Efficient expansion, folding, and unfolding of proteins

Erik D. Nelson* and Nick V. Grishin
Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, Texas 75235-9050, USA
(Received 19 February 2004; revised manuscript received 21 May 2004)

We consider a nonstatistical, computationally fast experiment to identify important topological constraints in folding small globular proteins of about 100–200 amino acids. In this experiment, proteins are expanded mechanically along a path of steepest increase in the free space around residues. The pathways are often consistent with folding scenarios reported in kinetics experiments and most accurately describe obligatory or mechanic folding proteins. The results suggest that certain topological "defects" in proteins lead to preferred, entropically favorable channels down their free energy landscapes.

DOI: XXXX PACS number(s): 87.15.Cc

As a protein unfolds, it encounters dynamic constraints that emerge as a consequence of its being folded into a particular low-resolution structure or topology. For example, it often occurs that parts of protein are entangled or wrapped within its interior, and for these "frustrated" parts to unfold requires the rest of the protein to reorganize and at least partially unfold first. At this level of resolution, topological constraints can impose a time order on unfolding events and occasionally this order can be recognized in a protein's actual nucleation process [1–6] or folding "pathway" despite the extreme complexity of its interactions.

Generally, it is recognized that a protein's native topology is a dominant factor affecting how it folds [7]. For example, the folding rates of small proteins of about 100 amino acids are well described by pure topological measures (such as relative contact order [7,8]). And although a protein's folding pathway(s) can depend sensitively on sequence [9], the nucleation features of most (including even large) proteins for which thorough nucleation kinetics data are available are described relatively accurately by simplistic, energetically nonfrustrated [10] models (or G$\bar{o}$ models [1–5,11]) where the topography of the free energy landscape is determined just by native topology.

Even these minimalist simulations are computationally intensive, but a number of recent results suggest simpler ways to explore the effects of topology on folding. In particular, elegant experiments that sample the chemically denatured states of proteins reveal a surprising amount of native order in protein unfolded ensembles [12–15]. All-atom simulations describe unfolded proteins in terms of secondary "mean structures" [16,17] that fluctuate locally about their native orientations on a time scale that is more rapid than that of tertiary organization [18]. It is natural to imagine folding in terms of such structures [15], and it seems worth asking whether anything useful could be learned just from a "mechanically efficient unfolding" of a topological model (such as a chain of flexible secondary structures or tubes [19–21])—for example, could such a model capture the basic order of events observed in topologically based folding models?

In this paper, we consider a simple topological experiment that, while not starting from the point of flexible secondary structures, effects them as a result.

To put this experiment into context, we first briefly review a statistical method developed some time ago by Galzitskaya and Finkelstein [22,23] to predict protein folding nuclei. In that work, proteins are considered to unfold in fixed steps in which a number of residues ($\sim$1) change from native to unfolded, dissolving all their energetically favorable native contacts and acquiring entropy dependent on their location in the surface loops or free ends. To predict the folding nucleus, an efficient search is performed for paths that minimize the free energy barrier at the folding transition temperature of the model. Although the model is purely structural, it does not include any dynamic constraints (e.g., of dihedral angle and excluded volume type) and any residue can be selected to unfold in an unfolding step [24]. Remarkably, the contact energy and loop entropy functions are able to select qualitatively accurate unfolding routes:

In this work, we explore a complementary problem in which *only* topological-dynamic constraints are considered. The dynamics of this model reflect the fact that minimal entropy loss (in the sense of loop closure [25]) and maximum entropy [26] amount to the same condition. In order to approximate this condition, we simply expand the native conformations of proteins to most efficiently increase free space around the residues [27]—specifically, we apply a repulsive potential between protein C^{α}'s and unfold along the path of steepest decent in the total potential subject to main-chain topological constraints.

To describe the effect of the constraints let us start with a set of disconnected α carbons located in the native positions of a protein. The unfolding path of this cluster is determined by its shape—depending on the potential, a spherical cluster dissolves from its surface, a needle like cluster dissolves from its ends, etc. Next, a protein has a much different unfolding path because its residues are (i) constrained to fixed distances along the chain and (ii) because the relative motion of a pair of α carbons gets, in effect, transmitted accross the chain to other parts of the protein (Fig. 1). For instance, a

