THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

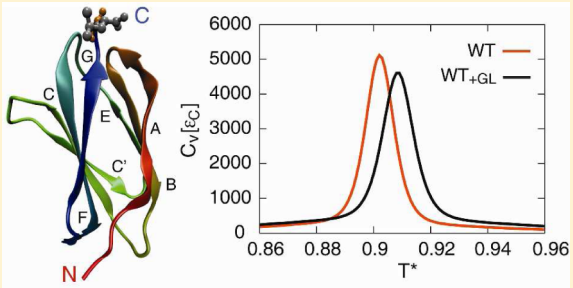
Backtracking due to Residual Structure in the Unfolded State Changes the Folding of the Third Fibronectin Type III Domain from Tenascin-C

Swarnendu Tripathi,[†,‡] George I. Makhatadze,[†,‡] and Angel E. Garcia*[,‡,§]

[†]Department of Biology, [‡]Center for Biotechnology and Interdisciplinary Studies, and [§]Department of Physics, Applied Physics, and Astronomy, Rensselaer Polytechnic Institute, 110 8th Street, Troy, New York 12180, United States

Supporting Information

ABSTRACT: Residual structure in the unfolded state of a protein may play a crucial role in folding and stability. In the present study, using an all (heavy)-atom structure based model and replica exchange molecular dynamics simulations, we explored the folding landscape of the third fibronectin type III domain from tenascin-C (TNfn3). Specifically, both the wild type (WT) and a variant with two additional amino acids, Gly-Leu (GL), at the C-terminus (WT_{+GL}) were studied. We found that, although both domains of TNfn3 are topologically frustrated, the early formation of the native contacts from the C-terminal end of WT_{+GL} causes more "backtracking" than in the WT. As a result, the WT exhibits a two-state folding behavior with a broad transition-state ensemble, whereas the WT_{+GL} folds through a metastable intermediate state. Furthermore, our study confirmed that the core of both proteins is conformationally heterogeneous and noncompact, and folds late mainly due to backtracking of the part of the core. Finally, in agreement with the previous experimental studies, our results clearly demonstrated distinct thermodynamic behavior of the two proteins with WT_{+GL} being more stable.



INTRODUCTION

The energy landscape theory[1] of protein folding states that, unlike a random heteropolymer, the sequence of a naturally occurring protein is (energetically) minimally frustrated and has a global minimum, the native state. On the basis of this principle, the folding mechanism of protein is defined by a funneled landscape[2,3] due to a strong energetic bias in favor of the native state topology. The physical implications of this funnel-like landscape have been extensively explored by applying a simplified native-state-biased, energetically unfrustrated Go-like (structure based) model, using different analytical[4] and simulation[5−11] approaches. Surprisingly, the predicted folding pathways from these studies agree well qualitatively (and even quantitatively) with site-directed mutagenesis experiments that probe the transition state ensemble (TSE).[12] The characterization of the ensemble of unfolded states of protein has received less attention mainly because of its traditional view as a random coil. Experimentally, the characterization of structures in the unfolded state becomes very challenging because the unfolded protein samples diverse conformations.[13,14] Nonetheless, results from several theoretical and experimental studies have suggested the existence of dominating conformational biases in the unfolded state, which may guide the protein to the native state.[15,16] Therefore, a detailed examination of the nature of local structural order (e.g., residual structure) in the unfolded state is of immense importance, since it can significantly affect the process of protein folding and stability.[17] In the present paper, by conducting a detailed comparative study of the folding mechanism of two variants of the third fibronectin type III domain from tenascin-C (TNfn3) using a structure based model approach, we confirmed that even a small fraction of local structural ordering in the unfolded state can substantially modify the folding energy landscape.

Human tenascin-C (TNC) is a large extracellular matrix glycoprotein composed of six monomers with suggested roles during development, and in tissue remodeling by regulating the cell adhesive and signaling properties. Each monomer of TNC consists of 8−15 fibronectin type III (FN−III) domains. The structure of the third FN−III domain of TNC, TNfn3, was determined by Leahy et al.[18] using X-ray crystallography. TNfn3 is an immunoglobulin-like (Ig-like) β-sandwich protein that has a Greek key strand topology (see Figure 1A and B). The Ig-like structural family comprises proteins from a number of different superfamilies that have no significant sequence identity and no detectable evolutionary relationship, yet they share the same fold and a common structural core.[19] The β strands of this family of proteins are commonly denoted by letters from A to G following a standard pattern (see Figure 1A). It has been suggested that a number of members of Ig-like fold share a common folding pathway[20,21] and have a folding nucleus composed of residues enclosing a hydrophobic core,

Received: October 10, 2012
Revised: December 19, 2012
Published: December 26, 2012

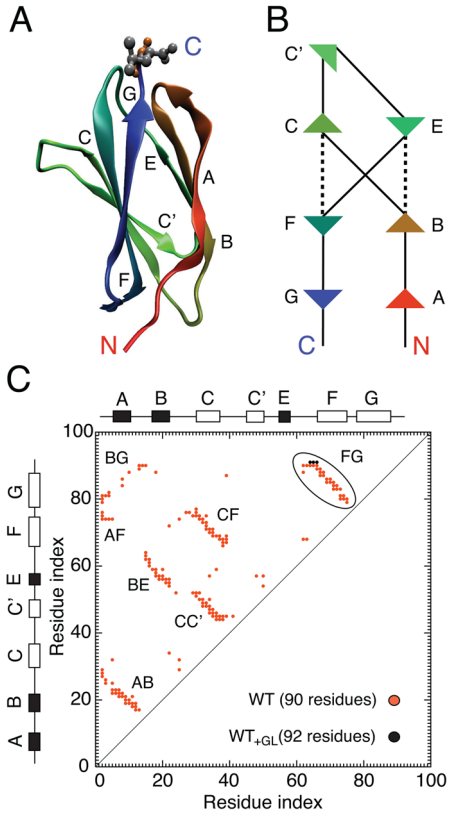


Figure 1. Native structure and contacts of TNfn3. (A) A three-dimensional structure of the WT_{+GL} TNfn3 is shown with the color changing from red (N-terminal) to blue (C-terminal). TNfn3 is an all β-sheet protein, and the secondary structure is defined as follows: A (5−10), B (17−22), C (30−37), C′ (45−50), E (55−58), F (66−75), and G (78−88). The last two residues Gly91 and Leu92 of WT_{+GL}, which are absent in the WT, are indicated by the ball-and-stick representation. All the three-dimensional images of proteins in this figure and other figures of this paper are generated using the Visual Molecular Dynamics (VMD) package.[66] (B) Schematic representation of the native topology of TNfn3. The dotted lines represent the contacts between BE and CF which are crucial to folding of this protein. (C) On the basis of the atom−atom contacts, a residue contact map of TNfn3 is shown for both the WT and WT_{+GL}. The difference in the residue−residue contacts between the proteins is circled. Native contacts in the different segments of the proteins are indicated inside the plot with the protein secondary structure at the side and the top.

drawn from the strands B, C, E, and F.[20−24] From earlier experimental studies, the folding, stability,[22] and dynamic properties[25,26] of the 90 amino acids long TNfn3 (WT) domain[18] have been well characterized. The protein has been described as moderately stable (~5 kcal mol^{-1}), with a simple two-state folding kinetics by Clarke et al.[22] In contrast, the refolding kinetics was described as biphasic with both slow and fast phases.[22] Later, the effect of domain boundary selection on the equilibrium folding and stability of the same WT domain was further examined by Clarke and co-workers[27] and it was inferred that the 90 amino acids domain is too short for optimal stability. In fact, nearly ~2.7 kcal mol^{-1} increase in the stability was achieved in their study by extending the WT TNfn3 domain with the two specific amino acids Gly and Leu (GL) at the C-terminal end at positions 91 and 92, respectively. Although this variant had a similar refolding rate relative to the WT protein, the unfolding rate was much slower.[27] Furthermore, the authors emphasized that a small number of additional long-range interactions might be important in changing the folding characteristics of WT_{+GL}.[27,28]

To elucidate the effect of additional interactions on the folding and stability of WT_{+GL} with atomistic detail, we performed all-atom structure based model simulations by employing the replica exchange molecular dynamics (REMD) method to study the folding/unfolding thermodynamics for both the WT and WT_{+GL} proteins. In this study, we are particularly interested in the following questions: (i) Which specific local interactions in WT_{+GL} are responsible for its increase in global stability? (ii) How do these interactions affect the overall folding process of WT_{+GL} relative to WT, such as the formation of the folding nucleus (hydrophobic structural core)? (iii) Are there any other effects on the folding mechanism of WT_{+GL} due to the boundary extension, accumulation of an intermediate state, or presence of a local structural order in the unfolded state that may affect the protein folding pathway?

