Article

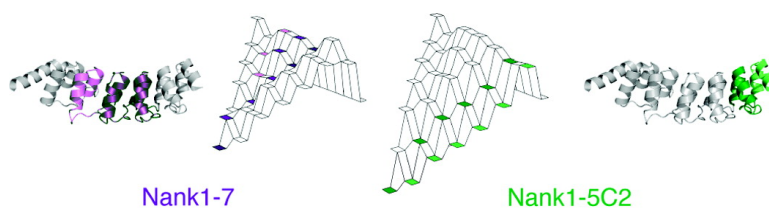# Rerouting the Folding Pathway of the Notch Ankyrin Domain by Reshaping the Energy Landscape

Katherine W. Tripp, and Doug Barrick

Nank1-7          Nank1-5C2

## More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 1 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

# Rerouting the Folding Pathway of the Notch Ankyrin Domain by Reshaping the Energy Landscape

Katherine W. Tripp[†] and Doug Barrick*

*The T.C. Jenkins Department of Biophysics, Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218*

Received August 22, 2007; E-mail: barrick@jhu.edu

***Abstract:*** The modular nature of repeat proteins has made them a successful target for protein design. Ankyrin repeat, TPR, and leucine rich repeat domains that have been designed solely on consensus information have been shown to have higher thermostability than their biological counterparts. We have previously shown that we can reshape the energy landscape of a repeat protein by adding multiple C-terminal consensus ankyrin repeats to the five N-terminal repeats of the Notch ankyrin domain. Here we explore how the folding mechanism responds to reshaping of the energy landscape. We have used analogous substitutions of a conserved alanine with glycine in each repeat to determine the distribution of structure in the transition state ensembles of constructs containing one (Nank1–5C$_1$) and two consensus (Nank1–5C$_2$) ankyrin repeats. Whereas folding of the wild-type Notch ankyrin domain is slowed by substitutions in its central repeats,[1] folding of Nank1–5C$_1$ and Nank1–5C$_2$ is slowed by substitutions in the C-terminal repeats. Thus, the addition of C-terminal stabilizing repeats shifts the transition state ensemble toward the C-terminal repeats, rerouting the folding pathway of the ankyrin repeat domain. These findings indicate that, for the Notch ankyrin domain, folding pathways are selected based on local energetics.

## Introduction

The process of protein folding is one of the more remarkable examples of spontaneous self-organization in nature. In most cases, the search for the native structure from a very large ensemble of denatured states occurs on a relatively short time scale. A variety of mechanisms have been proposed to describe how this very large search can be executed so quickly.[2–5] Some proposed mechanisms involve specific low energy pathways that restrict the search from the denatured to the native state.[2,6,7] In the past two decades, experimental studies have shown that specific pathways are common in the folding of globular proteins, as shown by both the formation of specific structures in transition state ensembles[8–12] and the formation of specific kinetic intermediates during folding.[13–18] However, the factors that determine which pathways are followed and their relative importance remain unresolved.

One factor that may specify folding pathways is the local stability of structural elements. As the native state is of lowest energy under folding conditions, individual segments of chain are likely to have low energies in their native configurations (in terms of both secondary structures and tertiary contacts), even when much of the chain remains unstructured. Thus, experimentally observed pathway preferences may reflect the most stable elements of structure. However, for globular proteins, it has been notoriously difficult to determine which elements of structure are most stable, which likely results from the highly interconnected structure of the native state.[19]

Another factor that clearly influences kinetics is the degree to which contacts in the native state are local in the primary sequence, often quantified by a parameter known as contact order. The influence of sequence-local interactions, a reflection of overall chain topology in accelerating folding, has been developed into several models for protein folding.[20,21] Topology

---

† Current address: Institute for Quantitative Biology, University of California, Berkeley, Berkeley, CA 94720.

