Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 1960–1964, April 1987
Cell Biology

On the transfer of integral proteins into membranes

(signal sequences/translocation/translocator proteins)

S. J. SINGER, PAMELA A. MAHER, AND MICHAEL P. YAFFE

Department of Biology, University of California at San Diego, La Jolla, CA 92093

Contributed by S. J. Singer, December 8, 1986

ABSTRACT **We have earlier proposed a molecular mechanism for the translocation of hydrophilic proteins across membranes that accounts for the experimental facts and meets the restrictions that we stipulate for such a mechanism. In particular, the restrictions are that translocation occurs by successive segments of the polypeptide chain and that the ionic groups of the polypeptide remain in contact with water throughout the translocation process. The evidence indicates that the transfer of integral proteins into membranes very likely uses the same molecular machinery as does the translocation of hydrophilic proteins across membranes. Here we show how the mechanism we have proposed for translocation can also be utilized in the intercalation of known types of integral proteins, accounting for their specific topologies in the membrane.**

In a recent paper (1), we dealt with the mechanism of the translocation of water-soluble proteins across membranes. Such translocations occur during the import of many polypeptides into one of several intracellular organelles; these polypeptides are synthesized in the cytoplasm and must traverse one, two, or sometimes three membranes to arrive within their specific organellar compartments. After a brief discussion of the relevant experimental facts, and an analysis of the restrictions to be met, we proposed a molecular mechanism for translocation which, with appropriate variations, could be applied universally. This mechanism stipulates that each translocation event is mediated by one of a set of special integral proteins in the membrane, termed translocator proteins (TPs). It is proposed that each TP is an aggregate consisting of n homologous but not identical transmembrane subunits that altogether form a membrane-spanning water-filled channel down the central axis of the aggregate. The special feature ascribed to all TP subunit aggregates is that at one of the n nonidentical interfaces between neighboring subunits within the membrane the two subunits are only weakly bound to one another. This interface provides the passageway for the translocation of the hydrophilic polypeptide. Intercalation of the signal sequence near the NH_2 terminus of the polypeptide into the TP interface on the cis side of the membrane initiates a process whereby successive segments of the polypeptide chain form "subdomains" of secondary structure in the interface; each subdomain is then displaced to the trans side of the membrane in an energy-dependent process. During chain intercalation, subdomain formation, and displacement of each successive segment, all of the ionic residues of the segment remain in contact with the water within the aqueous channel of the TP, while the hydrophobic residues are shielded within the hydrophobic portions of the interface.

Closely related to the problem of the complete translocation of a hydrophilic polypeptide across a membrane is the problem of the intercalation of integral proteins into membranes. Almost all known integral proteins span the membrane, with hydrophilic domains exposed to the aqueous phase on both the trans and cis faces of the membrane (see below). The intercalation of an integral protein into a membrane can be viewed as one of partial translocation of the polypeptide chain across the membrane (2): the hydrophilic domains that become exposed at the trans face are translocated, but the hydrophilic domains on the cis side are not translocated; in the process, hydrophobic domains of the polypeptide are embedded in the membrane bilayer. In this paper we show that the unitary mechanism we propose for the complete translocation of hydrophilic proteins across membranes can also account for the intercalation of the different kinds of integral protein structures known to exist in membranes.

Integral Proteins of Membranes

For our present purposes, there are four main classes of equilibrium configurations of transmembrane integral proteins that have so far been recognized (Fig. 1). In type I, which constitutes the large majority of all integral proteins known, the protein is anchored in the membrane bilayer by a single stretch of about 20 nonionic and predominantly hydrophobic amino acids, almost certainly in an α-helical configuration (3). This transmembrane helix in type I proteins is oriented in the membrane with its NH_2 terminus towards the trans face of the membrane and its COOH terminus towards the cis face. In this orientation the transmembrane stretch is called a *stop-transfer sequence* (2, 4). The hydrophilic domain protruding from the trans face is often very large, constituting the bulk of the polypeptide chain, and contains the NH_2 terminus of the chain. The hydrophilic domain protruding from the cis face is often small, and it contains the COOH terminus of the chain. The G protein of vesicular stomatitis virus is an example of type I proteins (5). Type II proteins are also anchored in the membrane by a single stretch of predominantly hydrophobic amino acids, but in this case the hydrophobic stretch is usually near the NH_2 terminus of the polypeptide chain and is oriented with its NH_2 terminus towards the cis face. The bulk of the type II protein is generally contained within the hydrophilic domain situated on the trans side of the membrane, which includes the COOH terminus of the chain. Examples of type II integral proteins are the transferrin receptor (6) and the neuraminidase of influenza virus (7).