The results in our study unveiled that the folding of both domains of TNfn3 has some degree of frustration and, hence, exhibits "backtracking" or local unfolding in the early stages of folding. Previously, the term backtracking has been used to illustrate a mechanism related to the incorrect ordering of two subsets of native contacts competing with each other, such that partial unfolding of one subset of contacts and their subsequent refolding could facilitate the formation of the contacts in the other part of the protein.[29] In the case of the Go-like model, since there is no energetic frustration, backtracking arises from geometrical constraints related to the specific fold and has been termed as "topological frustration".[29,30] From the analysis of the folding process of the two variants of TNfn3, we show that in WT_{+GL} additional contacts from the C-terminal end impose further constraints on the protein and cause more backtracking than in the WT. As a result, the WT TNfn3 folds through a two-state process with a broad transition state, whereas WT_{+GL} folds via an intermediate state resulting from a more pronounced backtracking. Moreover, we suggest that the presence of a slow refolding phase for WT TNfn3 observed in experiment[22] could also be due to the backtracking. As previous studies have pointed out, backtracking can significantly slow down the folding process[29,31−34] and, therefore, may also contribute to a metastable intermediate.[31,34]

On the basis of the experimental study of Hamill et al.,[21] folding of TNfn3 has been described by a nucleation condensation mechanism,[32] similar to that of other Ig-like proteins that share a common structural core.[20] In this protein family, formation of the folding nucleus occurs first from the central strands of two β-sheets, followed by packing of the surrounding β-sheets after the TS for folding.[20,21,23,24,35] From a detailed ϕ-value analysis of TNfn3, it has been further proposed that a four-residue "ring"-like structure from the structural core forms the folding nucleus of this protein.[21] Our study unveiled that the core of TNfn3 is partially unfolded in the TSE, as part of the core folds late due to backtracking and this is particularly pronounced in the WT_{+GL}. Overall, in agreement with the previous experiment,[27] our results clarify that the inclusion of GL in the C-terminal end of TNfn3 significantly increases the stability. Furthermore, we show that although the boundary extension does not change the key mechanism of the formation of the structural core of this protein it noticeably affects the folding dynamics. In particular, we found that the presence of some local structural ordering

dx.doi.org/10.1021/jp310046k | J. Phys. Chem. B 2013, 117, 800−810

from the C-terminal end of WT_{+GL} in the unfolded state guides this protein to fold via a metastable intermediate.

MATERIALS AND METHODS

Preparation of Structures. The structure of the extended protein WT_{+GL} was modeled from the known structure of the TNfn3 (PDB ID code 1ten) using Modeler version 7.7.[36] The residue number 802−891 from the original PDB file was renumbered from 1 to 90. Besides, the last two residues Gly91 and Leu92 and the incomplete residue Arg1 were also modeled in the WT_{+GL} TNfn3. To keep the native contacts consistent in both the WT_{+GL} and WT proteins, the latter was obtained by deleting the last two residues (Gly91 and Leu92) from the modeled structure of WT_{+GL}. Descriptions of the native states of WT and WT_{+GL} TNfn3 are shown in Table 1.

Table 1. Details of the Two Systems of TNfn3

			no. of native contacts	
protein	no. of residues	no. of Atoms	atom−atom	residue−residue
WT	90	707	529	164
WT_{+GL}	92	719	537	167

All-Atom Structure Based Model. We followed the approach of Clementi et al. to define an all-atom Go-like model.[5] In our all-atom structure-based model, the potential energy is defined by a designed potential that enforces a funneled energy landscape:

$$V_{AA} = V_{bond} + V_{nonbond} \quad (1)$$

The first term in the above equation describes the bonded interactions given by

$$V_{bond} = \sum_{bonds} K_r(r - r_o)^2 + \sum_{angles} K_\theta(\theta - \theta_o)^2 + \sum_{improper/planar} K_\chi(\chi - \chi_o)^2 + \sum_{proper} K_\phi[\cos(n\phi - \phi_o)] \quad (2)$$

In this equation, r, θ, and ϕ represent the protein configuration, $\Gamma(r, \theta, \phi)$. For the bonded parameters, K_r, K_θ, K_χ, K_ϕ, r_o, θ_o, and ϕ_o are taken from the united-atom force field ffg43a1 (GROMOS96)[37,38] and are not specific to the native state of the protein. The second term in eq 1 describes the nonbonded interactions given as

$$V_{nonbond} = \sum_{contacts} \varepsilon_C\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - 2\left(\frac{\sigma_{ij}}{r}\right)^{6}\right] + \sum_{noncontacts} \varepsilon_{NC}\left(\frac{\sigma_{NC}}{r}\right)^{12} \quad (3)$$

In the above equation, ε_C = 1 kcal mol^{-1} is the strength of a native contact defined by a cutoff distance 4 Å between the atoms i and j separated by the corresponding amino-acid residues $|i' - j'| > 3$, σ_{ij} is the distance between atoms i and j that are in the native contacts for the attractive Lennard-Jones potential. We did not use a shadowing scheme for defining interacting native pairs.[11] Interactions between the atoms which are not in the native contact are given a pure repulsive potential of strength, ε_{NC} = 0.01 kcal mol^{-1} and σ_{NC} = 3.5 Å. It is important to mention that our initial choice of σ_{NC} = 2.5 Å led the protein to get trapped into partially folded non-native states. To overcome this problem and determine the appropriate value of σ_{NC}, we ran several short MD simulations in the unfolded states of the WT_{+GL} TNfn3 at four different values of σ_{NC} = 2.5, 3.0, 3.5, and 4.0 Å. As shown in Figure S2 (Supporting Information), the Ramachandran (ϕ, ψ) plot at σ_{NC} = 3.5 Å properly represents the excluded volume of Ramachandran map. Sources of frustration in this model emerge from the shape of the backbone chain in the folded state (topological frustration) and from the dihedral angle energy terms. The balance between structure based native interactions and non-native followed Clementi et al. but using Gromos for the dihedral angle energy terms and σ_{ij} as the distance between pairs in the native structure.[5] Other implementations of all-atom Go models use different scaling of native and dihedral angle terms, cutoffs for defining native pairs, shadowing for defining interacting native pairs, and Gaussian energy functions for native, nonbonded interactions.[10,11]

Given the coarse grained nature of the model and the use of enhanced sampling methods, it is not straightforward to translate simulation times to experimental folding times. However, Kouza et al. have used the simulation of multiple proteins using C_α Go models to estimate the folding rates and have concluded that folding rates are predicted within an order of magnitude by $\kappa_F \approx \kappa_o \exp[-N^{1/2}]$, with $\kappa_o \approx 0.2\ \mu s^{-1}$.[39] Using this formula, with N = 90, we estimate the folding times for TNfn3 to be in the 100 ms time scale.

In the present study, we use the fraction of native contacts, Q, as the global reaction coordinate. Q is defined as the fraction of natively interacting residues that are in contact. Here, two residues are considered in contact if any two atoms from those residues are within 1.5 times their native distance σ_{ij}.

Simulation Details. All simulations in the present paper were performed using the Gromacs 4.0.7[40] without modifying the source code. We used Langevin dynamics with a time step of τ = 0.004 ps and a friction coefficient of 1 ps^{-1}. To study folding thermodynamics, we used the replica exchange molecular dynamics (REMD) method,[41] as implemented in Gromacs.[40] We chose 24 replicas close to their corresponding folding temperature, T_F, for both the WT_{+GL} and WT. In the reduced temperature unit, temperature ranges $0.898 \le T^* \le 0.921$ for WT_{+GL} and $0.892 \le T^* \le 0.915$ for WT with a temperature step of $\Delta T^* = 0.001$ were used for the consecutive replicas, where $T^* \equiv k_B T/\varepsilon_C$ is in the reduced temperature units. The range of temperatures of REMD simulations for both systems was decided on the basis of the observation that the proteins remain always folded at the lower range and unfolded at the higher end. Each replica was run with 3×10^8 steps of simulation that correspond to a total time of 28.8 μs. The data from the last 2.5×10^8 steps of each replica were used for all of the analysis. During the simulation, exchanges were attempted at 1000 integration step intervals, and configurations and energies were also stored every 1000 steps. All thermodynamic calculations in this work were obtained from the weighted histogram analysis method (WHAM).[42,43]

RESULTS AND DISCUSSION

WT_{+GL} is More Stable with an Intermediate State. We performed extensive replica exchange molecular dynamics (REMD) simulations to characterize the folding thermodynamics of the WT and WT_{+GL} TNfn3 proteins (see the

dx.doi.org/10.1021/jp310046k | J. Phys. Chem. B 2013, 117, 800−810

Materials and Methods section). Here, the focus is to differentiate the folding of the two proteins in their overall thermodynamic and structural properties. Figure 2A shows the

