WILEY InterScience® DISCOVER SOMETHING GREAT

PROTEINS STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

A model study of protein nascent chain and cotranslational folding using hydrophobic-polar residues

Hsiao-Mei Lu[1] and Jie Liang[1,2]*

[1] Department of Bioengineering, MC-063 University of Illinois at Chicago, Chicago, Illinois 60607

[2] Institute of Systems Biomedicine, Shanghai Jiaotong University Minghang District, Shanghai 200240, China

ABSTRACT

To study protein nascent chain folding during biosynthesis, we investigate the folding behavior of models of hydrophobic and polar (HP) chains at growing length using both two-dimensional square lattice model and an optimized three-dimensional 4-state discrete off-lattice model. After enumerating all possible sequences and conformations of HP heteropolymers up to length N = 18 and N = 15 in two and three-dimensional space, respectively, we examine changes in adopted structure, stability, and tolerance to single point mutation as the nascent chain grows. In both models, we find that stable model proteins have fewer folded nascent chains during growth, and often will only fold after reaching full length. For the few occasions where partial chains of stable proteins fold, these partial conformations on average are very similar to the corresponding parts of the final conformations at full length. Conversely, we find that sequences with fewer stable nascent chains and sequences with native-like folded nascent chains are more stable. In addition, these stable sequences in general can have many more point mutations and still fold into the same conformation as the wild type sequence. Our results suggest that stable proteins are less likely to be trapped in metastable conformations during biosynthesis, and are more resistant to point-mutations. Our results also imply that less stable proteins will require the assistance of chaperone and other factors during nascent chain folding. Taken together with other reported studies, it seems that cotranslational folding may not be a general mechanism of in vivo protein folding for small proteins, and in vitro folding studies are still relevant for understanding how proteins fold biologically.

Proteins 2008; 70:442–449.
© 2007 Wiley-Liss, Inc.

Key words: protein folding; nascent polypeptide; lattice model; off-lattice model; cotranslational folding.

INTRODUCTION

Biosynthesis of polypeptides is a critical event for cells. The speed of biosynthesis is astounding, as thousands of polypeptides can be produced by ribosomes per minute.[1,2] As each newly synthesized chain must fold into its unique three-dimensional structures to become functionally active,[3] how this brings to completion correctly in cells remains a fundamental problem.

Research over the last decade showed that many newly synthesized proteins need the help of complicated cellular machinery to reach the native states efficiently.[1,4–7] For these proteins, chaperone factors in cytosol play the roles of protecting nascent proteins from aggregation and helping them to fold into functional forms. It is estimated that upon release from the ribosome, about 30% of the proteins in a eukaryotic cell require the assistance of cellular chaperone proteins and other folding factors.[6] These proteins often contain multiple domains or have complicated native conformations. The remaining 70% of the proteins seem to be capable of folding autonomously. *In vitro* studies showed that although the refolding of complex proteins often encounter aggregation and misfolding occur frequently, the majority of smaller proteins can refold effectively and correctly.[1,7]

The high resolution structure of ribosomes reveals a detailed structure of the nascent tunnel.[8] It provides valuable information on how ribosome interacts with the newly synthesized polypeptide chain as it emerges to endoplasmic reticulum from a tunnel of approximately 10 nm in length and 1.5 nm in diameter, while still connected to the peptidyl transferase center.[6,9] A very short segment of the newly synthesized chain (less than 20–40 amino acids) can be accommodated completely inside the tunnel. Since this tunnel is narrow and long, the conformation of the polypeptide is restricted to only linear or helix formation. Experiments show that it seems very unlikely for significant amount of protein folding to occur inside ribosomal tunnel, because of the spatial restriction and the limited access to other parts of the protein.[8,9]

Grant sponsor: National Science Foundation; Grant number: CAREER DB10133856; Grant sponsor: National Institute of Health; Grant numbers: GM68958 and GM079804; Grant sponsor: Office of Naval Research; Grant number: N00014-06-1-0100.
*Correspondence to: Jie Liang, Department of Bioengineering, MC-063 University of Illinois at Chicago, Chicago, Illinois 60607. E-mail: jliang@uic.edu
Received 9 January 2007; Revised 3 April 2007; Accepted 11 April 2007
Published online 6 August 2007 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/prot.21575

 © 2007 WILEY-LISS, INC.

Although it takes more than a minute to synthesize a 300-residue protein in eukaryotes,[6] partially synthesized peptide has adequate time to start to fold after emerging from the nascent tunnel. It was shown by experiments and computer simulation that the folding time of a local structures is short.[10–14] For example, the *in vitro* folding time of a single α-helix, β-turns, and smallest proteins are less than 0.1, 1, and 50 μs, respectively.[10–14] These results provide evidence that proteins do have sufficient time to fold right after each of the newly synthesized residue is released from the exit of the ribosome tunnel. The cotranslational folding hypothesis postulates that a protein acquires its spatial structure in the course of translation through specific kinetic routes, in which newly grown nascent chain influences the folding of the rest of the chain.[15–18]

However, studying the folding of nascent chains experimentally is very challenging, as it is difficult to measure time-transient properties of nascent chains. Although intuitively appealing, whether cotranslational folding is the general mechanism of *in vivo* folding remains an open question.[19] In this regard, computational studies provide useful means to explore the cotranslational characteristics of nascent chain folding and to generate new hypotheses that may be eventually verified experimentally.