Proteins of type III have multiple hydrophobic stretches of the polypeptide chain embedded in the membrane. For our present purposes, we do not require that separate consideration be given to type III polypeptides with different numbers of transmembrane helices or different dispositions of their NH_2 and COOH termini. Examples of type III integral proteins are bacterial rhodopsin (8) and the L and M chains of the photosynthetic reaction center (9). Proteins of type IV are not usually considered as a separate class, but they have certain structural features that are unique (3, 10). They are

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: TP, translocator protein.

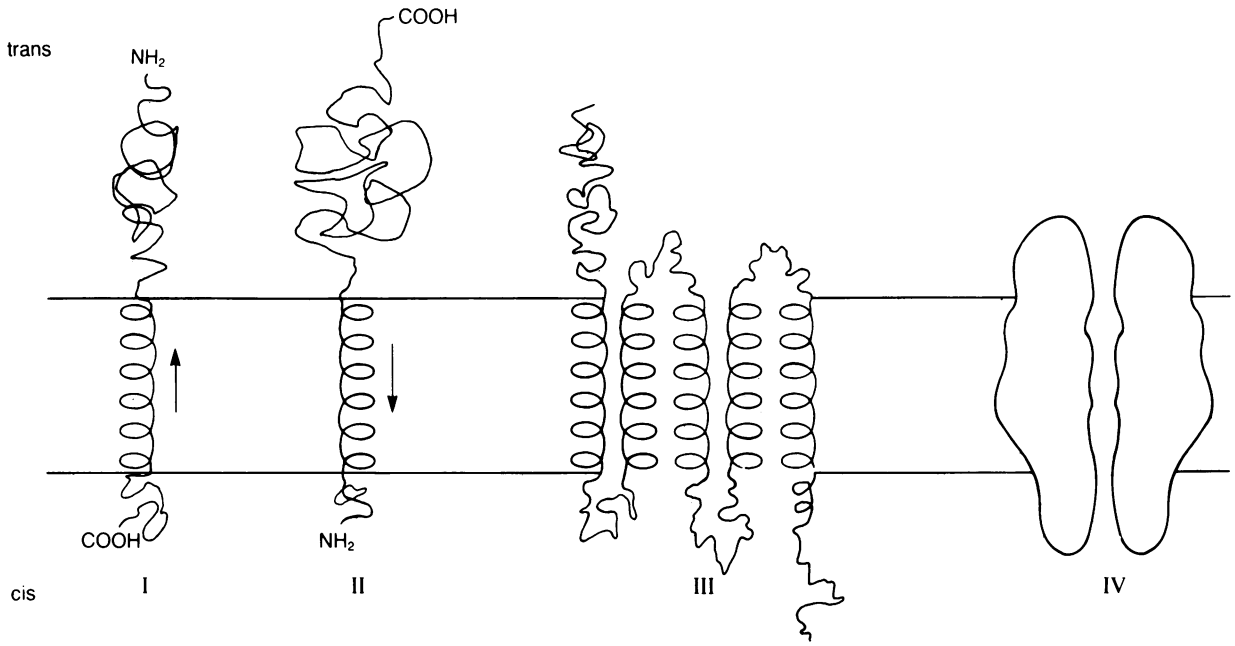

FIG. 1. Four types of known structures of transmembrane integral proteins in membranes. Type I and type II proteins have single membrane-spanning hydrophobic helices, but their chain orientations are opposite to one another as shown. Type III proteins have multiple membrane-spanning hydrophobic helices, there being several different ways that their NH_2 and COOH termini could be situated with respect to the cis and trans faces of the membrane. Type IV proteins are channel-forming subunit aggregates, with n identical or homologous subunits in the aggregate (here depicted with $n = 2$). A water-filled channel runs down the n-fold symmetry or pseudosymmetry axis. The conformations of the individual subunits are not shown in detail, although they may often contain multihelical transmembrane domains. Ionic residues of these domains would line the aqueous channel (3).