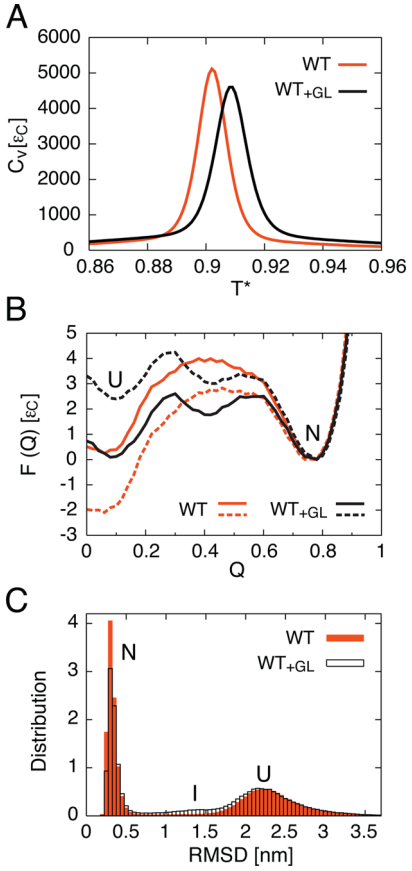


Figure 2. Folding thermodynamics of TNfn3. (A) Heat capacity (C_V) plots for the WT and WT_{+GL} TNfn3 at different temperatures, $T^* \equiv k_BT/\varepsilon_C$ (in the reduced units). (B) Free-energy profiles $F(Q)$ are plotted as a function of the global reaction coordinate, Q. The solid curves correspond to the free energies at $\sim T_F$ for both the WT and WT_{+GL}. The dotted curves are plotted to represent the free energies at a higher temperature, T_F (WT_{+GL}), for the WT and a lower temperature, T_F (WT), for the WT_{+GL}. (C) Distributions of RMSD for the WT and WT_{+GL} TNfn3 at the folding temperature, T_F.

calculated heat capacities, C_V, of the two systems. The folding temperatures T_F of the two proteins can be obtained from the peak maximum of their corresponding C_V profiles and are expressed in the reduced temperature units $T^* \equiv k_BT/\varepsilon_C$ (see eq 3). A comparison of the C_V profiles clearly shows an increase in the folding temperature for the WT_{+GL} relative to WT, $T_F(WT) = 0.902$ versus $T_F(WT_{+GL}) = 0.909$ (or ~3.5 K). This is consistent with the experimental study of Hamill et al.[27] that showed that the WT_{+GL} is thermodynamically more stable than the WT.

Figure 2B shows the free-energy profiles as a function of a global reaction coordinate, Q (fraction of native contacts), for both the WT and WT_{+GL}. Three states of WT_{+GL} are clearly seen from this plot, whereas the folding of the WT appears to be two-state. For the WT, the maximum of free energy defining the transition state (TS) is at $Q \sim 0.4$. The I-state for WT_{+GL} also occurs at $Q \sim 0.4$. These observations are also supported by the distributions of the total potential energy and the global reaction coordinate, Q (see Figure S1 in the Supporting Information), with a significant population of the I-state at $T \sim T_F$ for WT_{+GL}. On the other hand, the WT protein clearly exhibits a two-state behavior, judging by the distributions of potential energy and Q (Figure S1, Supporting Information). To calculate the difference in the stability between WT and WT_{+GL}, the free-energy profiles were compared at the same temperature (Figure 2B). The increase in stability due to the GL extension is $\sim 2.3\varepsilon_C$, in the reduced units, as estimated by comparing the free-energy difference $\Delta\Delta F_{U-N} = \Delta F_U(WT_{+GL} - WT) - \Delta F_N(WT_{+GL} - WT)$. This is in good agreement with the earlier experimental study that showed ~2.7 kcal mol^{-1} increase in the stability for WT_{+GL}.[27] Furthermore, an experimental equilibrium unfolding study of WT TNfn3 suggested a two-state transition,[22] while the refolding kinetics for the WT_{+GL} revealed a two-phase folding. On the basis of our results, we suggest that the I-state of WT_{+GL} could be a possible reason for the slow refolding phase. However, the presence of any I-state for WT_{+GL} was not detected from the folding experiment by Hamill et al.[27] Nevertheless, their experimental study also showed that for WT_{+GL} refolding kinetics did not change, but unfolding was slowed significantly.[27] Here, we propose that the extension of WT TNfn3 by GL stabilizes the native state of WT_{+GL} considerably relative to its unfolded state. This stabilization also occurs in the I-state and thus lowers the transition barrier of WT_{+GL} compared to WT (see the solid curves in Figure 2B). To identify the global structural difference between the two systems, we plot the distributions of their all (heavy)-atom root-mean-square deviation (RMSD) at $T \sim T_F$ in Figure 2C. The RMSD values were calculated with respect to their corresponding native-state structures. Here, we find that overall the two systems are structurally very similar in the folded state, while the unfolded state of the WT_{+GL} is somewhat more compact than the WT. In addition, we see a small population of an I-state at RMSD values of ~1.4 nm for WT_{+GL}, which as a result reduces the magnitude of the distribution in the folded state (see Figure 2C). In contrast, the RMSD distribution for the WT does not show the presence of an intermediate state, suggesting that the folding is two-state.

Folding through Backtracking. In order to understand the folding mechanism in detail, it is important to identify how different parts of a protein fold into the native-state conformation. Here, we are particularly interested to see how different parts of a protein fold with respect to the global reaction coordinate Q. By separating the native contacts of WT and WT_{+GL} into different segments (as shown in Figure 1C), we can calculate the partitioned local order parameter, $Q_p(Q)$, as a function of Q along the folding route for both proteins at $T \sim T_F$. Here, $Q_p(Q)$ refers to the fraction of contacts formed in the specific clusters on the contact map for the native state as a function of Q. These clusters are labeled AB, AF, BE, BG, CC′, CF, and FG in Figure 1C. Figure 3 illustrates that the overall folding mechanisms in terms of the different local $Q_p(Q)$ for both proteins are largely similar. This is particularly true for the segments AB, CC′, CF, AF, and BG. Moreover, we observe that, on average, for both proteins contacts in CC′ form early, whereas the contacts in the AB, AF, and BG segments form relatively late, only after the global $Q \sim 0.4$. The largest difference is observed for the formation of the native contacts in segments FG and BE in these proteins. We find that for WT_{+GL}, even in the unfolded states ($Q \sim 0.0-0.3$), contacts in FG form much earlier than for WT. A similar behavior is observed for the contacts in BE which form early in the unfolded state of the WT_{+GL} compared to the WT. However, in

dx.doi.org/10.1021/jp310046k | J. Phys. Chem. B 2013, 117, 800−810

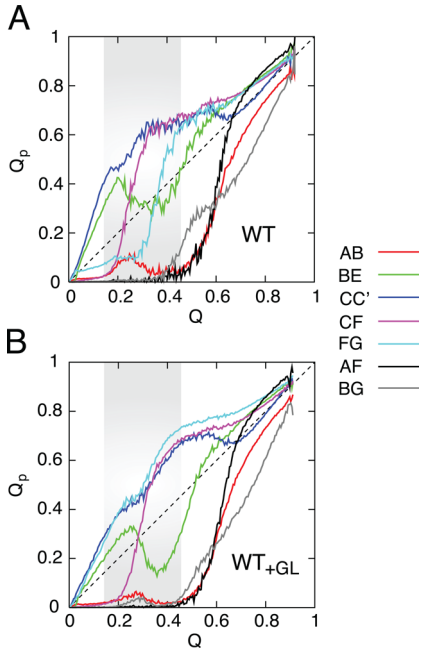


Figure 3. Folding of the different segments of the proteins. Parts A and B illustrate the formation of the partitioned fraction of native contacts (Q_p) for different segments of the protein, as shown in Figure 1C compared to the average Q for the WT and WT_{+GL}, respectively, at their corresponding $\sim T_F$. In each plot, the diagonal line is drawn as a reference for native contacts that ideally follow the average Q. The shaded regions in the plots indicate the range of $Q \sim 0.2–0.4$, where the backtracking occurs for both proteins through rise and fall of Q_p for BE (green line).

terms of folding, the most interesting feature in these proteins is folding of BE, that shows a transient increase and decrease of its $Q_p(Q)$ values (see the shaded regions in Figure 3A and B in the range $Q \sim 0.2–0.4$). Figure 3 also shows that while for WT the $Q_p(Q)$ value of BE decreases from 0.4 to 0.3, the change in $Q_p(Q)$ of BE for WT_{+GL} is from 0.3 to 0.1, as Q increases. This suggests that backtracking (transient unfolding of a particular protein segment in route to the folded state) in BE is more pronounced in WT_{+GL} than WT.