In this article, we study nascent chain folding using simple models. Lattice models and simplified off-lattice models have been widely used for studying protein folding, where the conformational space of proteins can be examined in detail.[20–27] For example, despite its simplistic nature, lattice model has provided important insights about proteins folding, including collapse and folding transitions,[26,28–31] influence of packing on secondary structure and void formation,[21,32–34] and the effects of chirality and side chains.[34] The connections between simple models and real proteins, in behavior such as topology-dependent folding rates, have also been studied comprehensively.[35,36] To model the spontaneous folding behavior of nascent chains emerging from the ribosomal tunnel, we study here the folding behavior of chains of growing length using both two-dimensional hydrophobic-polar (HP) lattice model and three-dimensional discrete 4-state off-lattice model. We enumerate all possible sequences and conformations of HP heteropolymers upto length $N = 18$ and $N = 15$, for two and three-dimensional models, respectively. We use a simple half-plane to mimic the excluded volume effect imposed by the body of ribosome as a first-degree approximation (Fig. 1). Our goal is to gain understanding of the following basic questions about nascent chain folding: does a protein nascent chain adopt a folded structure before the completion of the translation of the whole protein? Are partially folded structures similar to the final structures of the proteins? Do mutants affect the folding of nascent chains?

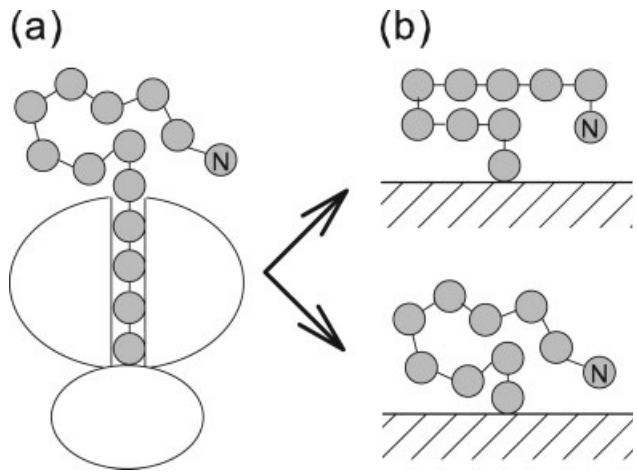

Figure 1

Simple models of nascent chain folding. ***(a)*** *A newly synthesized chain is excluded from the volume occupied by the body of the ribosome.* ***(b)*** *In our study, the body of the ribosome is approximated by a half-plane in two-dimensional space where a lattice model is used, and a half-space in three-dimensional space where a discrete 4-state off-lattice model is used.*

This article is organized as follows. In the next section, we describe our models and methods used to investigate the native structures and the folding intermediates of newly synthesized nascent chains. We then report our results using both two-dimensional square lattice model and a 4-state three-dimensional discrete off-lattice model. We conclude with a summary and remarks.

MODEL AND METHODS

HP model, protein-like sequences, and residue contact potential function

HP lattice models are heteropolymers with only two types of amino acids, hydrophobic (H) and polar (P) residues. Assuming proteins have unique ground state, we only study protein-like sequences, which are sequences with unique ground state conformation at full length. Following Ref. 31, the energy scheme adopted here is $E = \sum_{i,j \in \{H,P\}} E_{ij} n_{ij}$, where $E_{HH} = -3.8$, $E_{HP} = -2.5$, $E_{PP} = -1.5$, and n_{ij} is the number of contacts of residue type i and j. This energy scheme satisfies the conditions $E_{HH} < E_{HP} < E_{PP}$ and $(E_{HH} + E_{PP})/2 < E_{HP}$, which are characteristics of the Miyazawa-Jeernigan empirical statistical potential derived from real proteins. This energy scheme has been shown to be effective in constructing protein-like HP sequences that have unique ground states, and in generating HP sequences exhibiting non-trivial energy landscape, complex folding dynamics, and robustness against point mutations.[31]

During biosynthesis, either a hydrophobic or a polar new residue is attached to the existing nascent chain, and

this addition repeats itself until the chain reaches its full length. Similar to a full length protein, the stability of a nascent chain at any intermediate length is determined by its partial sequence. A nascent chain is stable if it has a unique conformation in its lowest energy state at this length.

Two-dimensional square lattice model

We enumerate all possible 5,808,335 conformations and 262,144 sequences of hydrophobic and polar residues heteropolymers of length $N = 18$ on a two-dimensional square lattice, with spatial restriction by the ribosome surface modeled as a half-plane. The nascent chain is only allowed in the upper half-plane (Fig. 1), which represents the space where the cytosol segment of nascent chains (from the the N-terminus to the last residue right at the exit of ribosome tunnel) is embedded in. The space for nascent chain folding is therefore limited because of the ribosome surface, but the chain is still anchored to the remaining part of protein residing inside the ribosome tunnel. Contacts are defined to exist for spatial nonbonded neighbors.

4-State three-dimensional discrete off-lattice model

For three-dimensional space, we use a reduced model of protein structures, where each residue only has the alpha carbon atom and a pseudo-atom placed at the side chain center. The radii are 1.5, 1.5, and 1.9 Å for C_α atoms, side chain centers of H and P residue, respectively. The bond length between C_α atoms is 3.8 Å, and the distance between the C_α atom and the side chain center is 1.5 Å for H residues, and 1.9 Å for P residues.

