Available online at www.sciencedirect.com
SCIENCE DIRECT®
Polymer 45 (2004) 557–571
polymer
www.elsevier.com/locate/polymer

The role of cotranslation in protein folding: a lattice model study

M.P. Morrissey[a], Z. Ahmed[b], E.I. Shakhnovich[b,*]

[a]*Division of Engineering and Applied Sciences, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA*
[b]*Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA*

Received 27 June 2003; received in revised form 3 September 2003; accepted 3 September 2003

Abstract

Computational studies of protein folding have implicitly assumed that folding occurs from a denatured state comprised of the entire protein. *Cotranslational folding* accounts for the linear production and release of a protein from the ribosome, allowing part of the protein to explore its conformation space before other parts have been synthesized. This gradual 'extrusion' from the ribosome can yield different folding kinetics than direct folding from the denatured state, for a lattice folding model. First, in model proteins containing chiefly short-ranged (local in sequence) contacts, cotranslational folding is shown to be significantly faster than direct folding from the denatured state. Secondly, for model proteins with two competing native states, cotranslational folding tilts the apparent equilibrium toward the state with a more local-contact dominant topology.

Keywords: Protein folding; Cotranslational folding; Lattice models

1. Introduction

Proteins find their native conformations in times which are many orders of magnitude shorter than that which is required to explore all possible conformations [1]. That proteins manage to evade this *Levinthal Paradox* indicates that there must be a statistical pathway from the denatured state (which is actually an ensemble of microstates) to the native state.[1] The energy landscape is not maximally 'rugged'; the process of free energy minimization is aided by the fact that conformations which are linked by small physical moves are similar in energy.

Since kinetic pathways are important in protein folding, the assumption that the observed native state is the global free energy minimum may be unnecessary. The scarcity of evidence to the contrary, however, indicates that the assumption of global minimization is reasonable in most cases. Metastable proteins[2] do exist, however: poliovirus [2,3] and plasminogen activation inhibitor [4,5] are two proteins for which metastability is not only evident, but is integral to function.

It has been shown in studies of protein statistics and in prior lattice-model studies that contact topology is an important determinant of folding rate. Plaxco et al. [6] observed that proteins with local contact topologies (local along the one-dimensional sequence) tend to fold more rapidly than proteins with more complex (nonlocal) contact topologies. Abkevich et al. [7] had previously demonstrated in a lattice model framework that backbones with nonlocal contact topologies (which tend to fold cooperatively) fold faster than those with local contact topologies *under the constraint that native state stability is independent of contact topology*. These two conclusions are compatible, as there is evidence that proteins with nonlocal contact topologies tend to be more stable than local-contact dominant proteins, under equivalent temperature and denaturant conditions, and assuming equal chain length. [8]

Cellular protein production occurs as *translation* of RNA

* Corresponding author. Tel.: +1-617-495-4130; fax: +1-617-496-5948.
E-mail address: eugene@belok.harvard.edu (E.I. Shakhnovich).

[1] The key descriptor is *statistical*, meaning a set of physical pathways which is more favorable than the set of all possible folding paths. The favorability of folding depends on temperature, since the free energy $F = E - TS$; all paths of a given length are equally likely at high temperature.

[2] Metastable proteins display a functional native state which is not a global free energy minimum; the global minimum state is kinetically unaccessible on the scale of the functional life of the protein.

0032-3861/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.polymer.2003.10.090

strands, which were in turn *transcribed* from cellular DNA. Translation occurs outside the nucleus in organelles called ribosomes. The peptide chain is synthesized linearly, beginning at its N-terminus, and 'extruded' piece by piece into the cytoplasm.[3] This allows for the portion of the chain which is produced first to begin the folding process.

The 'cotranslational' folding process is distinguished from 'denatured' folding in Fig. 1. Although the effect of cotranslation on folding kinetics has not been measured quantitatively, differences between contranslational and denatured folding have been elucidated. For instance, it has been shown that the N-terminal region of tryptophan synthase begins to fold long before synthesis is complete [10]. Other cases of cotranslational folding have been noted in the literature [11].

In the context of a simplified protein folding model, we will show that cotranslation can markedly accelerate the folding process, particularly in proteins with a large number of local contacts.

To date, no protein has been shown definitively to fold to one native state cotranslationally, yet to another state when folding from the fully denatured state, without the assistance of post-translational covalent modification. We will demonstrate a model system which behaves in such a manner based on the contact topologies of the two native states.

2. Results

2.1. Backbones

We studied a number of 36-mer backbones to differentiate the effect of cotranslational folding on structures rich in local contacts from that on structures with mostly long-ranged contacts. We define local contacts as contacts between residues i and $i+3$ or between i and $i+5$, denoted as $[i,j=i+3]$ and $[i,j=i+5]$, respectively. Nonlocal contacts are of form $[i,j\geq i+7]$. Recall that on a cubic lattice, contacts may form only for odd values of $|j-i|$; also note that interactions are prohibited for $|j-i|=1$ since neighboring residues are covalently bonded.

The folding kinetics of two backbones were studied in detail. The backbone *alpha* (see Fig. 2(a)) was designed by the authors to mimic a right-handed helix; it is a regular conformation with 33 contacts, including $9[i,i+3]$, $24[i,i+5]$, and no nonlocal contacts. The backbone *beta* shown in Fig. 2(b) was designed by Abkevich et al. to minimize the number of local contacts. Contact statistics for the two backbones are summarized in Table 1.

A useful measure for the predicted extrusion sensitivity of a backbone is its extrusion contact profile (ECP). As a protein is extruded, some residues become available to form contacts. Contacts which are entirely within the extruded region are considered to be 'extruded contacts'. The ECP is a plot of the normalized number of extruded contacts N_c^{extr} with respect to the number of extruded residues N^{extr}.

For a backbone rich in local contacts, the ECP should approximate a straight line from 0.0 to N_c.

$$\frac{dN_c^{extr}}{dN^{extr}} \approx \frac{N_c}{N} \tag{1}$$

A backbone with primarily long-ranged contacts would have a concave shape, with few contacts extruded with the first several residues

$$\frac{dN_c^{extr}}{dN^{extr}} < \frac{N_c}{N} \quad \text{when } N^{extr} \approx 0$$

but with many contacts extruded with the final residues

$$\frac{dN_c^{extr}}{dN^{extr}} > \frac{N_c}{N} \quad \text{when } N^{extr} \approx N$$

The ECPs of our two backbones are plotted in Fig. 3. In order to compare ECPs of conformations with different numbers of total contacts, each ECP curve is normalized to the total number of contacts in its respective conformation.

We can quantify the local-contact character of a structure by measuring the concavity of ECP, using Eq. (1)

$$\kappa = \sum_{N^{extr}=1}^{N} \left(\frac{N_c^{extr}}{N_c} - \frac{N^{extr}}{N} \right) = \sum_{N^{extr}=1}^{N} \left(\frac{N_c^{extr}}{N_c} \right) - \frac{N}{2}$$

A perfect local-contact structure would exhibit a $\kappa = 0$. Using this scale, $\kappa_{alpha} = -2.23$ and $\kappa_{beta} = -7.94$.

Folding data were gathered using alpha and beta as representatives of local contact-dominant and nonlocal contact-dominant conformations, respectively.

Table 1
Contact statistics for the primarily local backbone alpha and the primarily nonlocal backbone beta

Backbone name	N_c	$N_c^{nonlocal}$	N_c^{local}	% local	$N_c^{[i,i+3]}$	$N_c^{[i,i+5]}$	κ
Alpha	33	0	33	100.0%	9	24	−2.23
Beta	36	34	2	5.6%	0	2	−7.94

N_c is the total number of contacts in the structure; a maximally compact structure has 40 contacts. $N_c^{nonlocal}$ is the number of long-ranged contacts $[i,j\geq i+7]$. N_c^{local} is the number of $[i,i+3]$ and $[i,i+5]$ contacts. The parameter κ, defined in the text, should be near zero for local-contact rich structures, and strongly negative for long-range dominant structures.

[3] In many cases, proteins are excluded into chaperones rather than directly into the cytoplasm. Chaperones are proteins which assist in the folding of many classes of proteins; the assistance must be largely non-specific to sequence, given this many-to-one mapping. One proposed mechanism for chaperone function is as a 'dumb' storage bin, protecting a domain of the protein from extraneous cellular interactions while it attempt to fold [9]. Our model of cotranslational folding is demonstrated for proteins which are extruded directly into the cytoplasm, but is also valid for proteins which are extruded into chaperones in fragments smaller than the entire protein molecule.