Backtracking is a mechanism that causes local unfolding and refolding of some native contacts to facilitate folding in the other parts of a protein.[33] For WT_{+GL}, we notice that backtracking of BE assists the folding of CF. This can be seen from the crossing of the $Q_p(Q)$ curves of BE and CF in Figure 3B. On the other hand, Figure 3A shows in a similar way backtracking of BE facilitates the folding of the CF and FG segments in the WT. As a result, the contacts in both CF and FG form cooperatively in the WT, while for the WT_{+GL} the contacts only from CF form in a cooperative manner. Figure 3 also indicates some backtracking in segment AB for both proteins and in BG only for the WT_{+GL} at $Q \sim 0.3$. However, in comparison with BE, the backtracking in AB or BG occurs to a lesser extent. Additionally, AB has more backtracking in the WT than in the WT_{+GL}. This could be due to early formation of the contacts in BE, which also helps to form additional contacts in AB for WT and subsequently the backtracking in BE further forces the contacts in AB to break. Notice that, in order to fully form the contacts in AB and BG for both proteins, most of the contacts in AF need to be formed first (see the crossing of curve AF through AB and BG in Figure 3A and B at $Q \sim 0.6$), even though folding of BG starts early at $Q \sim 0.4$.

The analysis of folding using $Q_p(Q)$ gives further insights to the understanding of the free-energy profiles of the two proteins, shown in Figure 2B. The lowering of the transition barrier height of the WT_{+GL} can be recognized through the early folding of the FG segment that forms much later in the WT and results in a higher barrier for the transition. It also appears that early formation of the FG segment also stabilizes both the I-state and native state of the WT_{+GL} compared to the WT. On the other hand, the broad TS of WT can be understood in terms of the backtracking in BE and might be a plausible reason for the slow refolding phase that was previously observed near the TS of this protein.[22] Altogether, the results presented here suggest that the WT_{+GL} has more backtracking than the WT and therefore folds through an intermediate state (I-state) instead of a broad TS. Theoretical studies on the folding of interleukin-1β from the β-trefoil family have also indicated that backtracking could contribute to a broad TS barrier and to the experimentally observed intermediate state.[29,31]

Conformational Flexibility and Local Unfolding. In the previous section, we have shown how different parts of the WT and WT_{+GL} proteins fold but only in terms of their native contacts. In this section, we will provide a site-specific description of the folding mechanism by calculating the root mean square fluctuation (RMSF) of each residue for a given Q at $T \sim T_F$. The RMSF can also be viewed as an order parameter that defines the protein conformational flexibility (or rigidity) in terms of the degree of structural disorder (or order) for each residue during folding. In Figure 4A and B, the calculated RMSF at $Q \sim 0.2$ indicates that, for both the WT and WT_{+GL} proteins, strands B and E already have a low flexibility relative

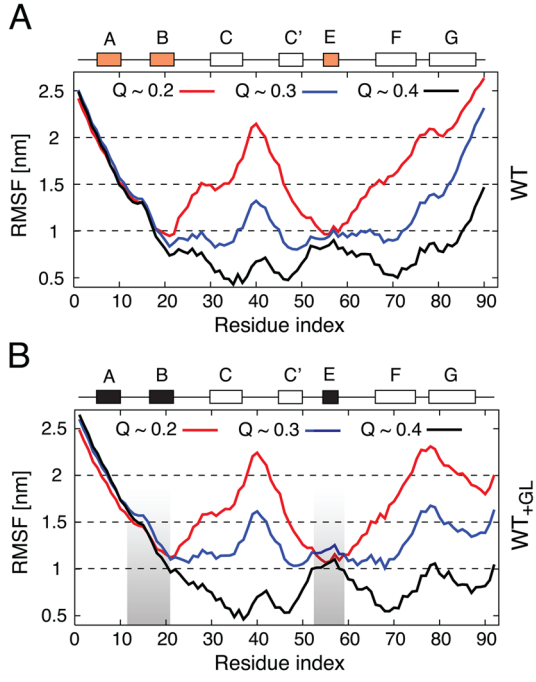


Figure 4. Root mean square fluctuations (RMSF) of the proteins. Parts A and B show the RMSF vs residue index calculated at different values of the fraction of native contacts $Q \sim 0.2$, 0.3, and 0.4 for the WT and WT_{+GL}, respectively, at their corresponding folding temperature T_F. The parts of the protein WT_{+GL} which show non-monotonic changes in the RMSF at $Q \sim 0.3$ due to backtracking in the BE sheet are denoted by the shaded regions in part B.

dx.doi.org/10.1021/jp310046k | J. Phys. Chem. B 2013, 117, 800–810

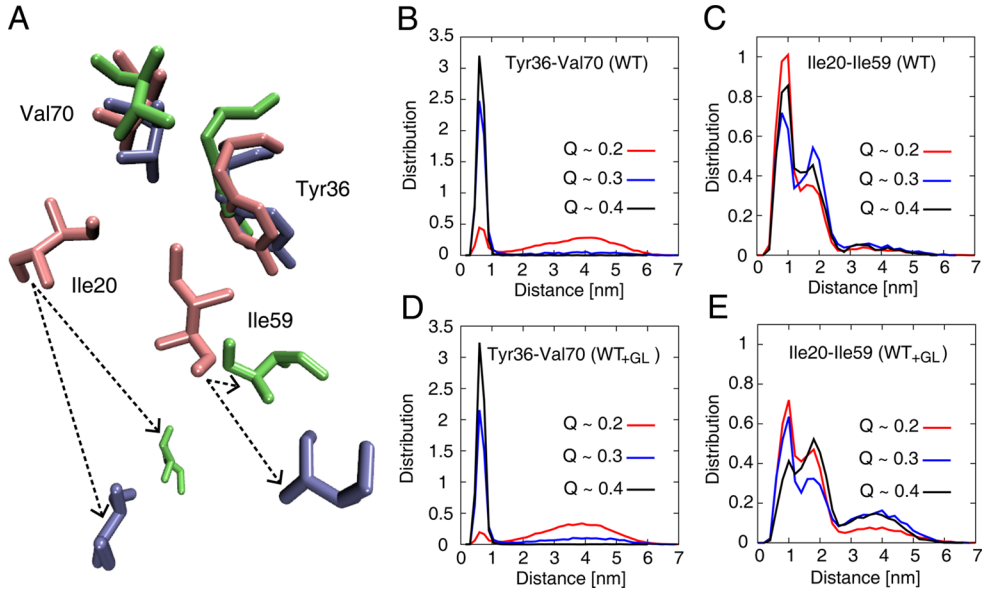


Figure 5. Formation of the four-residue ring structure of the TNfn3 core. (A) Comparison of the core structures in the native and partially folded states of WT and WT_{+GL}. The TNfn3 core consists of four residues Ile20, Tyr30, Val70, and Ile59 from strands B, C, F, and EF loop, respectively. Residues represented in pink are from the native-state conformation. The same residues represented in blue and green (also indicated by the arrows for Ile20 and Ile59) are from the WT and WT_{+GL}, respectively, for the most populated structures at $Q \sim 0.4$ (see also Figure 7C and F). Parts B and C are the distributions of the distances between the residue pairs Tyr36-Val70 and Ile20-Ile59, respectively, for the WT TNfn3 calculated at different values of $Q \sim 0.2$, 0.3, and 0.4. Parts D and E are the distributions of the distances between Tyr36-Val70 and Ile20-Ile59, respectively, for the WT_{+GL} TNfn3 calculated at the same values of Q as in parts B and C. All results are obtained from the data at $T \sim T_F$ for both the WT and the WT_{+GL}.

to the rest of the protein. This is in agreement with a NMR experiment by Meekhof and Freund[44] who found evidence of "nascent turn-like structures" for some specific residues from these two strands even in the fully denatured state of the WT. Other parts of these proteins that also show an early decrease in the fluctuations at $Q \sim 0.2$ are strands C, C′, and F. However, the most contrasting difference between the two proteins is the low fluctuations of the C-terminal end (including strand G) of WT_{+GL}, as can be seen in Figure 4. We suggest that the presence of a chemical exchange on a millisecond time scale for the residues from the C-terminal end of WT in the NMR relaxation experiments by Meekhof et al.[28] could be due to the high flexibility in this part of the protein. In contrast, a similar motion in the C-terminal end of the WT_{+GL} was not detected[28] probably due to the loss in flexibility of strand G, as can be seen in Figure 4B. In general, the results of RMSF analysis can be further understood by the $Q_p(Q)$ analysis introduced in the previous section (see also Figure 3). The higher structural order of the B−C and C′−E segments of both proteins can be realized from the early formation of native contacts in BE and CC′ at $Q \sim 0.2$. Similarly, lower fluctuations of strand G for WT_{+GL} can also be explained from the formation of some native contacts in FG at $Q \sim 0.2$.

