A Monte Carlo Model of fd and Pf1 Coat Proteins in Lipid Membranes

Mariusz Milik and Jeffrey Skolnick

The Scripps Research Institute, Department of Molecular Biology, La Jolla, California 92037 USA

ABSTRACT   A Monte Carlo Dynamics simulation was used to investigate the behavior of filamentous bacteriophage coat proteins in a model membrane environment. Our simulation agrees with the previous experimental observations that despite the low sequence similarity between the major coat proteins of Pf1 and fd bacteriophages, their structure in the membrane environment is very similar. These results support the hypothesis that the hydrophobic effect exerts an important influence on membrane protein structure. The model may also be used for modeling the insertion and transport processes in protein-membrane systems. The example of fd protein was also used as a test of sensitivity of our model to temperature, thickness of the hydrocarbon phase, and simulation time. In all cases, the results were independent (over the tested range) of the particular values of the parameters.

INTRODUCTION

In our previous work (Milik and Skolnick, 1993), we have proposed a simplified model of membrane-peptide systems. This model is based on the energy of transport of side chains from water into organic solvents; therefore, it emphasizes the importance of hydrophobic forces in membrane-containing systems. Starting from sequence information, this model is able to predict the spatial orientation of the peptides and their positions in the membrane. In the previous paper, we showed that our model gives results consistent with experiment for a number of peptides: magainin2, M2δ, melittin, and filamentous bacteriophage Pf1 major coat protein.

The present paper describes a continuation of this work. Here we present results of our simulation of the major coat protein of fd filamentous bacteriophage. To discuss some general features of membrane proteins, we compare our results with those obtained previously for the Pf1 bacteriophage major coat protein and with experimental data (McDonnell et al., 1993).

The goal of the present work is the formulation of a simple method for predicting the orientation and structure of membrane-bound peptides using information about their sequences. This simplified model may be used to predict more precisely the positions of transmembrane and membrane-adsorbed fragments in membrane protein sequences. We assume that the dominant interaction during membrane peptide insertion is the hydrophobic effect, and therefore, this term can be captured by a simplified model such as is presented here. Of course, the very schematic character of the model makes it difficult to address more complex problems of membrane-protein systems. If more detailed characteristics of the system (e.g., the membrane composition) become important, then more complex models and force fields should be used. Unfortunately, at present, such complex models of membrane-protein systems are computationally intractable.

Additionally, the present work tests the sensitivity of our simulation method to changes of simulation parameters such as temperature, thickness of the hydrocarbon phase, and simulation time. These tests are important to show that the results obtained in our simulations are not only artifacts of the method, but that they can give some insight into the nature of the membrane-peptide systems.

Pf1 and fd represent two classes of filamentous bacteriophages identified on the basis of their x-ray diffraction patterns (Nakashima et al., 1975; Clark and Gray, 1989). Virions of these phages consist of a single-stranded DNA ring coated by a protein layer (Ray, 1977). There are only two virion proteins, and the major protein (B) constitutes >90% of the protein coat. All of the major coat proteins for filamentous bacteriophages are relatively short (~50 amino acids), with large percentages (close to 100%) of α-helical secondary structure.

Although the Pf1 and fd bacteriophages have very similar overall structures, the sequences of their major coat proteins have low homology (McDonnell et al., 1993). The alignment using the "BESTFIT" program from the "GCG" package shows very low (16.67%) sequential identity between the fd and Pf1 coat proteins (see Fig. 1 a).

Fig. 1 b shows our attempt to align these sequences using only information about residue hydrophobicity. The alignment was obtained using a Genetic Algorithm (GA) minimization method (M. Milik and J. Skolnick, unpublished data). This algorithm uses a GA method to find the alignment (with gaps) of two sequences characterized by the minimum of a penalty function, f_p, defined in this case as

$$f_p = \sum_{i=1}^{N} (h_i^a - h_i^b)^2, \tag{1}$$

where h_i^a is value of hydrophobicity of the residue from the first sequence a, in the ith position of the alignment, h_i^b is the analogous value for the second sequence, b.