This model is parameterized by the dihedral angle τ between $C_{\alpha,i-1}$, $C_{\alpha,i}$, $C_{\alpha,i+1}$, $C_{\alpha,i+2}$, and the angle α between $C_{\alpha,i-1}$,$C_{\alpha,i}$, $C_{\alpha,i+1}$.[37] Originally introduced in Ref. 38, recent work showed that this 4-state model can generate conformations that are very native-like (on average 2.5 Å in RMSD to a native structure of protein).[37] By clustering the observed τ and α angles for each residues in the structures of the PDBSELECT database into discrete k classes, Zhang *et al.*[37] showed that this off-lattice model can reduce the complexity in computation, and yet retain many characteristics of protein structures. In this study, we use the 4-state model where $\{(\tau, \alpha)\} = \{(104.9, -112.3), (91.8, 52.1), (125.3, -175.7), (134.8, 86.0)\}$ for H residues, and $\{(\tau, \alpha)\} = \{(114.5, -103.3), (99.4, 49.3), (120.8, -163.8), (119.0, 122.1)\}$ for P residues.

To identify contacts in this model, we measure the distance between the α carbon of residue i and side chain center of residue j, the distance between the α carbon of residue j and side chain center of residue i, the distance between the α carbons of residue i and j, and the distance between the side chain centers. The contact distance thresholds are 6.7, 4.4, 4.7, 4.6, 6.1, and 4.7 Å for the α carbons, the α carbon and the side chain center of H residue, the α carbon and the side chain center of P residues, the side chain centers of H residues, the side chain centers of P residues, and the side chain centers of H and P residues, respectively. These thresholds have different values, as the radii of the atoms are different. If any one of the distances is less than the corresponding threshold, and if the residues i and j are separated by at least one residue, we declare that residues i and j are in contact.

In three-dimensional space, we again impose the spatial restriction of the ribosomal surface, which is modeled as a half-space. The upper half-space represents the cytosol space where the the three-dimensional peptide chain can grow into, whereas the lower half-space represents the body of the ribosome where the peptide is anchored to. We enumerate all possible 1,519,357 conformations and 32,768 sequences of hydrophobic and polar residues heteropolymers of length $N = 15$. Because three-dimensional space affords far more possible conformations, we have to restrict the length of nascent chain to $N = 15$ due to the limitation of computational resources.

Tolerance of protein-like sequences to mutations

A protein-like sequence may tolerate one or more point-mutations. A tolerated point-mutation is a sequence with one monomer different from the wild type sequence that takes exactly the same ground state conformation. We define the mutation tolerance at length k as the number of tolerated point-mutations a stable chain at length k has.

Structural similarity

Two types of conformation are used to describe how nascent chains may fold. The partial native conformation at length k, $c_n(k)$ is the conformation of the first k residues of the native state of the full length wild-type sequence. The stable partial conformation at length k, $c_s(k)$, is the ground state conformation of the first k residues when they take a unique ground state conformation. Figure 2 illustrates the difference in $c_n(k)$ and $c_s(k)$.

The measurement of conformational similarity $s(k)$ between $c_n(k)$ and $c_s(k)$ of a nascent at length k is the ratio of the number of common contacts between these two conformations to the number of native contacts in the partial native conformation $c_n(k)$:

$$s(k) = \frac{\#[c_s(k) \cap c_n(k)]}{\#c_n(k)},$$

where # represents the number count of native contact(s) of a conformation, $k \leq N$, and $0 \leq s(k) \leq 1$. If $s(k) = 1$ for a nascent chain at length k, this nascent chain of

DOI 10.1002/prot

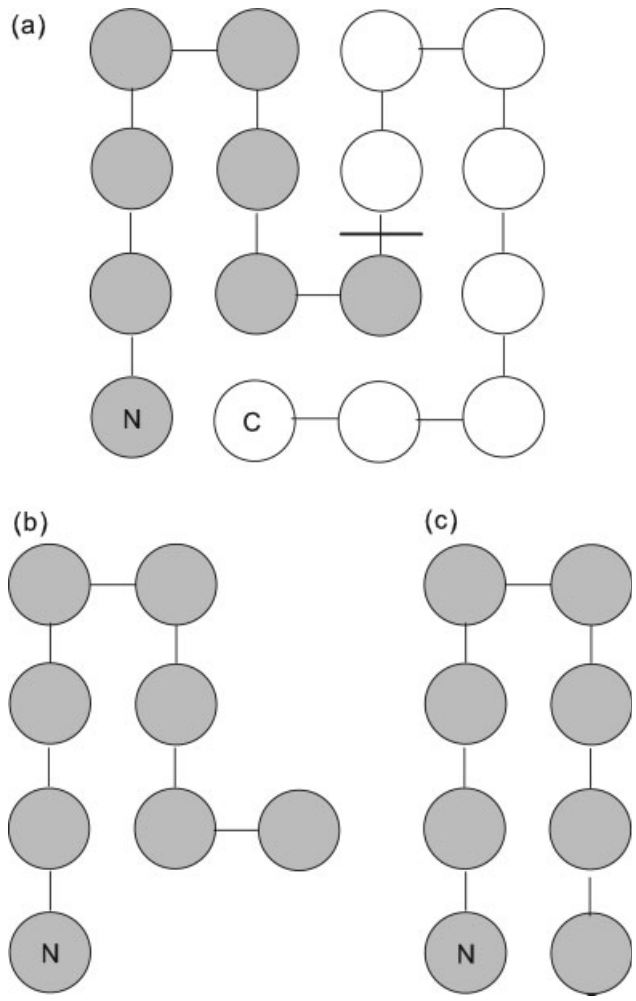

Figure 2

The conformational similarity of the partial native conformation and its final native structure. ***(a)*** *Native conformation of a full length sequence of N = 16,* ***(b)*** *partial native conformation $c_n(k)$ at length k = 8, and* ***(c)*** *stable conformation of nascent chain $c_s(k)$ at length k = 8.*