aggregates of n transmembrane subunits that are identical or closely homologous to one another, with a transmembrane water-filled channel running down the n-fold symmetry or pseudosymmetry axis of the aggregate. The acetylcholine receptor (for review, see ref. 11) and the gap junction connexon (12) are examples of type IV proteins. The individual subunits of such aggregates may generally have multiple helical stretches of predominantly hydrophobic residues embedded in the membrane, as with type III proteins. The aqueous channel that runs through the type IV aggregate, however, is the distinctive feature of this class of integral proteins. Our proposal, as indicated above, is that the TPs are type IV proteins (1).

While much less information is currently available about the process of intercalation of these four types of integral proteins into membranes than for the translocation of hydrophilic proteins across membranes, there is enough evidence from several systems (cf. refs. 2, 7, and 13) to indicate that initiation of the intercalation of an integral protein into a membrane involves a signal sequence, usually near the NH_2 terminus of the polypeptide chain, that is in all respects closely similar to the signal sequences of completely translocated polypeptides. Furthermore, the same protease that cleaves the signal peptide from translocated proteins also operates on the signal peptide of some integral membrane proteins (14); and for integral proteins inserted into the endoplasmic reticulum the same signal recognition particle is required as for translocated polypeptides (15–17). Finally, a similar energy dependence for intercalated and translocated proteins has been demonstrated (18).

In vitro experiments on the transfer of integral proteins into endoplasmic reticulum membranes have indicated that with some proteins, such as the G protein of vesicular stomatitis virus (2) and the erythrocyte anion transport protein (19), intercalation is obligatorily cotranslational, whereas with others, such as the human glucose transporter (15), intercalation can occur posttranslationally. We suggest that this difference, however, as in the case of the translocation of hydrophilic polypeptides (20), may be more apparent than real. Those integral proteins whose intercalation appears to be obligatorily cotranslational may contain disulfide bridges within their trans-side hydrophilic domains whose formation after translation *in vitro* may prevent the unfolding of the polypeptide chain that is required for posttranslational intercalation into the membrane (refs. 1 and 20 and see below).

These considerations suggest that a single mechanism should be able to account for the characteristic membrane intercalation of the four types of integral proteins described; furthermore, it should be the same one that applies to the complete translocation of hydrophilic polypeptides across membranes.

Membrane Intercalation of Integral Proteins

We begin where our previous paper left off (1). We invoke, without reiterating them here, the details of the TP-mediated scheme for the translocation of hydrophilic polypeptides embodied in the text and figures 1 and 2 of that paper.

Consider integral proteins of type I (Fig. 1). Such a polypeptide chain, in addition to its NH_2-terminal signal sequence, contains a single internal hydrophobic sequence of about 20 residues. Between the signal and hydrophobic sequences, and following the hydrophobic sequence, there are stretches of hydrophilic sequences of variable lengths. Intercalation is initiated, as previously described for entirely hydrophilic polypeptides (1), by the insertion of the signal sequence into the β–γ interface of the TP, which necessitates the intercalation of the following first hydrophilic subdomain into the interface as well. The ionic residues of the signal sequence and of the first subdomain are situated in the β–γ interface so as to face into the aqueous channel running through the TP, while the hydrophobic residues are generally positioned within the interface away from the aqueous channel. This allows the ionic residues to remain in contact with water during translocation, while the hydrophobic residues are generally removed from contact with water,

conditions that help keep the free energy of activation for translocation at a minimum.

The insertion of the signal and first hydrophilic sequences opens up the β–γ interface for all succeeding events. Energy is then utilized to effect a "quaternary rearrangement" (3) of the TP such that the first subdomain is displaced from the β–γ interface to the trans side of the membrane. The signal sequence, whether proteolytically cleaved at this stage or not, remains in the interface. This displacement process causes the second hydrophilic sequence of the polypeptide to be "pulled" into the β–γ interface of the TP, where it forms the second subdomain. The second subdomain is then displaced to the trans side of the membrane in another energy-requiring step, and so on.