In the light of evidence of backtracking presented here, the most intriguing part is the change in RMSF during folding as observed at $Q \sim 0.3$. Residues 12−21 (near strand B) and 53−59 (near strand E) for the WT_{+GL} clearly exhibit non-monotonic changes, i.e., crossover in the RMSF due to local unfolding and refolding (see the shaded regions in Figure 4B). In contrast, RMSF of these residues in the WT protein decreases monotonically in the range $Q \sim 0.2$−0.3 without significant change in the flexibility. Such behavior of residues in the WT_{+GL} can be understood from the transient breaking of the native contacts in BE through backtracking (compare to Figure 3). It is noteworthy to mention that, in terms of the change in conformational flexibility, the mechanism of backtracking is similar to "cracking",[45] associated with large conformational change in protein that also exhibits transient increase in the flexibility due to the local unfolding[46] that assists in relieving high stress in the specific residues.[45,47−49]

Overall, our results indicate that the residues in the range ∼20−80 in the WT have less conformational fluctuations than in the WT_{+GL}. Subsequently, at $Q \sim 0.4$, with the exception of strands A and B, the rest of the proteins show significant stabilization with lower fluctuations. Previously, large conformational fluctuations near the flexible AB turn (including parts of the strands A and B) in the μs to ms time scale for the WT have been reported in association with the presence of a β-bulge close to strand A.[26]

Formation of the Core Structure. It is well-known that for proteins with an Ig-like fold, which includes TNfn3, residues from strands B, C, E, and F form the central core.[19] Recently, the ϕ-value analysis of the WT_{+GL} has suggested that specifically the residues Ile20 (from strand B), Tyr36 (from strand C), Ile59 (from strand E), and Val70 (from strand F) constitute the folding nucleus with ϕ-values in the range 0.4−0.6.[21] To provide a more detailed description of the TSE of this protein, an all-atom model, with implicit solvent, was investigated by Paci et al.[50] In this study, the molecular dynamics sampling was biased to satisfy the experimental ϕ-values and a model for TSE was obtained by selecting only those subsets of structures for which the calculated average ϕ-values were close to the measured one. It was concluded that the experimental ϕ-values for four residues (Ile20, Tyr36, Ile59, and Val70) of the core that form an open ring-like structure in the folded state (see Figure 5A) are sufficient to give a rough interpretation of the structure of the TSE.[50]

In the current study, we focused mainly on the mechanism of the core formation of the WT and WT_{+GL} at their $T \sim T_F$. In particular, we analyzed the structures involving the core

dx.doi.org/10.1021/jp310046k | J. Phys. Chem. B 2013, 117, 800−810

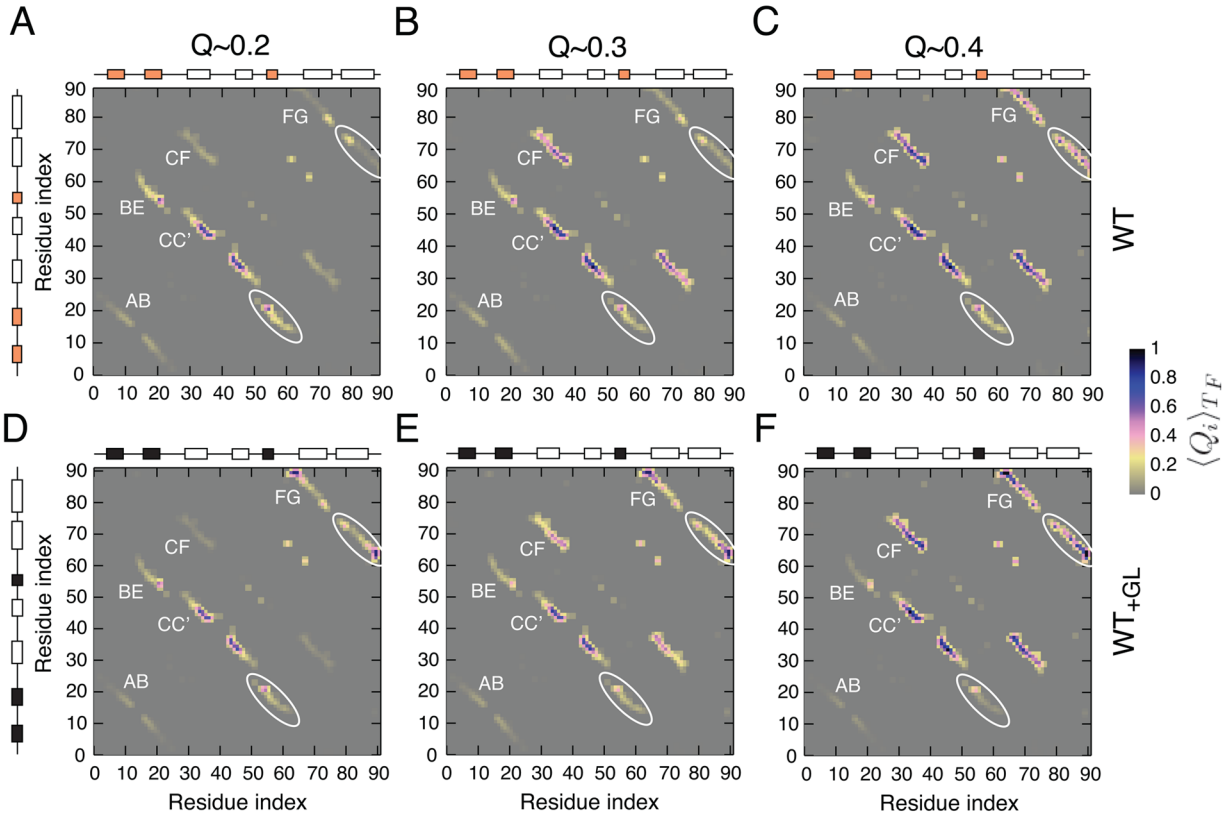


Figure 6. Probability of native contact formation. Parts A–C and D–F show the formation of native contacts along the reaction coordinate Q, obtained by calculating the probability of each contact, $\langle Q_i(Q) \rangle$, in the residue basis at different values of the fraction of native contacts, $Q \sim 0.2$, 0.3, and 0.4 for the WT and WT_{+GL}, respectively. $\langle Q_i(Q) \rangle$ were calculated at their corresponding folding temperature T_F. For both proteins, native contacts from BE and FG are circled in the lower half of the plots.

residues in the unfolded state of these proteins to envisage which residues initiate the formation of this four-residue ring-like scaffold. There are native contacts between strands B and E and also between the strands C and F, as can be seen from the residue contact map shown in Figure 1C. On the basis of this, we analyzed the formation of the core through rearrangement of the segments BE and CF, or more specifically how the distances between the residues Tyr36–Val70 and Ile20–Ile59 change in the unfolded state of these proteins at $Q \sim 0.2$–0.4. As shown in Figure 5B and D, the residues Tyr36 and Val70 initiate the core formation in the early stages of folding ($Q \sim 0.3$) for both the WT and WT_{+GL} proteins. This can be seen from the shift of the distributions of the distance between Tyr36 and Val70 from ~4.0 to ~0.5 nm for both proteins. The rapid development of the protein core through contacts between Tyr36 and Val70 can be further seen from the folding of segment CF. Figure 3 indicates that for both the WT and WT_{+GL} the native contacts between strands C and F form rapidly at $Q \sim 0.2$–0.4. On the other hand, the residue pair Ile20–Ile59 displays a very different mechanism for the core formation compared to the residue pair Tyr36–Val70 (see Figure 5C and E). For the WT, we find that the distance between Ile20–Ile59 varies between ~1.0 and 2.0 nm. However, in the case of WT_{+GL}, three distinct peaks can be seen at ~1.0, 2.0, and 4.0 nm. This suggests that, even though in both proteins some native contacts start to form early (at $Q \sim 0.2$) in BE, the residues Ile20 and Ile59 do not form the core in the unfolded state mainly because of the backtracking in this segment (see also Figure 3 for the Q_p curves of BE). The effect of backtracking in BE is also noticeable from the difference in the distributions of the distance between Ile20 and Ile59 for WT and WT_{+GL} (Figure 5C and E). In addition, due to more backtracking in WT_{+GL}, these two residues can be as far as 4.0 nm (see Figure 5E). Nonetheless, for the WT, we notice that, despite the closer distance between Ile20 and Ile59, these two residues remain far from the core. This suggests that in both proteins the residue pair Ile20–Ile59 is not part of the core in the unfolded state. This is also evident from the core structure for WT and WT_{+GL} at $Q \sim 0.4$ shown in Figure 5A. In fact, the entire core of the proteins is formed at $Q \sim 0.5$, as evidenced by the immediate formation of the contacts in BE after the backtracking (see Figure 3). These observations are in good agreement with studies that compared the properties of the structural core of TNfn3 with other Ig-like proteins and further indicated that the core of TNfn3 is relatively less folded in the TS and is structurally more heterogeneous.[50,51] As a result, the core of TNfn3 appears more dynamic and has a fluid-like nature, as observed by NMR spectroscopy.[52]

