PHYSICAL REVIEW E, VOLUME 63, 010901(R)

Lattice model of transmembrane polypeptide folding

C.-M. Chen
Physics Department, National Taiwan Normal University, Taipei, Taiwan, Republic of China
and Institute of Physics, Academia Sinica, Taipei, Taiwan, Republic of China

(Received 11 October 1999; revised manuscript received 14 August 2000; published 18 December 2000)

Folding of hydrophobic polypeptides into unique three-dimensional structures in a membrane is investigated by Monte Carlo simulations using the bond fluctuation model. Its ground state structure can be a helix or a double helix depending on the competition of hydrogen bonding and backbone bending energies. The folding pathway of hydrophobic polypeptides in a nonpolar environment is found to favor the helical structure over the double helix. The folding time of a transmembrane domain increases exponentially with the chain length. Folding at low temperatures exhibits an Arrhenius-like behavior. We discuss the kinetics of both random folding and channel complex assisted folding of a polypeptide chain. Our results suggest a significantly smaller energetic barrier in the folding pathway for channel complex assisted folding.

DOI: 10.1103/PhysRevE.63.010901 PACS number(s): 87.15.Aa, 87.14.Ee, 87.15.Cc, 87.15.Kg

The three-dimensional (3D) structures of proteins play an important role in determining their biological functions. Such unique structures are believed to be their thermodynamic ground state. Although tremendous efforts have been invested in studying the protein folding problem, it is so far unclear about the folding kinetics and also difficult in predicting the structure of proteins [1–3]. Much less is known about membrane proteins (MPs) than soluble proteins [4]. Until now, only a dozen or so MPs have known crystallographic structures. Transmembrane domains of MPs are buried in a nonpolar environment of lipid bilayers, where both the hydrophobic interaction and hydrogen bonding between amino acids and the environment are absent. In such an environment, the intrachain hydrogen bonding, instead of the hydrophobic interaction, plays an important role for the folding of MPs. It has been suggested that folding of many integral MPs can be understood by a two-stage model [5]: Independently stable helices are formed in lipid bilayers in the first stage, and the helices interact with others to form a functional membrane protein in the second stage. This model led to the expectation that the transmembrane region of MPs would consist of bundles of hydrophobic α-helices, which has been largely fulfilled except for few MPs.

Despite a qualitative understanding of MP folding from the two-stage model, the detail of the folding kinetics of MPs is still missing. A previous model using a full-backbone atom representation in a diamond lattice initiates an interesting study on the insertion of polypeptides into membranes [6]. However, its results are questionable due to a major drawback in its backbone hydrogen bonding potential which explicitly specifies that hydrogen bonding can only occur for $(i-4, i)$ and $(i, i+4)$ pairs, where i labels amino acids in the chain. Moreover, hydrophilic channels, such as the Sec61p complex in eukaryotes and the SecYEG complex in prokaryotes [7], seem to play an important role for the arrangement of many MPs into segments and their transport across the lipid bilayer [8]. It is also unclear how hydrophilic channels would assist the folding of MPs. Another interesting question is the effect of the nonpolar environment of the membrane on the folding pathway of MPs. Until now, many important questions about MP folding remain open.

In this Rapid Communication, we propose a simple lattice model for the folding of MPs, particularly on the folding of a transmembrane domain (TMD) consisting of 14–26 amino acids, based on the bond-fluctuation model [9,10]. The advantages of the bond-fluctuation model are to give reasonably good secondary structures and to simulate a more realistic diffusive kinetics than regular lattice models, while the computational cost is still quite limited compared to that of off-lattice models. In many cases, these polypeptide chains would fold into a helical structure mainly due to the backbone hydrogen bonding (the hydrogen bonding energy is then set to unity throughout this paper). The study of folding of a hydrophobic segment into a helix is important since transmembrane helices are regarded as autonomous folding domains. Moreover, the structures of many membrane polypeptides, such as alamethicin and gramicidin, are also helical. Our study could also be useful for the folding of many MPs assisted by channel complexes, in which cases MPs are arranged into segments assisted by the complex and each segment can then fold into a helix. To do so, we study the folding of polypeptide chains with two different types of initial conformation: (i) a random initial configuration near the membrane corresponding to the folding process of a polypeptide chain originally in water (nonconstitutive MPs), and (ii) an initial configuration perpendicularly penetrating the membrane corresponding to the folding of a polypeptide chain assisted by a hydrophilic channel complex (constitutive MPs). Our results indicate that it might be possible to know the folding of constitutive MPs without specifying the detail interaction involved between the polypeptide and the channel complex. We note that, to get statistically faithful results, the computational effort is quite substantial: All the data collected take about four months by using 16 dual CPU Linux workstations (16 Pentinum III 450 Mhz CPUs and 16 Pentinum II 333 Mhz CPUs).