Up to this point, the process for the type I protein is the same as that described for an entirely hydrophilic polypeptide (1). However, suppose that the *i*th hydrophilic sequence to be translocated, just ahead of the internal hydrophobic sequence, forms its subdomain secondary structure without including the hydrophobic sequence. The translocation of the *i*th subdomain would then "pull" the immediately following hydrophobic sequence into the β–γ interface. The important point here is that the hydrophobic sequence, unlike the usual hydrophilic sequences, would have no ionic residues to contribute to the TP aqueous channel. Without another hydrophilic chain sequence simultaneously present within the β–γ interface, it might therefore be energetically more favorable if this hydrophobic sequence slipped laterally out of the interface (in the direction away from the aqueous channel) to form a transmembrane helix within the lipid interior of the bilayer. The β–γ interface of the TP would then close up, particularly if the NH_2-terminal signal sequence had by then been cleaved off and ejected from the interface. In the absence of a following signal sequence in the polypeptide chain (see below for type III proteins) the β–γ interface would remain closed, and further translocation of the polypeptide chain would be arrested. At this stage, the entire type I molecule would be free in the bilayer. The internal hydrophobic sequence would thus have functioned as a stop-transfer sequence (2, 4). All of the polypeptide (minus the cleaved signal peptide) to the NH_2-terminal side of the internal hydrophobic sequence would be located on the trans side of the membrane, while all of the polypeptide to its COOH-terminal side would be on the cis side, and the stop-transfer sequence would be anchored in the bilayer with its NH_2 terminus facing the trans side. This description corresponds to the structure of type I integral proteins in membranes.

In order to account for type II integral proteins (Fig. 1), such polypeptides could have a sequence near the NH_2 terminus resembling the hydrophobic class of signal sequences (1) and no internal stop-transfer sequence; i.e., following the hydrophobic signal sequence, the remainder of the polypeptide chain is hydrophilic. We would propose that translocation is initiated by the intercalation of such a signal sequence into the β–γ interface of the TP as above, but that the signal sequence is not cleaved. However, the absence of cleavage does not affect the translocation of all of the remaining hydrophilic sequences of the polypeptide one subdomain at a time, as with entirely hydrophilic polypeptides (1). Throughout these successive subdomain translocations, the uncleaved signal sequence remains intercalated in the β–γ interface. The translocation of the last hydrophilic subdomain, however, leaves only the hydrophobic signal sequence intercalated in the β–γ interface. That sequence could then slip laterally out of the interface into the lipid bilayer, as in the type I case just described. The result of this process would therefore be a bilayer-intercalated protein (*i*) with its NH_2-terminal sequence on the cis side of the membrane; (*ii*) that was anchored in the lipid bilayer by a hydrophobic sequence whose NH_2 terminus was positioned towards the cis face; and (*iii*) with the entire chain to the carboxyl side of the anchor sequence located on the trans side. This topology corresponds to known type II integral proteins.

Type III integral proteins consist of polypeptide chains with multiple stretches of hydrophobic sequences, separated by hydrophilic sequences. The most NH_2-terminal hydrophobic sequence presumably serves as a signal sequence. The detailed topology of any one type III protein in the lipid bilayer would depend on a number of factors, including (*i*) whether the NH_2-terminal signal peptide was or was not cleaved; (*ii*) the numbers of internal hydrophobic sequences; and (*iii*) the lengths of the hydrophilic sequences between adjacent hydrophobic ones (21). To illustrate the operation of the unitary translocation mechanism for type III proteins, consider the L and M subunits of the photosynthetic reaction center complex of *Rhodopseudomonas viridis* (9), the only type III proteins whose complete structure and topology in the membrane are definitively established. The L and M subunits are homologous proteins, each having five hydrophobic transmembrane helices. There is no cleaved signal sequence. For the L subunit, which contains 270 residues, the topology of the chain is indicated in Fig. 2. The NH_2 terminus of the chain is located on the cis side of the membrane, the COOH terminus on the trans side. The five transmembrane helices are here labeled A through E, and the successive intervening sequences a/b, b/c, etc. The numbers of amino acid residues in each of these segments of the L polypeptide are indicated in the figure. The following successive stages in the intercalation of the chain into the membrane are postulated:

(*i*) Intercalation is initiated by the first hydrophobic sequence, A, functioning as part of an uncleaved signal sequence, as with type II proteins above. Sequence A intercalates at the β–γ interface of the TP, leaving the NH_2 terminus of the chain on the cis side. Intercalation of A pulls the hydrophilic a/b sequence into the interface. Its entire length (28 residues) is then translocated to the trans side as a single subdomain. The uncleaved signal sequence remains within the interface.

