Spontaneous insertion of polypeptide chains into membranes: A Monte Carlo model

(membrane proteins/lattice models/lipid–protein interactions/signal sequence/helical hairpin hypothesis)

MARIUSZ MILIK AND JEFFREY SKOLNICK

Department of Molecular Biology, The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, CA 92037

ABSTRACT   The Monte Carlo dynamics method was used to examine the process of protein insertion into model cell membranes. The water and lipid environments were taken into account via an effective medium approximation based on coordinate-dependent hydrophobic and hydrogen bond potentials. The polypeptide chain was represented in a full-backbone atom representation as a chain of diamond lattice vectors. The simulations support the idea that to a good approximation insertion may be depicted as a spontaneous thermodynamic process. The mechanism of membrane insertion of a simple lattice protein chain exhibits many features of theoretical predictions and is in good accord with experimental data. In the model, insertion begins with adsorption of the chain onto the interface, followed by the formation of helical fragments. These fragments, having partially saturated internal hydrogen bonds, can be transported into the lipid phase and then form transbilayer structures.

In order to function, cells must be able to exchange protein chains between their membrane-limited compartments (1, 2). This process requires the insertion and transport of amphiphilic polypeptide chains across the hydrophobic lipid membrane. Despite intense investigation (1–7), the details of this process remain controversial.

There is experimental evidence that the translocation process is coupled to the translation process and that a protein must be partially unfolded during the translocation process. When the native structure of a protein was stabilized by internal disulfide bridges, the protein was trapped in isolated mitochondria (8, 9). Conversely, denaturation with urea or destabilization of the native structure by a point mutation accelerated the translocation process (10–12). There is a supposition that the insertion process of a membrane protein into the bilayer is spontaneous (3, 5, 6) and mainly driven by the hydrophobic effect (4). In particular, Engelman and Steitz (3) emphasized the role of "helical hairpin" structures in the process of protein insertion. Jacobs and White (6) have added new features to the hypothesis of the thermodynamic basis for insertion. They suggest that protein insertion begins with adsorption on the membrane surface. This reduces the degrees of freedom of the chain and enhances the probability of forming intrapeptide hydrogen bonds leading to helical structure formation. The preformed helical fragments can then be transported into the bilayer without breaking hydrogen bonds.

The detailed mechanism of the insertion process still remains an open question. Based on previous work on water-soluble globular proteins and the phase transition in bilayers (13–15), it seems reasonable that simplified protein models employing dynamic Monte Carlo (MC) sampling can provide some useful insights into the mechanism of the insertion process.

Description of the Model

To model the two-phase environment, the MC cell is divided into "water" and "lipid" regions, according to the location along the z axis of the MC cell. The values of the hydrophobic and the hydrogen bond potentials depend on the z coordinate of the residue. This very schematic model environment appears to be sufficient to describe a substantial number of thermodynamic properties of the system.

The polypeptide molecule is represented by a diamond lattice chain. Every residue is composed of three united atoms, representing NH, C^α plus a side chain, and carbonyl groups, respectively. The geometrical properties of the diamond lattice restrict the number of possible torsional angles to three: $+60°$, $-60°$, and $180°$. As a result, the system can adopt a right- and left-handed square (four residues per turn) helix and an expanded zigzag state (equivalent to an α-helix and a β-strand in a real protein, respectively). The diamond lattice representation of the right-handed square helix and the extended conformation are presented in Fig. 1. Of course, this model cannot reproduce all the detailed structural aspects of a real peptide chain. However, our goal is to choose a minimal set of thermodynamic parameters that can describe the insertion process in the protein–membrane system, and then, to find its approximate mechanism.

The dynamics of the system were simulated using the usual set of moves (e.g., see ref. 15). Here, they are extended by one additional move that consists of the rotation of a randomly chosen bead with successive shifting of the appropriate chain fragment by a randomly chosen vector. This allows for movement of a chain fragment without destroying its internal substructure.