length k adopts a stable conformation, and it has all of the contacts of the first k residues appearing in the final native conformation. For example, even though the eighth residue adopts a different configuration for the partial chain [Fig. 2(c)] and the native structure [Fig. 2(b)], the conformational similarity $s(k)$ is $\frac{2}{2} = 1$ at $k = 8$, because both conformations have the same set of native contacts.

RESULTS

In this section, we examine (a) the number of stable nascent chains, (b) the stability, (c) the structural similarity of the stable nascent chain and its final native structure and how they relate to protein stability, and (d) their tolerance to mutations of the simplified proteins.

Stable proteins are less likely to have folded nascent chains

In our models, protein-like sequences with lower ground-state energy are more stable. In principle, there are two aspects about the stability of a protein: the value of the ground state energy and the value of the energy gap (the difference between the ground state and the state with the next lowest energy). In our models, we found that the values of the energy gap of the majority of the sequences in both two and three-dimensional models are very degenerate, compared to the diverse values of the ground state energy. The distributions of the values of energy gap in our models are highly left-skewed. In two-dimensional lattice model, the values of energy gap varies from 0.1 to 1.1, but more than 99.5% of sequences have energy gap ≤ 0.2. In three-dimensional off-lattice model, the energy gap varies from 0.1 to 2.6, but more than 97.8% of sequences have energy gap ≤ 1. On the other hand, the distributions of the values of ground state energy in both models follow roughly normal distribution and have a wide range from -16.0 to -34.2. For this model study, we use the ground state energy to represent the stability of protein sequences.

By this definition, we find that more stable proteins are less likely to fold before they reach full length. On two-dimensional lattice, stable proteins with lower ground-state energy have less stable nascent chains that maintain a stable partial conformation during the process of chain growth [Fig. 3(a), dashed lines]. For example, each sequence of the most stable proteins with ground state energy less than -32 (consisting of top 1.1% of protein-like sequences ranked by ground state energy) only has on average about two stable nascent conformation during the growing process. In contrast, the majority of proteins of less stability, for example, those with ground state energy ≥ -22 (bottom 4.7% of protein-like sequences) can form on average more than three stable conformations.

Conversely, we find that sequences with fewer stable nascent conformations while growing are more stable in their final ground state conformations [Fig. 3(b)]. That is, proteins which fold only when reaching full length are more stable on average and have lower ground state energy.

These observations also hold in three-dimensional 4-state off-lattice HP models (solid lines in Fig. 3). In this case, the difference between stable proteins and other proteins in the number of folded partial conformations is even more pronounced. For example, the most stable proteins with ground state energy less than -28 (top 0.2% sequences with lowest energy) have about two stable nascent conformations, while the least stable proteins with ground state energy greater than -18 (bottom 4.3% of foldable sequences) can form on average four stable nascent conformations. Overall, we find in these models

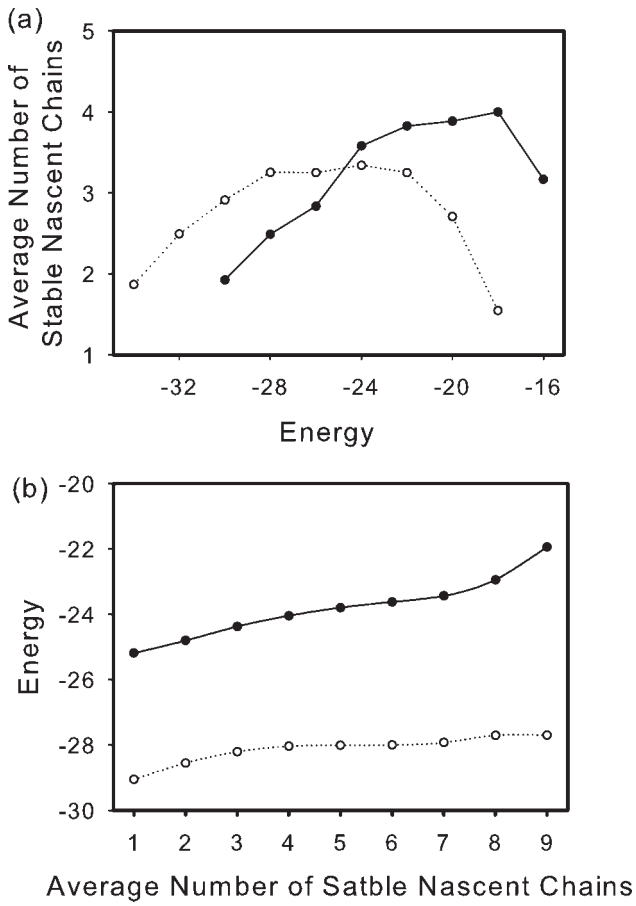

Figure 3

The relationship between protein stability and the number of folded conformations of nascent chains of two-dimensional HP lattice models (dashed line) and 4-state discrete three-dimensional off-lattice model (solid line). ***(a)*** *On average, stable proteins (with lower energy values) have fewer folded nascent chains.* ***(b)*** *Proteins with fewer folded nascent chains are more likely to be stable.*

that the number of stable nascent chains increases with the value of the ground state energy.