a. *Sequence alignment for fd and pfl filamentous bacteriophage coat proteins.*

```
fd   AEGDDPAKAA FDSLQASATE YIGYAWAMVV VIVGATIGIK LFKKFTSKAS
         |       |        ||    |     |          |
pfl  GVIDTSAVQSA ITDGQGDMKA IGGYIVGALV ILAVAGLIYS MLRKA
```

b. *Hydrophobicity alignment for fd and pfl filamentous bacteriophage proteins.*

```
fd   AEG**DDPA*KAAF*DSLQASATEYIG*YAWAMVVVIVGATI******GIKLFKKFTSKAS
pfl  **GVIDTSAVQSAITDG*QGDM*KAIGGYIVGALVILAVAGLIYSMLRKA
```

FIGURE 1   The sequential and "hydrophobicity-driven" alignments of sequences of the major coat proteins of fd and pfl bacteriophages. See text for details.

Fig. 2 plots the hydrophobicity of residues of fd (*solid line*) and pfl (*dotted line*) major coat proteins as a function of their position using the alignment from Fig. 1 *b*. From the point of view of the hydrophobic pattern, three main regions can be distinguished in both sequences. The first fragment from the N terminus to position 23 in the alignment is amphipathic. The second fragment running from position 23 to 32 in the alignment is very hydrophobic; its free energy of transport from water to membrane environment is always negative; this clearly defines a *trans* membrane fragment. The C termini of both proteins probably interact with DNA (Rowitch et al., 1988); therefore, their sequences are specific for a given virus and were excluded by the program from the alignment.

When two proteins with very low sequential similarity have such similar hydrophobic patterns, this suggests that their functions are related to their hydrophobicity rather than to other characteristics of their constituent amino acids.
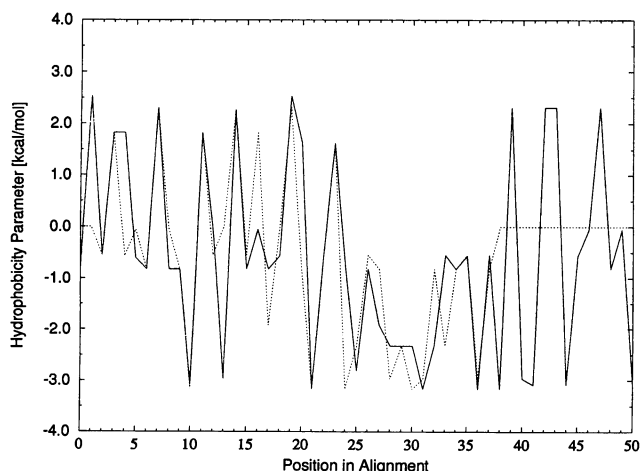

FIGURE 2   Plot of the hydrophobicity parameter as a function of position in the alignment for major coat proteins of fd (———) and pfl (·····) bacteriophages. Gaps in the aligned sequences are assigned the hydrophobicity parameter value of 0.

An analogous effect was established in the case of leader sequences (von Heijne and Abrahmsen, 1989).

In the present simulations, we have addressed the following questions. Are the similarities of hydrophobic patterns noticed by our model? Do the regions with a similar hydrophobic pattern in both proteins have similar secondary structure, orientation, and position in the model membrane?

MODEL

The Monte Carlo (MC) model of the protein-membrane system was presented in our previous paper (Milik and Skolnick, 1993). Here we only provide a brief summary for the reader's convenience. The protein is represented in this model by a chain of balls with centers at the C_α carbon positions. The radii of these balls are taken from a statistical analysis of a database of protein structures done by Gregoret and Cohen (1990). The Monte Carlo Dynamics method is used to calculate the equilibrium properties of this model. The set of elementary chain micro modifications used in this model contains "spike moves" for the ends and internal residues and a long range "sliding" move (see Milik and Skolnick (1994) for definitions and illustration). On the level of secondary structure, uniform (i.e., residue independent) helical propensities are assumed for all of the simulations and for all residues.

The membrane influence on the system is considered in this model as a z coordinate-dependent effective potential representing the different environment of the water, lipid, and lipid-water interfaces. Thus, the hydrophobic interaction energy is not only a function of the kind of amino acid but also its position in the system.

