of at least 75 residues (Freudl et al., 1989); shorter chains are not translocated at all. We have successively lengthened the periplasmic loop in the"inverted" Lep construct; so far, we have found that a 50 residue long loop is still translocated independently of the *sec* machinery (Andersson and von Heijne, 1993b).

These observations all seem to support the idea that the length of the translocated domain is an important factor that would distinguish *sec*-dependent and *sec*-independent translocation; yet recent results from an analysis of MalF–PhoA fusion proteins seem to suggest that length cannot be the only factor (McGovern and Beckwith, 1991), since translocation of the large PhoA moiety in these fusions apparently does not require the *sec* machinery (PhoA is *sec*-dependent when targeted for secretion by its own signal peptide).

Membrane Protein Topogenesis in the ER

von Heijne and Gavel (1988) have shown that the positive inside rule holds also for membrane proteins that inset into the ER membrane and traverse the secretory pathway, although the bias in the distribution of Arg and Lys residues is less extreme than in bacterial proteins. A more detailed statistical study has suggested that the charge bias across the first transmembrane segment is an even better predictor of the topology than the total charge bias over the entire molecule (Hartmann et al., 1989). This would be consistent with the fact that protein translocation across the ER membrane is obligatory cotranslational (as opposed to the situation in *E. coli*), which makes it more likely that insertion proceeds according to the sequential model discussed above, with the topology being determined by the first transmembrane segment(s).

There is as yet no good experimental data on this point, since the topological role of charged residues has only been tested on single-spanning proteins so far. For these proteins, however, it has been shown that positively charged residues play a similar role as in *E. coli* (von Heijne and Manoil, 1990). The role of negatively charged amino acids has not been systematically studied, but they do seem to have some effects, at least in particular contexts (Haeuptle et al., 1989; Parks and Lamb, 1991).

Membrane Protein Topogenesis in Chloroplasts and Mitochondria

Thylakoid membrane proteins follow the positive inside rule (Gavel et al., 1991), with the more highly charged loops facing the stromal compartment (i.e., not being translocated). Likewise, mitochondrial inner membrane proteins have a strong charge bias, with the matrix-facing parts having a high content of positively charged residues (Gavel and von Heijne, 1992). Considering the evolutionary relationships between chloroplasts, mitochondria, and bacterial cells, it thus seems highly likely that the mechanism of organellar membrane protein insertion is similar to that in *E. coli*. As far as insertion into the outer membranes of mitochondria and chloroplasts and into the inner envelope chloroplast membrane are concerned, there are very few data available, either from known sequences or from biochemical studies.

Conclusion: The Structure Prediction Problem

One important goal of research on membrane protein biogenesis is to improve our abilities to predict three-dimensional structures directly from sequence. Today, one can often make a good guess as to where the transmembrane segments are based on hydrophobicity analysis, and once these segments have been identified one can predict the orientation in the membrane with high confidence from the positive inside rule. A major difficulty, however, is that one often finds peaks on a hydrophobicity plot that cannot be unambiguously assigned as transmembrane (see Chapter 4); they fall in the "twilight zone" (von Heijne, 1986c). In such cases, it might be possible to proceed by constructing all possible models that either include or exclude each one of the "twilight" peaks and rank these according to how well they conform to the positive inside rule. From preliminary studies, it seems that such a strategy works reasonably well and can be used to improve the raw results from the hydrophobicity analysis. Another means to the same end would be to study experimentally the putative transmembrane segments that fall in the twilight zone to gain a better understanding of how stop-transfer function relates to amino acid sequence characteristics (Davis et al., 1985; Davis and Model, 1985; Doyle et al., 1986; Kuroiwa et al., 1991).

As discussed elsewhere in this volume (Chapters 1 and 3), the final step in membrane protein structure prediction would be to develop algorithms that can pack transmembrane α-helices together in some optimal way. This may well be within reach of our present-day modeling capabilities, provided we had a sufficiently large database of known structures to calibrate against and reasonably simple experimental methods for testing our predictions. This is clearly not the case, and few remedies are seen on the horizon.

