Model for coupled insertion and folding of membrane-spanning proteins

Andrew C. Hausrath*

Department of Chemistry and Biochemistry and Program in Applied Mathematics, University of Arizona, Tucson, Arizona 85721, USA

(Received 3 January 2014; revised manuscript received 6 March 2014; published 11 August 2014)

Current understanding of the forces directing the folding of integral membrane proteins is very limited compared to the detailed picture available for water-soluble proteins. While mechanistic studies of the folding process *in vitro* have been conducted for only a small number of membrane proteins, the available evidence indicates that their folding process is thermodynamically driven like that of soluble proteins. *In vivo*, however, the majority of integral membrane proteins are installed in membranes by dedicated machinery, suggesting that the cellular systems may act to facilitate and regulate the spontaneous physical process of folding. Both the *in vitro* folding process and the *in vivo* pathway must navigate an energy landscape dominated by the energetically favorable burial of hydrophobic segments in the membrane interior and the opposition to folding due to the need for passage of polar segments across the membrane. This manuscript describes a simple, exactly solvable model which incorporates these essential features of membrane protein folding. The model is used to compare the folding time under conditions which depict both the *in vitro* and *in vivo* pathways. It is proposed that the cellular complexes responsible for insertion of membrane proteins act by lowering the energy barrier for passage of polar regions through the membrane, thereby allowing the chain to more rapidly achieve the folded state.

DOI: 10.1103/PhysRevE.90.022707 PACS number(s): 87.14.ep, 87.14.et, 87.15.kt, 61.41.+e

I. INTRODUCTION

Water-soluble proteins take on a great diversity of three-dimensional structures [1] whereas the structural repertoire of integral membrane proteins consists of only two observed architectures. Helical bundles are comprised of membrane-spanning α helices, and β barrels contain a single closed β sheet in which successive strands cross the membrane, often forming an aqueous channel through the membrane on the interior of the sheet [2]. Individual proteins may have different numbers of transmembrane segments, with α helical proteins commonly composed of between 1 and 15 membrane-spanning segments, and β barrels observed to date having between 8 and 24 transmembrane β strands [3–6]. The availability of complete genetic sequences for many organisms reveals that roughly one-third of proteins are associated with membranes [7], so progress in understanding this class of proteins would have broad implications in biology.

Within cells the two types are trafficked to distinct membranes: outer membranes of bacteria contain primarily β barrels, and these are also found in chloroplast and mitochondrial membranes. The proteins in bacterial inner membranes and plasma membranes from eukarotic cells are primarily α helical [2]. Two primary cellular systems are responsible for insertion of the majority of membrane proteins, although examples exist that do not rely on these systems [8]. Helical membrane proteins are inserted by a multiprotein complex termed the translocon, and β barrels are inserted by the *bam* complex [9,10]. The central component of the translocon is the secY protein, an α helical bundle containing the protein-conducting pore [9]. The central component of the *bam* complex is the *bamA* protein, which forms a 16-stranded β barrel [11]. The mechanisms by which these complexes facilitate insertion are still under active investigation.

A central tenet in protein biophysics is the thermodynamic hypothesis, which states that a protein's amino acid sequence encodes both the information necessary to specify its structure and the folding pathway by which that structure is achieved [12,13]. This applies to membrane proteins (see evidence reviewed in Ref. [8]) so any proposed mechanism of these insertion complexes is constrained by the physical properties of the folding chain. Measurement of folding energetics and rates of membrane proteins is technically challenging, but ongoing efforts continue to add to the available information [2,8,14–17]. Because the folding process takes place in a membrane environment, the folding process is influenced by both interactions of the chain with the membrane lipids and with interactions of the chain with itself. The relative contributions of these factors have been the subject of study by diverse computational approaches including atomistic and coarse-grained molecular dynamics simulations of proteins or peptides in membranes, continuum models of the transmembrane segments or the membrane itself, and studies of lattice models [18–28]. Despite this body of work, in comparison to the present mature understanding of the folding process in water-soluble proteins, the physical properties and folding mechanisms of membrane proteins are poorly understood.

However, the comparatively simple structures of membrane proteins suggest some general attributes of their folding mechanisms. The sequences of membrane proteins can be readily identified by the presence of successive regions enriched alternately in hydrophobic and polar amino acids [5,29–31]. In the folded, functional states of membrane proteins, the hydrophobic regions span the hydrocarbon core of the membrane, and the polar loops protrude into solvent on alternate sides of the membrane. This energetically favorable embedding of the hydrophobic segments into the membrane can be achieved only by the energetically unfavorable process of passage of the polar regions through the membrane core. This work provides a simple model to investigate the competition between these conflicting attributes of the membrane protein folding process.

*hausrath@email.arizona.edu

II. EQUILIBRIUM MODEL

To examine the consequences of this competition, consider a simplified representation of an integral membrane protein as a coblock polymer with N identical hydrophobic segments, interspersed with polar segments. The chain achieves its lowest energy state when all hydrophobic segments traverse the core of the membrane. An ensemble of partially inserted states exists in which some but not all of the segments are buried, leaving some hydrophobic segments looped out from the core of the membrane (Fig. 1) but still associated with its periphery.

The properties of such a system may be described with an exactly solvable statistical thermodynamic model. The model postulates four states of each individual segment t,d,u, and b, for top, down, up, and bottom, respectively. A string of N symbols specifies each configuration of the chain, but not all possible combinations occur. The connectivity of the chain requires that a segment traversing the bilayer from the bottom to the top must be followed by either a segment that goes down or a segment that remains on the top. In the model, t or d can follow u or t, and b or u follow d or b. For example, a chain consisting of four segments buried in the membrane and with its N-terminus below the membrane can be represented $udud$, whereas the state $tbtb$ is prohibited. The symbols t, d, u, and b serve as multiplicative Boltzmann factors to assign the relative statistical weight for a given conformation. Each conformation for a chain with N segments will have a weight represented by the product of powers of the individual segment weights t, d, u, and b, with the sum of the powers being N. Summing over the statistical weights for these states yields the partition function Z_1, which can be expressed as the matrix product

$$Z_1 = \mathbf{e}\mathbf{M}^N\mathbf{e}^T, \quad M = \begin{pmatrix} t & d \\ u & b \end{pmatrix}, \quad \mathbf{e} = (1 \;\; 1), \tag{1}$$

where M is a transfer matrix and the end vectors $\mathbf{e}$ and $\mathbf{e}^T$ accomplish the sum over all starting and ending configurations. The rank 2 matrices reflect the fact that each segment other than the first can take only two states, resulting in a sum over 2^{N+1} states. The eigenvalues of M are $\lambda_i = ((t+b) \pm \sqrt{(t+b)^2 - 4(tb-ud)})/2$ with corresponding eigenvectors

$$v_i = \left(\frac{(t-b) \mp \sqrt{(t+b)^2 - 4(tb-ud)}}{2u} 1 \right). \tag{2}$$

Defining $\mathbf{E}$ as the matrix of eigenvectors $\{v_1 \;\; v_2\}$ the partition function can be written

$$Z_1 = \mathbf{e}\mathbf{E}(\mathbf{E}^{-1}\mathbf{M}\mathbf{E})^N\mathbf{E}^{-1}\mathbf{e}^T = \mathbf{e}\mathbf{E}\begin{pmatrix} \lambda_1^N & 0 \\ 0 & \lambda_2^N \end{pmatrix}\mathbf{E}^{-1}\mathbf{e}^T. \tag{3}$$

It remains to specify the particular form of the Boltzmann factors t, d, u, and b, which will allow the population of each state to be obtained. The simplest model postulates a free energy difference B between hydrophobic segments which are buried in the membrane versus those which are not. If we take the peripherally associated t and b states as the zero of energy, the relative weight of the u and d states becomes e^{-B/k_BT}, and a state i has energy $E_i = n_i B$ where n_i is the number of u or d segments. The t and b states are each assigned a relative weight of unity. The eigenvalues λ_i and matrix of eigenvectors $\mathbf{E}$ become

$$\lambda_i = 1 \mp e^{-B\beta} \quad \text{and} \quad \mathbf{E} = \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}, \tag{4}$$

where $\beta = 1/k_BT$. Substituting these expressions into Eq. (3) the partition function for the ensemble of partially inserted states is expressed as

$$Z_1 = 2(1 + e^{-B\beta})^N. \tag{5}$$

The model postulates an additional folding step between the fully embedded configuration and the mature, active state.