The Potential

The internal energy of the model is composed of torsional, Lennard–Jones, and hydrogen bond contributions. The torsional potential discriminates against the left-handed helix conformation for the φ and ψ angles (on a level of $0.5\ kT$ per torsional angle, where k is the Boltzmann constant and T is the absolute temperature) and gives preferences for the $\omega = 180°$ torsional angle on the level of $0.75\ kT$ per residue. Hence, the helix conformation is isoenergetic with the extended one. The hydrogen bond-type potential is

$$E_{Hb} = \begin{cases} \varepsilon_{Hb}, & \text{when } r_{ij} = \text{dist}_1 \\ 0, & \text{when } \text{dist}_1 < r_{ij} < \text{dist}_2 \\ 0, & \text{when } r_{ij} > \text{dist}_2 \text{ and atoms } i \text{ and } j \text{ are} \\ & \quad \text{in the "lipid" phase} \\ \varepsilon_{Hbw}, & \text{when } r_{ij} > \text{dist}_2 \text{ and at least one of } ij \text{ pair} \\ & \quad \text{of atoms is in the "water" phase,} \end{cases} \quad [1]$$
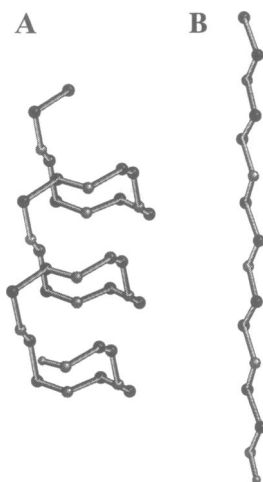
Abbreviation: MC, Monte Carlo.

FIG. 1. The right-handed square helix (A) and extended (B) conformations. Balls denote the C^α atoms.

where r_{ij} is the distance between the nitrogen or carbonyl atom of the residue i and the appropriate atom of residue i + 4 (for the carbonyl atom) or i − 4 (for the nitrogen atom), $dist_1$ is the length of the sums of two diamond-lattice vectors, and $dist_2$ is $19^{1/2}$ (the length of the sum of three diamond-lattice vectors in the "trans" conformation). $\varepsilon_{Hb} = -1.5\ kT$ is the energy of an intramolecular hydrogen bond interaction, and $\varepsilon_{Hbw} = -0.5\ kT$ is the energy of the hydrogen bond with water. The energy of a hydrogen bond depends on the environment of the residue, as well as on the local conformation of the polypeptide chain. When $dist_1 < r_{ij} < dist_2$, the potential is zero because between these distances, a water molecule cannot fit between the two atoms. When the residue is in an open, extended conformation ($r_{ij} > dist_2$ in Eq. 1), two bonds with water have to be broken to transfer it from the water to the lipid environment. When the residue is buried in a dense local globule ($dist_1 < r_{ij} < dist_2$ in Eq. 1), or the residue is in a helix conformation ($r_{ij} = dist_1$ in Eq. 1), the hydrogen bond contribution to the transfer free energy equals zero. This definition of the hydrogen bond interaction favors the transport into the lipid phase of "prefabricated" fragments of secondary structure, with partially saturated hydrogen bonds.

The intermolecular interaction energy is composed of one-body and two-body terms. The one-body term corresponds to the peptide–solvent interaction (hydrophobicity of a residue) and depends on the environment and residue type; this term is simulated by a coordinate-dependent mean force potential. A cuboid fragment of the MC cube is assumed to be the lipid phase, and the energy of a residue depends on its position in the MC box. In the present simulation, this "hydrophobic" effect is the main driving force for incorporation of a polypeptide chain into the lipid phase. Table 1 contains values of these parameters for the model set of residues.

The two-body term is a mean-force potential of interaction between two residues. This potential depends on the types of

Table 1. Values of one-body hydrophobic potential (E_{hph}) for the model residues

Symbol of residue	E_{hph}	Symbol of residue	E_{hph}
a	4.0	d	−1.5
b	0.0	e	−4.5
c	−1.5	f	−4.5

Positive values denote hydrophilic residues and negative values denote lipophilic residues. See text for additional information.