I would like to end on this note, with a plea for efforts to develop new techniques, e.g., for mapping the relative orientations of transmembrane helices or for determining distances between selected residues in different parts of a multispanning protein. These new techniques need not be high resolution: Since we know that we are dealing with α-helices that are oriented within some rather small angle of the membrane normal, already a few rough distances between helices or measures of relative helix orientations should suffice to give us the constraints we need for successful modeling. Thus we might be able to further our understanding of how transmembrane helices pack together to the point where we can finally solve the membrane protein folding problem: to go from sequence to structure at the touch of a button.

Acknowledgments

This work was supported by grants from the Swedish Natural Sciences Research Council and the Swedish Board for Technical Development.

References

Andersson, H., Bakker, E. and von Heijne, G. (1992) Different positively charged amino acids have similar effects on the topology of a polytopic transmembrane protein in *Escherichia coli*. *J. Biol. Chem.* 267: 1491–1495.

Andersson, H., and von Heijne, G. (1993a). Position-specific Asp-Lys pairing can affect signal sequence function and membrane protein topology. *J. Biol. Chem.* 268: 21398–21393.

Andersson, H., and von Heijne, G. (1993b). *Sec*-Dependent and *sec*-Independent Assembly of *E. coli* Inner Membrane Proteins—The Topological Rules Depend on Chain Length. *EMBO J.* 12: 683–691.

Bansal, A., and Gierasch, L. M. (1991) The NPXY internalization signal of the LDL receptor adopts a reverse-turn conformation. *Cell* 67: 1195–1201.

Baranski, T. J., Faust, P. L., and Kornfeld, S. (1990) Generation of a lysosomal enzyme targeting signal in the secretory protein pepsinogen. *Cell* 63: 281–291.

Bibi, E., Verner, G., Chang, C. Y., and Kaback, H. R. (1991) Organization and stability of a polytopic membrane protein—deletion analysis of the lactose permease of *Escherichia coli. Proc. Natl. Acad. Sci. USA* 88: 7271–7275.

Bonifacino, J. S., Cosson, P., and Klausner, R. D. (1990) Colocalized transmembrane determinants for ER degradation and subunit assembly explain the intracellular fate of TCR chains. *Cell* 63: 503–513.

Bosch, D., Scholten, M., Verhagen, C., and Tommassen, J. (1989) The role of the carboxy-terminal membrane-spanning fragment in the biogenesis of *Escherichia coli* K12 outer membrane protein PhoE. *Mol. Gen. Genet.* 216: 144–148.

Bosch, D., Voorhout, W., and Tommassen, J. (1988) Export and localization of N-terminally truncated derivatives of *Escherichia coli* K-12 outer membrane protein PhoE. *J. Biol. Chem.* 263: 9952–9957.

Boyd, D., and Beckwith, J. (1989) Positively charged amino acid residues can act as topogenic determinants in membrane proteins. *Proc. Natl. Acad. Sci. USA* 86: 9446–9450.

Boyd, D., and Beckwith, J. (1990) The role of charged amino acids in the localization of secreted and membrane proteins. *Cell* 62: 1031–1033.

Brändén, C., and Tooze, J. (1991) Introduction to Protein Structure. New York: Garland Publishing.

Breitfeld, P. P., Casanova, J. E., Simister, N. E., Ross, S. A., McKinnon, W. C., and Mostov, K. E. (1989) Sorting signals, *Curr. Opin. Cell Biol.* 1: 617–623.

Colley, K. J., Lee, E. U., Adler, B., Browne, J. K., and Paulson, J. C. (1989) Conversion of a Golgi apparatus sialyltransferase to a secretory protein by replacement of the NH2-terminal signal anchor with a signal peptide. *J. Biol. Chem.* 264: 17619–17622.

Dalbey, R. E. (1990) Positively charged residues are important determinants of membrane protein topology. *Trends Biochem. Sci.* 15: 253–257.

Davis, N. G., Boeke, J. D., and Model, P. (1985) Fine structure of a membrane anchor domain. *J. Mol. Biol.* 181: 111–121.

Davis, N. G., and Model, P. (1985) An artificial anchor domain: hydrophobicity suffices to stop transfer. *Cell* 41: 607–614.

Deisenhofer, J., and Michel, H. (1991) High-resolution structures of photosynthetic reaction centers. *Annu. Rev. Biophys. Biophys. Chem.* 20: 247–266.