In our model, we consider a polypeptide as a polymer chain with the following potential energy:

$$U = E_{\text{H-bond}} + E_{\text{bend}} + E_{\text{hydrophobic}}, \tag{1}$$

where $E_{\text{H-bond}}$ is the hydrogen bonding potential energy,

E_{bend} is the bending energy of the chain, and $E_{\text{hydrophobic}}$ is the hydrophobic interaction energy. In the bond-fluctuation model, each amino acid occupies a $2\times2\times2$ cube of sites on a cubic lattice. The set of allowed bond vectors is $\mathbf{B}=P(2,0,0)\cup P(2,1,0)\cup P(2,1,1)\cup P(2,2,1)\cup P(3,0,0)\cup P(3,1,0)$, where $P(a,b,c)$ stands for the set of all permutations and sign combinations of $\pm a$, $\pm b$, and $\pm c$. The total number of configurations for a N residue polypeptide chain in this model is about $\Gamma_N\simeq1.26\times(85.2)^{N-1}(N-1)^{1/6}$ (for example, $\Gamma_{18}\simeq10^{33}$) [10]. We note that, although the available configurations grow rapidly with chain length, it is not difficult to study the folding of a long membrane protein by this model since the whole sequence can be divided into several autonomous folding domains consisting of about 30 amino acids. A hydrogen bond can form if two amino acids are separated by a four lattice spacing, which means that one lattice spacing equals 1.35 Å (a quarter of a helix pitch). However, each amino acid can at most participate in two hydrogen bonds. For simplicity, we have explicitly excluded the possibility of forming 2_7 ribbons and 3_{10} helices. We note that hydrogen bonds between amino acids and water molecules can always form when amino acids are surrounded by water. The bending energy of two consecutive bonds is assumed to be proportional to $1-\cos\theta$ with a bending rigidity e relative to the hydrogen bonding strength, where θ is the angle between two bonds. In general, one expects that the value of e increases with the size of side group of each amino acid (for example, $e_{\text{Leu}}>e_{\text{Gly}}$) and therefore is heterogeneous along the chain. The hydrophobic interaction is switched off if the amino acids enter the bilayer whose thickness is 33 lattice spacing (or about 45Å). Experimentally, since the hydrophobic interaction and hydrogen bonding are estimated to be in the range of 1.5–2 kcal/mol and 3–6 kcal/mol respectively, we then take the relative strength of the hydrophobic interaction to the hydrogen bonding strength to be 0.66 [11]. Here we have ignored both the van der Waals interaction (less than 1 kcal/mol) and the electrostatic interaction for hydrophobic amino acids. In other words, we believe that the folding of a TMD can be treated as a homopolymer folding. Heterogeneity will be considered to study the absorption and insertion processes of MPs consisting of several TMDs in the future. Moreover, the insertion of a polypeptide chain into the membrane will disturb the integrity of the membrane and local lipid density around the chain, which increases the energy of the membrane. We model this effect by considering an effective lateral pressure (P) applied to the polypeptide chain by lipids to minimize the lateral area ($A=L^2$) of the polypeptide chain in the membrane, where L is the projected length of the inserted portion of the chain on the membrane. Therefore, the relevant physical quantity to be minimized in our model is the enthalpy $H=U+PA$. The dynamics of a chain is simulated by the Metropolis Monte Carlo (MC) algorithm in a cubic lattice at a constant temperature T using the bond-fluctuation model. At each instant, a monomer is picked up at random and attempts to move in any of the six directions by one lattice spacing. If any attempted move of monomers satisfies the excluded volume constraint and the new bond vectors are still in the allowed set, then the move is accepted with probability $w=\min[1,\exp(-\Delta H/T)]$, where ΔH is the enthalpy change of the system.