Topological Frustration and Folding Mechanism. So far, our results have provided an overall description of folding of the WT and WT_{+GL} TNfn3 proteins. In this section, we aim to underline some key features of the folding mechanism of these two proteins in terms of the native interactions and the overall structure formation. Here, our goal is to identify the specific native interactions in the unfolded state of these two proteins that make their folding and stability very distinct. In particular, we are interested in understanding how the topological frustration in these proteins induces backtracking

dx.doi.org/10.1021/jp310046k | J. Phys. Chem. B 2013, 117, 800–810

during folding. In Figure 6, we first compare the probability of the native contact formation $\langle Q_i(Q)\rangle_{TF}$ of the proteins at $T \sim T_F$ as a function of the reaction coordinate Q. As we see from Figure 6A and D, already at $Q \sim 0.2$ many contacts between strands C and C′ are formed, with high probabilities $\langle Q_i\rangle \sim 0.6-0.8$ for both WT and WT_{+GL}. However, the most significant difference between the two proteins at $Q \sim 0.2$ is observed for the contacts in the FG segment. Very few contacts are made in FG for the WT, while in the WT_{+GL} some contacts in FG, largely from the C-terminal end and near the EF loop (compare parts A and D of Figure 6), are well formed. In addition, several residues from strand B also make contacts to strand E in both proteins at $Q \sim 0.2$. Very similar observations have been made from the NMR experiments[44] that showed evidence of some nascent turn-like structure at residues 16−17 (from strand B) and 54−57 (from strand E), which could be related to an early alignment of these two strands even in the fully denatured state of the WT. Further insights supporting the above description can be gained from the representative structures of WT and WT_{+GL} at $Q \sim 0.2$, shown in Figure 7A and D, respectively.

The most remarkable structural rearrangements in both proteins during folding take place between $Q \sim 0.2$ and ∼0.3, and involve a very rapid formation of the native contacts between strands C and F (see Figure 6B and E). This can be seen more clearly from the representative structures presented in Figure 7B and E, which show that in both proteins the segment CC′ is formed first and then the CF contacts are formed. However, at this stage, the reorganization of the segment CC′ and strand F is topologically not favorable to facilitate BE contacts, because the CF contacts cannot be formed without repositioning of strand E. As a result, some contacts which formed earlier in the BE segment break transiently and result in backtracking to facilitate the formation of the contacts in CF. This is also evident from Figure 7B and E, where strand E is found to be away from strand F in the WT and WT_{+GL}. Interestingly, an earlier study that used a simple model for predicting folding routes by Weikl et al.[53] has suggested that the CF contacts in TNfn3 are crucial to balance the loop closure entropy. It is important to note that at $Q \sim 0.3$ the C-terminal end of strand G makes contacts with strand F in WT_{+GL} (see Figures 6E and 7E). On the other hand, for the WT, strand G (missing GL residues) remains predominantly unfolded, particularly at the C-terminal end. On the basis of this observation, we propose that the FG segment in the WT_{+GL} at $Q \sim 0.2$ becomes more rigid than that in the WT (see Figure 4) and as a result it becomes more difficult in the WT_{+GL} to form CF and BE contacts simultaneously. Accordingly, this causes more backtracking in the WT_{+GL} than in the WT. This is also evident from Figure 6F for WT_{+GL} at $Q \sim 0.4$, which shows some contacts in BE are formed initially at $Q \sim 0.2$ (compare to Figure 6D), and then break due to backtracking. In contrast, most of the contacts in BE for the WT remain intact because of less backtracking (see Figure 6C). The core structure of both proteins (made of the strands B, E, C, and F) appears more open-like and not well folded in the transition state (see Figure 7C and F for the WT and WT_{+GL}, respectively), in agreement with studies from other groups.[50−52] From our results, we further infer that, because of more backtracking in BE, the core of the WT_{+GL} appears more mobile and structurally heterogeneous than the core of the WT.

Overall, the structures for the WT and WT_{+GL} at $Q \sim 0.4$ look similar (see Figure 7C and F). In both proteins, the segments CC′ and CF are well structured with high contact probability $\langle Q_i\rangle$. The segment FG is found to be less structured, with lower $\langle Q_i\rangle$, than the segments CC′ and CF for both proteins. In contrast, few contacts between strands B and E are formed and strand A remains completely unfolded and hardly makes any contacts with strand B in both WT and WT_{+GL}. This description agrees well with the TS ensemble structures, which were obtained from simulations, biased with experimental ϕ-values.[50] This analysis has suggested that in most of the structures the A−B−E segment was largely disordered, with strand A almost completely disconnected and few hydrogen bonds existing between strands B and E. On the other hand, the C−C′−F−G segment was found to be predominantly more structured with a larger number of hydrogen bonds formed. It has been further suggested that strand G remains attached to strand F in most of the structures in the TS ensemble,[50] even though this particular strand was characterized by low (0.1−0.2) ϕ-values.[21] Such a description of TSE is in agreement with our results (Figure 7C and F) which also suggests the presence of some contacts in FG at $Q \sim 0.4$ (see also Figure 6C and F). One important point to note here is that, although the ϕ-value analysis in the experiments[21] was obtained from the WT_{+GL}, the simulation by Paci et al.[50] was performed for the 89-residue WT TNfn3 (from PDB ID 1ten) without Gly91-Leu92 and Arg1. Moreover, it should also be noted that the reported low ϕ-values of strand G from experiment did not contain the ϕ-

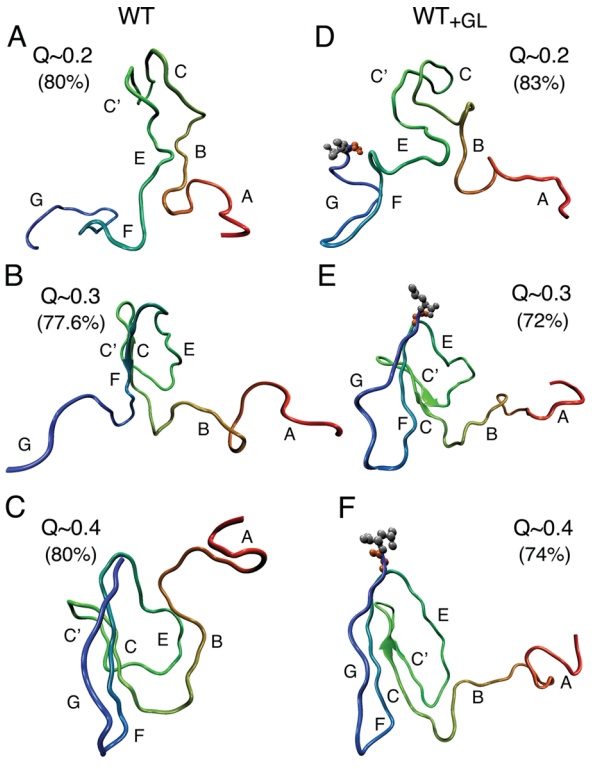


Figure 7. The most representative structures from clustering. The Daura clustering method, as implemented in Gromacs 4.0.7, was used to obtain the structures at different Q. Parts A−C and D−F show the structures at different values of the fraction of native contacts, $Q \sim 0.2$, 0.3, and 0.4, for the WT and WT_{+GL}, respectively, at their corresponding T_F. The population of each centroid structure is indicated in the brackets. The last two residues Gly91 and Leu92 of WT_{+GL} are shown with the ball-and-stick representation. All protein structures are colored from red (N-terminal) to blue (C-terminal).

dx.doi.org/10.1021/jp310046k | J. Phys. Chem. B 2013, 117, 800−810

value of Gly91. In the present study, using a 92-residue structural model for the WT_{+GL}, we observe that the formation of contacts in the FG segment in the unfolded state is initiated by the native interactions involving residues around Gly91 with the residues from strand F near the E−F loop. This further suggests that native interactions of Gly91 with other residues are crucial for the folding and stability of WT_{+GL} TNfn3.