Dill, K. A. (1990) Dominant forces in protein folding. *Biochemistry* 29: 7133–7155.

DiRienzo, J. M., Nakamura, K., and Inouye, M. (1978) The outer membrane proteins of Gram-negative bacteria: biosynthesis, assembly, and function. *Annu. Rev. Biochem.* 47: 481–532.

Doyle, C., Sambrook, J., and Gething, M. J. (1986) Analysis of progressive deletions of the transmembrane and cytoplasmic domains of influenza hemagglutinin. *J. Cell. Biol.* 103: 1193–204.

Eberle, W., Sander, C., Klaus, W., Schmidt, B., von Figura, K., and Peters, C. (1991) The essential tyrosine of the internalization signal in lysosomal acid phosphatase is part of a β-turn. *Cell* 67: 1203–1209.

Ehrmann, M., and Beckwith, J. (1991) Proper insertion of a complex membrane protein in the absence of its amino-terminal export signal. *J. Biol. Chem.* 266: 16530–16533.

Engelman, D. M., and Steitz, T. A. (1981) The spontaneous insertion of proteins into and across membranes: the helical hairpin hypothesis. *Cell* 23: 411–422.

Fourel, D., Mizushima, S., and Pages, J. M. (1992) Dynamics of the exposure of epitopes on OmpF, an outer membrane protein of *Escherichia coli. Eur. J. Biochem.* 206: 109–114.

Franzén, L. G., Rochaix, J. D., and von Heijne, G. (1990) Chloroplast transit peptides from the green alga *Chlamydomonas reinhardtii* share features with both mitochondrial and higher plant chloroplast presequences. *FEBS Lett.* 260: 165–168.

Freudl, R., Schwarz, H., Degen, M., and Henning, U. (1989) A lower size limit exists for export of fragments of an outer membrane protein (OpmA) of *Escherichia coli* K-12. *J. Mol. Biol.* 205: 771–775.

Gabathuler, R., and Kvist, S. (1990) The endoplasmic reticulum retention signal of the E3/19K protein of adenovirus type-2 consists of 3 separate amino acid segments at the carboxy terminus. *J. Cell. Biol.* 111: 1803–1810.

Gavel, Y., Steppuhn, J., Herrmann, R., and von Heijne, G. (1991) The positive-inside rule applies to thylakoid membrane proteins. *FEBS Lett.* 282: 41–46.

Gavel, Y., and von Heijne, G. (1990a) Cleavage-site motifs in mitochondrial targeting peptides. *Prot. Eng* 4: 33–37.

Gavel, Y., and von Heijne, G. (1990b) A conserved cleavage-site motif in chloroplast transit peptides. *FEBS Lett.* 261: 455–458.

Gavel, Y., and von Heijne, G. (1992) The distribution of charged amino acids in mitochondrial inner membrane proteins suggests different modes of membrane integration for nuclearly and mitochondrially encoded proteins. *Eur. J. Biochem.* 205: 1207–1215.

Gierasch, L. M. (1989) Signal sequences. *Biochemistry* 28: 923–930.

Gould, S. J., Keller, G.-A., Hosken, N., Wilkinson, J., and Subramani, S. (1989) A conserved tripeptide sorts proteins to peroxisomes. *J. Cell. Biol.* 108: 1657–1664.

Gould, S. J., Krisans, S., Keller, G. A., and Subramani, S. (1990) Antibodies directed against the peroxisomal targeting signal of firefly luciferase recognize multiple mammalian peroxisomal proteins. *J. Cell. Biol.* 110: 27–34.

Haeuptle, M. T., Flint, N., Gough, N. M., and Dobberstein, B. (1989) A tripartite structure of the signals that determine protein insertion into the endoplasmic reticulum membrane. *J. Cell. Biol.* 108: 1227–1236.

Halpin, C., Elderfield, P. D., James, H. E., Zimmermann, R., Dunbar, B., and Robinson, C. (1989) The reaction specificities of the thylakoidal processing peptidase and *Escherichia coli* leader peptidase are identical. *EMBO J.* 8: 3917–3921.

Hartl, F. U., Pfanner, N., Nicholson, D. W., and Neupert, W. (1989) Mitochondrial protein import. *Biochim. Biophys. Acta* 988: 1–45.