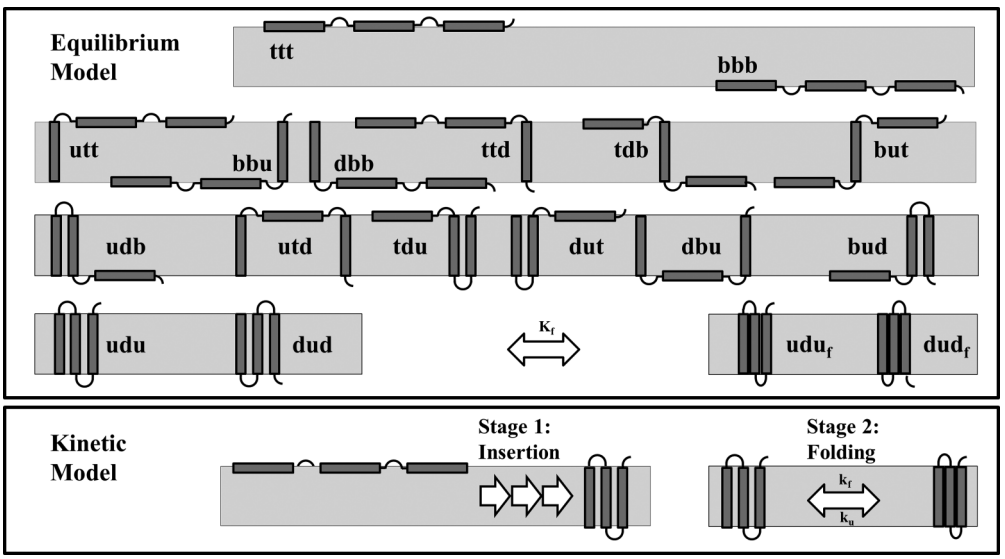

FIG. 1. Top: Schematic depiction of equilibrium states of a three-segment chain. The gray boxes representing the membrane separate the states with zero to three buried segments, and the folded states. Below: Schematic depiction of kinetic model, in which the boxes separate the insertion and folding processes. Thick rectangles represent hydrophobic segments, and the connecting thin lines represent the intervening polar segments.

There are two states with N transmembrane segments in the ensemble of membrane-associated states, with Boltzmann weight $e^{-NB\beta}$, and these fold to two final states, which have opposite orientations in the membrane. Let the constant K_f describe this equilibrium between states fully embedded in the membrane and the folded state in the same orientation. The partition function for these folded states is $Z_2 = 2K_f e^{-NB\beta}$, so that Z_1 describes interactions between the protein and the membrane, and Z_2 describes self-interactions of the chain. The partition function Z_{12} represents the entire ensemble:

$$Z_{12} = Z_1 + Z_2 = 2(1 + e^{-B\beta})^N + 2K_f e^{-NB\beta}. \quad (6)$$

The fraction of molecules which are in the folded or unfolded states are $f_f = Z_2/Z_{12}$ and $f_u = Z_1/Z_{12}$, and the probability of any given state is $P_i = e^{-n_i B\beta}/Z_{12}$ (see Fig. 2). The B parameter governs the equilibrium properties of the model.

III. KINETIC MODEL

The temporal behavior of the system may be understood by assigning transition rates between the states described above in the equilibrium model. These transition rates are governed by an Arrhenius rate law with an activation energy with two components, the relative energy of the initial and final state, and the energy barrier associated with passage of polar segments through the membrane. The energy B associated with burial of a single hydrophobic segment is the same as described above for the equilibrium model. An additional energy A is contributed to the activation energy for each polar linker that must pass through the bilayer. The equilibrium ratio of two states is controlled by a Boltzmann factor

$$\mathcal{B}_{ij} = \frac{P_i}{P_j} = \frac{e^{-(n_i B)\beta}}{e^{-(n_j B)\beta}} = e^{-(n_i - n_j)B\beta}, \quad (7)$$

where n_i and n_j are the number of buried segments in states i and j, and B is the energy associated with burial of a single hydrophobic segment in the bilayer. The other component of the transition rate constant is the Arrhenius factor

$$\mathcal{A}_{ij} = e^{-(m_{ij}A)\beta}, \quad (8)$$

where m_{ij} is the number of polar segments which must pass from one side of the membrane to the other in order to make a transition between the states i and j. The overall rate constant for a transition between states i and j is the product

$$k_{ij} = \kappa \begin{cases} \mathcal{A}_{ij}\mathcal{B}_{ij} & \text{if} \quad E_j \geqslant E_i \\ \mathcal{A}_{ij} & \text{if} \quad E_j < E_i \end{cases}. \quad (9)$$

In reactions where molecules must collide in order to react the preexponential factor κ is used to assign the fraction of collisions in which the molecules are in a productive orientation to enable them to react. In this model the transitions are conformational rearrangements of a single chain, so orientation effects are not applicable. Here all molecules which have sufficient energy are assumed able to surmount the activation barrier and enter state j. Therefore κ is assigned the value 1 and is assumed to be the same for all transitions.

As an example, the transition directly between states *dut* and *dbb* is illustrated in Fig. 3. An energy B is associated with every segment that is buried in the bilayer, so the two states differ by an energy of $-B$. Two hydrophilic segments must pass though the bilayer during this transition, contributing $2A$

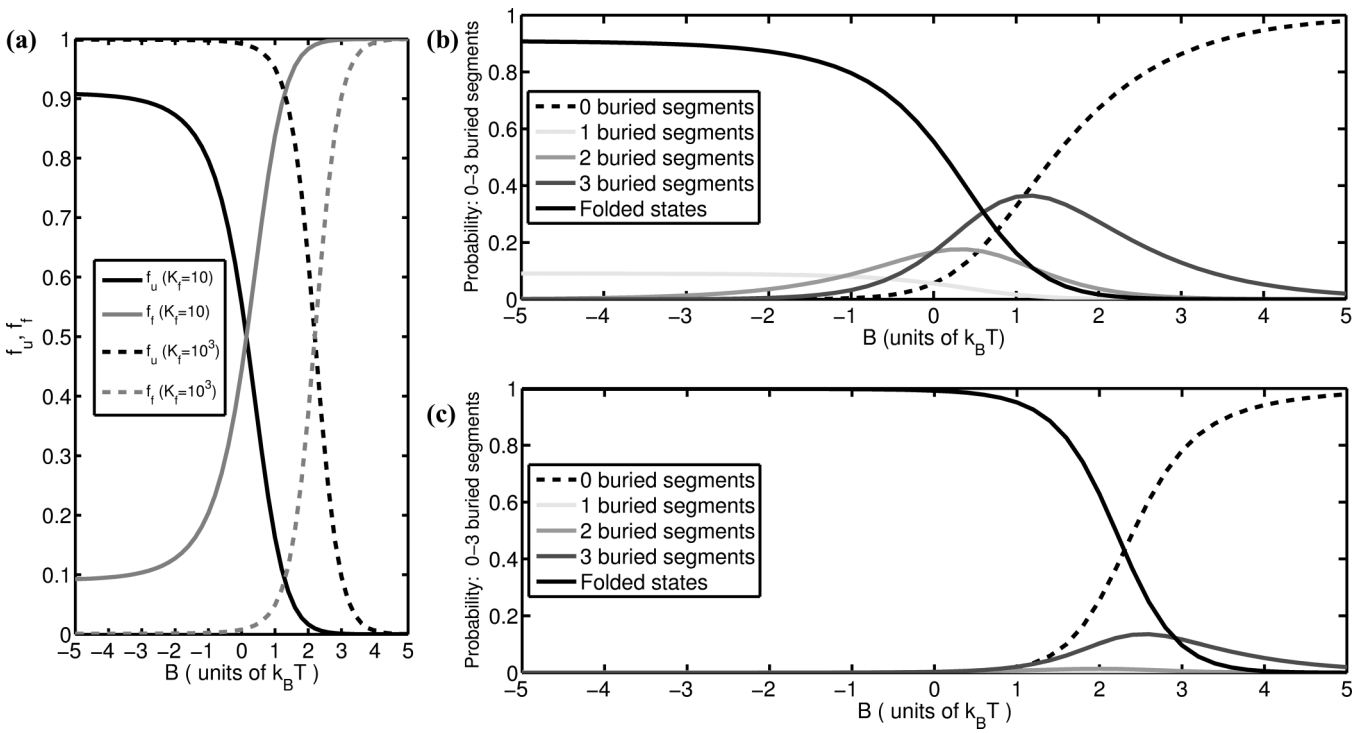

FIG. 2. (a) Dependence on segment burial energy B of f_u (gray) and f_f (black) for $K_f = 10$ (solid lines) and $K_f = 10^3$ (dashed lines.) (b, c) Equilibrium populations of states with 0-3 buried segments when $K_f = 10$ and $K_f = 10^3$. Partially inserted states accumulate to different extents depending on both the membrane and self-interactions of the chain.

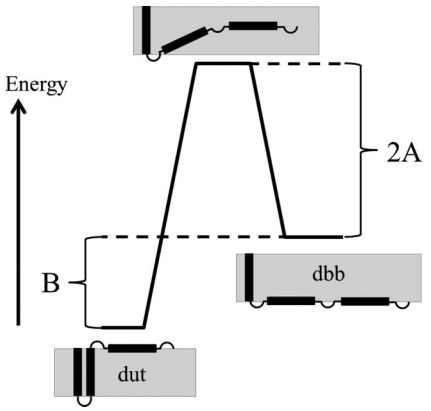

FIG. 3. Assignment of transition rate. The activation energy for a single transition $dut \rightarrow dbb$ has components from transient passage of two polar segments through the membrane and the transfer of one buried segment from the membrane core to the periphery.

to the activation energy. The transition $dut \rightarrow dbb$ will occur at a rate proportional to $e^{-(2A-B)\beta}$: the transition increases the state energy, and the relative energy of the states contributes to the rate of the transition. The transition $dbb \rightarrow dut$ will occur at a rate proportional to $e^{-2A\beta}$, as this transition decreases the energy, and so the energy difference does not contribute to the activation energy. The ratio of the two rate constants gives the equilibrium ratio between the two states, as required for detailed balance.

In addition, the model assumes that transitions to the fully folded state occur only from (and to) the fully membrane-embedded states. The forward rate constant for this step is denoted k_f and the rate constant for unfolding $k_u = k_f/K_f$. The rate constants k_{ij}, k_f, and k_u define the kinetics of the system. Therefore there are two primary time scales in this model. The first corresponds to the passage of polar segments through the membrane, and the other corresponds to the folding process involving association of the inserted segments.