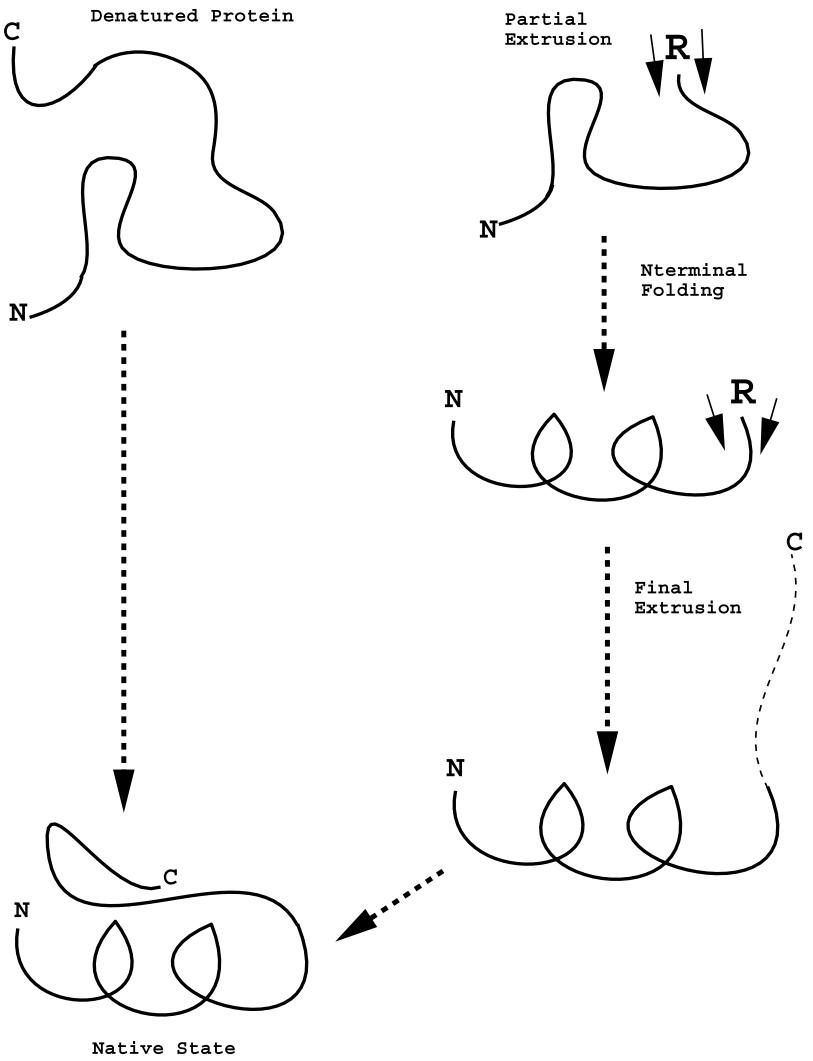

Fig. 1. Schematic of two pathways to a protein's native state. The *direct pathway* assumes that the entire protein is fully denatured before folding, and proceeds via some route to the native conformation. This is an appropriate model for in vitro folding, and is the process modeled by computer simulations to date. The *extrusion pathway* may be a more appropriate model for in vivo folding of many proteins. In vivo, a protein is synthesized on the ribosome (R), which releases the first elements of the chain (N-terminus) before synthesis of the final residues (C-terminus) is complete. If the time $\tau_{\text{synthesis}} \gg \tau_{\text{folding}}$, the N-terminal region may begin to fold without input from the C-terminal residues. In this schematic, folding occurs to the same native state with or without extrusion. Later, another possibility will be discussed: specifically, that extrusion may 'trap' the protein in a metastable alternative native state.

2.2. Folding temperature

Using the Go potential, the optimal folding temperatures for the two backbones were determined. Contact energies were set to -1.00 for each contact which was native in the corresponding conformation.

To determine appropriate simulation temperatures, 100 folding runs were performed, starting from the denatured state, for each of 6 temperatures, and for both backbones. Runs were cut off at 10^8 MC steps when folding was not achieved. Median first passage times are shown in Fig. 4. The optimal folding temperature for each backbone was slightly over 0.5. However, stability was observed to be poor for both backbones at this temperature. Particularly for the alpha conformation, the optimal-T MFPT represents a folding 'hit' which does not remain in the native state for a significant number of MC steps.

2.3. Effect of extrusion without competition

Median folding times were measured at the same array of temperatures, using the extrusion protocol detailed above, and $M = 1000$ steps per thawed residue. Data are presented in Fig. 5. The effect of extrusion is apparent at lower temperatures. At $T = 0.35$, folding is completed shortly after extruding the final residue: less than 1000 post-extrusion MC steps median for alpha and less than 100 MC steps post-extrusion for beta. The same is true for folding at $T = 0.5$, for the nonlocal backbone, indicating in both cases

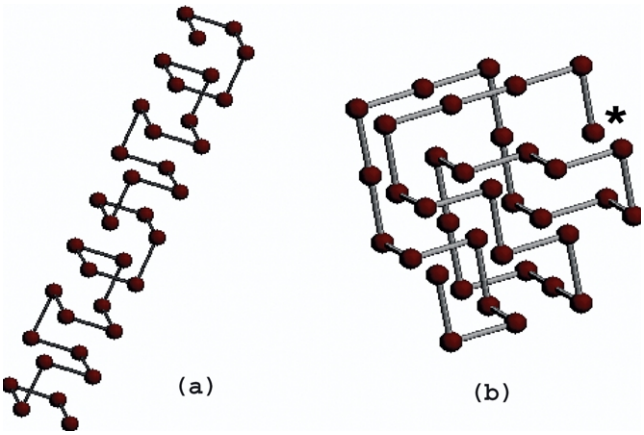

Fig. 2. Two backbones chosen for their distinctive contact topologies. Backbone (a) (alpha) is helical and thus has many local contacts, while backbone (b) (beta, see [12]) has mostly long-ranged contacts. For (b), residue #1 is denoted by the asterisk. Contact topology is identical starting from either direction for (a), since reversing the contact order forms an identical mirror image.

that folding occurs during the process of extrusion (translation).

Folding at higher temperatures ($T > 0.45$) was *not* aided by extrusion, however. The reason is straightforward: at this temperature, any structure formed during the extrusion process is destroyed before extrusion is complete. For instance, the alpha conformation is so unstable that most of the helix has already folded and unfolded several times before complete production of the chain.

A slightly different set of data was also extracted from the same set of runs. Defining 'foldicity' as the percentage of runs which fold within 2.6 million MC steps (arbitrarily chosen as 4 × the total extrusion time), the relationship in Fig. 6 is similar to that in Fig. 5. For alpha at $T = 0.35$, 70 out of 100 total runs folded with extrusion compared to only 5 without extrusion. For beta, the effect was less pronounced but still significant: at $T = 0.35$, 55 runs folded with extrusion compared to 35 without extrusion.

Sample low-temperature trajectories are presented in Fig. 7, with a trapped helical conformation shown in Fig. 8. Without extrusion, alpha collapses rapidly to $Q_{\text{alpha}} \approx 0.5$, and then climbs to $Q_{\text{alpha}} \approx 0.75$ where it remains trapped for the remainder of 50 million MC steps. With extrusion, however, this trapped state is avoided. Q_{alpha} rises (approximately) linearly to 1.0, beginning shortly after the first contact is extruded. Alternatively, it can be inferred that the ph relative quality is unity ($q_{\text{alpha}} \approx 1$) throughout the extrusion process and for the remainder of the simulation.

It is useful at this time to introduce the *cumulative folding histogram*, which plots the fraction of runs which have folded to the native state as a function of the number of elapsed MC steps. (The cumulative folding histogram represents data for one temperature only.) Such a diagram is particularly useful in distinguishing the effects of extrusion at times much greater than the extrusion time itself. The cumulative folding histograms for alpha and beta, without and with extrusion, and at $T = 0.35$, are shown in Fig. 9.

It is striking that at $T = 0.35$, the effect of extrusion (enabling folding, particularly of alpha) persists after more than 150 times the extrusion time has elapsed. (Extrusion time = 650,000 steps, while total run time is 10^8 steps.) At this relatively low temperature, the system exhibits a long-term 'memory' of the extrusion process.