Existence of folded nascent chain and protein final conformation

We find that some partial nascent chains can fold into stable conformations. In addition, we find that for those few stable proteins with folded nascent chains, their folded partial conformations are very similar to the final conformations of the full length proteins. In two-dimensional lattice models, the folded nascent chains of stable proteins have nearly 90% correct native contacts at intermediate lengths (Fig. 4). This conformational similarity is much lower (70%) for proteins with moderate ground state energy. This pattern is more pronounced for three-dimensional 4-state off-lattice HP models. In this case, the folded nascent chains have >97% similarity to folded full length proteins, while this similarity for less stable proteins is consistently below 80%.

Stable proteins are tolerant to point mutations

We also find that in general more stable proteins and proteins with native-like nascent chains are often more resistant to single point mutations. Furthermore, the full length conformations of these point mutants are usually exactly the same as the wide type sequence [Fig. 5(a)]. For example, in two-dimensional lattice HP models of length $N = 18$ [empty circles in Fig. 5(a)], proteins with ground state energy between -30 and -34 are stable, as all of which are from the top 4.4% of protein-like sequences ranked by ground state energy. These proteins can tolerate point mutations at about 5.3 positions, that is, they can tolerate mutations at about 30% of the positions in the full sequence. As an example of proteins of reduced stability, proteins whose ground state energy is between -20 and -16 are all from the bottom 0.6% of the sequences when ranked by ground state energy. They can only tolerate on average point mutations at 3.2 positions out of the full length of $N = 18$, representing 18% of the positions.

In the three-dimensional model [filled circles in Fig. 5(b)], stable proteins by the criterion of ground state energy less than -28 (top 0.2% sequences with lowest energy) can tolerate up to about 6 out of the full length $N = 15$ positions, namely, they can tolerate point mutations at about 40% of the positions. This is higher than

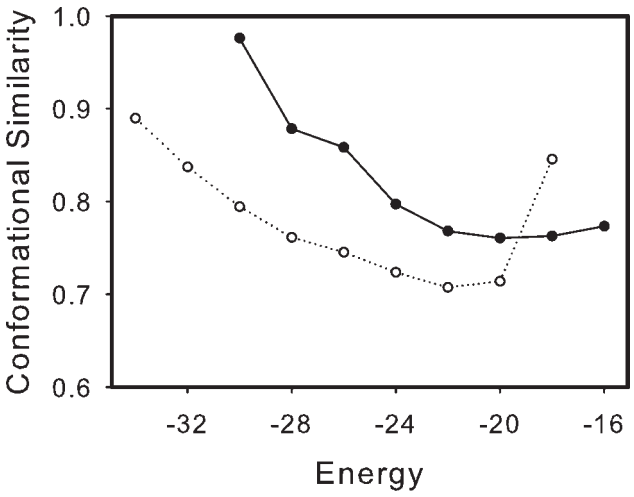

Figure 4

The relationship between protein stability and conformational similarity (s(k)) of nascent chains and final structures in the two-dimensional lattice (dashed line) and in the three-dimensional 4-state discrete off-lattice model (solid line).

DOI 10.1002/prot

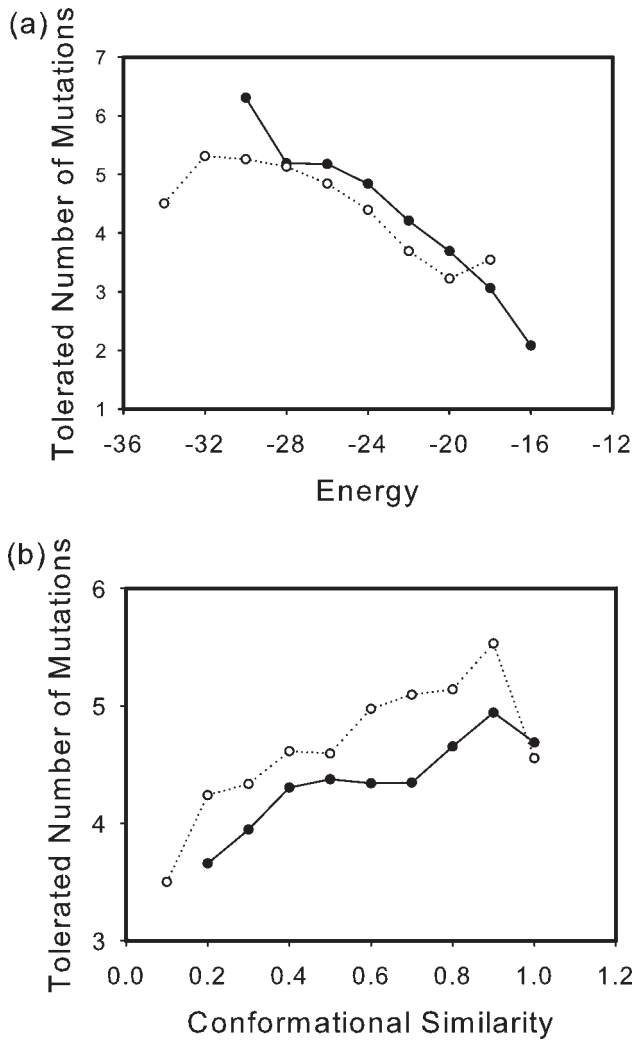

Figure 5

*The relationship between protein stability, tolerance to mutations, and conformational similarity of nascent chain in the two-dimensional lattice (dashed line) and the 4-state discrete three-dimensional off-lattice model (solid line). **(a)** On average, stable proteins with more negative ground state energy are more resistant to point-mutations. **(b)** Proteins with native-like stable nascent chains are also more resistant to point mutations.*