To formulate the dynamics, let vector $\mathbf{P}(t)$ describe the populations of the states, and define state $\mathbf{P}_1$ as the all-t state, $\mathbf{P}_2$ as the all-b state, $\mathbf{P}_{2^{N+1}-1}$ as the fully buried state starting with u (i.e., udu for a three-segment chain), and $\mathbf{P}_{2^{N+1}}$ as the fully buried state starting with d. Finally, state $\mathbf{P}_{2^{N+1}+1}$ and $\mathbf{P}_{2^{N+1}+2}$ are the folded states starting with d and u, respectively. This relaxation process is compactly expressed in the expressions below, where $\mathbf{K}$ contains rate constants defining the kinetics of insertion and $\mathbf{F}$ contains rate constants defining the kinetics of folding:

$$\frac{d\mathbf{P}}{dt} = (\mathbf{K} + \mathbf{F})\,\mathbf{T}, \tag{10}$$

where

$$\mathbf{K}_{ij} = \begin{cases} k_{ij} & \text{if} \quad i \neq j \\ -\sum_{l=1}^{N} k_{il} & \text{if} \quad i = j \\ 0 & i > 2^{N+1} \quad or \quad j > 2^{N+1} \end{cases} \tag{11}$$

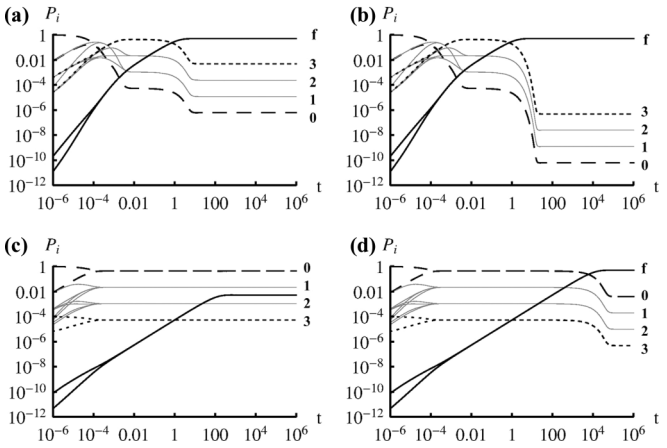

FIG. 4. Time courses of insertion and association for a three-segment chain. Each panel represents the populations P_i in a relaxation process in which the initial state is ttt. The traces for folded states udu_f and dud_f (indicated by the "f") are drawn with black solid lines. States udu and dud (three transmembrane segments) are indicated with black dashed lines, and traces for states ttt and bbb (zero transmembrane segments) are indicated with black dotted lines. States with one or two transmembrane segments are drawn wth gray lines. (a, b) parameters are indicated in units of β. Rate constants k_f are set to 10^3 s^{-1}, and K_f is unitless. (a) Time course for a low-stability folded state with favorable burial energy($A = 3$, $B = -3$, $K_f = 10^2$). (b) A high-stability folded state with favorable burial energy ($A = 3$, $B = -3$, $K_f = 10^6$). (c) A low-stability folded state with unfavorable burial energy ($A = 3$, $B = 3$, $K_f = 10^2$). (d) A high-stability folded state with unfavorable burial energy ($A = 3$, $B = 3$, $K_f = 10^6$).

and

$$\mathbf{F}_{ij} = \begin{cases} -k_f & \text{if} \quad i,j = 2^{N+1} - 1 \\ -k_f & \text{if} \quad i,j = 2^{N+1} \\ -k_u & \text{if} \quad i,j = 2^{N+1} + 1 \\ -k_u & \text{if} \quad i,j = 2^{N+1} + 2 \\ k_u & \text{if} \quad i = 2^{N+1} - 1 \quad \text{and} \quad j = 2^{N+1} + 1. \\ k_u & \text{if} \quad i = 2^{N+1} \quad \text{and} \quad j = 2^{N+1} + 2 \\ k_f & \text{if} \quad i = 2^{N+1} + 1 \quad \text{and} \quad i = 2^{N+1} - 1 \\ k_f & \text{if} \quad i = 2^{N+1} + 2 \quad \text{and} \quad i = 2^{N+1} \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

The rate constants are formulated to obey detailed balance, so the kinetics is a relaxation process and in the limit of long times will attain the distribution described by the equilibrium model. With fixed choice of model described by the four model parameters A, B, k_f, and K_f, and given an initial condition $\mathbf{P}(0)$, this system may be solved to obtain the state populations $\mathbf{P}(t)$ during subsequent times t. Figure 4 displays time courses of the relaxation process for a three-segment chain obtained by numerical integration.

IV. DISCUSSION

Proteins are heteropolymers of amino acids and carry out the majority of biological processes in living organisms. Most

processes are dependent on these molecules occupying a specific three-dimensional configuration. The particular sequence of monomers contains both the information necessary to specify this folded state and the information to specify the path by which it is attained [12,13]. With few documented exceptions (e.g., Ref. [32]) the folded, functional states are at equilibrium, and so proteins are often described as self-organizing polymers. The primary driving force for folding is the hydrophobic effect, the partitioning of hydrophobic amino-acid side chains away from aqueous solvent to the hydrophobic volume at the interior of the folded configuration (hydrophobic collapse), leaving the amino acids with polar side chains exposed to the solvent on the exterior [33]. However, an additional strong constraint is that buried hydrogen bonding groups along the chain must have interaction partners [34]. These properties characterize the folded structures of proteins, but there are myriad ways in which these fundamental requirements can be satisfied.

For soluble proteins, current understanding of the energetics of the folded state [35] and the rates and pathways of the folding process [36–38] is well advanced. However, while structures of the folded states are known in great detail for many proteins, it is important to note that the properties of proteins cannot be attributed solely to the folded state, but also depend on the ensemble of unfolded states [39]. However, many properties can be understood using exactly solvable models which utilize simplified representations, but thereby enable a complete treatment of the entire conformational ensemble [40–42].

It is well documented that development of a comparable body of folding data for membrane proteins has been hindered by the technical difficulties associated with their study [8]. However, a fruitful approach has been to make use of short peptides consisting of individual transmembrane segments of membrane proteins (e.g., Refs. [43,44]) or synthetic peptides (e.g., Ref. [45].) These simple models are typically more experimentally tractable than natural proteins, and have provided measures of energies which are difficult to make in intact natural proteins. Similar physical forces are operating in small peptides and in proteins, but there is an additional complexity in the folding kinetics of the latter imposed by the connectivity of chains with multiple membrane-spanning segments. One motivation for this work is to understand the influence on the behavior of an extended chain due to its connectivity, in terms of the properties of its constituent segments.

A. Model parameters

The polypeptide backbone is intrinsically polar, and where the chain passes through the membrane, its hydrogen bonding groups must be satisfied by self-interactions with other parts of the chain. In membrane-spanning α helical bundles, this is accomplished by interactions within each transmembrane segment, as the helical configuration of the chain forms hydrogen bonds between residues separated by four residues. In transmembrane β barrels, the hydrogen bonding potential of the chain is satisfied by interactions with distinct segments. Thus the model parameter B, describing the free energy of burial of a segment, may be considered the resultant of two competing effects: the favorable transfer of hydrophobic side chains into the hydrocarbon core of the membrane versus the unfavorable transfer of the polar backbone. For α helical membrane proteins, the classical view is that the individual transmembrane helices are independently stable within membranes [43,44,46] although there are exceptions [47]. Transmembrane β strands would not be be expected to form in isolation. As a simplifying assumption, in this work it is assumed that for transmembrane α helices, burial is favorable, and these segments may be described with $B < 0$. For transmembrane β strands, the opposite is the case, and for such segments, $B > 0$.

Estimates of values of B for hydrophobic helices were obtained with an ingenious assay based on the detection of glycosylation patterns [48], yielding estimates with an upper limit on the order of 3 kcal/mol. This value corresponds to $B \sim -5k_BT$ at 298 K. Although the energetics of strand insertion is not directly measurable, an estimate of the energy for a transmembrane β strand can be obtained from studies which determined the transfer energy from an aqueous environment to the membrane interior [45]. For an isolated strand, an estimate of B may be obtained as the sum of group transfer energies, yielding $B \sim 3k_BT$. Values for A may be estimated similarly to yield $A \sim 3$–$5k_BT$.

The parameter K_f represents the association equilibrium between transmembrane segments after complete insertion, and k_f represents the forward rate constant describing this process. These model parameters are intertwined with both the folding equilibrium and the folding kinetics. Although for modeling purposes it is convenient to separate the distinct stages, distinguishing the energetics and rates of these within experimental data requires careful consideration.

B. Equilibrium properties of the ensemble of states

The stability of a protein conventionally refers to the difference in free energy between its folded state and unfolded state. Where known, measured stabilities of helical bundle and β barrel membrane proteins are comparable to values observed for globular proteins (1.5–10 kcal/mol [15,49], corresponding to a folding equilibrium constant of ~ 10–10^8). Although in a two-state folding transition, the folded and unfolded states are considered to be of fixed energies, a more realistic interpretation is to consider these as ensembles. Thus the energy spectrum commonly attributed to globular proteins includes a folded state ensemble consisting of a single or a small number of states with a well-defined energy, and an unfolded state ensemble containing a large number of states with a range of energies. Here the ensemble of partially inserted states described by Z_1 may be compared to the unfolded state ensemble, and the ensemble of folded states may compared to Z_2. The spectrum of energies in the membrane protein model therefore has an energy distribution similar to that of globular proteins, albeit for different reasons, and will have thermodynamic properties that reflect that similarity.