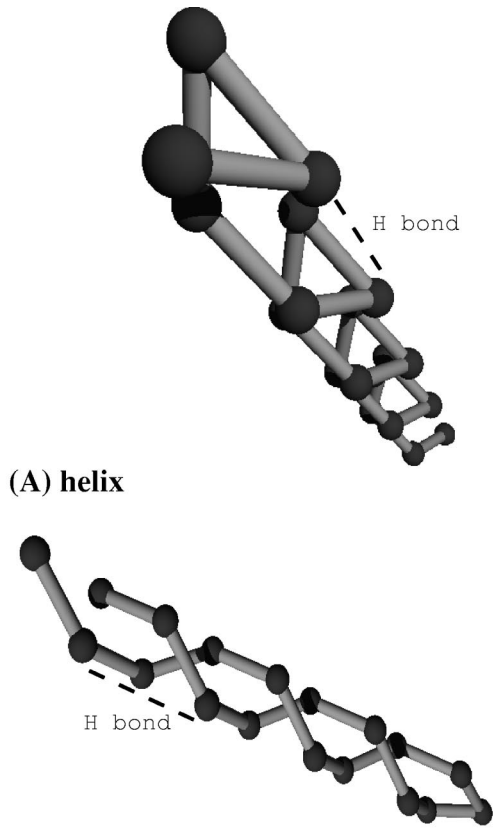

FIG. 1. Ground state structures of an 18 residue hydrophobic polypeptide in a bilayer membrane: (a) a helix for $e<0.11$ and (b) a double helix (or a twisted β strand) for $e>0.11$. Only one hydrogen bond is shown in each structure. The effective lateral pressure P is set to be 0.4.

First, we study the thermodynamic ground state structure of a polypeptide chain of 18 hydrophobic residues for various values of the bending rigidity e. The effective lateral pressure P is set to be 0.4 throughout this paper. Five long Monte Carlo simulations with different initial configurations of the chain near the membrane were performed to find the conformation with the lowest energy. For $e<0.11$, the lowest energy conformation is a helix, as shown in Fig. 1(a), while the double helix structure, as shown in Fig. 1(b), has the lowest energy for $e>0.11$. In the helix, there are four residues per turn and a hydrogen bond can form between n and $n+4$ residues, as shown in Fig. 1(a), where n is the index of amino acids in the chain. The helical structure has a larger number of hydrogen bonds than the double helix, while its bending energy is also higher than that of the double helix. Both structures have the same lateral area. An exact calculation of the enthalpy for the two structures confirms the simulation results.

The helical structure in Fig. 1(a) has been seen in many MPs and hydrophobic polypeptides, but the double helix structure predicted in this model has rarely been reported

before. Nevertheless, a metastable double helix structure has been observed for gramicidin dimers in lipid membranes, which can be converted to a stable helical structure by incubating the samples for several days at 68°C [12]. We note that such a double helix structure is in fact a β strand, which is twisted in order to maximize the number of hydrogen bonds without changing the overall geometry. For MPs, α helices are much more abundant than β strands. The reason of this phenomenon has been conjectured to relate to the folding pathway of MPs. To examine this conjecture, we study the folding of the 18 residue chain in the regime where the double helix is the ground state ($e=0.13$) such that only folding kinetics is possible to favor the formation of α helices. Two different initial configurations are used, and for each case we study the folding time t [or the mean first passage time (MFPT)] to the ground state for 100 different runs at a temperature at which the folding rate is the fastest. One set of simulations starts with a random initial configuration near the membrane, while the other starts with an initial configuration perpendicularly penetrating the membrane. In both cases, the polypeptide chain folds into the helical structure almost every time (96/100 in the first case, and 97/100 in the second case), and the MFPT is about 1.5×10^8 MC steps to the double helix but only 10^6 MC steps to the helix. The substantial difference in the MFPT for the two structures implies a strong bias in the folding pathway as conjectured. It is reasonable to conjecture a funnellike structure in the free energy landscape corresponding to the helix state. We note that, in the absence of hydrophobic interaction, the compact structure of a MP is finally reached by the association of autonomous folding domains due to the van der Waals–London interaction, suggesting that its intermediate structure is less compact. The above bias is then due to the fact that hydrogen bonding is a local interaction in an α helix, while the interaction is nonlocal in a β strand.