*Corresponding author.
Electronic address: enelson@spirit.sdsc.edu

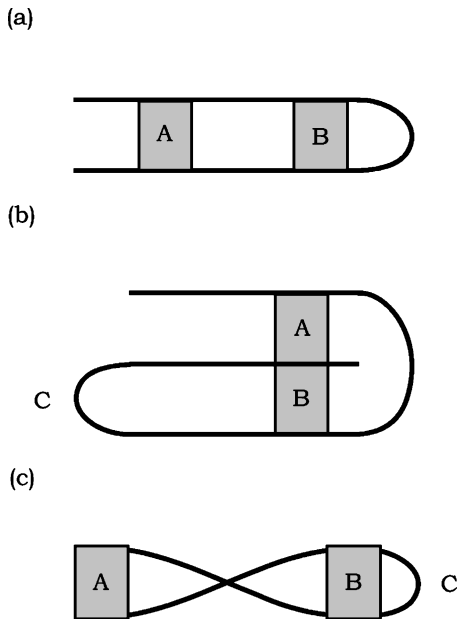

FIG. 1. Schematics of mechanical frustration in protein topologies (after Toulouse [28]). The shaded boxes represent repulsive interactions between the segments they connect. The hairpin (a) unfolds from its ends inward, and the distancing of any pair of chain segments (at *A* for example) by a repulsive potential is consistent (cooperates) with the others (i.e., such as those at *B*). However, for the enclosed hairpin (b) and the hairpin braid (c), the segment pairs at *A* and *B* send conflicting signals to dihedrals in the turn at *C*. All of these situations occur in the proteins described in the figures below.

hairpin [Fig. 1(a)] is confined to unfold from its free ends since it is connected at the turn. The separation of each cross-interacting pair of C^{α}'s is, however, consistent with motion that would separate (i.e., lower the energy of) the others. By contrast, in the topology of Fig. 1(b) the cross-interacting C^{α}'s on opposite sides of the enclosed segment send conflicting signals [28] through the chain to the dihedrals in the small turn—the chain is confined to unfold ("unwrap") from its exterior end. Finally, in Fig. 1(c) the separation of cross-connected C^{α} pairs in the ends is inconsistent with pairs on the turn side of the hairpin braid. Again, the braid of the hairpin causes conflicting signals to be sent to the dihedrals in the turn region and the chain is topologically frustrated from unfolding except near its free ends.

Accordingly, when applied to proteins the method identifies "defect" regions that are either confined or frustrated from unfolding, and our results suggest that these regions are the source of entropically favorable routes [22] down the free energy landscape. As in Fig. 1(a), less frustrated topologies lead to more delocalized routes because the chain is able to unfold more cooperatively. This is similar to the description of Shea *et al.* [1] where topological frustration is characterized by the shape of the distribution of ϕ values [29] in the transition ensemble. Then, less frustrated topologies lead to unimodal distributions of ϕ values consistent with more delocalized (less obligatory [30,31]) folding as in our mechanical description above.

In the rest of the paper, we study a large group of proteins comprising most nonhomologous folds found in the folding literature for which kinetics data are available [1–5,31–45], focusing on proteins that have been studied by topologically based simulation methods [1–5]. The main chains of these proteins are unfolded recursively, and the decay of native tertiary contacts is compared to experimental folding chronologies. The unfolding paths are nonchaotic [46] as long as the unfolding rate is much less than that needed to cause collisions during individual steps of the algorithm. Similar topologies (such as src and α-spectrin SH3) lead to similar temporal patterns of native contact dissociation, and generally such patterns qualitatively agree with the kinetic folding chronologies. In the next section, we outline the methods used in the unfolding experiments and in subsequent sections we discuss the results and explain when and why the model is expected to work.

Methods. We unfold protein crystal and NMR solution structures [47] by the following prescription: At each site n along the chain we calculate partial derivatives ($\delta F_n/\delta\phi_n$ and $\delta F_n/\delta\psi_n$) of the parameter

$$F_n = \sum_{i<n,\ j>n} w(\boldsymbol{r}_i - \boldsymbol{r}_j), \tag{1}$$

where $w(\boldsymbol{r}_i - \boldsymbol{r}_j)$ is the repulsive potential acting between C^{α} atoms (at positions $\boldsymbol{r}_j$). To complete a step in the recursion, the dihedral angles are given a simultaneous kick $\Delta\phi_n, \Delta\psi_n$ proportional to the partial derivatives where the proportionality constant is adjusted to keep the step sizes relatively small so that locally the system unfolds smoothly and continuously [48].

The unfolding paths are self-avoiding, and although we make no attempt to model amino acid (local) dihedral restrictions, the deflections between native and unfolded dihedral angles are small enough ($\lesssim 10°$) that these restrictions are at least marginally satisfied (as long as they do not begin in a forbidden region of Ramachandran space [30]). Native contacts dissociate roughly linearly as a function of the mean energy per residue. The chains unfold primarily by unwinding loops and turns that connect secondary structures together since it is these dihedral groups that lead to the dominant response in terms of increasing C^{α} separations. Although the method does not unfold helices in a satisfactory way (without eventually encountering steric conflicts), this turns out to have a negligible affect on the results because tertiary contacts dissolve on a so much faster time scale [49].