Hartmann, E., Rapoport, T. A., and Lodish, H. F. (1989) Predicting the orientation of eukaryotic membrane proteins. *Proc. Natl. Acad. Sci. USA* 86: 5786–5790.

Henderson, R., Baldwin, J. M., Ceska, T. A., Zemlin, F., Beckmann, E., and Downing, K. H. (1990) A model for the structure of bacteriorhodopsin based on high resolution electron cryo-microscopy. *J. Mol. Biol.* 213: 899–929.

Jackson, M. R., Nilsson, T., and Peterson, P. A. (1990) Identification of a consensus motif for retention of transmembrane proteins in the endoplasmic reticulum. *EMBO J.* 9: 3153–3162.

Jing, S. Q., Spencer, T., Miller, K., Hopkins, C., and Trowbridge, I. S. (1990) Role of the human transferrin receptor cytoplasmic domain in endocytosis—localization of a specific signal sequence for internalization. *J. Cell. Biol.* 110: 283–294.

Klausner, R. D., and Sitia, R. (1990) Protein degradation in the endoplasmic reticulum. *Cell* 62: 611–614.

Kornfeld, S., and Mellman, I. (1989) The biogenesis of lysosomes. *Annu. Rev. Cell. Biol.* 5: 483–525.

Ktistakis, N. T., Thomas, D., and Roth, M. G. (1990) Characteristics of the tyrosine recognition signal for internalization of transmembrane surface glycoproteins. *J. Cell. Biol.* 111: 1393–1407.

Kuhn, A. (1988) Alterations in the extracellular domain of M13 procoat protein make its membrane insertion dependent on *secA* and *secY*. *Eur. J. Biochem.* 177: 267–271.

Kuhn, A., Kreil, G., and Wickner, W. (1986a) Both hydrophobic domains of M13 procoat are required to initiate membrane insertion. *EMBO J.* 5: 3681–3685.

Kuhn, A., Wickner, W., and Kreil, G. (1986b) The cytoplasmic carboxy terminus of M13 procoat is required for the membrane insertion of its central domain. *Nature* 322: 335–339.

Kuroiwa, T., Sakaguchi, M., Mihara, K., and Omura, T. (1991) Systematic analysis of stop-transfer sequence for microsomal membrane. *J. Biol. Chem.* 266: 9251–9255.

Machamer, C. E., and Rose, J. K. (1987) A specific transmembrane domain of a coronavirus E1 glycoprotein is required for its retention in the Golgi region. *J. Cell. Biol.* 105: 1205–1214.

McGovern, K., and Beckwith, J. (1991) Membrane insertion of the *Escherichia coli* MalF protein in cells with impaired secretion machinery. *J. Biol. Chem.* 266: 20870–20876.

McGovern, K., Ehrmann, M., and Beckwith, J. (1991) Decoding signals for membrane protein assembly using alkaline phosphatase fusions. *EMBO J.* 10: 2773–2782.

Miyazawa, S., Osumi, T., Hashimoto, T., Ohno, K., Miura, S., and Fujiki, Y. (1989) Peroxisome targeting signal of rat liver acyl-coenzyme A oxidase resides at the carboxy terminus. *Mol. Cell. Biol.* 9: 83–91.

Nilsson, I. M., and von Heijne, G. (1990) Fine-tuning the topology of a polytopic membrane protein. Role of positively and negatively charged residues. *Cell* 62: 1135–1141.

Parks, G. D., and Lamb, R. A. (1991) Topology of eukaryotic type-II membrane proteins—importance of N-terminal positively charged residues flanking the hydrophobic domain. *Cell* 64: 777–787.

Pelham, H.R.B. (1990) The retention signal for soluble proteins of the endoplasmic reticulum. *Trends Biochem. Sci.* 15: 483–486.

Peters, C., Braun, M., Weber, B., Wendland, M., Schmidt, B., Pohlmann, R., Waheed, A., and von Figura, K. (1990) Targeting of a lysosomal membrane protein—a tyrosine-containing endocytosis signal in the cytoplasmic tail of lysosomal acid phosphatase is necessary and sufficient for targeting to lysosomes. *EMBO J.* 9: 3497–3506.

Pugsley, A. P. (1989) Protein Targeting. San Diego: Academic Press.