the case of two-dimensional lattice HP model. Less stable proteins with ground state energy greater than -16 (bottom 0.2% of foldable sequences), on the other hand, can only have single point mutations at 2 out of all $N = 15$ positions, namely, only 13.33% of all positions. The difference in the tolerance to mutations for stable and unstable proteins is more significant in the three-dimensional model.

In addition, we find that for proteins with stable nascent chains, if their nascent conformations are similar to the native conformations of the full length proteins, these proteins are more resistant to point mutations [Fig 5(b)]. In two-dimensional lattice models, proteins whose stable nascent chains retaining 90% native contacts of the corresponding parts of the final protein structures can tolerate single point mutations at up to 5.5 different positions (31% of the positions). This is higher than proteins with non-native-like stable nascent chains (3.5 positions, about 19%) . In three-dimensional space, proteins whose stable nascent chains are similar to the native structure of the full length proteins are also more resistant to mutations. Those proteins whose stable nascent chains retaining more than 90% corresponding contacts of the native structures at full length can tolerate single point mutations at 4.9 different positions (33% of the positions). Again, this is higher than the proteins with non-native-like stable nascent chains can.

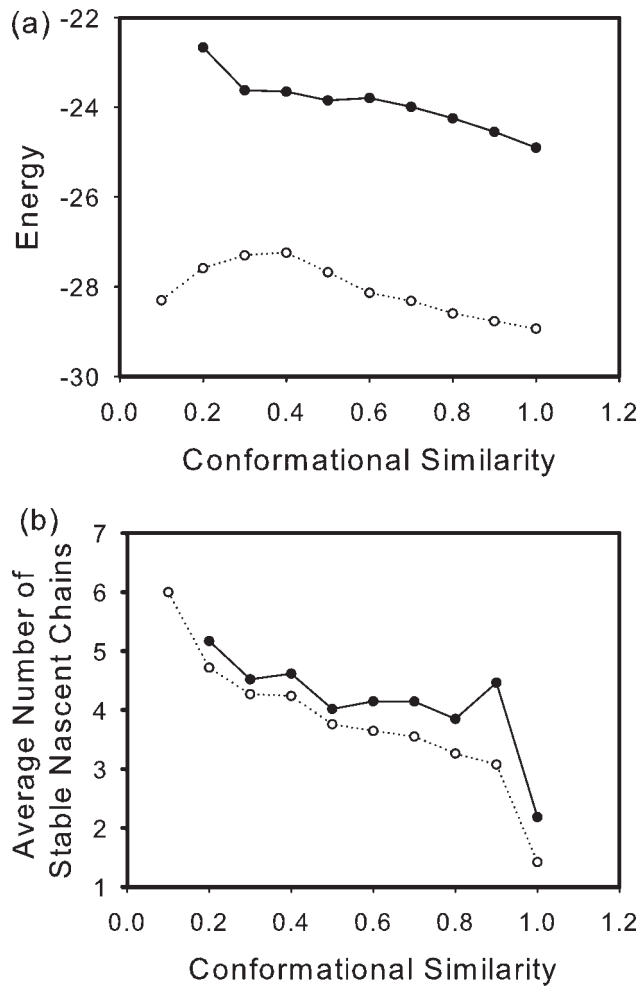

Figure 6

*The relationship between the conformational similarity of folded nascent chains and the final full length conformations, protein stability, and the number of stable nascent chains in two-dimensional lattice model (dashed line) and in 4-state discrete three-dimensional off-lattice model (solid line). **(a)** Proteins with native-like nascent chains are more stable. **(b)** Proteins with native-like nascent chains have fewer stable nascent chains.*

Folded conformation of nascent chains are native-like

It is interesting to examine the properties of protein-like sequences by the similarity of their folded nascent chains to their final full-length conformations. As shown in Figure 6(a), proteins with native-like folded nascent chains are more stable, that is, they have lower ground state energy, and have fewer folded nascent chains [Fig. 6(b)]. For two-dimensional models, proteins with native-like nascent chains are stable in a lower than two units of ground state energy and have fewer than three stable nascent chains on average. In contrast, proteins whose folded nascent chains are very different from their native structures are less stable. On average they also have more folded nascent chains (>5). These observations also hold for three-dimensional models.

CONCLUSION

In this article, we have examined the folding behavior of protein nascent chains emerging from the ribosomal tunnel using simplified models. We study simple protein-like sequences of HP residues, with the spatial restriction of allowing conformations residing only in a half-space. This half-space model provides a first-degree approximation of the excluded volume effects of the ribosomal surface. We use both two-dimensional square lattice and 4-state three-dimensional discrete off-lattice to model protein conformations. By enumerating all possible conformations of chains of different lengths while growing the protein nascent chains by adding residues one by one for different HP sequences, we are able to study the folding behavior of the nascent chains.