The hydropathy scale used here and in the previous paper (Milik and Skolnick, 1994) was derived from the experimental data of Jacobs and White (1987) for the interaction of tripeptides with model lipid bilayers and from Roseman's partition data for model compounds in octanol/water systems (Roseman, 1988). Additionally, the self-solvation effect was used to scale the thermodynamic data according to Roseman's (1988) proposition. The methods of calculation of the hydrophobic parameters and their values are presented in the previous paper.

The energy of the model chain is a sum of the uniform helical propensities, hydrogen bonding, and hydrophobic terms. The excluded volume of the chain is implemented by using the "hard core" repulsion between residues. The effect of the lipid membrane is considered using the z coordinate-dependent hydrophobic effect and the "ordering effect." This last term is included to consider the anisotropy of the lipid bilayers and is equal to $C_{ord} \sin^2(\theta)$, where C_{ord} is a parameter ranging from 0.05 to 0.15 kcal/mol and θ is the angle between the end-to-end vector of a polypeptide fragment and the z axis (normal to the membrane surface).

RESULTS

Fig. 3 contains a series of representative snapshots of the process of insertion of the model chain into the membrane. The figure shows a side view of the MC box, where the borders of the hydrocarbon phase are represented by the walls of a cuboid. For most of the simulations used in the present paper, the parameters were the same as in our previous work and are presented in Table 1.

Most of simulations ran for 5×10^6 MC steps, where one MC step is defined as the time when the algorithm tries to move every residue of the model chain. Information about the system was stored every 2500 MC steps into a trajectory file. This file was then used for the analysis of the behavior of the model system during the simulation. Depicted below is a typical trajectory of the model insertion process.

Starting from a random conformation outside the membrane, the peptide chain adsorbs very quickly onto the surface of the membrane and forms a slightly distorted helical structure (Fig. 3 a). This process is almost instantaneous in all tested trajectories. Then the chain that is adsorbed on the interface reorganizes its structure and waits for an energy fluctuation sufficient to transport the hydrophilic C terminus of the protein across the membrane (Fig. 3b-d). The chain spends a large fragment of the simulation time (3 to 4.5×10^6 Monte Carlo steps) in this surface-adsorbed state. In all 15 simulations of fd protein chain, the C terminus of the chain finally crossed the membrane. In principle, if we would have run even more trajectories, some would not have folded during the 5×10^6 Monte Carlo steps. Therefore, all structures and topology predictions should be based on conclusions gleaned from a large number of independent trajectories.

When the C terminus crosses the membrane (Fig. 3 e), the final structure is formed very quickly and remains stable during the remainder of the simulation (Fig. 3 f). The final structure is very similar to that obtained by NMR experi-

TABLE 1 Values of parameters used in the present simulations

Parameter	Value
Temperature	305.0 (K)
Minimum of the bond potential ($E_{bond}(q_0)$)	2.00 (kcal/mol)
Equilibrium bond angle (q_0)	89.5 (deg)
Minimum of the torsional potential ($E_{bond}(r_0)$)	1.5 (kcal/mol)
Equilibrium torsional angle (r_0)	52.1 (deg)
Nematic potential coefficient	0.05 (kcal/mol)
Thickness of the hydrocarbon phase	27.0 Å
Thickness of the interface	4.5 Å

ments (McDonell et al., 1993) and contains two helices connected by a less ordered link.

These simulations provide the important information that the C terminus crosses the membrane as a formed helical fragment and that insertion of a randomly coiled fragment of chain into the membrane phase is energetically very unfavorable.