We study 17 different protein topologies, including escherichia coli CheY, villin, staphylococcal nuclease, RNase H, lysozyme, ADA2h, FKBP12, chymotripsin inhibitor, barnase, proteins B and E, α-spectrin SH3, its circular permutants, and structural analogs and homologs—all of these examples are discussed and several are illustrated below. To explore sensitivity to the conditions in Eq. (1), we unfold structural homologs of these topologies using alternate rates of unfolding and different ranges of two different (inverse power and screened charge) pair potentials. The paths are not always sensitive to the choice of range but for a number of (even mechanic folding) proteins obvious disagreements with kinetics results appear (such as no apparent nucleus) when the range is decreased. Consequently, we select the longest range potential $w(r) = r^{-1}$ to represent the examples

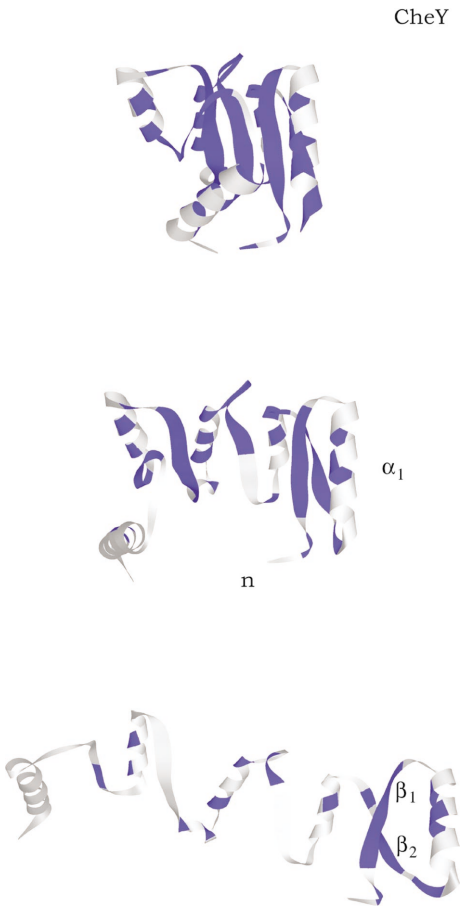

FIG. 2. (Color) Unfolding of CheY (pdb id 3chy). To avoid repetition in Figs. 2–5, we refer to panels of progessively unfolded states by (a), (b), (c), etc., even though they are not labeled as such. (a),(b) The native topology initially unwraps against its cylindrical curvature and the large domain (connected to the nucleating domain $\beta_1\alpha_1\beta_2$) begins to unfold. (b),(c) The nucleating domain remains structured until only isolated, short-range contacts exist between all other turns joining secondary structure segments.

[50]. In a few cases it is necessary to add a short-range repulsive part $\sim r^{-6}$ (significant in the range <4 Å) to enforce excluded volume constraints, and we note these situations when they appear below.

In continuous unfolding paths, proteins unfold by dissociating individual native contacts, and here a residue is considered "unfolded" once all its native tertiary contacts have dissolved. To identify native contacts we apply the following condition [1] to the initial pdb file: A contact is registered between two residues if (i) the side chains (any pair of heavy atoms) are less than 4.5 Å apart and (ii) their C^α atoms are less than 8 Å apart [if either residue is glycene, the contact is calculated from condition (ii) only]. When the main chain unfolds, native contacts are considered to have dissolved when condition (ii) is no longer satisfied. In the figures, we represent only the behavior of contacts between residues separated by $\geqslant 6$ units along the chain (about one bonding unit longer than needed to form a β turn). Sites that participate (do not participate) in native contacts are colored blue (light gray). When all contacts originally formed with a par-

FIG. 3. (Color) Unfolding of villin 2vik. (a),(b) The aromatic core formed by helix α_3 and stands $\beta_4\beta_5$ unfolds, releasing the left edge of the β sheet. (b)–(d) The β sheet unfolds while contacts in the alipatic core $\alpha_2\beta_4$ and the enclosed hairpin $\beta_1\beta_2$ persist. In the transition state, the aliphatic core is expected to be much more structured than the aromatic core. The small hairpin nucleus predicted in the unfolding path was not sampled in the protein engineering experiments.

ticular site have dissolved the color of the site changes from blue to light gray.

Basic examples. To explain the results of this experiment, we select several large proteins to discuss in detail (CheY, villin, SNase, and RNase H). In subsequent sections, we describe what one can expect to recover from this type of experiment and then briefly summarize the results for other proteins from the list above.