Comparison of stability measurements between different membrane proteins is complicated by the diverse conditions found necessary to obtain reversible folding [8,50]. Depending on the experimental protocol employed the measured energy difference may include contributions from partitioning between solvent and the membrane, conformational transitions

on the membrane surface, insertion into the membrane, and association of transmembrane segments. Note that in the model, the parameter K_f describes the equilibrium between the fully buried states and the folded state, and does not include the energy associated with protein-lipid interactions. In several experimental protocols for measurement of the stability of membrane proteins, the unfolded reference state is obtained by chemical denaturation of protein, and folding into the membrane is induced by dilution of denaturant. Therefore the transition measured is between states with no segments embedded in the membrane, and the fully folded state. In this situation the stability (as described by the model) will be the sum of the energy associated with transmembrane segment burial and the additional energy of folding:

$$\Delta G_f = \beta \ln K_f + NB. \tag{13}$$

For molecules with a favorable segment burial energy, where $B < 0$, the measured stability ΔG_f represents the maximum energy difference between a member of the folded state and a member of the ensemble of partially inserted states. (This would typically be case for an α helical membrane protein, where it is often found that the individual transmembrane helices can fold and remain stably integrated in isolation [43,44,46].) In this case the embedding of segments into the membrane contributes to stability. For comparison, if $B > 0$ (as expected for β barrel membrane proteins), then the embedding of individual segments into the membrane does not contribute to the stability as defined above. The ΔG_f in this case represents the minimum energy difference between a member of the folded state and a member of the ensemble of partially inserted states. The fact that the ΔG_f has different interpretations for the two classes of proteins is a somewhat unsatisfactory situation, underscoring the difficulty of comparing the stabilities due to the differences in properties of their unfolded states.

In the case of globular proteins, the tendency of the chain to select a specific folded state (e.g., the biologically functional state) is governed by the energy gap Δ between this folded state, and the member of the unfolded state ensemble *nearest* in energy [40,51]. For globular proteins, increasing this "energy gap" may contribute more effectively to the biological activity of a protein than increasing the overall stability, as the activity of a molecule with a small gap is subject to interruption due to small fluctuations in energy, whereas large fluctuations in energy that can globally unfold a protein (and also interrupt activity) are rare. The relationship between ΔG_f and Δ frames the distinction between overall stability and structural specificity. Proteins with a very small value of Δ may easily access partially folded states and display the characteristics of molten globules. The membrane protein model provides an explicit representation of the ensemble of partially inserted states, which is not typically available for globular proteins. Therefore a structural rationale for this energy gap may be described for the membrane protein model, although in general its basis remains elusive.

Within the membrane protein model, the parameter K_f represents the equilibrium between the fully inserted state and the folded state, and the associated energy difference has magnitude $\beta \ln K_f$ (see Fig. 5). For α helical proteins, where $B < 0$, the states nearest the folded states in energy

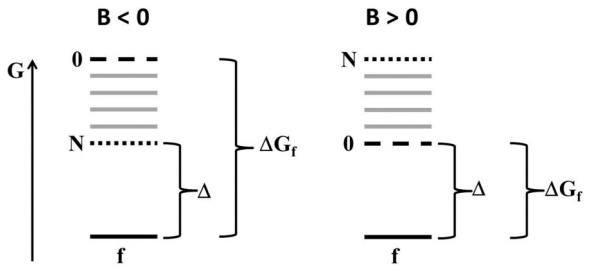

FIG. 5. Schematic depiction of the energy spectrum of the model for $B < 0$ (left), and $B > 0$ (right). The energy of the folded state is denoted f. The energy levels of the partially inserted states are depicted with dotted lines for the fully inserted states (indicated with N), with dashed lines for the states with no inserted segments (indicated with 0), and in gray for those in between. The stability ΔG_f is the energy difference between the states with no inserted segments and the folded states. The energy gap Δ is the difference in energy between the lowest energy state in the ensemble of partially inserted states and the folded state.

are the fully inserted states. In this case, the energy gap $\Delta = \beta \ln K_f$. Because much of the overall stability of α helical proteins derives from the insertion of the individual helices, the ratio $\Delta/\Delta G_f$ may be relatively small. At equilibrium, some molecules will tend to occupy these non-native states, as they are close in energy to the folded state. Indeed, many α helical membrane proteins are found to be relatively mobile within the membrane, and the existence of alternative conformations with the membrane likely contributes to their biological functions [52].

In contrast, for β barrel proteins, where $B > 0$, the states nearest to the folded state in energy are those with no buried segments. In this case, the energy gap Δ coincides with the definition of ΔG_f, and to access the state closest in energy to the folded state, the chain must not only fully unfold, but also cross the membrane. The β barrel architecture represents an extreme case where the stabilization accorded by transmembrane segment association must fully overcome the unfavorable energetics of insertion.

The quantity Δ defines the energy required for a fluctuation to access the nearest non-native state. The mean-square fluctuation in energy may be obtained from the partition function using the standard relation

$$\langle \Delta E^2 \rangle = \frac{\partial^2}{\partial \beta^2} \ln Z_{12}, \tag{14}$$

which is illustrated in Fig. 6. For very stable proteins where K_f is large, the fluctuations are much less than Δ, but fluctuations become more significant for marginally stable proteins. To illustrate this effect, Fig. 6 shows fluctuations for the case $K_f = 10^1$, for chains with different numbers of segments. The magnitude of the fluctuations is small for large positive or negative values of B and has two peaks at $B \sim 3\beta$ and $B \sim -1\beta$, with a minimum near $\beta = 0$. This behavior can be rationalized because for large values of B, the separation between energy levels corresponding to partially buried states is large, and the fluctuations do not approach Δ. At very small values of B, the separation in energy levels between partially buried states is insignificant compared to the energy difference between the folded and unfolded state, so again,

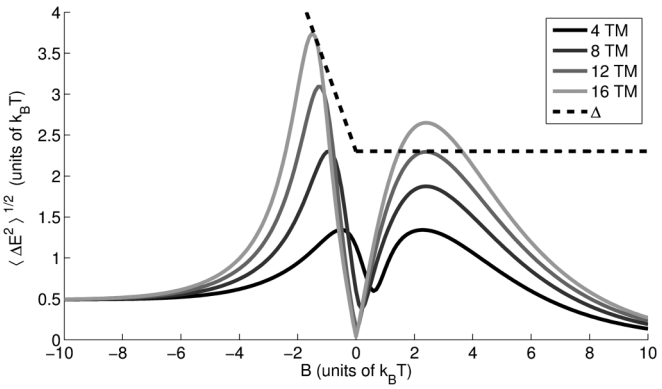

FIG. 6. Energy fluctuations as a function of burial energy B. The quantity Δ, representing the gap between the folded state and the nearest higher-energy state, is depicted with the dashed line. The region $B > 0$ corresponds to the case where the individual segment burial is unfavorable (β barrels) and the region $B < 0$ corresponds to the case where segment burial is favorable (α helical bundles.) The mean-square fluctuations in energy for a chain with 4, 8, 12, and 16 transmembrane segments are shown in shades of gray. The value $K_f = 10^1$ for these traces.

the fluctuations do not approach Δ. However, for intermediate values of B, both positive and negative, the fluctuations are of similar magnitude to Δ, such that the low-lying unfolded states are accessed by the chain. The graph also shows that for this system $\langle \Delta E^2 \rangle$ increases with the number of segments.

C. Folding kinetics schemes

Now, consider the folding process by which the equilibrium distribution is achieved. The two-stage model of Popot and Engelman [53] suggested the separation of the insertion and assembly stages, and the folding kinetics described in the model follows these two stages. The first stage, membrane insertion, is governed by the A parameter which represents the energy of a transient, high-energy state with a single polar segment passing through the membrane. However, the kinetic model does not assume a fixed pathway and considers parallel pathways in which distinct steps result in different numbers of segments crossing the membrane. As there are many possible pathways for folding, it is useful to consider the folding process as traversing an energy landscape, schematically depicted in Fig. 7. The energies of the distinct states are defined by the number of inserted segments, and so these states differ by the burial energy B, resulting in a funnel-like landscape. The energy barriers separating states which can be interconverted by passage of a single polar segment through the membrane will have height A. But transitions involving passage of multiple polar segments also occur, and these will have barrier heights corresponding to multiples of A, resulting in the presence of multiple time scales in the relaxation process.