In the case of the fd coat protein, the transfer free energy of side chains into the membrane constitutes ~25% of the stabilization energy. However, despite the apparently modest contribution, being the only sequentially dependent energetic term in our model, the transfer free energy decides the global configuration of a peptide fragment. According to our simulations, predominantly hydrophilic peptide fragments and hydrophobic peptide fragments with random patterns do not form stable membrane-bound structures. Thus, the hydrophobic effect makes the most important contribution to the energy of formation of the structure of small membrane proteins. Hydrogen bonding in the membrane phase contributes ~40% of the energy of stabilization. The secondary structure propensities, which are superimposed to be uniformly helical for all the residues, have a contribution of ~33%. The contribution of the ordering term was marginal and was never >0.3%.
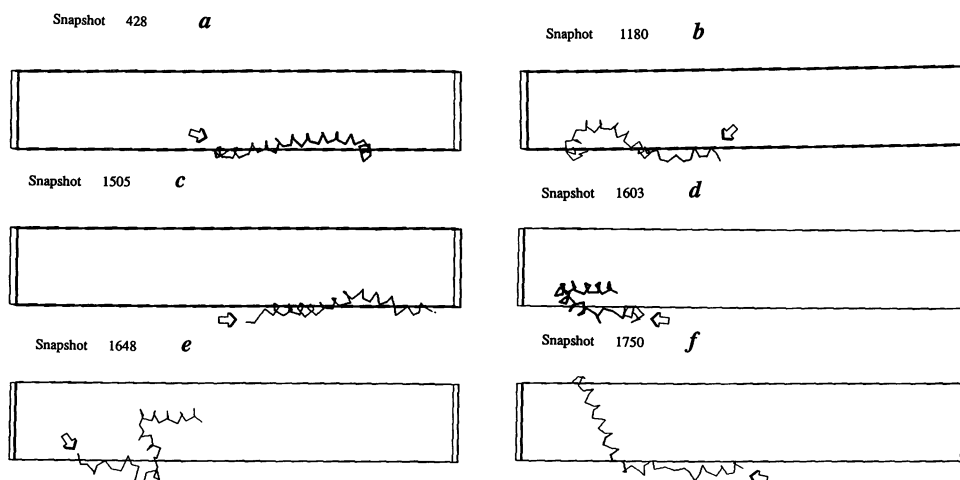

FIGURE 3 Snapshots from a typical simulation of the membrane insertion mechanism of the fd filamentous bacteriophage coat protein. The arrows indicate the N terminus of the protein. Arrows point to the C terminus of the model fd protein. See text for additional details.

To check the stability of the final state, we have run very long simulations (5×10^7 MC steps), starting from the final state with all parameters as in Table 1. Both the orientation of the helices and location of the link were stable during this simulation, despite the very fast diffusion of the peptide in the membrane phase. This argues that the final structure obtained during the simulation is in a deep minimum of free energy.

The sensitivity of the model to temperature was tested by simulating the peptide-membrane system at a temperature of 270 K, instead of the usual 305 K. The behavior of the model system at the lower temperature is analogous to that at higher temperature. According to our observations, the final structure of model chains does not depend on temperature over the examined temperature range. Obviously, considering the very simplified representation of the membrane phase in our simulations, these low temperature simulations should be understood only as a test of the method for prediction of the structure of membrane peptides. With the current level of complication, our model cannot be used to examine the effect of temperature on the kinetics of membrane peptide insertion.

One of the important features of our model is the possibility of obtaining information about the position of the transbilayer helices, adsorbed fragments, and turns. This information is not assumed in our model, because we use uniform helix propensities for the entire chain. However, it is possible that the predicted protein structure is only an artifact of the geometry of the model membrane. A set of runs of our model with different values of the model membrane thickness was used to examine this possibility. The results of simulations of systems with linear dimensions of the hydrocarbon phase ranging from 21 to 27 Å showed that the final structure of the model membrane protein is not correlated with the membrane thickness. Fig. 4 shows a comparison of the mean values of the z coordinates of fd
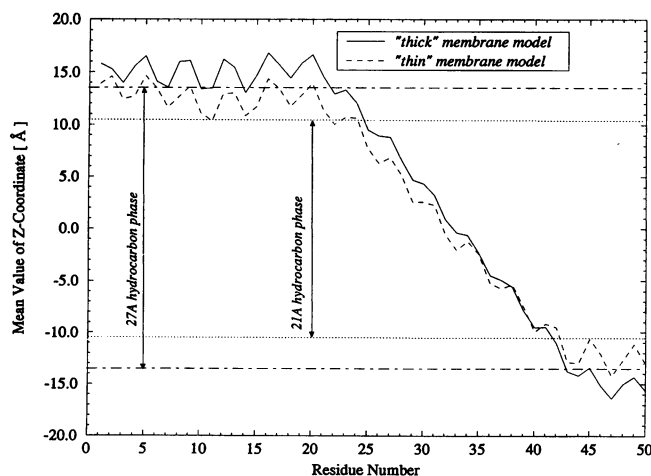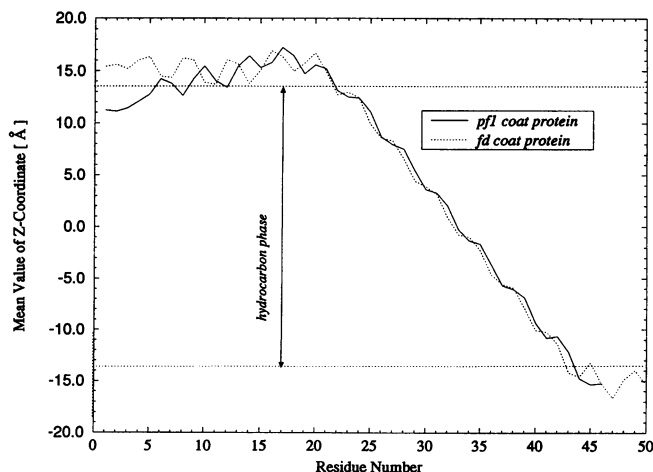