In vitro, CheY (Fig. 2) folds by nucleation and condensation. The $\beta_1\alpha_1\beta_2$ domain is structured in the transition state and is believed to be the nucleus around which the rest of the protein folds [2,32]. Part of this nucleating domain (the strand β_1) is wrapped into the core [similar to Fig. 1(b)], and in the unfolding path the second domain is required to unwrap from around β_1 (dissolving the first hydrophobic core) before the nucleating β_1-β_2 domain can start to unfold. When the protein unfolds, β_1-β_2 remains structured until the second (outer) hydrophobic domain is unfolded and only isolated contacts exist between all other turns joining secondary structure segments, in qualitative agreement with both experiment and Gō-model molecular dynamics simulations [2,32].

The ordering of two hydrophobic cores in the folding reaction of villin (Fig. 3) were studied by protein engineering methods, and the folding nucleus was was found to include one of the cores preferentially [33]. The unfolding path for the NMR minimized average structure of villin clearly identifies the correct (aliphatic) core nucleus [33] between helix α_2 and β strand β_4 (the aromatic core formed by α_3 and β_4 unfolds first). This path also predicts an additional small "nucleus" in the hairpin $\beta_1\beta_2$ which was not probed by the mutations in these experiments and could be structured along with the aliphatic core.

Staphylococcal nuclease (Fig. 4) is predicted to fold through a three-state mechanism (the intermediate state is rate limiting [34,35]). Folding begins by formation of the N terminal β sheet, $\beta_2\beta_3$, and continues with the acquisition of α-helical structure throughout the rest of the molecule which results in a nativelike domain that docks to β_1-β_3 as a final step. In the unfolding path, the most prominent feature is the persistence of the β domain containing $\beta_2\beta_3$, followed by the persistance of contacts near the turn region of $\beta_4\beta_5$—these regions persist due to mutual confinement or complementarity (they form halves of the β barrel region of the protein). In the first stage of unfolding, these two regions decouple as units along a line connecting them. This stage is followed by decay of α helical contacts in the second domain (in which the sheet $\beta_4\beta_5$ remains partly structured) and the first domain β_1-β_3 which remains mostly structured. In the final stage, β_1-β_3 is mostly structured, $\beta_4\beta_5$ is partly structured, and the loop connecting the first helix to the first β domain has a few isolated contacts—all other (single) contacts being isolated at turns connecting secondary structure elements. The results identify two of the main features of SNase folding (nucleation and docking), and since the decay of tertiary contacts closely follows the folding scenario *in vitro* [34,35], we beleive the rate limiting stage of folding may involve formation of the $\beta_4\beta_5$ sheet [34]. We note that the unfolding path for a quadruple mutant of SNase (1ey9) is basically the same as that in Fig. 4 except that the loop segment is less persistent (one of the mutations in 1ey9 affects this loop).

Folding of RNase H (Fig. 5) has been studied by both hydrogen exchange and Gō-model molecular dynamics simulations [2]. Again, part of the protein (the sheet β_1-β_3) is wrapped into the interior, here by the C-terminal α-helix α_5. The domain formed by these ends is connected to the helical region α_2-α_4 through a braid [see Fig. 1(c)] near a solvent protected region of the protein (including $\alpha_1, \beta_4, \alpha_2$, and α_4). The protein unfolds [51] by expanding about this protected part; as it expands, the ends (α_5 and β_1-β_3) pull apart, allowing the sheet to break contact with the helix α_1 and the strand β_4. Finally, the protected part of the protein is allowed to expand and the most persistent contacts connect α_1, β_4, and α_4 (the internal β-sheet contacts persist throughout this process). Our results agree with the scenario interpreted from Gō-model folding simulations and suggest that the critical folding event is formation of the $\alpha_1, \beta_4, \alpha_4$ (braid) region.

However, we note that although the helical contacts with α_2 decay, the helical domain is frustrated from expanding until after the braid, and to be consistent, this should be interpeted as "helical and protected regions fold together." This result (and, to some extent, that in Fig. 3) begins to uncover the limits of this approach related to the neglect of side-chain volume restrictions (see the Discussion section) and the fact that secondary structures move as cooperative units. In RNase H, the helical region is relatively large and more dynamically independent from the rest of the protein than our model suggests; hence, it may not be accurate to describe the folding process in terms of a single linked sequence, or channel, of topological causes and effects. What does the method recover in such cases?

A structurally similar situation occurs in lysozyme [52,53] (not shown) which folds along two pathways corresponding to preferential ordering of either its β sheet or α-helical domains (in lysozyme, the β sheet acts similar to a large loop or turn connecting the two α-helical ends). In the dominant folding path, (i) the α-helical ends of lysozyme join to fold the α domain first, while in the alternate path (ii) the three stranded β sheet folds first. The dominant route (i) is not the fastest folding pathway (extensive optimization for folding rate does not seem to be a high priority in protein evolution [54]) but the unfolding experiment still selects the fastest route.