The second stage of the two-stage model is the association of transmembrane segments. The kinetics of the second stage are governed by the k_f and K_F (or equivalently, k_u) parameters. When k_f is slower than the insertion steps, the unfolded but fully embedded configurations will accumulate, and thus insertion will temporally precede folding. When k_f is faster, folding will proceed as the inserted states are produced,

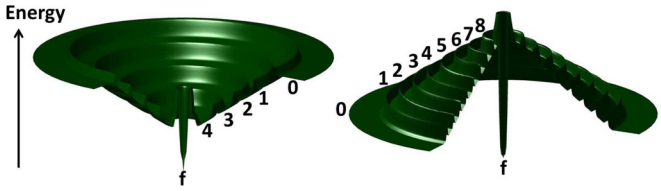

FIG. 7. (Color online) Schematic depiction of energy landscapes for the kinetic model. The outer rim corresponds to the high-energy states with no inserted segments (labeled 0), and the folded state (f) at the center has the lowest energy. The intervening levels represent states with intermediate numbers of inserted segments, with each level differing by the insertion energy B. The barriers between each level represent the activation energy A. Left: Landscape for α helical bundle with four segments, with $B < 0$. The insertion of successive segments results in successively lower energy states. Right: Landscape for β barrel with eight segments, with $B > 0$.

so there is little accumulation of the inserted intermediates. *In vitro* studies of membrane protein folding kinetics find that in most cases the unassisted folding is a very slow process, requiring hours [54,55]. When insertion and folding can be kinetically resolved, the spontaneous insertion of transmembrane segments into the membrane is quite slow (e.g., Ref. [56]). However, the association of segments once inserted into the membrane is not itself an intrinsically fast process either [57,58]. Moreover, the observed rates and energies of insertion and folding are strongly dependent on the composition of the lipids used in these studies [59,60]. Therefore, while the assumption that insertion precedes folding is a useful simplifying principle for modeling purposes, the actual kinetics of each stage must be considered on a case-by-case basis. Although individual proteins appear to vary widely in their properties, the marked differences between α helical and β barrel proteins suggest some simplifying generalizations.

1. *Folding pathways for α helical bundles*

The transitions with the largest rate constants are those which involve only a single polar segment traversing the membrane, thereby embedding two adjacent hydrophobic segments. Assuming $B < 0$ for α helical proteins, this is an energetically favorable transition. Therefore the most likely pathway for α helical bundles will involve successive transitions of this type. This folding mechanism was originally proposed by Engelmann and Steitz as the *helical hairpin hypothesis* [61]. Described within the model, insertion of a hairpin requires passage of a single polar segment through the membrane and buries two hydrophobic segments. Because each insertion event results in a lower energy state, the result is a downhill folding reaction with barriers of height A, and with each successive intermediate state lower in energy by $2B$ [see Fig. 11(a)]. The forward rate constant for insertion (k_{ins}) or extraction (k_{out}) of a single hairpin will be

$$k_{\text{ins}} = e^{-A\beta} \quad k_{\text{out}} = e^{-(A-2B)\beta}. \tag{15}$$

Stepwise pathways, consisting of successive insertions of adjacent pairs of segments, do occur within the kinetic model as a limiting case among the many possible folding pathways. As a specific example, a stepwise insertion pathway for the

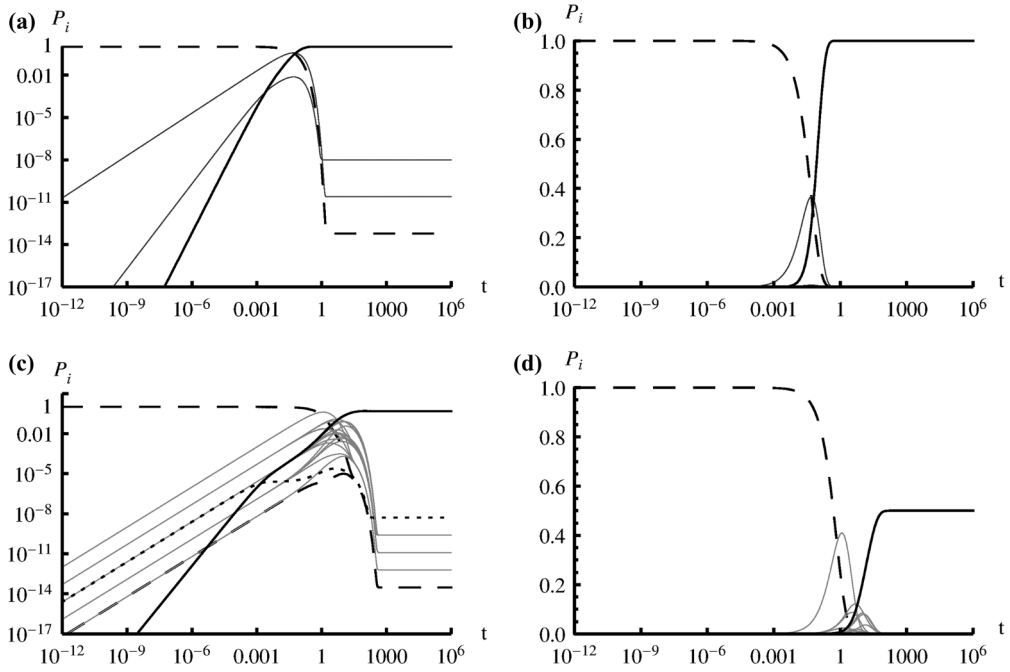

FIG. 8. Comparison of timecourses from a stepwise insertion mechanism with the kinetic model, for a four-segment chain. (a, b) Populations of the states in the stepwise mechanism are $tttt$ (dashed line), $ttud$ and $udud$ (gray lines), and $udud_f$ (solid line). (c, d) Populations of the states of the full kinetic model are indicated in black, with the states with zero inserted segments indicated by dotted lines, states with one to three inserted segments with thin black lines, the fully inserted states with dashed lines, and the folded states with thick black lines. Model parameters are $A = 3$, $B = -3$, $k_f = 10^3$, and $K_f = 10^8$.

folding of a four-segment helical bundle (where the initial and unfolded state is $tttt$, and the final folded state is $udud_f$) may be written

$$tttt \underset{k_{\text{out}}}{\overset{k_{\text{ins}}}{\rightleftharpoons}} ttdu \underset{k_{\text{out}}}{\overset{k_{\text{ins}}}{\rightleftharpoons}} dudu \underset{k_u}{\overset{k_f}{\rightleftharpoons}} dudu_f. \tag{16}$$

Figure 8 shows a comparison of the time course of this single pathway to the kinetics of the whole relaxation. Despite the use of identical energies in assignment of the kinetics of the elementary steps of the two mechanisms, the stepwise pathway proceeds more quickly.

Folding kinetics studies have been carried out for only a handful of helical membrane proteins, and typically these utilize a partially folded starting state where the protein is solubilized in detergents. However, for the bacterial enzyme diacylglycerol kinase, Lorch and Booth showed it was possible to accomplish insertion and folding from fully denatured diacylglycerol kinase into lipid vesicles [56]. The interpretation of the kinetics of this process is complicated by aggregation of the unfolded state and oligomerization of the folded state, but under conditions where a process consistent with formation of monomers could be observed, an estimate for an insertion or folding rate constant with a value of 0.24 s^{-1} was obtained. This enzyme contains three transmembrane segments, so a stepwise pathway for this protein would require passage of two successive polar segments through the membrane, as described in Eq. (16). For this stepwise pathway each insertion step is assumed equivalent, and the rate-limiting step is likely to be a single insertion step described by k_{ins}. Comparing the value 0.24 s^{-1} to the estimated rate constant k_{ins} from Eq. (15), one obtains the estimate that $A \sim +1.4k_BT$. This is smaller than the value of 3–5k_BT estimated above but of a similar magnitude. Unfortunately, the data did not resolve the two processes of insertion and folding. Therefore the measured rate constant should be considered as an upper bound for the insertion rate constant.

2. *Folding pathways for β barrels*

For comparison, consider a stepwise pathway for the insertion of individual segments of a β barrel, which are described here as having transmembrane segments with $B > 0$. Successive insertion of each segment would result in a states with increasing energy, resulting in an "uphill" energy landscape for insertion (Fig. 7):

$$\begin{aligned} tttttttt &\underset{k_{\text{out}}}{\overset{k_{\text{ins}}}{\rightleftharpoons}} ttttttud \underset{k_{\text{out}}}{\overset{k_{\text{ins}}}{\rightleftharpoons}} tttudud \underset{k_{\text{out}}}{\overset{k_{\text{ins}}}{\rightleftharpoons}} ttududud \\ &\underset{k_{\text{out}}}{\overset{k_{\text{ins}}}{\rightleftharpoons}} udududud \underset{k_u}{\overset{k_f}{\rightleftharpoons}} udududud_f. \end{aligned} \tag{17}$$

Formalizing the kinetics using the scheme above allows comparison of the folding rate to other proposed mechanisms. Since this particular pathway occurs within the kinetic model, it is of interest to compare the two.