FIGURE 5   Mean (over one simulation cycle) values of the z coordinates of the residues of fd and pf1 coat proteins plotted as a function of residue number. The figure indicates considerable similarity between the structure of pf1 coat protein and the predicted structure of the fd coat protein.

coat protein obtained for thicknesses of the model hydrocarbon phase equal to 21 and 27 Å. Over the examined range, the position of the predicted turn and orientation of the helical fragments does not depend on the membrane geometry.

Thus, we conclude that the structures of the membrane proteins determined here depend strongly on the pattern of hydrophobic residues of the protein and not on the parameters of the simulation (over the investigated range).

Fig. 5 shows the average values of the z coordinates of the α-carbons of Pf1 and fd coat proteins, obtained using the present method. Despite the low sequential homology, the similarity of the hydrophobic patterns for both these proteins implies a very similar structure. Both proteins have an interfacial adsorbed fragment, which is 17 residues long (from 3 to 20), and a transbilayer fragment from residue 22 to residues 42 for Pf1 and 43 for fd.
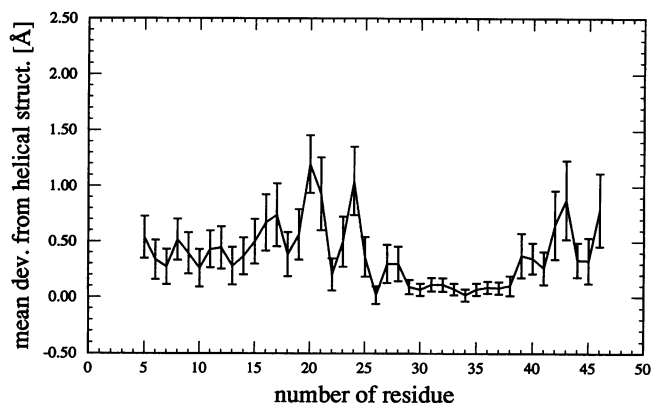
Fig. 6 presents the mean deviations from the "ideal"

FIGURE 4   Mean values of the z coordinates of the residues of fd coat protein obtained during simulations using different values of the model membrane thickness.

FIGURE 6   The mean (over one simulation cycle) deviations from an "ideal" helix structure for residues in the fd model. The bars denote values of the mean absolute deviation of the data.

helical structure for the model of fd proteins. The deviations are defined as a sum of distances in Å between the mean values of $r_{i,i+3}$ and $r_{i,i+4}$ from simulation and the "ideal" values from the literature (Barlow and Thornton, 1988); $r_{i,i+k}$ denotes the distance between residues i and $i + k$. Values of $r_{i,i+4}$ for residues 1–4 and 46–50 are unavailable; therefore, these residues are excluded from the plot. These deviations are presented here to show how external conditions (membrane geometry, hydrophobic forces) distort the secondary structure of the inserted protein molecule. According to the figure, the fd molecule can be divided into four fragments:

- from 5 to 19, first, surface-adsorbed helix;
- from 20 to 24, a distorted linker;
- from 25 to 41, second, transbilayer helix; and
- from 42 to 46, mobile C terminus.