We find that: (1) the nascent chains of more stable proteins tend to remain unfolded during the biosynthesis process, (2) for stable proteins whose nascent chains do fold, the folded partial structures are very similar to the corresponding parts of the final native structures at full length. In addition, we find (3) stable proteins are often resistant to point-mutations.

We find conversely, sequences with fewer folded nascent chains are more stable and have lower ground states at full length. For the cases in which partial chains of stable proteins fold, these partial conformations are very similar to the corresponding partial conformations of the final structures at full length. In addition, we find proteins with native-like folded nascent chains are likely to be more stable and have fewer of folded nascent chains during biosynthesis. Furthermore, proteins whose stable nascent chains are very different from their partial native structures in full length have higher ground state energy and less number of stable nascent chains.

Our results suggests that stable proteins are less likely to be trapped in metastable conformations during biosynthesis. They can fold with more ease and may not require the help of chaperones. An interesting extrapolation is that less stable proteins are more likely to need the help of chaperones and other factors. It remains to be seen if these predictions can be confirmed by experimental studies.

After biosynthesis, proteins have to be transported to their targeting compartment or organelle in a cell. It was shown in several systems that a steady folded protein is not able to translocate to its destination.[39,40] Nascent chain residues contribute the recognition as a tag of the final destination of the mature protein.[41] However, there is no sequence conservation of nascent chains.[41] It indicates that the recognition is not from sequence information but from more general biophysical or structural properties. There are common characteristics in those targeting sequences, including a hydrophobic core of about 10 residues, which makes nascent chain to form possibly a short secondary structure that are unstable.[41,42] This is consistent with our results that nascent chains have unstable conformations in general. More work needs to be done to relate the structural motif of nascent chains and the recognition of translocation.

After examining over 400 representative nonhomologous protein structures in the Protein Data Bank, Laio and Micheletti found that the C-terminal regions of natural proteins are significantly more compact, and they are also organized more locally than the N-terminal regions.[19] The N-terminal regions have fewer contacts, and among these contacts, a higher fraction is nonlocal. This is consistent with our finding that the nascent chains of stable proteins in general have unstable conformations before maturation because of lack of contacts, and these nascent chains fold upon completion of the translation of the rest of polypeptide chains. Although cooperativity would significantly affect the kinetic process of folding, and it is well-known that the HP model is far less cooperative than natural proteins,[43] our finding based on HP models is consistent with that of Laio and Micheletti, which is based on statistical analysis of real proteins. This coincidence may suggest that the lack of well-folded nascent chains for stable proteins originates from the fundamental nature of hydrophobic interactions, regardless of the degree of cooperativity of the folding process. Taken together, it seems that cotranslational folding may not be a general mechanism of *in vivo* folding for small single domain proteins, and *in vitro* folding studies therefore may be very relevant for understanding biological folding of stable proteins.

In conclusion, our study examined several interesting questions about nascent chain protein folding, including whether there exist folded structures in newly synthesized nascent chains, the effects of protein stability on partial chain folding, and how stability and nascent chain folding related to protein tolerance to point mutations. As it is still very difficult to provide direct experimental evidence to answer these important questions, computational stud-

 DOI 10.1002/prot

ies provide an effective means for exploration and for generating hypothesis for future verifications. Although our models are simplistic, the three-dimensional 4-state discrete models can already model real protein structures accurately.[37] The HP residues types can capture the most fundamental forces driving protein folding, namely, the hydrophobic interactions.[22] The agreement of our results using both two- and three-dimensional models suggest that our observations provide some useful information about the nascent chain folding behavior of real proteins. This is consistent with the fact that studies of simple HP models in the past have provided rich insights on protein folding.[26,28–31,44]

ACKNOWLEDGMENTS

We thank Dr. Patricia Clark for suggesting studying the problem of nascent chain folding, both Dr. Patricia Clark and Dr. Nathan Baker for stimulating discussions.