(*ii*) The translocation of a/b pulls the immediately following hydrophobic sequence, B, into the β–γ interface. As with type I proteins, B then acts as a stop-transfer sequence. Without displacing A from the interface* and in the absence of a hydrophilic sequence simultaneously present within the interface, B slips out laterally into the adjacent lipid bilayer, forming an α-helix. Although A is still bound within the TP, the sequence b/c is now physically separated from the TP and cannot be translocated.

(*iii*) It has previously been proposed (4) that a suitable hydrophobic sequence *following* a stop-transfer sequence can act as an *internal* signal sequence, and reinitiate translocation of a partially translocated polypeptide chain. Applying this proposal to the L chain, we suggest that the hydrophobic C sequence now acts as part of an internal signal sequence. It inserts into the β–γ interface of the same TP, displacing A into the lipid bilayer. (C has the same chain orientation within the interface as does A and may have a larger affinity than A for the interface.) The intercalation of C into the TP initiates the insertion and the subsequent translocation of the c/d sequence either as one large subdomain or as two successive smaller subdomains.

*It is possible that the hydrophobic sequences A, C, and E are discriminated from B and D in their binding to the β–γ interface because of their different helix orientations. The α-helix has a strong net dipole moment along the helix axis (22, 23), which might be involved in such discrimination.

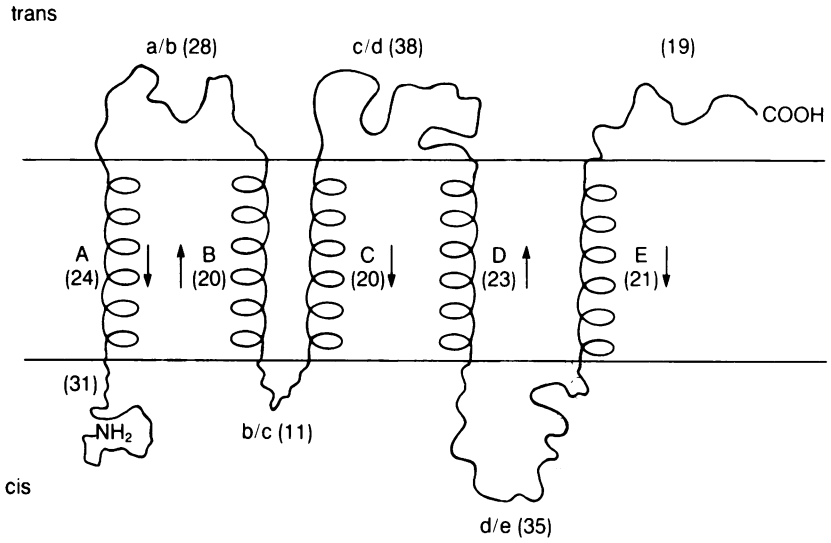

FIG. 2. A representation of the known membrane conformation of the L subunit of the photosynthetic reaction center complex of *R. viridis* (9), with its five transmembrane hydrophobic helices (A–E), and NH_2-terminal, COOH-terminal, and interhelical hydrophilic sequences (the last labeled a/b, b/c, etc.). The numbers of amino acid residues in each region of the sequence is given in the parentheses, but they are somewhat arbitrary because the precise locations of the polar headgroups of the lipids are not determined from the x-ray data. The orientation of each transmembrane helix with respect to the NH_2 terminus of the chain is shown by an arrow.

(*iv*) The translocation of the entire c/d sequence pulls the hydrophobic sequence D into the β–γ interface, where it has the properties (including chain orientation) to act as a stop-transfer sequence. That is, it slips laterally out of the interface into the adjacent lipid bilayer, forming an α-helix. This prevents the translocation of the d/e sequence.