Table 2. Values of ε parameters of the two-body interaction for pairs of residues in a water environment

	a	b	c	d	e	f
a	0.0	0.2*	0.5*	0.5*	2.0*	2.0*
b		0.2*	0.5*	0.5*	0.5*	0.5*
c			0.5*	0.5*	0.5*	0.5*
d				0.5	0.5*	0.5
e					0.5*	0.5*
f						0.5

Values without asterisks are related to the Lennard–Jones-type potential defined in Eq. 2. Values with asterisks denote a repulsive interaction, defined in Eq. 3.

residues, the distance between them, and their environment. The two-body term is represented by the Lennard–Jones type (8 − 4) potential centered on the C^α atoms:

$$v_{LJ}(r) = 4\varepsilon[(\sigma/r)^8 - (\sigma/r)^4].\qquad[2]$$

For repulsive interactions, only the repulsive part of a Lennard–Jones potential is used:

$$v_{LJ}^*(r) = 4\varepsilon(\sigma/r)^8.\qquad[3]$$

The parameters ε are given in Tables 2 and 3 (the asterisks denote a repulsive interaction, where Eq. 3 is used). The value of the σ parameter is $(3)^{1/2}/(2)^{1/4}$. This way, the minimum of the attractive potential occurs at a distance equal to one diamond-lattice vector ($\sqrt{3}$).

Six types of residues denoted by letters from a to f are used. Type a residues are very hydrophilic, represent charged amino acids such as lysine or arginine, and are used to build the hydrophilic linker (ends) in sequence 1 (sequence 2) (see Table 4). A type b residue is inert; its transport free energy from the water into the lipid phase equals zero and models the out-of-membrane fragments of the chain that should be transported through the membrane (ends in the sequence 1 and linker in the sequence 2). Using the a and b residues, we can switch from the linker-first to the ends-first insertion mechanisms, as shown below. Types c and d are slightly lipophilic (Table 1); c residues interact attractively among themselves and with e residues, while the d residues interact repulsively with every residue in the lipid phase. The c and e residues are used for building the transbilayer fragments in sequence 1 and one transbilayer fragment in sequence 2 (Table 4). Attractive interactions between c and e residues help to form the supersecondary, hairpin structure. Types e and f are very hydrophobic (Table 1) and their two-body potential is identical to that of c and d residues. These residues form the very hydrophobic transbilayer structure in sequence 2 used to model the effect of the leader sequence in real proteins.

The polypeptide–lipid interaction is partially incorporated into the model by the use of an ordering free energy, which depends on the orientation of the fragment relative to the

Table 3. Values of ε parameters of the two-body interaction for pairs of residues in a lipid environment

	a	b	c	d	e	f
a	0.0	0.2	1.0	0.5*	2.0*	2.0*
b		0.2	0.0	0.5*	0.5*	0.5*
c			1.5	0.5*	1.5	0.5*
d				0.5*	0.5*	0.5*
e					1.5	0.5*
f						0.5*

Values without asterisks are related to the Lennard–Jones-type potential defined in Eq. 2. Values with asterisks denote a repulsive interaction, defined in Eq. 3.

Table 4. Sequences of model chains

Sequence	N end	Transbilayer fragment 1	Linker	Transbilayer fragment 2	C end
1	*bb*	*ccddccddccddccddccdd*	*aaaaaa*	*ccddccddccddccddccdd*	*bb*
2	*aa*	*eeffeeffeeffeeffeeff*	*bbbbbb*	*ccddccddccddccddccdd*	*aa*

Sequence 1 contains two slightly hydrophobic transbilayer fragments connected by a hydrophilic linker; the N and C ends are inert. The first transbilayer fragment of sequence 2 is very lipophilic to model the effect of a leader sequence in real proteins. The second transbilayer fragment of this sequence is the same as in sequence 1. Both fragments are connected by an inert linker; the ends are very hydrophilic.