To illustrate the difference in the kinetics of random folding and assisted folding, we calculate the MFPT of TMD folding with two different initial conditions at various temperatures. Figure 2 shows the dependence of folding time t on temperature T for both nonconstitutive and constitutive polypeptides consisting of 18 residues. Here, we choose the bending rigidity $e=0.067$ and the helix is the expected ground state. It has a clear minimum at some optimal temperature ($T_o\sim0.3$), which is about the same for both cases. At low temperatures, the folding in both cases shows an Arrhenius-like behavior [2], i.e., the logarithm of folding time depends linearly on inverse temperature or $t\sim\exp(E_b/T)$. This suggests that folding at low temperatures is an activated process in which an energetic barrier (E_b) must be overcome to find the ground state. For nonconstitutive polypeptides the energetic barrier E_b is about 3.35, while E_b is about 2.56 for constitutive polypeptides. This demonstrates that the channel complex assisted folding actually selects a folding pathway with a significantly smaller energetic barrier (the difference is about 2.6 kT_o). At high temperatures, the folding time is about the same for both cases since the folding process searches for most of the conformation space to find the ground state structure. We note that the insertion time of a nonconstitutive chain into the membrane has been subtracted from the folding time since we limit our discussion to the folding process of a TMD. Moreover, additional interfacial regions of thickness about 10–15 Å between the bulk water and the bilayer should be introduced in order to calculate the insertion time properly [4].

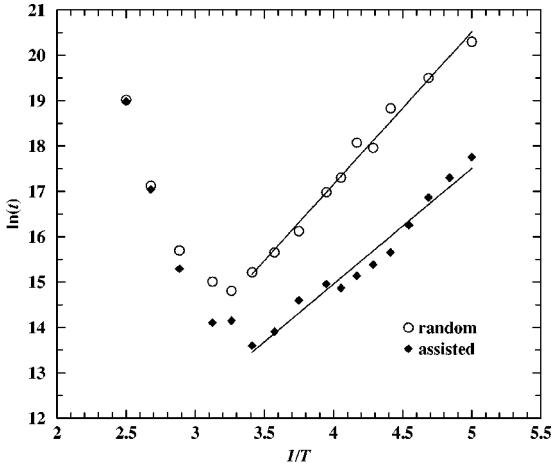

FIG. 2. Temperature dependence of the folding time t of an 18 residue hydrophobic polypeptide for both random folding and assisted folding. The ground state structure is a helix for $e=0.067$ and $P=0.4$. Straight lines are linear fits of the low temperature part of the dependence. The slope of the lines is the energy barrier E_b.

For the folding of single domain proteins in water, a three stage multipathway mechanism (TSMM) has been proposed and observed in a class of lattice and off-lattice models [13]. The length dependence of folding time for these three stages is proposed to be proportional to N^2 for nonspecific collapse, N^3 for diffusive search, and $e^{0.6\sqrt{N}}$ for activated transitions. For assisted membrane protein folding, the length dependence of folding time to the ground state and to all helical states at the optimal temperature is shown in Fig. 3. Our simulation data indicate that the MFPT to the ground state increases exponentially with chain length and the MFPT to helical states is roughly proportional to N^9. Such a strong length dependence is because, unlike TSMM in which the number of compact structures increases slowly with N, available configurations increase much faster ($\propto85.2^{N-1}$) than the rim of the funnel of the ground state as length increases. Therefore, folding time of proteins in a nonpolar environment is expected to be much longer than that in water. This shows another major difference between a hydrogen bonding dominated folding and a hydrophobicity dominated folding. Note that, in addition to the ground state, there are many other helical states whose energy is close to the ground state energy in our model. The existence of these other low energy helical states is biologically important since they provide the flexibility for polypeptides to be biologically functional. For example, a voltage gated gramicidin A can thus adjust its cross section area to transport ions across the membrane. Clearly, if the ground state of a protein is too stable to other biologically related structures, it would not be biologically functional. In fact, several inherited neurodegenerative disor-

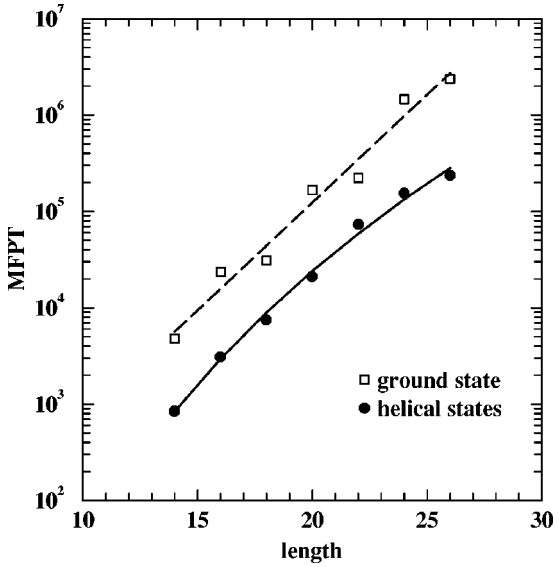

FIG. 3. Folding time of a transmembrane domain to the ground state or to all helical states. The ground state structure is a helix for $e=0.067$ and $P=0.4$.

ders are now known to be caused by the poly-$_L$-glutamine helix which has a long-lived open state [14].