In conclusion, at temperatures for which native states are stable, cotranslational folding speeds folding for both

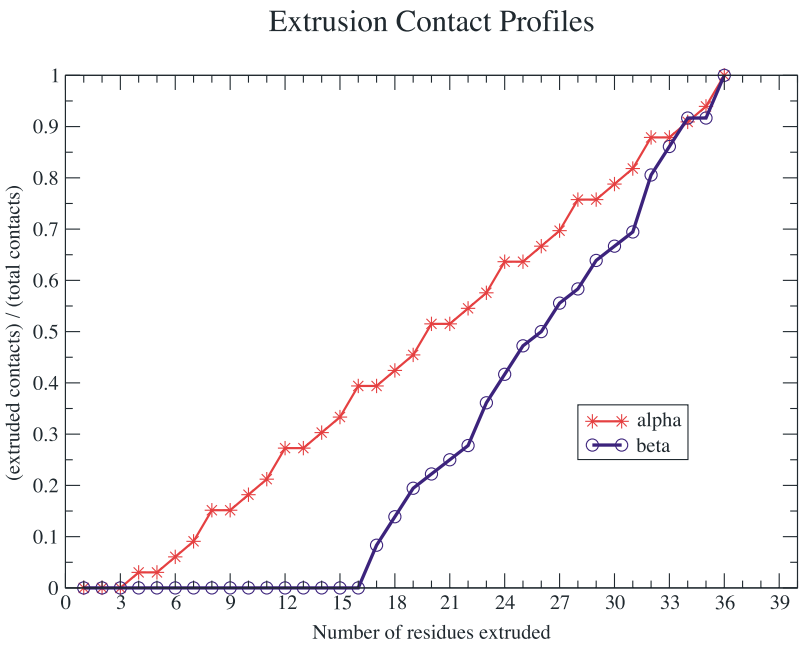

Fig. 3. Extrusion contact profiles of two local and four nonlocal backbones. See text for a description of the ECP parameter.

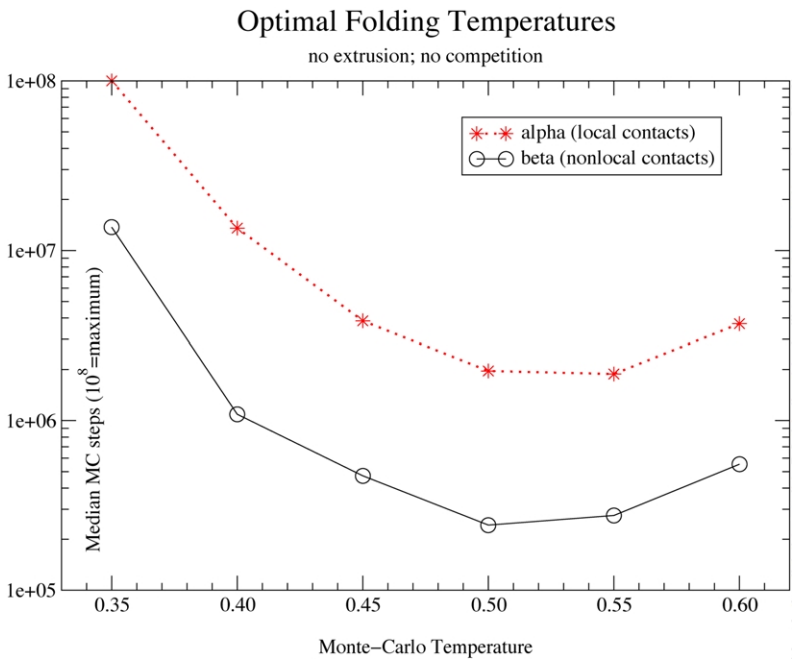

Fig. 4. Median first passage times (MFPT) for two 36-mer backbones, alpha and beta. For the both backbones, optimal folding occurs near $T = 0.5$. Note that stability is poor at optimal folding temperature, particularly for the helical form alpha. The slower folding and lower stability of alpha are due to alpha's smaller number of native contacts, and hence its less favorable native energy. For two backbones with equal overall stability, the backbone relationship would reverse, with the local-contact form folding fastest [6,7].

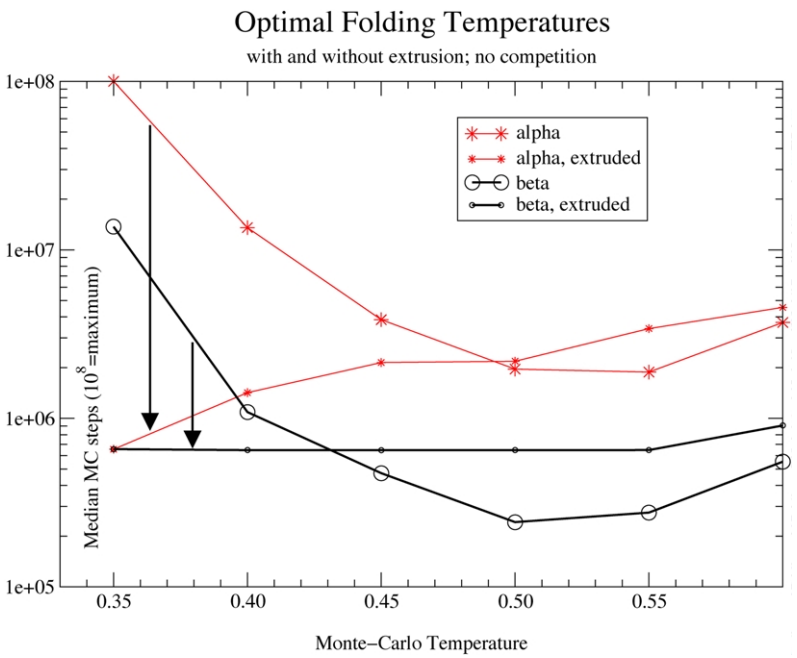

Fig. 5. Median first passage times (MFPT) for two 36-mer backbones, alpha and beta, folded with the extrusion protocol detailed in the text. Data without extrusion is repeated for comparison. Folding times for extrusion systems include the total time for extrusion; thus the minimum folding time is approximately 650,000 MC steps. At low temperature ($T < 0.4$), it is clear that extrusion accelerates folding markedly for both backbones, though the effect is stronger for alpha than for beta. Arrows represent the effect of 'turning on' extrusion.

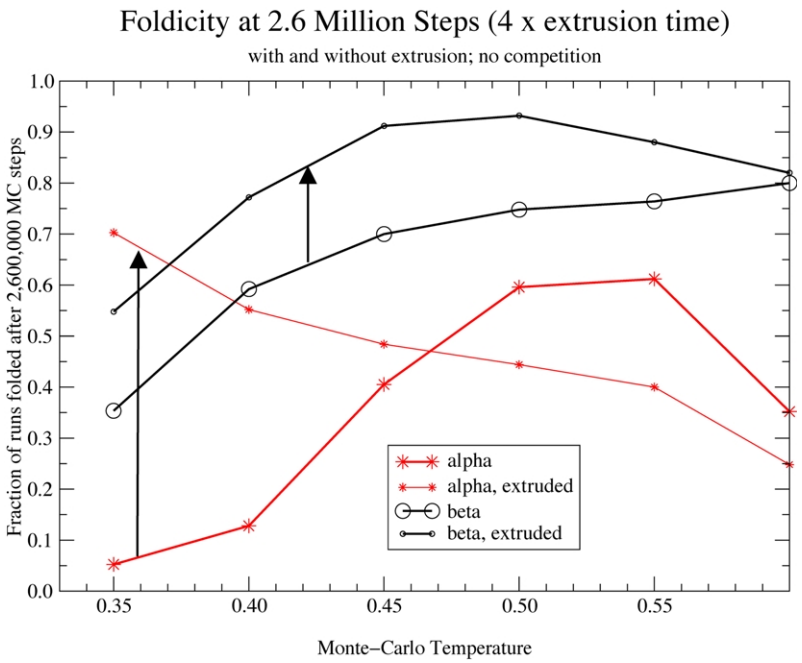

Fig. 6. Fraction of runs folded in 2.6 million MC steps. Each data point represents 100 runs. Folding times for extruded systems include the total extrusion time (approximately 650,000 steps). Once again, it is clear that extrusion aids folding at low temperature; the arrows represent the effect of 'turning on' extrusion for each backbone. The folding acceleration provided by extrusion is eliminated at higher temperatures.

backbones. However, the effect is markedly stronger for alpha, which has numerous local contacts.