main ordering axis in the lipid (which is perpendicular to the surface). This energy is

$$E_{ord} = \varepsilon_{ord} \sin^2(\theta),\qquad [4]$$

where ε_{ord} is a coefficient and θ is the angle between the end-to-end vector of a polypeptide fragment and the normal to the membrane surface. ε_{ord} is small, assuming values from 0.15 to 0.3 kT per residue. For comparison, the hydrogen bond interaction in the model varies from 1.5 to 2 kT per bond. Thus, E_{ord} is a weak interaction that favors structures which orient parallel to the main ordering axis. The value of ε_{ord} is very important. A too-small value of this parameter generates disordered conformations in the lipid phase. If ε_{ord} is too large, it drastically slows the dynamics of the chain in the bilayer. This effect is quantitatively consistent with experimental data, where either increasing or decreasing the fluidity of the bilayer drastically decreases the secretion of proteins (16, 17).

Results

Insertion of a hydrophobic hairpin having a hydrophilic linker. The model polypeptide consists of two hydrophobic fragments connected by a hydrophilic linker. The chain ends are slightly hydrophilic. The sequence of the model chain is given in Table 4, sequence 1, and was chosen to keep the linker on the starting side of the membrane.

A representative trajectory of the insertion process is shown in Fig. 2, where the contributions of the torsional (a), hydrophobic (b), and hydrogen bond (c) energies to the overall energy of the system are plotted versus time, measured in MC steps. The same process is shown in Fig. 3 as a sequence of snapshots allowing visualization of the accompanying structural changes.

The system starts from an initial random conformation in the water phase. After several MC steps, as shown in Fig. 3 *A* and *B*, the chain partially adsorbs onto the lipid surface. Because the ends of the chain are only slightly hydrophilic, they can be inserted into the lipid phase and form trans-bilayer structures (Fig. 3 *C–E*). This process experiences an
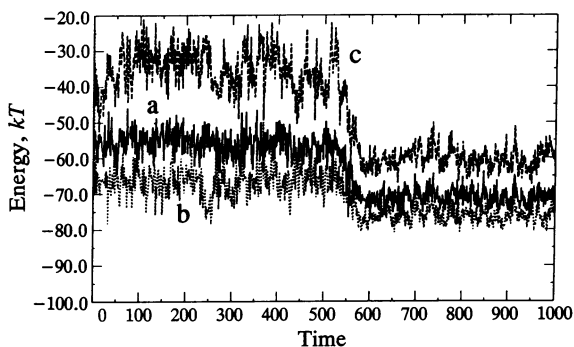

FIG. 2. Representative trajectory depicting the insertion process of a helical hairpin having a hydrophobic turn region. Curve a, contribution of the torsional energy; curve b, contribution of hydrophobic and side chain–side chain interactions; curve c, contribution of hydrogen bond energy. Time scale is in MC time steps.

energetic barrier of about 20 kT resulting from the breaking of peptide–water hydrogen bonds. The mostly disordered chain adsorbed on one surface of the bilayer (as in Fig. 3*B*) appears to be in local equilibrium.

The possibility of crossing this barrier is related to the probability that both chain ends jump across the bilayer. When this happens (Fig. 3*F*), both the side chain–side chain interaction energy and the hydrogen bond energy decrease and the system reaches the global energy minimum (Fig. 3*G*). The model polypeptide chain forms a regular, stable transbilayer hairpin (Fig. 3*H*).

The insertion of a hairpin with transportation of the ends across the membrane is kinetically favored. The ends are more mobile; thus they have a possibility to insert into the lipid phase parallel to the main ordering axis. However, this mechanism is "wrong" for the process of protein transportation across the membrane, according to Engelman and Steitz (3). A very important aspect of their hypothesis is the transport of the central turn across the membrane. The simplest way to force the system to insert the central turn is to make the ends more hydrophilic. Our hypothesis is that for this reason, the first part of the signal sequence of secreted proteins contains one or two very hydrophilic residues. Thus, the next portion of the present work contains an analysis of turn-first insertion of the helical hairpin.

Insertion of a hairpin having hydrophilic ends. The analogous sequence with a slightly hydrophilic linker and strongly hydrophilic ends could not cross the energetic barrier and could not form the hairpin. The time when the central linker stays in the lipid phase is too short for it to diffuse to the opposite side of the bilayer, and the hydrophobic interaction is too weak to stabilize the high-energy transition state.