REFERENCES

1. Frydman J. Folding of newly translated proteins in vivo: the role of molecular chaperones. Annu Rev Biochem 2001;70:603–647.
2. Thulasiraman V, Yang CF, Frydman J. In vivo newly translated polypeptides are sequestered in a protected folding environment. EMBO J 1999;18:85–95.
3. Bukau B, Deuerling E, Pfund C, Craig EA. Getting newly synthesized proteins into shape. Cell 2000;101:119–122.
4. Ellis RJ, Hemmingsen SM. Molecular chaperones: proteins essential for the biogenesis of some macromolecular structures. Trends Biochem Sci 1989;14:339–342.
5. Hartl FU. Molecular chaperones in cellular protein folding. Nature 1996;381:571–579.
6. Hartl F, Hayer-Hartl M. Molecular chaperones in the cytosol: from nascent chain to folded protein. Science 2002;295:1852–1858.
7. Clark P. Protein folding in the cell: reshaping the folding funnel. Trends Biochem Sci 2004;29:527–534.
8. Ferbitz L, Maier T, Patzelt H, Bukau B, Deuerling E, Ban N. Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. Nature 2004;431:590–596.
9. Nissen P, Hansen J, Ban N, Moore P, Steitz TA. The structural basis of ribosome activity in peptide bond synthesis. Science 2000;289:920–930.
10. Dobson C. Protein folding and misfolding. Nature 2003;426:884–890.
11. Eaton W, Munoz V, Thompson P, Henry E, Hofrichter J. Kinetics and dynamics of loops, α-helices, β-hairpins, and fast-folding proteins. Acc Chem Res 1998;31:745–753.
12. Snow C, Nguyen H, Pande V, Gruebele M. Absolute comparison of simulated and experimental protein-folding dynamics. Nature 2002; 420:102–106.
13. Yang WY, Gruebele M. Folding at the speed limit. Nature 2003; 423:193–197.
14. Mayor U, Guydosh NR, Johnson CM, Grossmann JG, Sato S, Jas GS, Freund SM, Alonso DO, Daggett V, Fersht AR. The complete folding pathway of a protein from nanoseconds to microseconds. Nature 2003;421:863–867.
15. Fedorov A, Baldwin T. Contribution of cotranslational folding to the rate of formation of native protein structure. Proc Natl Acad Sci USA 1995;92:1227–1231.
16. Komar A, Kommer A, Krasheninnikov I, Spirin A. Cotranslational folding of globin. J Biol Chem 1997;272:10646–10651.
17. Fedorov A, Baldwin T. Cotranslational protein folding. J Biol Chem 1997;272:32715–32718.
18. Kolb V. Cotranslational protein folding. Mol Biol Mosk 2001; 35:682–90.
19. Laio A, Micheletti C. Are structural biases at protein termini a signature of vectorial folding? Proteins 2006;62:17–23.
20. Lau KF, Dill KA. A lattice statistical mechanics model of the conformation and sequence spaces of proteins. Macromolecules 1989;22: 3986.
21. Chan HS, Dill KA. Compact polymers. Macromolecules 1989;22: 1559–1573.
22. Dill KA. Dominant forces in protein folding. Biochemistry 1990;29: 7133–7155.
23. Shakhnovich EI, Gutin A. Enumeration of all conformations of copolymers with random sequence of links. J Chem Phys 1990;93: 5967–5971.
24. Camacho CJ, Thirumalai D. Kinetice and therodynamics of foding in model proteins. Proc Natl Acad Sci USA 1993;90:6369–6372.
25. Pande VS, Grosberg AY, Joerg C, Tanaka T. Enumeration of the Hamiltonian walks on a cubic sublattice. J Phys A Math Gen 1996; 29:4753–4756.
26. Socci ND, Onuchic JN. Folding kinetics of protein like heteropolymers. J Chem Phys 1994;101:1519.
27. Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD, Chan HS. Principles of protein folding–a perspective from simple exact models. Protein Sci 1995;4:561–602.
28. Sali A, Shakhnovich EI, Karplus M. How does a protein fold? Nature 1994;369:248–251.
29. Shrivastava I, Vishveshwara S, Cieplak M, Maritan A, Banavar JR. Lattice model for rapidly folding protein-like heteropolymers. Proc Natl Acad Sci USA 1995;92:9206–9209.
30. Klimov DK, Thirumalai D. Criterion that determines the foldability of proteins. Phys Rev Lett 1996;76:4070–4073.
31. Melin R, Li H, Wingreen N, Tang C. Designability, thermodynamic stability, and dynamics in protein folding: a lattice model study. J Chem Phys 1999;110:1252.
32. Chan HS, Dill KA. The effects of internal constraints on the configurations of chain molecules. J Chem Phys 1990;92:3118–3135.
33. Liang J, Zhang J, Chen R. Statistical geometry of packing defects of lattice chain polymer from enumeration and sequential Monte Carlo method. J Chem Phys 2002;117:3511–3521.
34. Zhang J, Chen Y, Chen R, Liang J. Importance of chirality and reduced flexibility of protein side chains: a study with square and tetrahedral lattice models. J Chem Phys 2004;121:592–603.
35. Chan H, Shimizu S, Kaya H. Cooperativity principles in protein folding. Methods Enzymol 2004;380:350–79.
36. Wallin S, Chan H. A critical assessment of the topomer search model of protein folding using a continuum explicit-chain model with extensive conformational sampling. Protein Sci 2005;14:1643–60.
37. Zhang J, Chen Y, Chen R, Liang J. Emprirical protein function for simplified protein models: combining contact and local sequence-structure descriptors. Proteins 2006;63:949–960.
38. Park B, Levitt M. The complexity and accuracy of discrete state models of protein structure. J Mol Biol 1995;249:493–507.
39. Eilers M, Schatz G. Binding of a specific ligand inhibits import of a purified precursor protein into mitochondria. Nature 1986;322:228–232.
40. Neupert W, Hartl FU, Craig EA, Pfanner N. How do polypeptides cross the mitochondrial membranes? Cell 1990;63:447–450.
41. Landry SJ, Gierasch LM. Recognition of nascent polypeptides for targeting and folding. Trends Biochem Sci 1991;16:159–163.
42. von Heijne G. Signal sequences. The limits of variation. J Mol Biol 1985;184:99–105.
43. Chan H. Modeling protein density of states: additive hydrophobic effects are insufficient for calorimetric two-state cooperativity. Proteins 2000;40:543–571.
44. Gillespie B, Plaxco KW. Using protein folding rates to test protein folding theories. Annu Rev Biochem 2004;73:837–859.