(*v*) The following hydrophobic sequence, E, now acts as part of another internal signal sequence. It inserts into the β–γ interface of the same TP, displacing C into the lipid bilayer. This initiates the insertion and subsequent translocation of the COOH-terminal sequence as a single subdomain. After this occurs, in the absence of any hydrophilic sequence simultaneously intercalated in the β–γ interface, E slips laterally out of the interface into the lipid bilayer. After this final stage, the entire L chain is left intercalated in the bilayer independent of the TP.

This scenario would therefore result in the observed topology of the L chain in the bilayer, using the same TP and unitary translocation mechanism previously applied to entirely hydrophilic proteins (1) as well as to type I and type II integral proteins. The function of successive hydrophobic sequences as alternating stop-transfer and internal signal sequences (4) can in a similar way generate other complex topologies for different type III proteins.

The membrane insertion of type IV proteins (including TPs themselves, if our model is correct) is probably a complex process involving multiple stages of chain translocations and channel integrations (10). If the individual subunits each contains multiple transmembrane hydrophobic stretches, the monomeric subunits are probably each transferred into the membrane by a process resembling that just depicted for a type III protein. These monomers must then undergo aggregation in the bilayer, and probably simultaneously a concerted intercalation of their channel-forming domains to span the membrane (10). A more detailed model for this process does not seem to be warranted at this time.

DISCUSSION

In this paper, schemes for the intercalation of the several different known types of integral proteins into membranes have been put forward in considerable detail. The purpose is to show that the same unitary mechanism, mediated by the same specific TPs, can function in the membrane integration of these amphipathic polypeptides (whatever their topologies in the bilayer) as well as for the complete translocation of hydrophilic polypeptides across the bilayer (1). The only additional feature of the mechanism that is required with the integral proteins is the thermodynamic discrimination by the TP of the long (≈20 residues) hydrophobic sequences within these proteins from the usual hydrophilic sequences of completely translocated polypeptides; we show how these internal hydrophobic sequences then function as either stop-transfer or internal signal sequences. It should be emphasized that, with the proposed mechanisms of intercalation of integral proteins into a membrane, at no stage is any part of the sequence of the protein inserted directly into the lipid bilayer. All intercalation stages are via the β–γ interface of the TP. With the type III L protein (Fig. 2), for example, each of the five transmembrane hydrophobic helices is pictured as first entering the TP interface and only subsequently being laterally displaced into the lipid bilayer. Any mechanism that postulated the direct insertion of these helices into the bilayer would also require the direct transfer of hydrophilic sequences through the hydrophobic interior of the bilayer. For example, the direct insertion of the A and B helices as a helical hairpin (24) into the bilayer (Fig. 2) would also require the direct translocation of the a/b sequence, with its four ionic residues, through the bilayer. Our mechanism avoids this unlikely event, and keeps these and all other ionic residues in contact with water throughout the intercalation process.

An important consequence of the proposals discussed in this paper is the prediction that the final topology of an integral protein in the membrane is determined in large part by the mechanism of its intercalation into the membrane. In particular, which hydrophilic sequences of a transmembrane integral protein become located on the trans side of the membrane and which on the cis side is determined primarily by the properties of the TP and its interactions with successive sequences of the polypeptide chain of the integral protein. In this picture, the gross topology and orientation of the integral protein in the membrane has nothing directly to do, for example, with the asymmetrical distribution of the lipids and other proteins (except the TP) existing in that membrane. Once intercalated, the integral protein maintains its asymmetric orientation in the membrane, not because it is in equilibrium with the bilayer but because its rotation across the membrane is forbidden by the very large free energy of

activation required to translocate the hydrophilic domains through the lipid interior (3, 25).