To aid in insertion, we changed the sequence, making it similar to the N-terminal fragment of secreted proteins, before cleavage of the signal peptide (5). On the schematic level, this sequence is composed of the following fragments: [charged fragment]–[hydrophobic fragment (very often a trans-bilayer helix)]–[hydrophilic linker]–[amphipathic protein chain]. To model this pattern, the hydrophobicity parameter of one helix of the model hairpin was increased to simulate the signal sequence of the protein. The sequence is given in Table 4, sequence 2. Fig. 4 contains representative snapshots of the insertion process.

As before, insertion begins with adsorption of the chain onto the lipid surface phase, but now the hydrophilic helix is favored because of its differential hydrophobicity along the sequence. The amphiphilic fragment of the chain still tends to remain adsorbed on the surface, but the hydrophobic fragment is consistently buried in the lipid phase, with a high level of secondary structure (Fig. 4 *A–C*). A quasistable conformation associated with hairpin insertion into the bilayer is one where the amphiphilic fragment forms an adsorbed helix on the lipid surface, while the hydrophobic fragment forms a helix that grows perpendicular to the bilayer surface. Both helices are connected via an extended linker. The length of the hydrophobic helix varies with time (Fig. 4 *A–D*) up to the moment when the turn fragment reaches the opposite side of the bilayer. The fragment adsorbed on the surface is then pulled into the lipid. This fragment partially retains its secondary

A time = 10

E time = 500

F time = 530

B time = 260

G time = 550

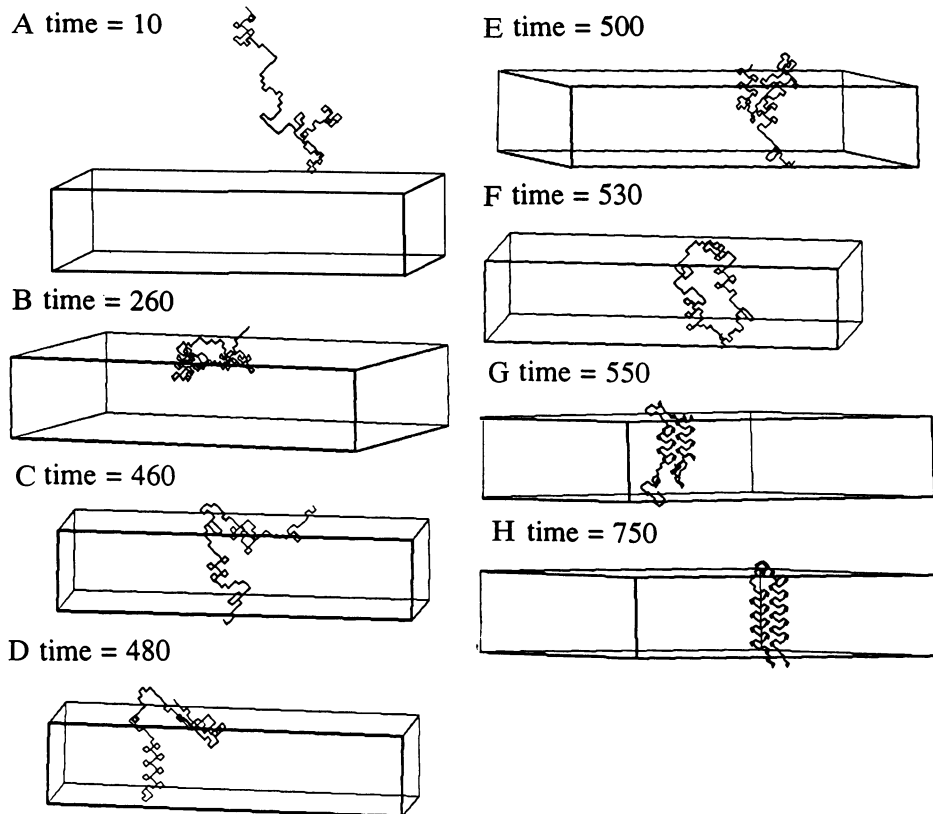C time = 460

H time = 750

D time = 480

FIG. 3. Representative snapshots of the insertion process of the model hairpin with a hydrophilic turn region.

structure, therefore reducing the transfer energy (Fig. 4*F*). The second helix then rapidly rebuilds in the neighborhood of the first (Fig. 4*G*). The transbilayer hairpin structure, which is additionally stabilized by side chain–side chain interactions, can fluctuate and diffuse in the lipid phase but does not show any tendency to dissolve during the simulation (Fig. 4*H*).