It is interesting to compare lysozyme to the two-dimensional (2D) lattice (Gō) model [55] used by Ozkan *et al.* to explain "nonclassical" ϕ values in protein engineering [29] experiments. The native state of this model also consists of two domains (a lattice helix and sheet "sandwiched" together), and there are two folding pathways which correspond to (i) nucleation by the loop joining the domains or (ii) independent nucleation of the helix and sheet. The two paths emerge because the lattice moves do not allow the domains to move accross the lattice without at least partially dissolving their native shapes. In path (i) this problem is resolved in the initial nucleation event, while in path (ii) independent domain folding leads the protein into a free energy trap (see Fig. 3 of Ref. [55]). Thus, path (i) is the fastest folding path of the model [note that in path (i) of lysozyme, partial unfolding of the β domain is not required to reach the native state [56]].

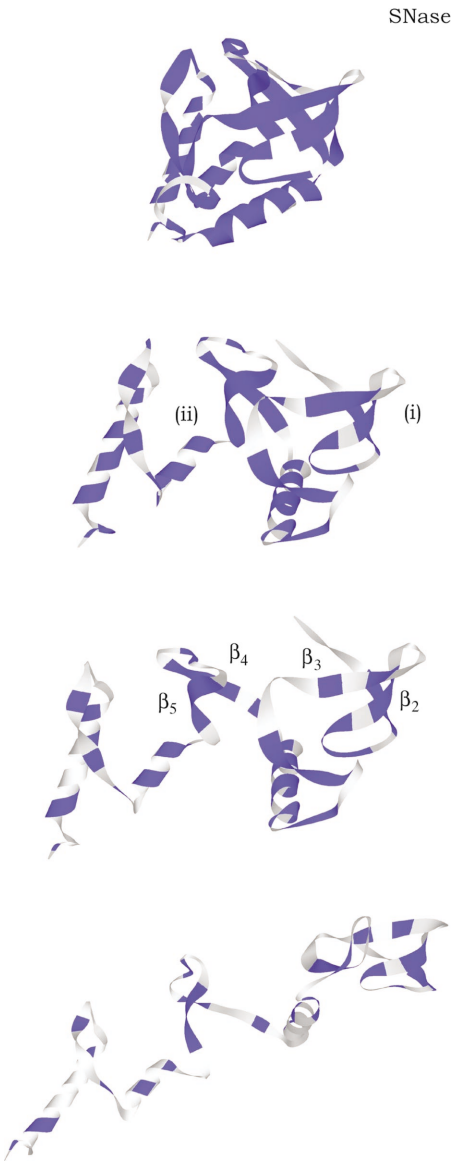

FIG. 4. (Color) Unfolding of a P117G mutant of SNase 1ez6. (a), (b) In the first stage of the unfolding process, the β sheet (i) and α-helical (ii) domains decouple as units along a line joining them. (b)–(d) This stage is followed by the decay of α helical contacts in domain (ii) and further exposure of the core. Finally, in (d) the nucleating domain is still structured, $\beta_4\beta_5$ is partly structured, and the loop connecting the first helix to β_3 has a few isolated contacts—all other (single) contacts being isolated at turns connecting secondary structure units.

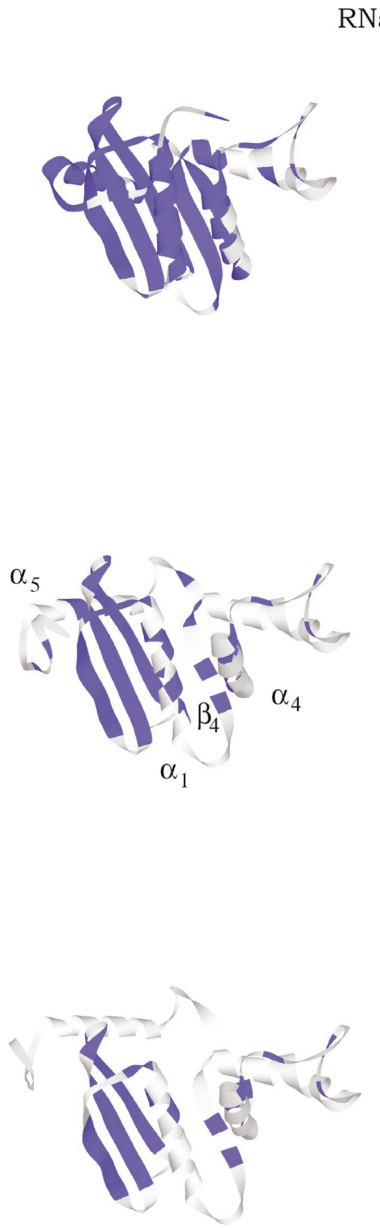

FIG. 5. (Color) Unfolding of RNase H 2rn2. (a),(b) Initially, the C-terminal α-helix α_5 unwraps from around the β sheet β_1-β_3, releasing the protected (braid) region of the protein ($\alpha_1, \beta_4, \alpha_2$, and α_4). (b),(c) The β sheet breaks contact with the protected region, which is then allowed to expand. The most persistent contacts are interior to the sheet β_1-β_3 and in the protected part including α_1, β_4, and α_4.