CONCLUDING REMARKS

In the present paper, we used an all-atom, structure-based model and carried out extensive REMD simulations for two variants of an all β-sheet protein, TNfn3, which has an Ig-like fold. Our study reveals that some early native interactions in the unfolded state together with major structural rearrangements are the key to the formation of the folding nucleus of this protein. By conducting a detailed comparative analysis of the folding mechanism of the WT and WT_{+GL} (extended by two residues Gly91-Leu92 in the C-terminal) proteins, we find a significant increase in the stability of the later one, in good agreement with the experimental studies from the Clarke laboratory.[27] This further indicates that the boundary selection of TNfn3 is crucial for its thermal stability, as previously proposed by Hamill et al.[27]

Experimental studies based on the ϕ-value analysis have suggested that initiation of the structural core of TNfn3 proceeds through the early formation of β sheets BE and CF.[21] It was, therefore, concluded that in the TSE the two terminal strands A and G remain predominantly unstructured. A computational study by Paci et al.[50] clarified that although, in the TSE, strand A remains almost unfolded and strand G is significantly structured, in contrast to the experimental results by Hamill et al.[21] and in agreement with our results. These observations further indicate that fractional ϕ-values are difficult to interpret and might not be good predictors of the TSE structures.[54] In addition, due to the limitations in current experimental approaches and techniques, it still remains a challenge to characterize the unfolded state of proteins and to detect the presence of any residual structure that may result in local unfolding and refolding through backtracking.

Finally, we want to emphasize that, besides the influence from the native interactions in the unfolded state, non-native interactions, such as electrostatics, are also crucial for the folding and stability of proteins, as indicated by several studies.[15,16,55−57] In fact, several experimental studies have also demonstrated that by optimizing the charge−charge interactions on the protein surface the stability can be increased substantially.[58−65]

ASSOCIATED CONTENT

Supporting Information

Two figures with distributions of potential energy and Ramachandran maps. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: angel@rpi.edu. Fax: 518-276-2955.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by grants from the National Science Foundation MCB-0110396 to G.I.M., MCB-1050966 to A.E.G., and MCB-1051344.

REFERENCES

(1) Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. Theory of Protein Folding: The Energy Landscape Perspective. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545−600.

(2) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. Funnels, Pathways, and the Energy Landscape of Protein-Folding - a Synthesis. *Proteins: Struct., Funct., Genet.* **1995**, *21*, 167−195.

(3) Veitshans, T.; Klimov, D.; Thirumalai, D. Protein Folding Kinetics: Timescales, Pathways and Energy Landscapes in Terms of Sequence-Dependent Properties. *Folding Des.* **1997**, *2*, 1−22.

(4) Portman, J. J.; Takada, S.; Wolynes, P. G. Variational Theory for Site Resolved Protein Folding Free Energy Surfaces. *Phys. Rev. Lett.* **1998**, *81*, 5237−5240.

(5) Clementi, C.; Garcia, A. E.; Onuchic, J. N. Interplay among Tertiary Contacts, Secondary Structure Formation and Side-Chain Packing in the Protein Folding Mechanism: All-Atom Representation Study of Protein L. *J. Mol. Biol.* **2003**, *326*, 933−954.

(6) Clementi, C.; Nymeyer, H.; Onuchic, J. N. Topological and Energetic Factors: What Determines the Structural Details of the Transition State Ensemble And "En-Route" Intermediates for Protein Folding? An Investigation for Small Globular Proteins. *J. Mol. Biol.* **2000**, *298*, 937−953.

(7) Nymeyer, H.; Garcia, A. E.; Onuchic, J. N. Folding Funnels and Frustration in Off-Lattice Minimalist Protein Landscapes. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5921−5928.

(8) Shimada, J.; Kussell, E. L.; Shakhnovich, E. I. The Folding Thermodynamics and Kinetics of Crambin Using an All-Atom Monte Carlo Simulation. *Protein Folding, Evolution and Design* **2001**, *333*, 161−171.

(9) Shimada, J.; Shakhnovich, E. I. The Ensemble Folding Kinetics of Protein G from an All-Atom Monte Carlo Simulation. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11175−11180.

(10) Whitford, P. C.; Noel, J. K.; Gosavi, S.; Schug, A.; Sanbonmatsu, K. Y.; Onuchic, J. N. An All-Atom Structure-Based Potential for Proteins: Bridging Minimal Models with All-Atom Empirical Force-fields. *Proteins* **2009**, *75*, 430−441.

(11) Noel, J. K.; Whitford, P. C.; Onuchic, J. N. The Shadow Map: A General Contact Definition for Capturing the Dynamics of Biomolecular Folding and Function. *J. Phys. Chem. B* **2012**, *116*, 8692−8702.

(12) Li, L.; Shakhnovich, E. I. Constructing, Verifying, and Dissecting the Folding Transition State of Chymotrypsin Inhibitor 2 with All-Atom Simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 13014−13018.

(13) Meier, S.; Blackledge, M.; Grzesiek, S. Conformational Distributions of Unfolded Polypeptides from Novel Nmr Techniques, *J. Chem. Phys.* **2008**, *128*.

(14) Meier, S.; Grzesiek, S.; Blackledge, M. Mapping the Conformational Landscape of Urea-Denatured Ubiquitin Using Residual Dipolar Couplings. *J. Am. Chem. Soc.* **2007**, *129*, 9799−9807.

(15) Azia, A.; Levy, Y. Nonnative Electrostatic Interactions Can Modulate Protein Folding: Molecular Dynamics with a Grain of Salt. *J. Mol. Biol.* **2009**, *393*, 527−542.

(16) Cho, J. H.; Sato, S.; Horng, J. C.; Anil, B.; Raleigh, D. P. Electrostatic Interactions in the Denatured State Ensemble: Their Effect Upon Protein Folding and Protein Stability. *Arch. Biochem. Biophys.* **2008**, *469*, 20−28.

(17) Dill, K. A.; Shortle, D. Denatured States of Proteins. *Annu. Rev. Biochem.* **1991**, *60*, 795−825.

(18) Leahy, D. J.; Hendrickson, W. A.; Aukhil, I.; Erickson, H. P. Structure of a Fibronectin Type-Iii Domain from Tenascin Phased by Mad Analysis of the Selenomethionyl Protein. *Science* **1992**, *258*, 987−991.