Schatz, P. J., and Beckwith, J. (1990) Genetic analysis of protein export in *Escherichia coli. Annu. Rev. Genet.* 24: 215–248.

Shackleton, J. B., and Robinson, C. (1991) Transport of proteins into chloroplasts—the thylakoidal processing peptidase is a signal-type peptidase with stringent substrate requirements at the -3-position and -1-position. *J. Biol. Chem.* 266: 12152–12156.

Silver, P. A. (1991) How proteins enter the nucleus. *Cell* 64: 489–497.

Struyve, M., Moons, M., and Tommassen, J. (1991) Carboxy-terminal phenylalanine is essential for the correct assembly of a bacterial outer membrane protein. *J. Mol. Biol.* 218: 141–148.

von Heijne, G. (1985) Signal sequences. The limits of variation. *J. Mol. Biol.* 184: 99–105.

von Heijne, G. (1986a) The distribution of positively charged residues in bacterial inner membrane proteins correlates with the transmembrane topology. *EMBO J.* 5: 3021–3027.

von Heijne, G. (1986b) Mitochondrial targeting sequences may form amphiphilic helices. *EMBO J.* 5: 1335–1342.

von Heijne, G. (1986c) Net N-C charge imbalance may be important for signal sequence function in bacteria. *J. Mol. Biol.* 192: 287–290.

von Heijne, G. (1986d) Towards a comparative anatomy of N-terminal topogenic protein sequences. *J. Mol. Biol.* 189: 239–242.

von Heijne, G. (1988) Transcending the impenetrable: how proteins come to terms with membranes. *Biochim. Biophys. Acta* 947: 307–333.

von Heijne, G. (1989) Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. *Nature* 341: 456–458.

von Heijne, G. (1990a) Protein targeting signals. *Curr. Opin. Cell. Biol.* 2: 604–608.

von Heijne, G. (1990b) The signal peptide. *J. Membrane Biol.* 115: 195–201.

von Heijne, G. (1992) Membrane protein structure prediction—hydrophobicity analysis and the positive-inside rule. *J. Mol. Biol.* 225: 487–494.

von Heijne, G., and Blomberg, C. (1979) Trans-membrane translocation of proteins: the direct transfer model. *Eur. J. Biochem.* 97: 175–181.

von Heijne, G., and Gavel, Y. (1988) Topogenic signals in integral membrane proteins. *Eur. J. Biochem.* 174: 671–678.

von Heijne, G., and Manoil, C. (1990) Membrane proteins—from sequence to structure. *Prot. Eng.* 4: 109–112.

von Heijne, G., and Nishikawa, K. (1991) Chloroplast transit peptides—the perfect random coil? *FEBS Lett.* 278: 1–3.

von Heijne, G., Steppuhn, J., and Herrmann, R. G. (1989) Domain structure of mitochondrial and chloroplast targeting peptides. *Eur. J. Biochem.* 180: 535–545.

von Heijne, G., Wickner, W., and Dalbey, R. E. (1988) The cytoplasmic domain of *Escherichia coli* leader peptidase is a "translocation poison: sequence. *Proc. Natl. Acad. Sci. USA* 85: 3363–3366.

Weiss, M. S., Kreusch, A., Schiltz, E., Nestel, U., Welte, W., Weckesser, J., and Schulz, G. E. (1991) The structure of porin from *Rhodobacter capsulata* at 1.8 Å resolution. *FEBS Lett.* 280: 379–382.

Wickner, W. (1988) Mechanisms of membrane assembly: general lessons from the study of M13 coat protein and *Escherichia coli* leader peptidase. *Biochemistry* 27: 1081–1086.

Wickner, W., Driessen, A.J.M., and Hartl, F. U. (1991) The enzymology of protein translocation across the *Escherichia coli* plasma membrane. *Annu. Rev. Biochem.* 60: 101–124.

Wolfe, P. B., Rice, M., and Wickner, W. (1985) Effects of two *sec* genes on protein assembly into the plasma membrane of *Escherichia coli. J. Biol. Chem.* 260: 1836–1841.

Yamane, K., Akiyama, Y., Ito, K., and Mizushima, S. (1990) A positively charged region is a determinant of the orientation of cytoplasmic membrane proteins in *Escherichia coli. J. Biol. Chem.* 265: 21166–21171.