In each of these examples, the fastest folding path is initiated by a constrained or frustrated fold region whose formation tends to expedite folding the rest of the molecule—our results suggest that this situation is typical. However, lysozyme unfolding is cooperative [see Fig. 1(a)], so it would seem more appropriate to view its fast folding path as an entropic (loop closure) effect. This effect is intertwined with topological frustration but the two ideas can be distinguished. For instance, in each schematic in Fig. 1, native contacts dissociate in order of sequence separation, but less cooperative topologies should lead to greater dispersion among contact unfolding times (see Fig. 6).

To study these ideas carefully, it may be useful to think of dividing larger proteins into cooperative substructures (hairpins, sheets, etc.). If one considers that proteins are grown by joining cooperative substructures topologically, the potential field around a substructure is proportional to the number of residues it contains, and the loop lengths within substructures are typically smaller than the loop length connecting them; consequently, the same rule for unfolding a hairpin can be, roughly speaking, propagated to "renormalized" levels of

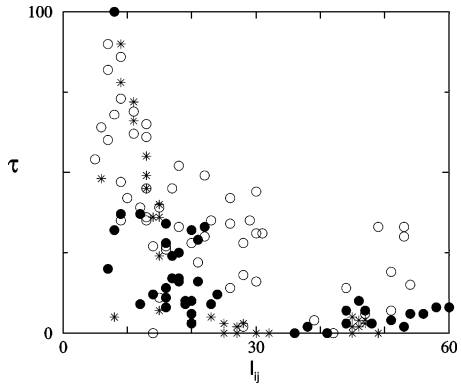

FIG. 6. Anticorrelation between the lengths of loops $l_{ij}=|j-i|$ joined by native contacts and the recursion step τ_{ij} at which native contacts (i,j) dissolve for three small proteins: protein G (stars), SH3 (open circles), and CI2 (solid circles). Larger and less cooperative unfolding proteins (see Fig. 1) lead to more disperse anticorrelation.

protein structure [57,58]. It may then be worthwhile to consider folding rates in terms of an expansion or decomposition in terms of these dynamically cooperative substructures that are natural to describe by loop closure ideas in order to include frustration effects separately.

"Extremal" examples. The four examples presented in the previous section are representative of the "usual" features we observe. Here we briefly summarize the results for proteins with more mechanical, mixed, diffuse, or frustrated folding mechanisms: (i) As expected, obligatory ("mechanic nucleus" [31]) folders such as SH3 and its circular permutants all unfold in excellent agreement with previous work [4,31,36,37]. For example, clipping a hairpin that is unstructured in the transition state does not affect the persistance of the (distal hairpin) nucleus, while clipping the nucleating hairpin leads to a qualitatively different unfolding path in which an alternate (n-src) hairpin becomes the new nucleus of the fold [4,31]. (ii) The three-helical, homologous B and E domains of staphylococcal protein A [1,38–40,59] fold into a symmetric bundle, but despite this symmetry, one of the helices α_1 unfolds preferentially (i.e., is the first helix to dissociate its long-range contacts) in agreement with experiment. (iii) Although chymotripsin inhibitor folds through a diffuse transition state, good agreement is obtained with one of the main pathways observed in (un)folding simulations [4,42,43]. Likewise, barnase unfolds consistent with its kinetic pathway except at the C terminal edge of its β sheet. (iv) More "globally frustrated" topologies (such as ADA2h and FKBP12 [5,41,45]) are not as well described by the model. For these folds, the method recovers roughly the correct dynamic features of the protein but the unfolding of residue contacts is in much poorer agreement with kinetics data than for "mechanic" folders. Conversely, the folding paths of small, "cooperative" folds (such as protein G) can be well descibed by the model, but appear to be more susceptible to mutations (as in protein L) than most of the folds studied in this paper [9]. A recent prediction method which includes side-chain effects [60] is relatively successful in detecting these variations.

Discussion. A basic dynamic feature of our model is that secondary structures, while flexible, keep their shape. The efficient expansion paths of this system identify regions of a protein's main chain that are topologically confined or frustrated from unfolding, and the results suggest that these regions are the source of rapid, entropically favorable routes [22] down the free energy landscape.