Discussion

Engelman and Steitz's helical hairpin hypothesis (3) emphasizes the importance of an antiparallel helix hairpin formed prior to peptide insertion into the bilayer. Subsequently,

Jacobs and White (6, 18) have further extended the hairpin hypothesis. The insertion begins with adsorption of the unfolded chain onto the membrane interface of the membrane. A polypeptide chain partially anchored to the interface has a greater chance of saturating its internal hydrogen bonds and forming helices. There are examples of polypeptides having a random coil structure in the water phase and adopting the α-helical conformation after binding to the bilayer surface (19, 20). According to Jacobs and White (6), these fragments of helical structure have a greater possibility of diffusing into the lipid phase.
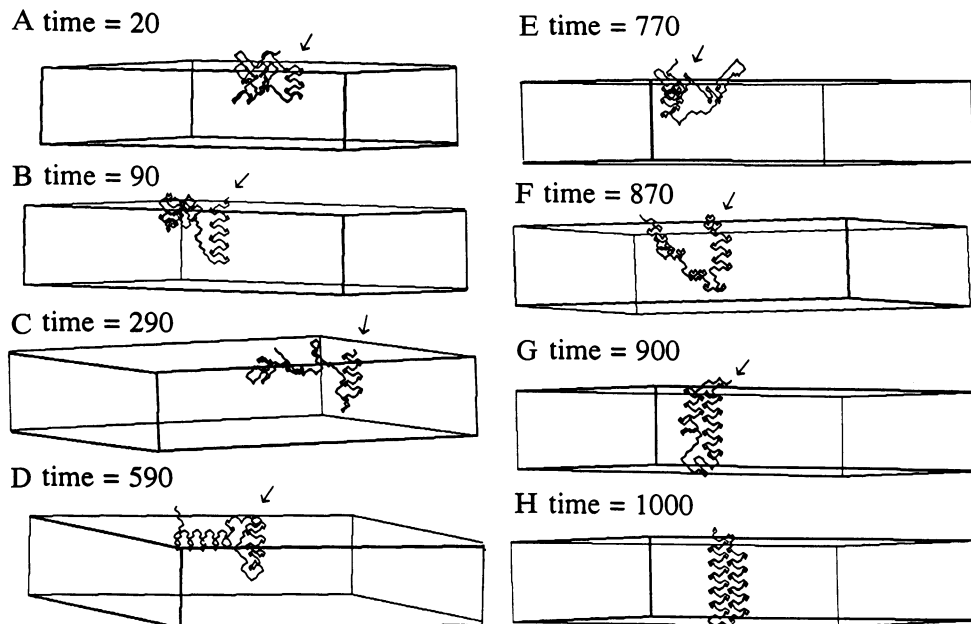
A time = 20

E time = 770

B time = 90

F time = 870

C time = 290

G time = 900

D time = 590

H time = 1000

FIG. 4. Snapshots of the insertion process of the model hairpin with hydrophilic ends. Arrow points to the N terminus of the model chain.

Biophysics: Milik and Skolnick

Proc. Natl. Acad. Sci. USA 89 (1992)   9395

Our MC model exhibits many elements of the Jacobs and White mechanism. In our model, the first stage of the insertion is always adsorption of the polypeptide chain onto the surface of the model lipid phase. This is very fast and driven mostly by the hydrophobic effect (Fig. 2). The adsorbed chain can then adopt a very ordered, but thermodynamically unstable, helical structure (see, for example, Fig. 4D). The hydrophobic helix is parallel to the main order axis of the bilayer, and the amphipathic one lies on the surface. The helices are connected by a fragment of the chain in an extended conformation. This extended fragment is thermodynamically disfavored because of its unsaturated hydrogen bonds.