(19) Bork, P.; Holm, L.; Sander, C. The Immunoglobulin Fold - Structural Classification, Sequence Patterns and Common Core. *J. Mol. Biol.* **1994**, *242*, 309−320.
(20) Clarke, J.; Cota, E.; Fowler, S. B.; Hamill, S. J. Folding Studies of Immunoglobulin-Like Beta-Sandwich Proteins Suggest That They Share a Common Folding Pathway. *Structure* **1999**, *7*, 1145−1153.
(21) Hamill, S. J.; Steward, A.; Clarke, J. The Folding of an Immunoglobulin-Like Greek Key Protein Is Defined by a Common-Core Nucleus and Regions Constrained by Topology. *J. Mol. Biol.* **2000**, *297*, 165−178.
(22) Clarke, J.; Hamill, S. J.; Johnson, C. M. Folding and Stability of a Fibronectin Type Iii Domain of Human Tenascin. *J. Mol. Biol.* **1997**, *270*, 771−778.
(23) Fowler, S. B.; Clarke, J. Mapping the Folding Pathway of an Immunoglobulin Domain: Structural Detail from Phi Value Analysis and Movement of the Transition State. *Structure* **2001**, *9*, 355−366.
(24) Wright, C. F.; Lindorff-Larsen, K.; Randles, L. G.; Clarke, J. Parallel Protein-Unfolding Pathways Revealed and Mapped. *Nat. Struct. Biol.* **2003**, *10*, 658−662.
(25) Akke, M.; Liu, J.; Cavanagh, J.; Erickson, H. P.; Palmer, A. G. Pervasive Conformational Fluctuations on Microsecond Time Scales in a Fibronectin Type Iii Domain. *Nat. Struct. Biol.* **1998**, *5*, 55−59.
(26) Carr, P. A.; Erickson, H. P.; Palmer, A. G. Backbone Dynamics of Homologous Fibronectin Type Iii Cell Adhesion Domains from Fibronectin and Tenascin. *Structure* **1997**, *5*, 949−959.
(27) Hamill, S. J.; Meekhof, A. E.; Clarke, J. The Effect of Boundary Selection on the Stability and Folding of the Third Fibronectin Type Iii Domain from Human Tenascin. *Biochemistry* **1998**, *37*, 8071−8079.
(28) Meekhof, A. E.; Hamill, S. J.; Arcus, V. L.; Clarke, J.; Freund, S. M. V. The Dependence of Chemical Exchange on Boundary Selection in a Fibronectin Type Iii Domain from Human Tenascin. *J. Mol. Biol.* **1998**, *282*, 181−194.
(29) Gosavi, S.; Chavez, L. L.; Jennings, P. A.; Onuchic, J. N. Topological Frustration and the Folding of Interleukin-1 Beta. *J. Mol. Biol.* **2006**, *357*, 986−996.
(30) Shea, J. E.; Onuchic, J. N.; Brooks, C. L. Exploring the Origins of Topological Frustration: Design of a Minimally Frustrated Model of Fragment B of Protein A. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 12512−12517.
(31) Capraro, D. T.; Roy, M.; Onuchic, J. N.; Jennings, P. A. Backtracking on the Folding Landscape of the Beta-Trefoil Protein Interleukin-1 Beta? *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 14844−14848.
(32) Fersht, A. R. Optimization of Rates of Protein-Folding - the Nucleation-Condensation Mechanism and Its Implications. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 10869−10873.
(33) Gosavi, S.; Whitford, P. C.; Jennings, P. A.; Onuchic, J. N. Extracting Function from a Beta-Trefoil Folding Motif. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 10384−10389.
(34) Hills, R. D.; Brooks, C. L. Subdomain Competition, Cooperativity, and Topological Frustration in the Folding of Chey. *J. Mol. Biol.* **2008**, *382*, 485−495.
(35) Hamill, S. J.; Cota, E.; Chothia, C.; Clarke, J. Conservation of Folding and Stability within a Protein Family: The Tyrosine Corner as an Evolutionary Cul-De-Sac. *J. Mol. Biol.* **2000**, *295*, 641−649.
(36) Marti-Renom, M. A.; Stuart, A. C.; Fiser, A.; Sanchez, R.; Melo, F.; Sali, A. Comparative Protein Structure Modeling of Genes and Genomes. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 291−325.
(37) Scott, W. R. P.; Hunenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Kruger, P.; van Gunsteren, W. F. The Gromos Biomolecular Simulation Program Package. *J. Phys. Chem. A* **1999**, *103*, 3596−3607.
(38) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hunenberger, P. H.; Kruger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: The Gromos96 Manual and User Guide*; Hochschuleverlag AG an der ETH Zurich: Zurich, Switzerland, 1996.
(39) Kouza, M.; Li, M. S.; O'Brien, E, P., Jr.; Hu, C. K.; Thirumalai, D. Effect of Finite Size on Cooperativity and Rates of Protein Folding. *J. Phys. Chem. A* **2006**, *110*, 671−676.
(40) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. Gromacs 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(41) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(42) Ferrenberg, A. M.; Swendsen, R. H. New Monte-Carlo Technique for Studying Phase-Transitions. *Phys. Rev. Lett.* **1988**, *61*, 2635−2638.
(43) Ferrenberg, A. M.; Swendsen, R. H. Optimized Monte-Carlo Data-Analysis. *Phys. Rev. Lett.* **1989**, *63*, 1195−1198.
(44) Meekhof, A. E.; Freund, S. M. V. Probing Residual Structure and Backbone Dynamics on the Milli- to Picosecond Timescale in a Urea-Denatured Fibronectin Type Iii Domain. *J. Mol. Biol.* **1999**, *286*, 579−592.
(45) Miyashita, O.; Onuchic, J. N.; Wolynes, P. G. Nonlinear Elasticity, Proteinquakes, and the Energy Landscapes of Functional Transitions in Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12570−12575.
(46) Tripathi, S.; Portman, J. J. Inherent Flexibility Determines the Transition Mechanisms of the Ef-Hands of Calmodulin. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 2104−2109.
(47) Tripathi, S.; Portman, J. J. Conformational Flexibility and the Mechanisms of Allosteric Transitions in Topologically Similar Proteins. *J. Chem. Phys.* **2011**, *135*, 075104−075109.
(48) Whitford, P. C.; Miyashita, O.; Levy, Y.; Onuchic, J. N. Conformational Transitions of Adenylate Kinase: Switching by Cracking. *J. Mol. Biol.* **2007**, *366*, 1661−1671.
(49) Whitford, P. C.; Onuchic, J. N.; Wolynes, P. G. Energy Landscape Along an Enzymatic Reaction Trajectory: Hinges or Cracks? *HFSP J.* **2008**, *2*, 61−64.
(50) Paci, E.; Clarke, J.; Steward, A.; Vendruscolo, M.; Karplus, M. Self-Consistent Determination of the Transition State for Protein Folding: Application to a Fibronectin Type Iii Domain. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 394−399.
(51) Geierhaas, C. D.; Paci, E.; Vendruscolo, M.; Clarke, J. Comparison of the Transition States for Folding of Two Ig-Like Proteins from Different Superfamilies. *J. Mol. Biol.* **2004**, *343*, 1111−1123.
(52) Best, R. B.; Rutherford, T. J.; Freund, S. M. V.; Clarke, J. Hydrophobic Core Fluidity of Homologous Protein Domains: Relation of Side-Chain Dynamics to Core Composition and Packing. *Biochemistry* **2004**, *43*, 1145−1155.
(53) Weikl, T. R.; Dill, K. A. Folding Rates and Low-Entropy-Loss Routes of Two-State Proteins. *J. Mol. Biol.* **2003**, *329*, 585−598.
(54) Naganathan, A. N.; Munoz, V. Insights into Protein Folding Mechanisms from Large Scale Analysis of Mutational Effects. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 8611−8616.
(55) Pace, C. N.; Alston, R. W.; Shaw, K. L. Charge-Charge Interactions Influence the Denatured State Ensemble and Contribute to Protein Stability. *Protein Sci.* **2000**, *9*, 1395−1398.
(56) Zarrine-Afsar, A.; Zhang, Z. Q.; Schweiker, K. L.; Makhatadze, G. I.; Davidson, A. R.; Chan, H. S. Kinetic Consequences of Native State Optimization of Surface-Exposed Electrostatic Interactions in the Fyn Sh3 Domain. *Proteins: Struct., Funct., Bioinf.* **2012**, *80*, 858−870.
(57) Zhou, H. X. A Gaussian-Chain Model for Treating Residual Charge-Charge Interactions in the Unfolded State of Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 3569−3574.
(58) Gribenko, A. V.; Makhatadze, G. I. Role of the Charge-Charge Interactions in Defining Stability and Halophilicity of the Cspb Proteins. *J. Mol. Biol.* **2007**, *366*, 842−856.
(59) Gribenko, A. V.; Patel, M. M.; Liu, J.; McCallum, S. A.; Wang, C. Y.; Makhatadze, G. I. Rational Stabilization of Enzymes by Computational Redesign of Surface Charge-Charge Interactions. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 2601−2606.
(60) Loladze, V. V.; Makhatadze, G. I. Removal of Surface Charge-Charge Interactions from Ubiquitin Leaves the Protein Folded and Very Stable. *Protein Sci.* **2002**, *11*, 174−177.
(61) Makhatadze, G. I.; Loladze, V. V.; Gribenko, A. V.; Lopez, M. M. Mechanism of Thermostabilization in a Designed Cold Shock

Protein with Optimized Surface Electrostatic Interactions. *J. Mol. Biol.* **2004**, *336*, 929−942.
(62) Sanchez-Ruiz, J. M.; Makhatadze, G. I. To Charge or Not to Charge? *Trends Biotechnol.* **2001**, *19*, 132−135.
(63) Schweiker, K. L.; Zarrine-Afsar, A.; Davidson, A. R.; Makhatadze, G. I. Computational Design of the Fyn Sh3 Domain with Increased Stability through Optimization of Surface Charge-Charge Interactions. *Protein Sci.* **2007**, *16*, 2694−2702.
(64) Spector, S.; Wang, M. H.; Carp, S. A.; Robblee, J.; Hendsch, Z. S.; Fairman, R.; Tidor, B.; Raleigh, D. P. Rational Modification of Protein Stability by the Mutation of Charged Surface Residues. *Biochemistry* **2000**, *39*, 872−879.
(65) Strickler, S. S.; Gribenko, A. V.; Gribenko, A. V.; Keiffer, T. R.; Tomlinson, J.; Reihle, T.; Loladze, V. V.; Makhatadze, G. I. Protein Stability and Surface Electrostatics: A Charged Relationship. *Biochemistry* **2006**, *45*, 2761−2766.
(66) Humphrey, W.; Dalke, A.; Schulten, K. Vmd: Visual Molecular Dynamics. *J. Mol. Graphics Modell.* **1996**, *14*, 33−38.