While extensive optimization for folding speed does not appear to be a high priority in protein evolution [54], the folding routes of "mechanic proteins" are relatively insensitive to energetic frustration [30]. And the fact that the G$\bar{o}$ model describes protein nucleation features qualitatively accurately suggests that proteins usually rely more on the shapes of their native folds than on any unique crystalline ordering or interactions between their side-chain atoms to guide their assembly [61,62]. Evolutionary mutations can qualitatively change the folding mechanisms of cooperative topologies [9] and are noticable even for mechanic folders [60] but generally seem to influence the rate of protein folding and the stability of intermediates more than the preferred order of steps along the folding pathway [15].

Conversely, more globally frustrated proteins could require and, perhaps, evolve some type of special side chain order to compensate their topologies. As in the G$\bar{o}$ model, inaccuracies in our results can often be traced to the absence of side-chain excluded volume restrictions, and it would be interesting to conduct this same type of experiment with explicit side chains to see if agreement improves and whether the paths can be still be computed rapidly.

The authors would like to thank Andres Colubri and Ariel Fernandez for helpful comments during the completion of this work.

[1] J. E. Shea, J. N. Onuchic, and C. L. Brooks III, Proc. Natl. Acad. Sci. U.S.A. **96**, 12512 (1999).
[2] C. Clementi, H. Nymeyer, and J. N. Onuchic, J. Mol. Biol. **298**, 937 (2000).
[3] C. Clementi, P. A. Jennings, and J. N. Onuchic, Proc. Natl. Acad. Sci. U.S.A. **97**, 5871 (2000).
[4] C. Clementi, P. A. Jennings, and J. N. Onuchic, J. Mol. Biol. **311**, 879 (2002).
[5] N. Koga and S. Takada, J. Mol. Biol. **313**, 171 (2001).
[6] M. S. Cheung, A. E. Garcia, and J. N. Onuchic, Proc. Natl. Acad. Sci. U.S.A. **101**, 511 (2002).
[7] D. A. Baker, Nature (London) **405**, 39 (2000).
[8] O. V. Galzitskaya, S. O. Garbuzynskiy, D. N. Ivankov, and A. V. Finkelstein, Proteins **51**, 162 (2003).
[9] C. Clementi, A. E. Garcia, and J. N. Onuchic, J. Mol. Biol. **326**, 933 (2003).