2.4. Effect of extrusion with competition

Next, a Go contact potential was created to allow folding to either alpha or beta. This was accomplished by making each contact present in *either* alpha or beta attractive. The double native state Go potential may be generalized as

$$E(\vec{x}) = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \min(\omega_{\text{alpha}} \Delta_{ij}^{N_{\text{alpha}}}, \omega_{\text{beta}} \Delta_{ij}^{N_{\text{beta}}}) \Delta_{ij} \tag{2}$$

where $\Delta_{ij} = 1$ is residues i and j are currently in contact, and 0 otherwise. $\Delta_{ij}^{N_a} = 1$ if residues i and j are in contact in native state a. The ωs are weighting factors for the two

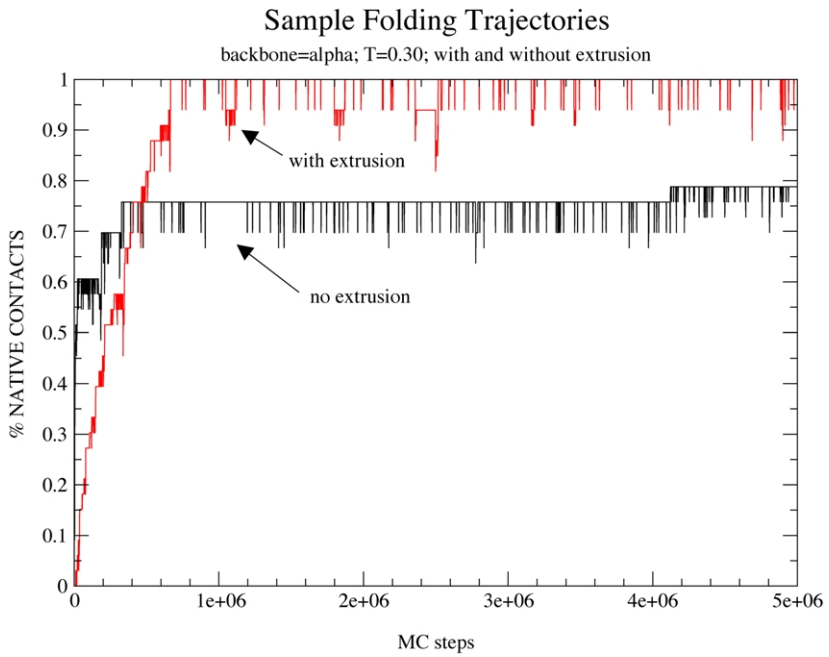

Fig. 7. Sample low-temperature ($T = 0.30$) folding trajectories for the local-contact dominated backbone conformation alpha. Extrusion enables the backbone to avoid traps such as the one encountered in the non-extrusion trajectory.

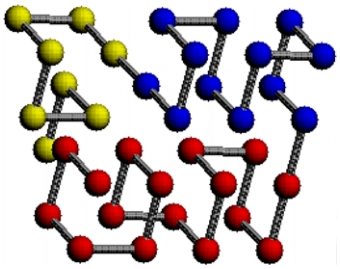

Fig. 8. A trap encountered in lieu of folding to conformation alpha (without extrusion). The trapped conformation exhibits a significant number of native contacts (26/33), and is organized into three helix-like regions, denoted, respectively, in red, blue, and yellow (in web version). Folding to the native state from this trap requires significant disruption of these native contacts, making passage to the native state extremely slow.

native states. If a contact is present in both native states, the lower energy $\min(\omega_1, \omega_2)$ is assigned to that contact.

The most straightforward case was tested first, setting $\omega_{\text{alpha}} = \omega_{\text{beta}} = -1.0$; that is, any contact present in either native state is assigned energy -1.0. The native energies of the two backbones remained $E^N_{\text{alpha}} = -33$ and $E^N_{\text{beta}} = -36$. Runs were initiated, one set without extrusion (stating from a random state) and another set with extrusion. Each run ended with one of 3 events: folding to alpha, folding to beta, or reaching 100,000,000 steps without folding to either state.

It would seem that the beta would dominate at lower temperatures due to its lower energy, and this must indeed be true for sufficiently long runs. However, previous studies [6,7,12] showed that folding occurs much more rapidly to local-contact rich backbones due to greater *kinetic accessibility*.[4]

The folding data in Fig. 10 bolsters the kinetic partitioning hypothesis: at $T = 0.35$, folding occurs only to the alpha form. And consistently with expectation, extrusion increased the fraction of folds to the native form, from 0.21(out of 2231 runs) to 0.45 (out of 4130 runs). Note that, once again, the effect of extrusion persisted for more than 100,000,000 MC steps, or about 150 × the extrusion time (about 650,000 steps).

The same set of experiments was also run at $T = 0.45$, and is shown in Fig. 11. These conditions produced a small but significant number of folds to conformation beta (approximately 5%, with or without extrusion). The effect of extrusion on folding to alpha was small even at relatively short simulation times, and was nearly nonexistent near the cutoff time of 10^8 MC steps. It is apparent that the benefit of extrusion is greatly reduced due to the conformational fluidity at this higher temperature.

The next attempt was to distill the effect of extrusion at low temperature from the general effects of contact topology on folding. It is already clear that kinetics favor folding to conformation alpha rather than beta, even if thermodynamics are tilted in the opposite direction. Since there were no folds to beta at $T = 0.35$, we designed a system for which there would be a significant number of folds to beta by setting $\omega_{\text{beta}} = -1.1$ and leaving $\omega_{\text{alpha}} = -1.0$. While these choices further tilt the system (thermodynamically) toward beta; it was hoped that they would also alter kinetics as measured by first-folding times. Fig. 12 is a cumulative folding histogram for such a system, at $T = 0.385$. Once again, the run cutoff time was 10^8 MC steps. Without extrusion, the fraction of states folded to alpha and beta were 0.42 and 0.07, respectively, out of 2803 completed runs.

It is not surprising that extrusion once again increased folding to alpha from 0.42 to 0.52. (This is a familiar story by now.) What *is* surprising is that extrusion reduced folding to beta from 0.07 to 0.01 out of 2500 runs. In Fig. 9, it was shown that extrusion aided folding to both alpha and beta, though it helped the local-contact rich alpha backbone more strongly. With competition, folding to alpha was aided, while folding to beta was markedly reduced. The result may be summarized as follows.

At sufficiently low temperature: for sequences designed for one native structure, extrusion is a general aid to folding, though the effect on local-contact rich topologies is strongest; for sequences which favor two structures with distinct contact topologies, extrusion aids folding to the local contact-rich topology, but hinders folding to the backbone which is rich in long-range contacts. The result is a significant tilt in folding kinetics toward the local contact-rich state.

In both cases, the effect of extrusion lasts for orders of magnitude longer than the extrusion time itself.

Yet another shift was made in thermodynamic equilibrium, setting $\omega_{\text{beta}} = -1.25$ while leaving $\omega_{\text{beta}} = -1.0$. Data from runs at $T = 0.385$ are shown in Fig. 13. In this case, the effect of extrusion on kinetic 'equilibrium' is quite clear. Without extrusion, the system prefers folding to beta rather than alpha by a ratio of 1.3 to 1; with extrusion, the ratio shifts to less than 0.3 to 1. By turning on extrusion, we have changed the preferred folded state! (See Fig. 14). Thus one more statement may be added to the above observations:

Under certain conditions, extrusion may tilt folding equilibrium to such a degree that it changes the dominant metastable state.

2.5. Conformational interconversion

The cumulative folding histograms presented thus far have been monotonically increasing, as a run was considered to have ended once it reached one of the two target states. The implicit assumption has been that both native states are highly stable; i.e. to visit a native state is to remain there forever. The validity of this assumption was

[4] By native state, we mean the exact native microstate. When including similar states which are conformationally 'near' the native state, entropy must be considered.

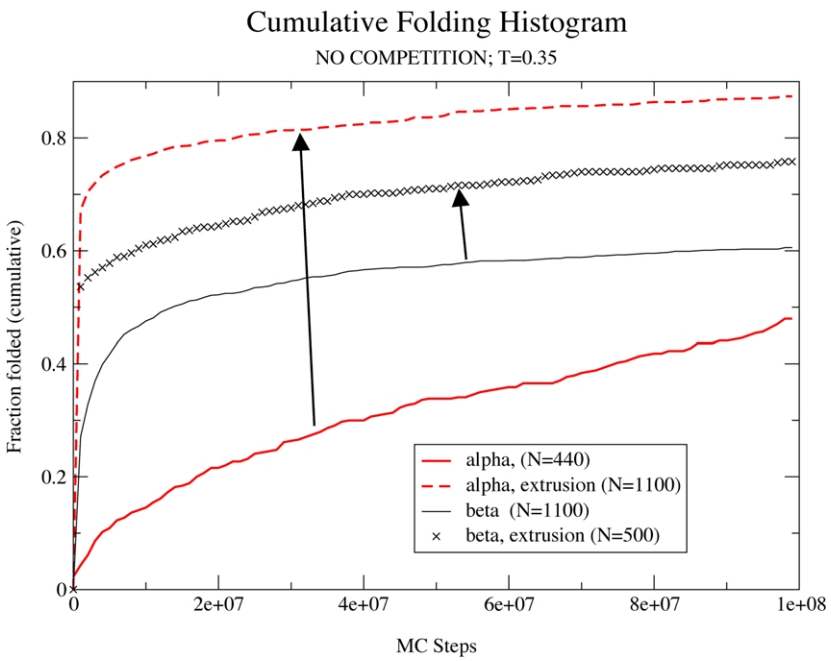

Fig. 9. Cumulative folding histograms for alpha and beta at $T = 0.35$. As before, arrows represent the 'turning on' of extrusion. The effect of extrusion on the alpha backbone is particularly pronounced, allowing nearly 1000 out of 1100 runs to fold within 10^8 MC steps, compared to less than half of the runs without extrusion.