However, for some β barrel proteins, experimental evidence (see below) suggests an alternative pathway where a folding event precedes membrane insertion [15]. (This is at odds with the assumption that insertion precedes folding, as used in the kinetic model as well as the stepwise pathway above.) This

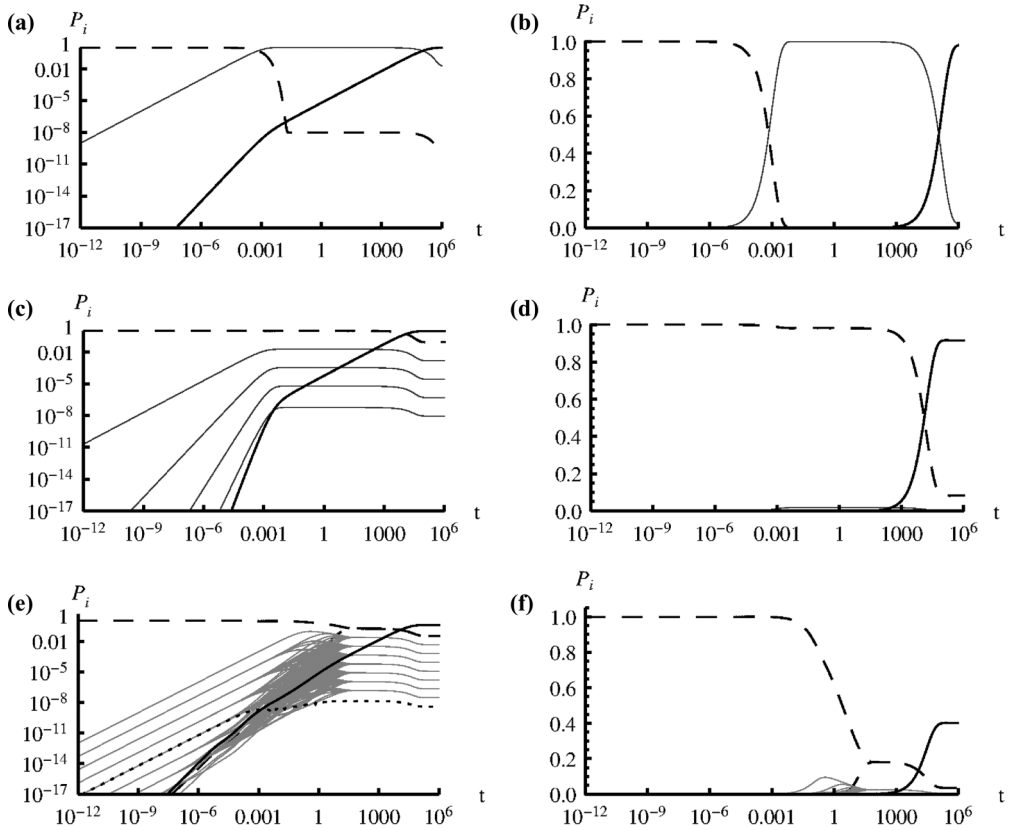

FIG. 9. Comparison of time courses from concerted and stepwise insertion mechanisms with the kinetic model, for an eight-segment chain. (a, b) Populations of the states in the concerted mechanism are *tttttttt* (dashed line), F^* (gray dashed line), and $udududud_f$ (solid line.) (c, d) For a stepwise insertion mechanism., populations of states *tttttttt* (dashed line) *ttttttdu*,*tttdudu*, and *ttdududu* (gray lines), and $dudududu_f$ (solid line.) (e, f) Populations of the states of the full kinetic model. Initial state *tttttttt* is indicated with a dashed line, partially inserted states with one to seven inserted segments are in gray, and the folded states are shown with black lines. For the concerted mechanism (a, b) $A = 3$, $k_f = 10^3$, $K_f = 10^8$, and $B = -0.5$, reflecting an overall favorable driving force for insertion if the hydrogen bonds in the strands are satisfied due to formation of the partially folded state F^*. For panels c to f, model parameters are $A = 3$, $B = 2$, $k_f = 10^3$, and $K_f = 10^8$.

concerted mechanism avoids the high-energy intermediate states with multiple transmembrane segments. However, the insertion process of a largely folded barrel necessitates simultaneous transfer through the bilayer of multiple polar segments, which is expected to be kinetically inefficient.

The concerted pathway may also be described within the present framework. To embed an intact barrel with N transmembrane segments into a membrane, $N/2$ polar segments must traverse the membrane simultaneously, resulting in a high activation barrier. In this case, the rate constants for insertion and extraction are

$$k_{\text{ins}} = e^{-NA\beta/2} \quad k_{\text{out}} = e^{-N(A/2-B)\beta}. \tag{18}$$

To define a pathway where the folding event precedes insertion, denote the folded but not inserted state as f^*, with the rate constants by which this state is achieved being k_f and k_u as above. The kinetics for the concerted pathway may be defined with the scheme

$$U \underset{k_u}{\overset{k_f}{\rightleftharpoons}} f^* \underset{k_{\text{out}}}{\overset{k_{\text{ins}}}{\rightleftharpoons}} f, \tag{19}$$

where k_{ins} is the rate constant for insertion [Eq. (18)].

As an example, consider the kinetics of folding for an eight-segment chain, representing an eight-stranded β barrel, the simplest transmembrane β barrel fold. The stepwise pathway and concerted pathway may be considered as limiting cases of the possible paths by which the protein may be inserted in the membrane. In the stepwise pathway, a series of steps with smaller activation barriers is used, whereas the concerted pathway uses a single step with a large activation barrier. Figure 9 compares the predictions of the folding process described by the concerted pathway [Eq. (19)] and the stepwise pathway [Eq. (17)], with the predictions of the kinetic model.

For the concerted pathway the initial state U is *tttttttt* and the folded state f is $dudududu_f$. The state f^* does not correspond to a state in the relaxation process, but the rate constants for its formation are k_f and $k_u = k_f/K_f$, as above. In this scheme the folding process by which f^* is formed precedes membrane insertion. The forward rate constant for insertion (k_{ins}) has a high energy barrier, resulting in a slow insertion process. If the rate of this step is slower than the rate of the folding process (i.e., if $k_f < k_{\text{ins}}$), the state f^* will accumulate [see Fig. 8(b)]. In the stepwise pathway [Fig. 8(c) and 8(d)] the individual states do not accumulate. However, as the formation of the folded state can only occur from the high-energy fully folded state, there is a persistent kinetic

phase where the concentration of the intermediates exists at an approximate steady state, during which time the folded state slowly builds up. The time course for the full kinetic model shows similar behavior following a complex earlier phase, with the net result that the kinetics of formation of the folded state is very similar to the case of the stepwise model.

An example of an eight-stranded β barrel is the protein *OmpA* from the bacterium *Escherichia coli*, which was the first transmembrane β barrel for which reversible folding conditions were established [55,62]. This protein has subsequently served as a valuable model system for study of both the thermodynamics and kinetics of β barrel membrane proteins. For this protein, the actual folding process most closely resembles the concerted pathway.

The folding process of *OmpA* may be measured by following the folding transition after dilution of unfolded protein from denaturant solutions. The kinetics of insertion in membranes and folding were resolved into three kinetic phases, described in Ref. [63], of which the latter two phases correspond to the concerted model depicted in Eq. (19). (The first phase displayed a time constant of about 15 min and corresponded to unfolded protein binding to the membrane, which is not described in the model.) The second phase, interpreted as a partial folding event with partial membrane penetration, corresponding to k_f in the concerted pathway, and had a time constant of between 15 min and 3 h. The third phase was assigned to the insertion of the barrel into the membrane and displayed a time constant of about 2 h, or a rate constant of about $1.5 \times 10^{-4}\ \mathrm{s}^{-1}$. The rate constant k_{ins} is most appropriate for comparison to this value. For a barrel of eight strands, using $k_{\mathrm{ins}} = 1.5 \times 10^{-4}\ \mathrm{s}^{-1}$ in Eq. (18) results in a value of $A = +2.2 k_B T$. (The polar segments in *OmpA* vary in length and composition, so this value represents an estimate of the average activation energy contributed by a single polar segment passing through the membrane.) Given the many simplifying assumptions in the model, the predicted value compares well with the value of $A = 3 - 5 k_B T$ estimated above based on group transfer energies.

While the refolding pathway of chemically denatured *OmpA* protein appears to be best described by the concerted process, it is interesting to note that mechanical unfolding of the related protein *KpOmpA* from *Klebsiella pneumoniae* results in a stepwise unfolding process where individual β hairpins unfold successively [64]. Similar behavior has also been observed in β barrels *OmpG* and *FhuA* [65,66]. Moreover, in the case of *OmpG*, refolding was also observed and occurred in a stepwise fashion involving insertion of individual β hairpins [65]. These results suggest that there may be a greater diversity of folding and unfolding mechanisms for β barrels than has been previously appreciated.

D. Energetic constraints on cellular pathways for membrane insertion

Now consider how the cellular machinery installs these proteins in membranes. Pathways within bacteria are schematically depicted in Fig. 10, but similar systems operate in eukaryotic cells. Proteins are synthesized in the cytoplasm on the ribosome in a chemically driven process using energy derived from ATP, used to activate the amino acid monomers before polymerization into proteins, and GTP, which is consumed during protein elongation. The overall mechanism by which proteins are installed in membranes involves sequential recognition by a series of factors in the cell.

Ribosomes synthesizing proteins destined for the inner membrane are bound by the signal recognition particle, a protein-RNA complex which recognize the initial hydrophobic segment in the emerging chain. This complex facilitates attachment of the ribosome to the translocon, a multiprotein assembly which is the central component of the cellular machinery for membrane incorporation of α helical proteins. In bacteria this complex is located in the inner membrane and serves both to incorporate proteins into the membrane, and as a channel through which periplasmic and outer membrane proteins can traverse it. In eukaryotes a similar complex exists on the membrane of the endoplasmic reticulum. As successive hydrophobic segments emerge from the ribosome, they form into helices within the translocon pore, which then successively releases the helical segments laterally into the membrane [67]. Transmembrane segments are released through a partitioning process based on the hydrophobicity of the chain [48,68,69]. The released helical segments then associate within the membrane to form the final folded configuration.

Nascent outer membrane β barrels are maintained after synthesis in an unfolded state by the *secB* protein to permit trafficking and prevent aggregation or degradation [10,15]. The chain is delivered by *secB* to *secA*, which associates with the translocon. The *secA* protein is a motor protein which uses energy from ATP hydrolysis to drive the chain through the translocon [10]. Inside the periplasm, the chain associates with the chaperones *skP* and *surA*, which maintain it in a unfolded state and enable it to cross this intermembrane space [70]. At the outer membrane, the chain associates with the bam complex, which facilitates chain insertion and folding and releases the folded β barrel into the membrane [10].