A further corollary of our proposal is that only after the intercalation of an integral protein into the membrane had occurred, with its gross topology and orientation determined by the mechanism of its intercalation, would the final structure and function of the integral protein be attained. In the case of a type III protein, only after all of the hydrophilic trans-side subdomains had been successively translocated would they assemble and rearrange into a conformation corresponding to the lowest free energy state of the entire domain in the aqueous medium. This lowest free energy conformation would also be influenced by any additional interactions of the hydrophilic domain with other proteins and the polar head-groups of the membrane lipids. Similarly, the hydrophilic sequences of the protein that had been retained on the cis side of the membrane during intercalation would then assemble and adopt an equilibrium conformation for the entire domain in that aqueous compartment. In addition, the several hydrophobic sequences that had become embedded in the membrane would assemble into a collective equilibrium conformation within the hydrophobic interior of the bilayer.

The fluid mosaic model of membrane structure (3, 25, 26) predicted correctly on thermodynamic grounds that there exist proteins integral to the membrane that have an amphipathic structure, with hydrophilic domains exposed in the aqueous phases and a hydrophobic domain embedded in the bilayer. The model did not consider, however, the central problem of how such structures become intercalated into membranes. The mechanisms discussed in the present paper directly address this problem, and therefore they provide an essential complement to the fluid mosaic model.

1. Singer, S. J., Maher, P. A. & Yaffe, M. P. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 1015–1019.
2. Rothman, J. E. & Lodish, H. F. (1977) *Nature (London)* **269,** 775–780.
3. Singer, S. J. (1971) in *Structure and Function of Biological Membranes*, ed. Rothfield, L. I. (Academic, New York), pp. 145–222.
4. Blobel, G. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 1496–1500.
5. Rose, J. K. & Gallione, C. J. (1981) *J. Virol.* **39,** 519–528.
6. Schneider, C., Owen, M. J., Banville, D. & Williams, J. G. (1984) *Nature (London)* **311,** 675–678.
7. Blok, J., Air, G. M., Laver, W. G., Ward, C. W., Lilley, G. G., Woods, E. F., Roxburgh, C. M. & Inglis, A. S. (1982) *Virology* **119,** 109–121.
8. Khorana, H. G., Gerber, G. E., Herlihy, W. C., Gray, C. P., Anderreg, R. J., Nihei, K. & Biemann, K. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 5046–5050.
9. Deisenhofer, J., Epp, O., Miki, K., Huber, R. & Michel, H. (1985) *Nature (London)* **318,** 618–624.
10. Singer, S. J. (1982) in *Membranes and Genetic Disease*, eds. Sheppard, J. R., Anderson, V. E. & Eaton, J. W. (Liss, New York), pp. 3–24.
11. Anholt, R., Lindstrom, J. & Montal, M. (1984) in *Enzymes of Biological Membranes*, ed. Martinosi, A. (Plenum, New York), 2nd Ed., Vol. 3, pp. 335–401.
12. Unwin, P. N. T. & Ennis, P. D. (1983) *J. Cell Biol.* **97,** 1459–1466.
13. Hase, T., Müller, V. M., Riezman, H. & Schatz, G. (1984) *EMBO J.* **3,** 3157–3164.
14. Dalbey, R. E. & Wickner, W. (1985) *J. Biol. Chem.* **260,** 15925–15931.
15. Mueckler, M. & Lodish, H. F. (1986) *Cell* **44,** 629–637.
16. Yost, C. S., Hedgpeth, J. & Lingappa, V. R. (1983) *Cell* **34,** 759–766.
17. Spiess, M. & Lodish, H. F. (1986) *Cell* **44,** 177–185.
18. Mueckler, M. & Lodish, H. F. (1986) *Nature (London)* **322,** 549–552.
19. Braell, W. A. & Lodish, H. F. (1982) *Cell* **28,** 23–31.
20. Maher, P. A. & Singer, S. J. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 9001–9005.
21. Coleman, J., Inukai, M. & Inouye, M. (1985) *Cell* **43,** 351–360.
22. Wada, A. (1976) *Adv. Biophys.* **9,** 1–63.
23. Hol, W. G. J. (1985) *Prog. Biophys. Mol. Biol.* **45,** 149–195.
24. Engelman, D. M. & Steitz, T. A. (1981) *Cell* **23,** 411–422.
25. Singer, S. J. & Nicolson, G. L. (1972) *Science* **175,** 720–731.
26. Lenard, J. & Singer, S. J. (1966) *Proc. Natl. Acad. Sci. USA* **56,** 1828–1835.