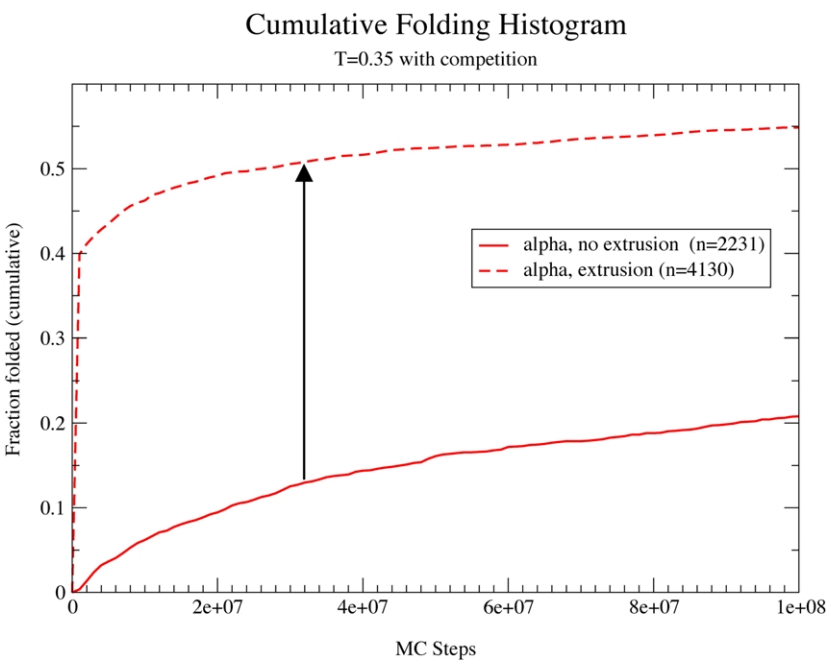

Fig. 10. Cumulative folding fraction is plotted as a function of MC time, for a system with competition between two native states. $\omega_{\text{alpha}} = \omega_{\text{beta}} = -1.0$. Consistently with Abkevich et al. [7,12], folding was not observed to state beta due to the preponderance of local contacts in conformation alpha. Extrusion greatly increases folding to alpha, an effect which persists to more than 150 × the extrusion time itself (650,000 steps).

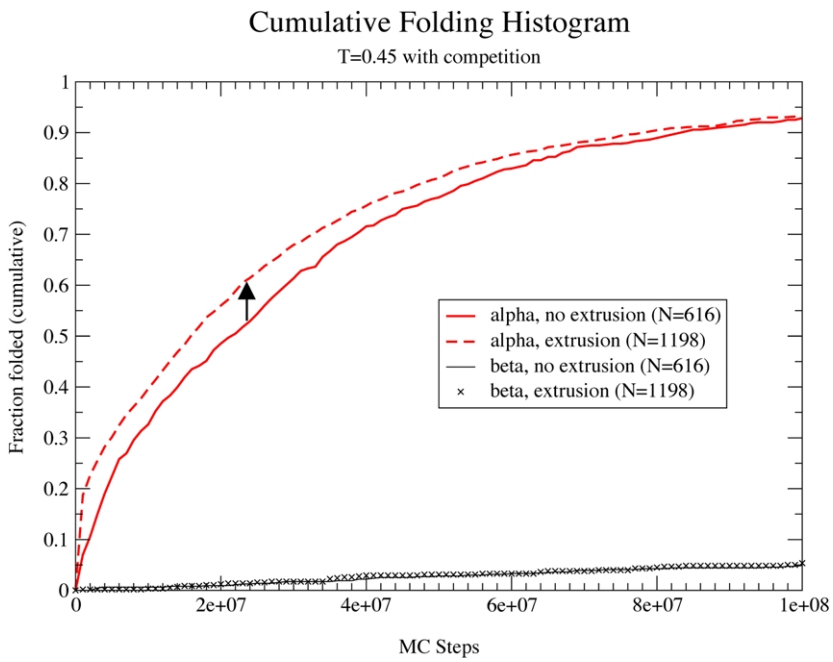

Fig. 11. Cumulative folding fraction at $T = 0.45$; $\omega_{\text{alpha}} = \omega_{\text{beta}} = -1.0$. The effect of extrusion is markedly less than at $T = 0.35$ (see Fig. 10), as the structure which is produced during extrusion may destroyed at the higher temperature. Nevertheless, folding to alpha is improved slightly, while folding to beta remains essentially unchanged.

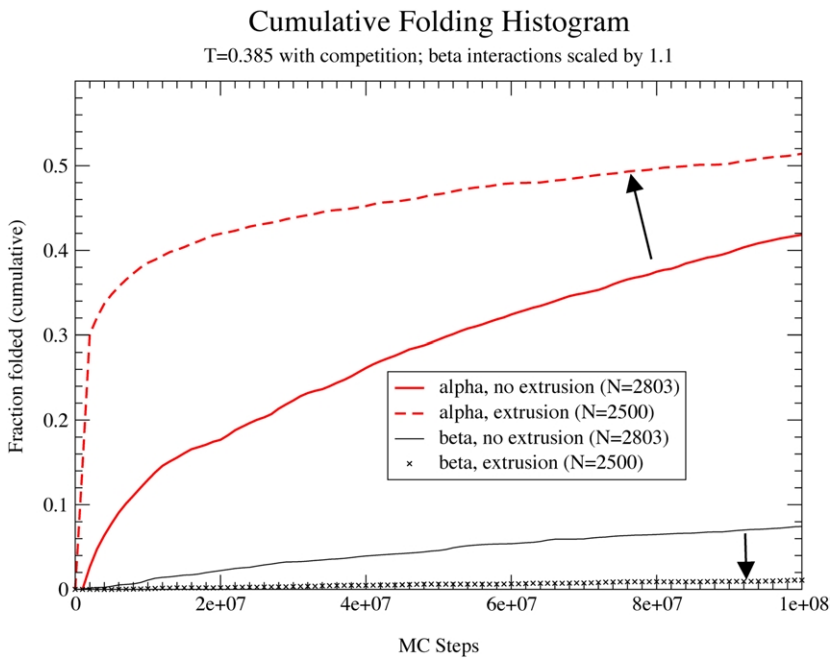

Fig. 12. Cumulative folding fraction at $T = 0.385$; $\omega_{\text{alpha}} = -1.0$; $\omega_{\text{beta}} = -1.1$. Extrusion assists in folding to state alpha, while markedly hindering folding to state beta. See text for discussion.

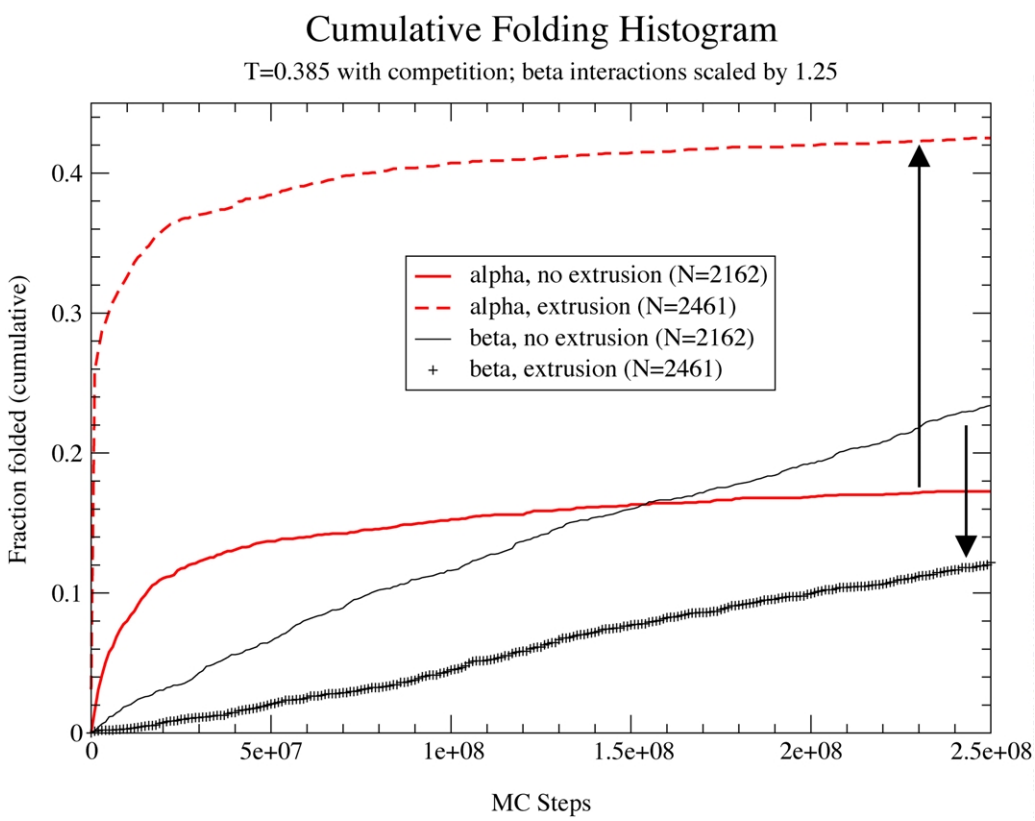

Fig. 13. Cumulative folding fraction at $T = 0.385$; $\omega_{\text{alpha}} = -1.0$; $\omega_{\text{beta}} = -1.25$. The cutoff time was extended to 250,000,000 (2.5e + 8) MC steps. As in Fig. 12, extrusion assists in folding to state alpha while disrupting folding to beta. In this case, the relationship is even more dramatic: extrusion changes the dominant state from beta to alpha!