Both pathways consume energy, although the net folding reaction itself is believed to be a thermodynamically favorable one. The primary energy input is in the form of chemical energy associated with ATP and GTP hydrolysis. This takes place in the cytoplasm, but the proteins must be physically trafficked to their final locations. For α helical proteins the process of synthesis is directly coupled to translocation through the inner membrane. However, upon dissociation of the chain from the translocon, the remaining steps in the folding process proceed without known energy input. Therefore the action of the translocon accomplishes the insertion phase for α helical proteins, but the folding of the chain appears to be under thermodynamic control [69]. For β barrel proteins the energy required for directional motion across the inner membrane is supplied by ATP consumed by *secA*. However, in the periplasm, there is no energy source to drive subsequent stages in the insertion and folding process. Therefore both the insertion and folding appear to be under thermodynamic control [71]. The endpoint of the cellular trafficking processes and the unassisted folding processes are the same folded structures, encoded in the protein sequence itself.

Kinetic studies of the incorporation of α helical model substrates by the translocon show the process occurs in minutes [69,72,73]. In the cell, the characteristic time for the phoE β barrel to be inserted can be less than 30 seconds [74].

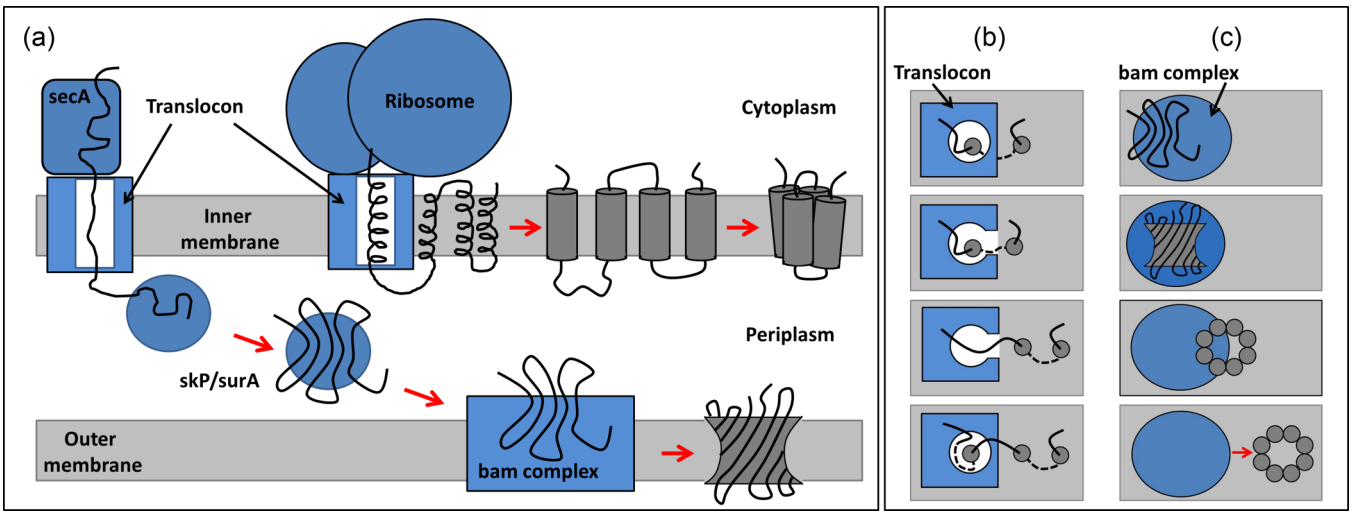

FIG. 10. (Color online) (a) Simplified depiction of pathways for membrane insertion in bacteria. Proteins are synthesized on ribosomes located in the cytoplasm. α helical membrane proteins are inserted into the inner membrane cotranslationally, and the emerging chain is extruded into an aqueous channel in the translocon complex. The individual helices form successively in the channel and then partition into the membrane, where they can associate to form the folded state. β barrel proteins must traverse the inner membrane through the translocon, driven by the *secA* chaperone. They then cross the periplasmic space bound to chaperones *skP* and/or *surA*, and associate with the *bam* complex, which facilitates insertion and folding. (b) Schematic depiction of translocon insertion mechanism (top view, cytoplasmic side). From top, the nascent chain forms a transmembrane helix within the aqueous channel of the translocon. The presence of a sufficiently hydrophobic segment within the channel results in opening of the channel to the membrane, allowing the transmembrane segment to partition laterally into the bilayer. As the chain is lengthened, additional transmembrane segments are incorporated in a sequential process. (c) Proposed facilitated insertion mechanism of the *bam* complex (top view, periplasmic side). Initially, the chain delivered by *skP* or *surA* binds on the periplasmic surface of the bam complex. The chain undergoes a partial folding event before the insertion step facilitated by *bamA*, in which the orientation of all transmembrane strands with respect to the membrane is established in a concerted process.

In vitro experiments with purified components show that the presence of its central component *bamA* in isolation is able to accelerate the membrane incorporation of the model substrate *OmpA* in a process requiring on the order of 1–10 min [70,75]. More generally, the times required for different stages of trafficking and insertion of membrane proteins are bounded by the cellular generation time. This can be as little as 20 min for *E. coli*, during which the cell's full complement of membrane proteins must be produced. Therefore the rates observed *in vitro* for the unassisted folding reaction of both α helical and β barrel membrane proteins are exceedingly slow compared to the rates of the more complex trafficking, insertion, and folding process in the cell.

During the installation of helical bundles in membranes by the translocon, successive transmembrane segments thread through and form helices inside a channel within the complex before release into the membrane [9]. Studies have found that individual segments are released sequentially into the membrane [67,76], suggesting that the translocon-assisted pathway is a stepwise pathway. Selection of the transmembrane segment is driven by partitioning between an aqueous environment within the translocon and the more hydrophobic environment of the membrane outside the translocon [68,77,78]. By utilizing an aqueous channel through which a nascent protein chain can pass, translocon-mediated insertion avoids the energetically unfavorable passage of polar segments through the hydrophobic membrane.

In the case of transmembrane β barrels, the cellular apparatus (the *bam* complex in this case) may act to decrease this energy barrier, possibly by creating a defect in the membrane or by forming polar interactions with the chain during the transfer [11]. In this sense, both the translocon and bam complex act to facilitate the folding process by decreasing the energy barrier for motion of polar segments through the membrane (Fig. 11). Action of these complexes by

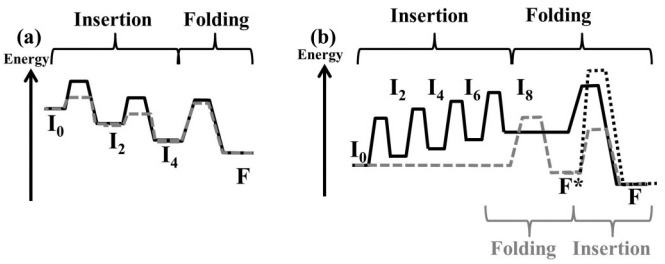

FIG. 11. (a) Schematic reaction profile for the folding process of a four-segment α helical bundle (solid line), where the insertion of each successive pair of segments is energetically favorable. A proposed mechanism of translocon-assisted insertion is to facilitate transfer of segments into the membrane by lowering the activation energy for insertion (gray dashed line.) (b) Comparison of a hypothetical reaction profile for the successive insertion and folding of an eight-segment β barrel (solid line), which results in very high-energy intermediate states, with a process in which formation of a structured intermediate (f^*) precedes insertion. The spontaneous insertion requires transfer of all polar segments simultaneously (dotted line) which would be prohibitively slow. The *bam* complex may facilitate folding by lowering the activation energy for this transfer (gray dashed line.) Partially inserted intermediate states with i buried segments are labeled $\mathbf{I}_i$ and the folded states are labeled $\mathbf{f}$.

this mechanism may be described within the model as acting on the A parameter, while preserving the equilibrium governed by the B parameter. Through different mechanisms, the cellular complexes can thereby speed the folding process to produce the folded configuration of membrane proteins on a time scale faster than is conferred by the intrinsic properties of the chain.

ACKNOWLEDGMENTS

The author thanks the Department of Chemistry and Biochemistry and the Program in Applied Mathematics at the University of Arizona for support, and an anonymous reviewer for insightful suggestions during the review process.