The final transition begins with dissolution of the adsorbed helix and the insertion of the fragments of this helix (with partially saturated internal hydrogen bonds) into the lipid phase (Fig. 4F). This insertion gives a thermodynamically stable transbilayer structure.

Simulations for both sequences were repeated 15 times with consistent qualitative results. Sequence 1 leads to ends-first insertion, and sequence 2 to linker-first insertion. The exact values of the parameters are not critical. Moderate changes modify the detailed kinetics rather than the final structure and the overall mechanism of the process.

Conclusion

The present MC model of polypeptide chain insertion into a model membrane can provide a number of insights into the mechanism(s) of insertion, transport, and secretion of proteins. The simulations show on the schematic level that these processes can be modeled on a thermodynamic basis as the sum of the hydrophobic effect, hydrogen bond interactions, and protein–lipid interactions. The model cannot exclude the possibility that an active membrane transport mechanism (with consumption of ATP) also contributes to the insertion process (5); but, as pointed out by Jacobs and White (6), an active mechanism probably serves to catalyze a spontaneous thermodynamic process.

Additional insight into the thermodynamics of the membrane–protein system can be obtained by combining our diamond lattice model of the lipid bilayer (15) with the present model of the peptide chain. This model, having both elements of the membrane–protein system (the lipid bilayer and the protein) at the same level of detail, may provide important insights into the nature of the very cooperative processes in the system. Another question is whether the membrane can be treated as an effective medium or whether its internal structure must be explicitly considered in order to understand the behavior of membrane proteins. A natural extension of these models will be to an off-lattice model of the membrane–protein system.

1. Wickner, W. T. & Lodish, H. F. (1985) *Science* 230, 400–407.
2. Wickner, W. T. (1988) *Biochemistry* 27, 1081–1086.
3. Engelman, D. M. & Steitz, T. A. (1981) *Cell* 23, 411–422.
4. Jahnig, F. (1983) *Proc. Natl. Acad. Sci. USA* 80, 3691–3695.
5. Briggs, M. S. & Gierasch, L. M. (1986) *Adv. Protein Chem.* 38, 109–180.
6. Jacobs, R. E. & White, S. H. (1989) *Biochemistry* 28, 3421–3437.
7. Popot, J.-L. & Engelman, D. M. (1990) *Biochemistry* 29, 4031–4037.
8. Verstweber, D. & Schatz, G. (1988) *J. Cell Biol.* 107, 2037–2043.
9. Verstweber, D., Brunner, J., Baker, A. & Schatz, G. (1989) *Nature (London)* 341, 205–209.
10. Verstweber, D. & Schatz, G. (1988) *EMBO J.* 7, 1147–1151.
11. Eilers, M., Hwang, S. & Schatz, G. (1988) *EMBO J.* 7, 1147–1151.
12. Lazdunski, C. J. & Benedetti, H. (1990) *FEBS Lett.* 268, 408–414.
13. Skolnick, J. & Kolinski, A. (1990) *Science* 250, 1121–1125.
14. Godzik, A., Kolinski, A. & Skolnick, J. (1992) *Proc. Natl. Acad. Sci. USA* 89, 2629–2633.
15. Milik, M., Kolinski, A. & Skolnick, J. (1992) *J. Phys. Chem.* 96, 4015–4022.
16. Pages, J.-M., Piovant, M., Varenne, S. & Lazdunski, C. (1978) *Eur. J. Biochem.* 86, 589–602.
17. Pages, J.-M. & Lazdunski, C. (1981) *FEMS Microbiol. Lett.* 12, 65–69.
18. Jacobs, R. E. & White, S. H. (1986) *Biochemistry* 25, 2605–2612.
19. Schwytzer, R. (1986) *Biochemistry* 25, 4281–4286.
20. Sargent, D. F. & Schwytzer, R. (1986) *Proc. Natl. Acad. Sci. USA* 83, 5774–5778.