[10] J. D. Bryngelson and P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **84**, 7524 (1987).
[11] N. Gō, Adv. Biophys. **18**, 149 (1984).
[12] K. W. Plaxco and M. Gross, Nat. Struct. Biol. **8**, 659 (2001).
[13] D. Shortle, M. S. Ackerman, Science **293**, 487 (2001).
[14] S. W. Englander, L. Mayne, and J. N. Rumbley, Biophys. Chem. **101**, 57 (2002).
[15] S. W. Englander, Annu. Rev. Biophys. Biomol. Struct. **29**, 213 (2000).
[16] B. Zagrovic *et al.*, J. Mol. Biol. **323**, 153 (2002).
[17] B. Zagrovic and V. S. Pande, Nat. Struct. Biol. **10**, 955 (2003).
[18] D. E. Makarov and K. W. Plaxco, Protein Sci. **12**, 17 (2003).
[19] T. X. Hoang, A. Trovato, F. Seno, J. R. Banavar, and A. Maritan, Proc. Natl. Acad. Sci. U.S.A. **101**, 7960 (2004).
[20] J. R. Banavar, A. Maritan, C. Micheletti, and A. Travato, Proteins **47**, 315 (1998).
[21] A. G. Murzin and A. V. Finkelstein, J. Mol. Biol. **204**, 749 (1988).
[22] O. V. Galzitskaya and A. V. Finkelstein, Proc. Natl. Acad. Sci. U.S.A. **96**, 11299 (1999).
[23] A. V. Finkelstein and A. Y. Badretdinov, Mol. Biol. **31**, 391 (1997).
[24] See also, R. Zwanzig, Proc. Natl. Acad. Sci. U.S.A. **92**, 9801 (1995).
[25] A. Fernandez, H. Arias, and D. Guerin, Phys. Rev. E **52**, R1299 (1995).
[26] C. Micheletti, J. Banavar, A. Maritan, and F. Seno, Phys. Rev. Lett. **82**, 3372 (1999).
[27] This resembles rapid denaturation or high-temperature unfolding as a means to describe folding (see the references to Williams *et al.* and Lazardis and Karplus below).
[28] G. Toulouse, Commun. Phys. (London) **2**, 115 (1977); in *Spin Glass Theory and Beyond*, edited by M. Mezard, G. Parisi, and M. Virasoro (World Scientific, Englewood Cliffs, NJ, 1987).
[29] A. Fersht, *Structure and Mechanism in Protein Science* (Freeman, New York, 1999).
[30] M. Gruebele and P. G. Wolynes, Nat. Struct. Biol. **5**, 662 (1998).
[31] A. R. Viguera, L. Serrano, and M. Wilmanns, Nat. Struct. Biol. **3**, 874 (1996).
[32] E. Lopez-Hernandez and L. Serrano, Folding Des. **1**, 43 (1996).
[33] S. E. Choe, L. Li, P. Matsudaira, G. Wagner, and E. I. Shakhnovich, J. Mol. Biol. **304**, 99 (2000).
[34] W. F. Walkenhorst, J. A. Edwards, and J. L. Markley, Protein Sci. **11**, 82 (2002).
[35] W. F. Walkenhorst, S. M. Green, and H. Roder, Biochemistry **36**, 5795 (1997).
[36] J. C. Martinez, M. T. Pisabarro, and L. Serrano, Nat. Struct. Biol. **5**, 721 (1998).
[37] V. P. Grantcharova, D. S. Riddle, J. V. Santiago, and D. Baker, Nat. Struct. Biol. **5**, 714 (1998).
[38] Y. Bai, A. Karimi, H. J. Dyson, and P. E. Wright, Protein Sci. **6**, 1449 (1997).
[39] D. O. V. Alonso and V. Daggett, Proc. Natl. Acad. Sci. U.S.A. **97**, 133 (2000).
[40] J. K. Myers and T. G. Oas, Nat. Struct. Biol. **8**, 552 (2001).
[41] V. Villegas, J. C. Martinez, F. X. Avilles, and L. Serrano, J. Mol. Biol. **283**, 1027 (1998).
[42] T. Lazardis and M. Karplus, Nature (London) **278**, 1928 (1997).
[43] V. Dagget *et al.*, Proc. Natl. Acad. Sci. U.S.A. **98**, 4349 (2001).
[44] C. J. Bond, K. Wong, J. Clarke, A. Fersht, and V. Daggett, Proc. Natl. Acad. Sci. U.S.A. **94**, 13 109 (1997).
[45] K. F. Fulton, E. R. G. Main, V. Daggett, and S. E. Jackson, J. Mol. Biol. **291**, 445 (1999).
[46] E. N. Lorentz, *The Essence of Chaos* (University of Washington Press, Seattle, 1993).
[47] For most examples cited in this paper, we fit planar unit chains [G. Nemethy and H. A. Scheraga, Biopolymers **3**, 155 (1965)] to the proteins' pdb coordinates using a current parameter set [R. A. Engh and R. Huber, Acta Crystallogr., Sect. A: Found. Crystallogr. **47**, 392 (1991)] and unfold using the two-point derivative formula. In some cases, we unfold the exact crystal structure with the exact dihedral derivative [S. He and H. A. Scheraga, J. Chem. Phys. **108**, 271 (1998)] using software provided in "Protein Library" written by Andres Colubri at the University of Chicago. Because the fitting procedure is very accurate (0.05–0.2 Å root-mean-square derivative) the results of these methods are essentially the same.
[48] This could also be viewed as a sum or superposition of local moves [see S. Cahill, M. Cahill, and K. Cahill, J. Comput. Chem. **24**, 1364 (2003)].
[49] *In vitro*, the time scales are reversed. Then it is thought that amino acid steric constraints focus the unfolded ensembles of helical segments around the helical region of Ramachandran space (see, for example, the references to Zagrovic and Pande above).
[50] The $1/r$ form is similar to the entropically repulsive potential derived in a simple model of dna unfolding [see T. Garel, C. Monthus, and H. Orland, Europhys. Lett. **55**, 132 (2001)].
[51] For RNase H, ADA2h, and FKBP12 it was necessary to include the short-range repulsive term noted earlier in the text.
[52] C. M. Dobson, P. A. Evans, and S. E. Radford, Trends Biochem. Sci. **19**, 31 (1994).
[53] M. A. Williams, J. M. Thornton, and J. M. Goodfellow, Protein Eng. **10**, 895 (1997).
[54] B. Gillespie *et al.*, J. Mol. Biol. **330**, 813 (2003).
[55] S. B. Ozkan, I. Bahar, and K. A. Dill, Nat. Struct. Biol. **8**, 765 (2001).
[56] Y. Bai, Protein Sci. **9**, 194 (2000).
[57] P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **94**, 6170 (1997).
[58] P. G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, NY, 1979).
[59] See also Y. Zhou and M. Karplus, Nature (London) **401**, 400 (1999).
[60] R. Guerois and L. Serrano, J. Mol. Biol. **304**, 967 (2000).
[61] P. Lindgard and H. Bohr, Phys. Rev. Lett. **77**, 779 (1996).
[62] P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **93**, 14 249 (1996).