tested, and a modification made to Fig. 13 to correct for this result.

It is interesting that folding occurs at all to alpha at low temperatures. Under the conditions of Fig. 10 ($T = 0.35$ and $\omega_{\text{alpha}} = \omega_{\text{beta}} = 1.0$), the thermodynamic equilibrium ratio of α to β occupancy is

$$\frac{[alpha]}{[beta]} = e^{(36*(-1.0)-33*(-1.0))/0.35} \approx 10^{-4}$$

while for the system in Fig. 13 the equilibrium ratio is

$$\frac{[alpha]}{[beta]} = e^{(36*(-1.25)-33*(-1.0))/0.385} \approx 10^{-14}$$

In both cases, alpha must be considered metastable, although it is certainly long-lived on the scale of the above simulations.

A set of runs was performed at $T = 0.385$, starting from conformation alpha and ending either with folding to beta or at 250,000,000 MC steps. The interconversion histogram is presented in Fig. 15. As a control (for errors in implementation), the same experiment was run starting from beta and attempting folding to alpha, though detailed balance would indicated that this folding rate is 10^{14} times slower than *alpha* → *beta*. No folds were observed.

The rate of *alpha* → *beta* interconversion was about 0.0005 per million MC steps, after a lag of about 10^8 steps. Linear regression to this Poisson-like process displayed a near-perfect correlation of 0.998. Folding to beta from alpha is slower than folding to beta from the unfolded state (which was about 0.0009 per million MC steps), confirming that alpha is stable with respect to the unfolded state.

3. Discussion

3.1. Cumulative remaining folds

The folding histograms presented thus far have been nondecreasing, as the number of folds as a function of time can only increase. Each run was stopped at the point of folding, disallowing the possibility that a run which reaches alpha in time t_1 will convert from alpha to beta in the remaining time $t_{\max} - t_1$. It is possible to combine data from Figs. 13 and 15 to form a histogram of *cumulative remaining folds*. Under the assumption that beta will never (practically) convert to alpha, the cumulative remaining alpha folds, denoted by $N_{\alpha,\text{rem}}$, is the fraction of runs which have reached alpha (N_α) *minus* that fraction of alpha folds which subsequently converted to beta. Correspondingly, the cumulative remaining beta folds $N_{\beta,\text{rem}}$ is the fraction of runs folded directly to beta from the unfolded state (N_β) plus that fraction of alpha folds which

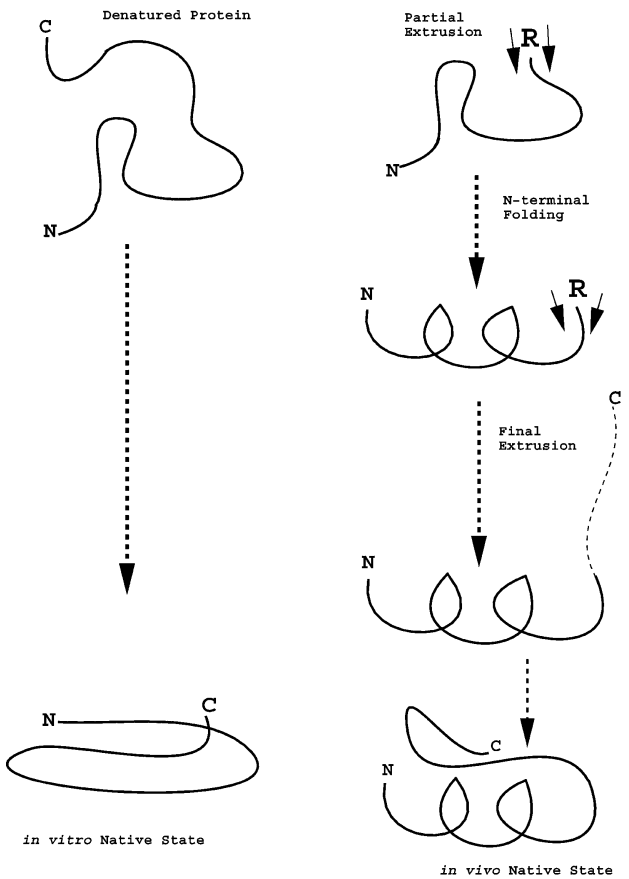

Fig. 14. Schematic for the data in Fig. 13. Cotranslational folding changes the kinetically accessible 'native' state from the long-range contact-rich conformation beta to the local contact-dominant conformation alpha.

subsequently converted to to beta. It is the second term in each of these expressions which must now be computed.

Taking $N_\alpha(t)$ and $N_\beta(t)$ from Fig. 13 and $N_{\alpha\to\beta}(t)$ from Fig. 15, one can define

$$H_\alpha(t) = \frac{dN_\alpha(t)}{dt} \text{ and } H_\beta(t) = \frac{dN_\beta(t)}{dt}$$

as the rate of folding to the respective states at time t, and

$$H_{\alpha\to\beta}(t) = \frac{dN_{\alpha\to\beta}}{dt}$$

as the instantaneous rate of interconversion. Then the conversion correction is a convolution of $H_\alpha(t)$ and $H_{\alpha\to\beta}(t)$, yielding

$$H_{\alpha,\mathrm{rem}}(t) = H_\alpha(t) - \int_{t_1=0}^{t_{\max}} H_\alpha(t_1)H_{\alpha\to\beta}(t_{\max} - t_1)dt_1 \quad (3)$$

and

$$H_{\beta,\mathrm{rem}}(t) = H_\beta(t) + \int_{t_1=0}^{t_{\max}} H_\alpha(t_1)H_{\alpha\to\beta}(t_{\max} - t_1)dt_1. \quad (4)$$

The cumulative remaining folds may be expressed as

$$N_{\alpha,\mathrm{rem}}(t) = \int_{t_1=0}^{t_{\max}} H_{\alpha,\mathrm{rem}}(t_1)dt_1 \quad (5)$$

and

$$N_{\beta,\mathrm{rem}}(t) = \int_{t_1=0}^{t_{\max}} H_{\beta,\mathrm{rem}}(t_1)dt_1 \quad (6)$$

Fig. 16 plots $N_{\alpha,\mathrm{rem}}$ and $N_{\beta,\mathrm{rem}}$ as a function of MC time,

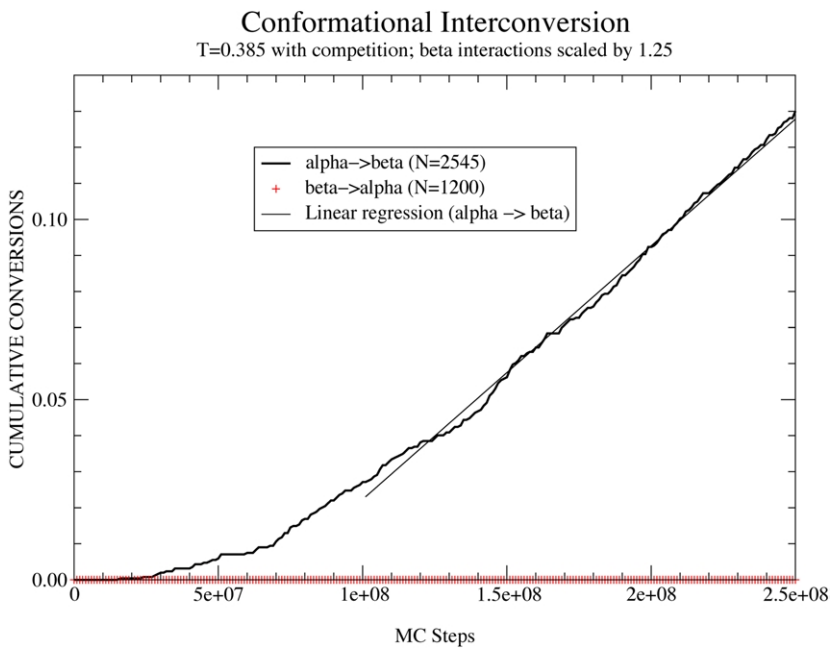

Fig. 15. Interconversion histogram, starting from conformation alpha and awaiting conversion to beta. ($T = 0.385$; $\omega_{\mathrm{alpha}} = -1.0$; $\omega_{\mathrm{beta}} = -1.25$.) After a lag time, interconversion becomes Poisson-like, or roughly linear with MC time. (For the linear regression shown, the correlation $\mu = 0.998$.) Conversion from alpha to beta is markedly slower than from the unfolded state to beta (see Fig. 13), indicating that the alpha backbone maintains significant stability at this temperature. As a control, interconversion from beta to alpha was tested. No folds were expected due to the high stability of the beta form; no folds were observed out of 1200 runs.