To conclude, we have used the bond-fluctuation model to study the structure and folding kinetics of polypeptide chains consisting of 14–26 residues in membranes. This model is able to predict the structure of an autonomous folding domain, such as a helix or a double helix. The helix state is found to be both energetically and kinetically favored, which is consistent with recent experimental results demonstrating significant membrane-promoting helix formation by using circular dichroism spectroscopy [15]. Our simulation results show that the folding process favors the formation of a helix over a double helix: In the case where the double helix is the ground state, the folding time of the double helix is 100 times longer than that of the helix. This result suggests a single funnel-like structure in the free energy landscape [3] corresponding to the helix state for various values of e. We also investigate the folding time of both nonconstitutive and constitutive polypeptide chains and find that channel complex assisted folding selects a folding pathway, which has a lower energetic barrier than that of a nonconstitutive chain. The difference in the free energy barrier is about 2.6 kT_o. The folding time of a TMD is found to increase exponentially with chain length. Several helical states whose energy is comparable with the ground state energy are found in our model, which provides the flexibility for the polypeptides to be biologically functional. Finally, we note that this model can be applied to study the folding of MPs by including the partition process, which divides the chain into several autonomous folding domains, and the association process, in which those folding domains aggregate to form a compact structure, as suggested by the two-stage model. In this case, the heterogeneity in the interactions has to be included in our model.

The author is grateful to Professor C.-K. Hu for providing his computing facility (the Pentinum II PC cluster). We thank C. Tang, Y.-C. Sun, and C.-K. Hu for stimulating discussions. This work is supported, in part, by the National Science Council of Taiwan under Grant No. NSC 88-2112-M-003-012.

[1] H. Li, R. Helling, C. Tang, and N. Wingreen, Science **273**, 666 (1996); J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes, Annu. Rev. Phys. Chem. **48**, 545 (1997); H. S. Dill, and K. A. Dill, Annu. Rev. Biophys. Biomol. Struct. **26**, 425 (1997); Y. Duan and P. A. Kollman, Science **282**, 740 (1998).

[2] J. D. Bryngelson and P. G. Wolynes, J. Phys. Chem. **93**, 6902 (1989); A. M. Gutin, V. I. Abkevich, and E. I. Shakhnovich, Phys. Rev. Lett. **77**, 5433 (1996).

[3] P. E. Leopold, M. Montal, and J. N. Onuchic, Proc. Natl. Acad. Sci. U.S.A. **89**, 8721 (1992); P. G. Wolynes, J. N. Onuchic, and D. Thirumalai, Science **267**, (1995).

[4] For a review, see S. H. White and W. C. Wimley, Annu. Rev. Biophys. Biomol. Struct. **28**, 319 (1999).

[5] J.-L. Popot and D. M. Engelman, Biochemistry **29**, 4031 (1990).

[6] M. Milik and J. Skolnick, Proc. Natl. Acad. Sci. U.S.A. **89**, 9391 (1992).

[7] For a review, see T. A. Rapoport, B. Jungnickel, and U. Kutay, Annu. Rev. Biochem. **65**, 271 (1996).

[8] K. E. S. Matlack, W. Mothes, and T. A. Rapoport, Cell **92**, 381 (1998).

[9] I. Carmesin and K. Kremer, Macromolecules **21**, 2819 (1988).

[10] C.-M. Chen and P. G. Higgs, J. Chem. Phys. **108**, 4305 (1998).

[11] D. M. Engelman and T. A. Steitz, Cell **23**, 411 (1981).

[12] S. Arumugam *et al.*, Proc. Natl. Acad. Sci. U.S.A. **93**, 5872 (1996).

[13] C. J. Camacho and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **90**, 6369 (1993); D. Thirumalai and Z. Guo, Biopolym. Res. Commun. **35**, 137 (1995); D. Thirumalai, J. Phys. I (France) **5**, 1457 (1995).

[14] H. Monoi *et al.*, Biophys. J. **78**, 2892 (2000).

[15] C. M. Deber and S.-C. Li, Biopolymers **37**, 295 (1995); C. M. Deber and N. K. Goto, Nat. Struct. Biol. **3**, 815 (1996).