[1] C. Chothia, Nature (London) **357**, 543 (1992).
[2] S. H. White and W. C. Wimley, Ann. Rev. Biophys. Bio. **28**, 319 (1999).
[3] S. Neumann, H. Hartmann, A. J. Martin-Galiano, A. Fuchs, and D. Frishman, Proteins **80**, 839 (2012).
[4] S. Neumann, A. Fuchs, B. Hummel, and D. Frishman, PLoS ONE **8**, e77491 (2013).
[5] T. C. Freeman and W. C. Wimley, Bioinformatics **26**, 1965 (2010).
[6] G. E. Schulz, BBA-Biomembranes **1565**, 308 (2002).
[7] J. F. Liu and B. Rost, Prot. Sci. **10**, 1970 (2001).
[8] A. M. Stanley and K. G. Fleming, Arch. Biochem. Biophys. **469**, 46 (2008).
[9] J. Luirink, G. von Heijne, E. Houben, and J. W. de Gier, Ann. Rev. Microbiol. **59**, 329 (2005).
[10] T. J. Knowles, A. Scott-Tucker, M. Overduin, and I. R. Henderson, Nat. Rev. Microbiol. **7**, 206 (2009).
[11] N. Noinaj, A. J. Kuszak, J. C. Gumbart, P. Lukacik, H. Chang, N. C. Easley, T. Lithgow, and S. K. Buchanan, Nature (London) **501**, 385 (2013).
[12] C. B. Anfinsen, Science **181**, 223 (1973).
[13] P. S. Kim and R. L. Baldwin, Ann. Rev. Biochem. **51**, 459 (1982).
[14] F. W. Lau and J. U. Bowie, Biochemistry-US **36**, 5884 (1997).
[15] L. K. Tamm, H. H. Hong, and B. Y. Liang, Biomembrane **1666**, 250 (2004).
[16] J. U. Bowie, Proc. Natl. Acad. Sci. USA **101**, 3995 (2004).
[17] P. Curnow and P. J. Booth, Proc. Natl. Acad. Sci. USA **104**, 18970 (2007).
[18] W. Im and C. L. Brooks, Proc. Natl. Acad. Sci. USA **102**, 6771 (2005).
[19] A. C. Johansson and E. Lindahl, Biophys. J. **91**, 4450 (2006).
[20] S. Dorairaj and T. W. Allen, Proc. Natl. Acad. Sci. USA **104**, 4943 (2007).
[21] P. J. Bond and M. S. P. Sansom, J. Am. Chem. Soc. **128**, 2697 (2006).
[22] P. J. Bond, C. L. Wee, and M. S. P. Sansom, Biochemistry-US **47**, 11321 (2008).
[23] J. Gao and J. Chen, J. Phys. Chem. B **117**, 8330 (2013).
[24] D. Bucher, Y.-H. Hsu, V. D. Mouchlis, E. A. Dennis, and J. A. McCammon, PLoS Comput. Biol. **9**, e1003156 (2013).
[25] L. J. D. Frink and A. L. Frischknecht, Phys. Rev. Lett. **97**, 208701 (2006).
[26] S. Park, T. Kim, and W. Im, Phys. Rev. Lett. **108**, 108102 (2012).
[27] C.-M. Chen, Phys. Rev. E **63**, 010901(R) (2000).
[28] C. M. Chen and C. C. Chen, Biophys. J. **84**, 1902 (2003).
[29] J. Kyte and R. F. Doolittle, J. Mol. Biol. **157**, 105 (1982).
[30] A. Krogh, B. Larsson, G. von Heijne, and E. L. L. Sonnhammer, J. Mol. Biol. **305**, 567 (2001).
[31] B. Rost, R. Casadio, P. Fariselli, and C. Sander, Prot. Sci. **4**, 521 (1995).
[32] D. Baker and D. A. Agard, Biochemistry-US **33**, 7505 (1994).
[33] C. Tanford, J. Am. Chem. Soc. **84**, 4240 (1962).
[34] P. J. Fleming and G. D. Rose, Prot. Sci. **14**, 1911 (2005).
[35] W. A. Baase, L. Liu, D. E. Tronrud, and B. W. Matthews, Prot. Sci. **19**, 631 (2010).
[36] D. N. Ivankov, S. O. Garbuzynskiy, E. Alm, K. W. Plaxco, D. Baker, and A. V. Finkelstein, Prot. Sci. **12**, 2057 (2003).
[37] J. K. Myers and T. G. Oas, Nat. Struct. Biol. **8**, 552 (2001).
[38] S. E. Radford, C. M. Dobson, and P. A. Evans, Nature (London) **358**, 302 (1992).
[39] W. J. Becktel and J. A. Schellman, Biopolymers **26**, 1859 (1987).
[40] K. A. Dill, S. Bromberg, K. Z. Yue, K. M. Fiebig, D. P. Yee, P. D. Thomas, and H. S. Chan, Prot. Sci. **4**, 561 (1995).
[41] H. S. Chan, Z. Zhang, S. Wallin, and Z. Liu, Ann. Rev. Phys. Chem. **62**, 301 (2011).
[42] R. Zwanzig, Proc. Natl. Acad. Sci. USA **92**, 9801 (1995).
[43] T. W. Kahn and D. M. Engelman, Biochemistry-US **31**, 6144 (1992).
[44] J. L. Popot, Curr. Opin. Struct. Biol. **3**, 532 (1993).
[45] W. C. Wimley, T. P. Creamer, and S. H. White, Biochemistry-US **35**, 5109 (1996).
[46] T. Marti, J. Biol. Chem. **273**, 9312 (1998).
[47] L. E. Hedin *et al.*, J. Mol. Biol. **396**, 221 (2010).
[48] T. Hessa, H. Kim, K. Bihlmaier, C. Lundin, J. Boekel, H. Andersson, I. Nilsson, S. H. White, and G. von Heijne, Nature (London) **433**, 377 (2005).
[49] J. P. Schlebach, D. Peng, B. M. Kroncke, K. F. Mittendorf, M. Narayan, B. D. Carter, and C. R. Sanders, Biochemistry-US **52**, 3229 (2013).
[50] Y. C. Chang and J. U. Bowie, Proc. Natl. Acad. Sci. USA **111**, 219 (2014).
[51] R. B. Hill and W. F. DeGrado, Struct. Fold. Des. **8**, 471 (2000).
[52] X. Deupi and B. K. Kobilka, Physiology **25**, 293 (2010).
[53] J. L. Popot and D. M. Engelman, Biochemistry-US **29**, 4031 (1990).
[54] J. P. Schlebach, Z. Cao, J. U. Bowie, and C. Park, Prot. Sci. **21**, 97 (2012).
[55] J. H. Kleinschmidt and L. K. Tamm, Biochemistry-US **35**, 12993 (1996).
[56] M. Lorch and P. J. Booth, J. Mol. Biol. **344**, 1109 (2004).
[57] J. Tang, H. Yin, J. Qiu, M. J. Tucker, W. F. DeGrado, and F. Gai, J. Am. Chem. Soc. **131**, 3816 (2009).
[58] J. H. Kleinschmidt, P. V. Bulieris, J. Qu, and M. Dogterom, and T. den Blaauwen. J. Mol. Biol. **407**, 316 (2011).

[59] A. M. Seddon, M. Lorch, O. Ces, R. H. Templer, F. Macrae, and P. J. Booth, J. Mol. Biol. **380**, 548 (2008).
[60] C. P. Moon, S. Kwon, and K. G. Fleming, J. Mol. Biol. **413**, 484 (2011).
[61] D. M. Engelman and T. A. Steitz, Cell **23**, 411 (1981).
[62] T. Surrey and F. Jahnig, Proc. Natl. Acad. Sci. USA **89**, 7457 (1992).
[63] H. D. Hong and L. K. Tamm, Proc. Natl. Acad. Sci. USA **101**, 4065 (2004).
[64] P. D. Bosshart, I. Iordanov, C. Garzon-Coral, P. Demange, A. Engel, A. Milon, and D. J. Mueller, Structure **20**, 121 (2012).
[65] M. Damaghi, S. Koester, C. A. Bippes, O. Yildiz, and D. J. Mueller, Angew. Chem. Intl. Edit. **50**, 7422 (2011).
[66] J. Thoma, P. Bosshart, M. Pfreundschuh, and D. J. Mueller, Structure **20**, 2185 (2012).
[67] H. Sadlish, D. Pitonzo, A. E. Johnson, and W. R. Skach, Nat. Struct. Mol. Biol. **12**, 870 (2005).
[68] K. Ojemalm, T. Higuchi, Y. Jiang, U. Langel, I. Nilsson, S. H. White, H. Suga, and G. von Heijne, Proc. Natl. Acad. Sci. USA **108**, E359 (2011).
[69] S. H. White and G. von Heijne, Ann. Rev. Biophys. **37**, 23 (2008).
[70] G. J. Patel and J. H. Kleinschmidt, Biochemistry-US **52**, 3974 (2013).
[71] C. P. Moon, N. R. Zaccai, P. J. Fleming, D. Gessmann, and K. G. Fleming, Proc. Natl. Acad. Sci. USA **110**, 4285 (2013).
[72] I. Taufik, A. Kedrov, M. Exterkate, and A. J. M. Driessen, J. Mol. Biol. **425**, 4145 (2013).
[73] Z. Cheng and R. Gilmore, Nat. Struct. Mol. Biol. **13**, 930 (2006).
[74] C. Jansen, M. Heutink, J. Tommassen, and H. de Cock, Eur. J. Biochem. **267**, 3792 (2000).
[75] D. Gessmann, Y. H. Chung, E. J. Danoff, A. M. Plummer, C. W. Sandlin, N. R. Zaccai, and K. G. Fleming, Proc. Natl. Acad. Sci. USA **111**, 5878 (2014).
[76] H. Dale, C. M. Angevine, and M. P. Krebs, Proc. Natl. Acad. Sci. USA **97**, 7847 (2000).
[77] T. Hessa, N. M. Meindl-Beinker, A. Bernsel, H. Kim, Y. Sato, M. Lerch-Bader, I. Nilsson, S. H. White, and G. von Heijne, Nature (London) **450**, 1026 (2007).
[78] T. Junne, L. Kocik, and M. Spiess, Mol. Biol. Cell. **21**, 1662 (2010).