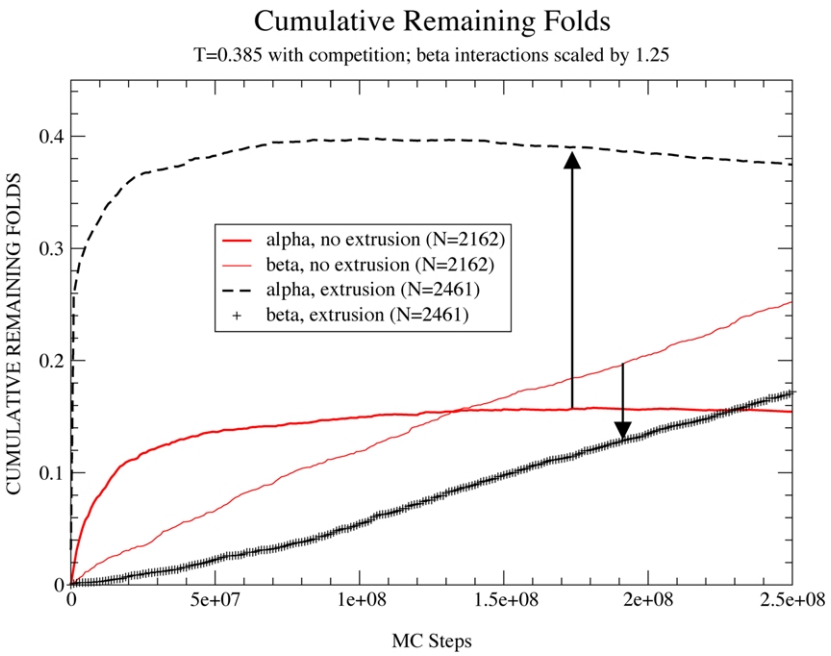

Fig. 16. Cumulative remaining folding fraction at $T = 0.385$; $\omega_{\text{alpha}} = -1.0$; $\omega_{\text{beta}} = -1.25$, with cutoff time of 250,000,000 MC steps. Cumulative remaining folds to alpha at time t are those runs which folded to alpha and *did not* convert to beta by this time. Cumulative remaining folds to beta are those runs which folded to beta either directly from the unfolded state, or to beta after previously folding to alpha. Data are presented with and without extrusion. The data is nearly identical to that of Fig. 13, in that (a) folding to alpha is assisted by extrusion; (b) folding to beta is inhibited by extrusion; and (c) extrusion shifts the folding from beta to alpha, even after 400 × the characteristic extrusion time.

up to 250,000,000 steps. Note that $N_{\alpha,\text{rem}}$ is no longer nondecreasing, as the curve now reflects the number of runs which folded to backbone alpha minus those which folded to alpha and then converted to beta. The new results thoroughly uphold the conclusions from Fig. 13: extrusion shifts the preferred folding state from beta to alpha.

The cumulative remaining folds data (Fig. 16) demonstrate that, even at 400 × the extrusion time, there are still twice as many remaining alpha folds as beta folds. This rules out the possibility that visiting alpha is a brief affair *en route* to settling at beta, over the time scales studied. Of course, at much longer times, all runs are expected to settle at the enormously stable beta conformation.

3.2. Biological ramifications and conclusions

In developing and testing an extrusion model of lattice protein folding, it was shown that extrusion can speed folding, especially to conformations with many local contacts. In addition, it is possible to induce folding of a single model protein sequence to different native states by turning cotranslation on and off. This is of course a kinetic phenomenon, as over a sufficiently long time period the system would be expected to achieve thermodynamic equilibrium. However, times needed to reach equilibrium may be quite long, making the 'metastable' state highly relevant.

It was also demonstrated that extrusion plays an important role in folding only for temperatures at which the native state is stable. At less stable (higher) temperatures, there is rapid equilibration among states, and no chance for a persisting 'memory' of extrusion. (For the preceding Go-model simulations, stable temperatures were somewhat below optimal folding temperatures, though this may reflect the unusually strong kinetic traps induced by the Go potential. [13] In sequence model simulations and for real proteins, fast-folding often occurs under conditions of considerable native-state stability; thus extrusion may be relevant at fast-folding temperatures as well as in the low-temperature regime.) Specifically, for structures which fold cotranslationally, native contacts formed during extrusion must be stable enough to persist until extrusion is complete.

The typical time scale for ribosomal translation is 1–10 residues per second, while folding from the denatured state is known to occur in the range of milliseconds to minutes [14] for proteins studied to date. Cotranslation cannot affect folding kinetics when folding times are much shorter than translation times, except in the trivial sense that folding cannot finish until translation is complete. The low-temperature regime, in which folding is slowed markedly compared to its optimal rate, is one of slow folding with respect to translation. It is this regime which shows the largest shift in folding kinetics toward states rich in local contacts.

The kinetic effects of cotranslation were long-lived, shifting the temporary, 'metastable equilibrium' toward one conformation over a time more than 400 × that of cotranslation itself. Functional metastability is a known property of several proteins, including poliovirus [2,3] and

plasminogen activation inhibitor [4,5]; thus a thermodynamically stable state need not be the functional state of a protein.

A protein which is known to have two biologically relevant conformations is the mammalian prion protein, which has an endogenous form PrP^C and a largely uncharacterized pathogenic form PrP^{Sc}. (PrP^{Sc} has been shown to be the causal agent in diseases such as scrapie and mad cow disease [15]). Since the endogenous form is known to be more helical [16] (local-contact dominant) than the pathogenic form, an interesting hypothesis is that extrusion guides folding toward PrP^C and away from PrP^{Sc}. Experiment has ruled out this hypothesis as a complete explanation for prion behavior, since denaturing and refolding PrP^C yields a non-active form which is not PrP^{Sc} (and may indeed not be a specific native conformation) [17], and since there is evidence that PrP^{Sc} may fundamentally be an aggregative state [18,19]. Nevertheless, cotranslational folding may be a factor in the reliable folding of the prion protein to its endogenous form, and may still contribute to the avoidance of endogenous PrP^{Sc} production.

The model of extrusion implemented in these studies may not be optimal for studying all cotranslational phenomena. For instance, proteins may be extruded a domain at a time, rather than one monomer at a time. Nevertheless, there is no reason to suspect a fundamental divergence in behavior based on the extrusion schedule. The elimination of trapping by extrusion is not scale-dependent, and likely supports the folding of proteins by domain (a 'local' contact topology on a larger scale than the one we have discussed), in comparison to folding into an undesired enormous single domain with or without a global energy minimum.

The demonstration of the strong role of cotranslation on model protein folding kinetics raises a number of interesting biological questions, not the least of which is the global free-energy minimum hypothesis. Extrusion, combined with the known influence of contact topology on folding, may give rise to proteins which are metastable, but with characteristic conversion times longer than the useful life of the protein, even without the influence of post-translational covalent modification.

4. Materials and methods

4.1. The lattice model

Proteins are represented as self-avoiding chains on a cubic lattice, with each residue occupying one lattice site. Residues which are nearest neighbors on the lattice but are *not* covalently bonded along the chain interact with a potential which will be described below.

The folding algorithm was Metropolis Monte Carlo [20], which is known to yield a canonical distribution of states and is thus an ideal tool for studying thermal processes. [21]. The move set was that of Hilhorst [22], utilizing three standard moves: crankshaft, corner flip, and tail moves, plus diffusion moves for chains of one or two monomers. One Monte Carlo (MC) step consists of one randomly chosen move, whether or not that move is accepted according to the Metropolis criterion.