This is qualitatively consistent with experimental data (Mc-Donnell et al., 1993), where the first, amphipathic helix contains residues from 7 to 20 and the second, transbilayer helix is located between residues 23 and 42.

The vertical bars in Fig. 6 represent values of mean absolute deviations of distribution of deviation from the ideal helical structure. The size of the bar is proportional to the mobility of the peptide residue (intuitively, a more mobile residue has a broader distribution). According to this interpretation, the linker residues (20–24) are more mobile than residues in adsorbed and transbilayer helices; this is also consistent with experimental results (McDonnell et al., 1993).

CONCLUSION

The Monte Carlo Dynamics model of the major coat protein of the fd bacteriophage in the membrane environment can provide insights into the supersecondary structure and putative position of the transbilayer region. Despite the lack of sequence identity, the model structure of the fd bacteriophage major coat protein is very similar to the model structure of Pf1 bacteriophage major coat protein (Milik and Skolnick, 1993) and is in accord with experimental data (McDonnell et al., 1993). According to our model, the structure results from the similarity of hydrophobic patterns for both proteins. Following this line of reasoning, the similarity of structures implies similarity in the processes of penetration and assembly for both of these phages.

Additionally, the example of fd coat protein was used to test the sensitivity of the present model to changes of temperature, thickness of the hydrocarbon phase, and simulation time. The tests showed that the final structure obtained in the process of modeling does not depend on the model parameters over the tested range. The very long run (5×10^7 MC steps) of the fd coat protein model showed that (despite the large energy fluctuations) the final structure

is stable over the course of the simulation and remains unchanged. The results of these tests encourage us to believe that our predictions of supersecondary structures and orientation of membrane proteins are reasonable and can give insight into the membrane protein structure and dynamics.

One of the important simplifications of the model of membrane proteins presented above was the representation of amino acids as single balls with centers at the C_α position. For a more realistic analysis of the structure of membrane-bound peptides and proteins, a more detailed representation of backbone and side chains must be used. Obviously, the shape and chemical character of side chains are very important for membrane-protein systems, especially in the case of large, amphipathic side chains (e.g., tryptophan). Additionally, according to the experimental data, positively charged side chains have a smaller probability of crossing the cell membrane (the "positive inside" rule) (Gavel et al., 1991). This effect is very important for the prediction of the structure of membrane proteins and will be considered in future work.

REFERENCES

Barlow, D. J., and J. M. Thornton. 1988. Helix geometry in proteins. J. Mol. Biol. 201:601–619.

Clack, B. A., and D. M. Gray. 1989. A CD determination of the α-helix contents of the coat proteins of four filamentous bacteriophages: fd, IKe, Pf1 and Pf3. Biopolymers. 28:1861–1873.

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

Gregoret, L. M., and F. E. Cohen. 1990. Novel method for the rapid evaluation of packing in protein structures. J. Mol. Biol. 211:959–974.

von Heijne, G., and L. Abrahmsen. 1989. Species-specific variation in signal peptide design: implications for protein secretion in foreign hosts. FEBS Lett. 244:439–446.

Jacobs, R. E., and S. H. White. 1987. Lipid bilayer perturbations induced by simple hydrophobic peptides. Biochemistry. 26:6127–6134.

McDonnell, P. A., K. Shon, Y. Kim, and S. J. Opella. 1993. fd coat protein structure in membrane environments. J. Mol. Biol. 233:447–463.

Milik, M., and J. Skolnick. 1993. Insertion of peptide chains into lipid membranes: an off lattice Monte Carlo Dynamics model. Proteins. 15:10–25.

Nakashima, Y., R. L. Wiseman, W. Konigsberg, and D. A. Marvin. 1975. Primary structure of side chain interactions of PFI filamentous bacterial virus coat protein. Nature. 253:68–71.

Ray, D. S. 1977. Replication of filamentous bacteriophages. In Comprehensive Virology, Vol. 7. H. Frankel-Conrat and R. R. Wagner, editors. Plenum Press, New York. 105 pp.

Roseman, M. A. 1988. Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds. J. Mol. Biol. 200:513–522.

Rowitch, D. H., G. J. Hunter, and R. N. Perham. 1988. Variable electrostatic interaction between DNA and coat protein in filamentous bacteriophage assembly. J. Mol. Biol. 204:663–674.