4.2. Go potential

In order to easily tune the parameters of folding, we use the potential of Go [23,24]. The Go model was one of the first potentials used in simplified protein modeling, and is still extremely useful in that same field. Recall that in our previous lattice simulations the total energy was expressed as the pairwise sum of contact energies:

$$E = \sum_{i=1}^{N} \sum_{j<i} B(\xi_i, \xi_j)\Delta_{ij} \tag{7}$$

The term ξ_i is the sequence identity of residue i; $\Delta_{ij} = 1$ if residues are in contact, $\Delta_{ij} = 0$ if not. $B(\xi_i, \xi_j)$, abbreviated as B_{ij}, is the parameter set matrix. For *sequence models* such as that of Miyazawa and Jernigan, [25] B_{ij} is an $L \times L$ matrix, where L is the number of distinct amino acids (usually 20). The Go model is a specific type of sequence model, which sets $L = N$ (the number of residues in the protein chain). The primary advantage of the Go model in this study is that sequence design becomes trivial: simply set $B_{ij} = \Delta_{ij}^N$, where $\Delta_{ij}^N = -1$ if residues i and j are in contact *in the native state*, and 0 otherwise. While sequence models are more useful in studying protein design and evolution, fundamental folding properties are quite similar between Go and sequence models [26].

4.3. Implementation; time measurement

The implementation of extrusion requires a time table for the sequential extrusion of monomers. Our implementation extrudes one residue at a time, though it would be a trivial modification to extrude several residues at a time.

The time constant M is introduced as the number of simulation steps *per extruded residue* between the extrusion of sequential monomers. The total number of Monte Carlo steps between the extrusion of monomers n and $n+1$ is nM, rather than simply M, in order to properly count time. Normally in Monte Carlo simulations, time may be counted as either the total number of attempted moves (as is done here) or as the number of attempted moves *per monomer*. The convention is merely a matter of preference when the system is of fixed size, as time may be rescaled by a constant. With extrusion, however, the situation is different as the system size changes with time. Attempting ten moves for a one-monomer system has a different meaning than attempting ten total moves for a ten monomer system.[5] If

[5] The vibration and rotation of bonds in physical systems occur with time constants which are roughly independent of system size; therefore the total number of vibrations or rotations scales with system size.

extrusion is to be modeled as linear with time, then the average number of moves *per extruded residue* must remain constant.

Extrusion simulations are run as follows.

1. Begin with only one residue (the N-terminus). Allow this residue to move M times. This step is trivial, as only one move is allowed (diffusion of the monomer from one lattice location to a neighboring lattice location).
2. Add a residue in one of the available sites adjacent to residue 1. (This site is selected randomly among all available neighboring sites.) Simulate the 2-residue system for $2M$ moves.
3. Continue to add residues, simulating the chain for nM steps after the extrusion of residue n.
4. Once the chain is completely extruded, continue the simulation for a fixed number of steps, or until whatever desired event (such as folding to a particular state) occurs.

We chose the value $M = 1000$, meaning that approximately 650,000 total simulation steps were executed before the chain was fully extruded. This value allows a moderate level of equilibration after each monomer is extruded, but such that the total extrusion time is small compared to low-temperature folding events.

Folding without extrusion was begun in each case from a new random denatured conformation, generated by 10,000 Monte Carlo steps of high-temperature simulation ($T = 500$).

4.4. Quality and relative quality

It is usual to measure the progress of a folding simulation, with respect to Monte Carlo time, using the 'quality' Q:

$$Q = \frac{n_c}{N_c} = \frac{\text{Current number of native contacts}}{\text{Total number of native contacts}}$$

Since we are interested in two backbones, we must now define two qualities:

$$Q_1 = \frac{n_{1,c}}{N_{1,c}}$$

$$Q_2 = \frac{n_{2,c}}{N_{2,c}}$$

The two target backbones need not possess the same number of native contacts. Note that $Q_1 + Q_2 \geq 1$, where equality exists if and only if the two backbones have no contacts in common.

Folding to state A is confirmed when $Q_A = 1$.

We introduce a new quantity, called *relative quality* q, which is the number of native contacts formed *relative to the number of extruded native contacts* N_c^{extr}. For a given backbone i,

$$q_i(N^{extr}) = \frac{n_{i,c}}{N_{i,c}^{extr}} \tag{8}$$

Relative quality is useful as a measure of folding relative to the extrusion process. If folding occurs cotranslationally (along a linear ECP) to conformation A, we would expect $q_A \approx 1$ throughout the extrusion process. This would not be the case if the ECP is highly concave, discouraging cotranslational folding.

References

[1] Levinthal C. Are there pathways for protein folding? J Chem Phys 1968;65:44–5.
[2] Ansardi D, Morrow C. Encapsidation and serial passage of a poliovirus replicon which expresses an inactive 2a proteinase. J Virol 1995;69:1359–66.
[3] Wien M, Chow M, Hogle J. Poliovirus—new insights from an old paradigm. Structure 1996;4:763–7.
[4] Sharp A, Stein P, Pannu N, Carrell R, Berkenpas M, Ginsburg D, Lawrence D, Read R. The active conformation of plasminogen activator inhibitor 1, a target for drugs to control fibrinolysis and cell adhesion. Structure 1999;7:111–8.
[5] Im H, Seo E, Yu M. Metastability in the inhibitory mechanism of human alpha(1) antitrypsin. J Biol Chem 1999;274:11072–7.
[6] Plaxco K, Simons K, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. J Mol Biol 1998;277:985–94.
[7] Abkevich V, Gutin A, Shakhnovich EI. Impact of local and non-local interactions on thermodynamics and kinetics of protein folding. J Mol Biol 1995;252:460–71.
[8] Lindberg M, Tangrot J, Otzen D, Dolgikh D, Finkelstein A, Oliveberg M. Folding of circular permutants with decreased contact order: general trend balanced by protein stability. J Mol Biol 2001;314: 891–900.
[9] Buchner J. Supervising the fold: functional principles of molecular chaperones. FASEB J 1996;10:10–19.
[10] Fedorov A, Friguet B, Djavadi-Ohaniance L, Alakhov Y, Goldberg M. Cotranslational protein folding. J Mol Biol 1992;228:351–8.
[11] Fedorov A, Baldwin T. Cotranslational protein folding. J Biol Chem 1997;272:32715–8.
[12] Abkevich V, Gutin A, Shakhnovich E. Theory of kinetic partitioning in protein folding with possible applications to prions. Proteins 1998; 31:335–44.
[13] Kaya H, Chan H. Origins of chevron rollovers in non-two-state protein folding kinetics. Phys Rev Lett 2003;90:258104.
[14] Creighton TE. Proteins: structures and molecular properties, 2nd ed. San Francisco, CA: W.H. Freeman and Co.; 1993.
[15] Liemann S, Glockshuber R. Transmissible spongiform encephalopathies. Biochem Biophys Res Commun 1998;250:187–93.
[16] Billeter M, Riek R, Wider G, Hornemann S, Glockschuber R, Wuthrich K. Prion protein NMR structure and species barrier for prion disease. Proc Natl Acad Sci 1997;94:7281–5.
[17] Prusiner S. Prions. Proc Natl Acad Sci 1998;95:13363–83.
[18] Lansbury P, Caughey B. The chemistry of scrapie infection: implications of the 'ice 9' metaphor. Chem Biol 1995;2:1–5.
[19] Morrissey M, Shakhnovich E. Evidence for the role of PrP_C helix-1 in the hydrophilic seeding of prion aggregates. Proc Natl Acad Sci 1999; 96:11293–8.
[20] Metropolis N, et al. Equation of state calculations by fast computing machines. J Chem Phys 1953;21:1087–92.

[21] Binder K, Heerman DW. Monte Carlo simulation in statistical physics, 2nd ed. New York: Springer; 1992.
[22] Hilhorst H, Deutch J. Analysis of Monte Carlo results on the kinetics of lattice polymer chains with excluded volume. J Chem Phys 1975; 63:5153–61.
[23] Go N. Noninteracting local-structure model of folding and unfolding transitions in globular proteins. i. formulation. Biopolymers 1981;20: 991–1011.
[24] Abe H, Go N. Noninteracting local-structure model of folding and unfolding transitions in globular proteins. ii. applications to two-dimensional lattice proteins. Biopolymers 1981;20:1013–31.
[25] Miyazawa S, Jernigan R. Estimation of effective interresidue contact energies from protein crystal structures: Quasi-chemical approximation. Macromolecules 1985;18:534–52.
[26] Clementi C, Vendruscolo M, Maritan A, Domany E. Folding lennard-jones proteins by a contact potential. Proteins: Struct, Funct, Genet 1999;37:544–53.