(1) Bradley, C. M.; Barrick, D. *Structure (Cambridge, MA, U.S.)* **2006**, *14*, 1303–12.
(2) Kim, P. S.; Baldwin, R. L. *Annu. Rev. Biochem.* **1990**, *59*, 631–60.
(3) Dill, K. A.; Chan, H. S. *Nat. Struct. Biol.* **1997**, *4*, 10–9.
(4) Bilsel, O.; Matthews, C. R. *Adv. Protein Chem.* **2000**, *53*, 153–207.
(5) Onuchic, J. N.; Nymeyer, H.; Garcia, A. E.; Chahine, J.; Socci, N. D. *Adv. Protein Chem.* **2000**, *53*, 87–152.
(6) Bashford, D.; Weaver, D. L.; Karplus, M. *J. Biomol. Struct. Dyn.* **1984**, *1*, 1243–55.
(7) Myers, J. K.; Oas, T. G. *Annu. Rev. Biochem.* **2002**, *71*, 783–815.
(8) Grantcharova, V. P.; Riddle, D. S.; Santiago, J. V.; Baker, D. *Nat. Struct. Biol.* **1998**, *5*, 714–20.
(9) Kim, D. E.; Fisher, C.; Baker, D. *J. Mol. Biol.* **2000**, *298*, 971–84.
(10) Northey, J. G.; Di Nardo, A. A.; Davidson, A. R. *Nat. Struct. Biol.* **2002**, *9*, 126–30.
(11) Mayor, U.; Guydosh, N. R.; Johnson, C. M.; Grossmann, J. G.; Sato, S.; Jas, G. S.; Freund, S. M.; Alonso, D. O.; Daggett, V.; Fersht, A. R. *Nature* **2003**, *421*, 863–7.
(12) Li, Y.; Gupta, R.; Cho, J. H.; Raleigh, D. P. *Biochemistry* **2007**, *46*, 1013–21.
(13) Briggs, M. S.; Roder, H. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 2017–21.
(14) Roder, H.; Elove, G. A.; Englander, S. W. *Nature* **1988**, *335*, 700–4.
(15) Udgaonkar, J. B.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 8197–201.
(16) Miranker, A.; Radford, S. E.; Karplus, M.; Dobson, C. M. *Nature* **1991**, *349*, 633–6.
(17) Krishna, M. M.; Lin, Y.; Mayne, L.; Englander, S. W. *J. Mol. Biol.* **2003**, *334*, 501–13.
(18) Bycroft, M.; Matouschek, A.; Kellis, J. T., Jr.; Serrano, L.; Fersht, A. R. *Nature* **1990**, *346*, 488–90.
(19) Kloss, E.; Courtemanche, N.; Barrick, D. *Arch. Biochem. Biophys.* **2008**, *469*, 83–99.
(20) Weikl, T. R.; Dill, K. A. *J. Mol. Biol.* **2003**, *329*, 585–98.

has also been invoked as a major factor behind the successful prediction of folding pathways in globular proteins by simple funneled Go-models,[22] in which all native contacts are treated as uniformly favorable.[23-25] The success of Go-models in predicting the folding pathways of several proteins without energetic bias for specific contacts calls into question the importance of local stability in determining folding pathways, although a few recent simulations have indicated that folding is influenced by essential energetic components not accounted for by overall topological features.[24,26,27]

In comparison to globular proteins, the structures of repeat proteins are highly conducive to testing for the existence of preferred pathways and for determining how such pathways are specified. Repeat proteins are constructed of a linear array of structural elements, giving rise to a regular topology lacking contacts that are distant in the primary sequence. The large (and open-ended) size of repeat proteins, along with high structural similarity from repeat to repeat, might be expected to result in multiple distributed folding pathways. Folding of repeat proteins through multiple parallel pathways is especially likely if folding is guided by topology rather than local stability. In contrast, folding of repeat proteins by specific pathways would highlight the importance of additional factors in pathway selection. The relationship between pathway selection and local stability can be experimentally determined by taking advantage of the modular architecture and structural redundancy of repeat proteins to map the folding energies of different fragments at single-repeat resolution.[28]

Previous studies on the folding pathway of several ankyrin repeat proteins suggest a limited number of pathways during folding.[1,29,30] For the ankyrin domain from the Drosophila Notch receptor, the subject of this study, folding appears to proceed through a short-lived on-pathway kinetic intermediate.[31,32] Φ-Value analysis demonstrates that folding of this intermediate is limited by structure formation in the central repeats (primarily repeats 3–5) and that the C-terminal repeats become structured in a fast step that converts the intermediate to the native state.[1] The experimentally determined energy landscape suggests that, compared to a more C-terminal folding route, the observed folding route is low in energy.

To stringently test whether the observed folding route is selected based on its relatively low energy, we replaced the C-terminal repeats of the Notch ankyrin domain with highly stabilizing consensus repeats.[33] These consensus repeats substantially stabilize the entire Notch ankyrin domain, decreasing the free energy of folding by approximately 6 kcal/mol.[34] To
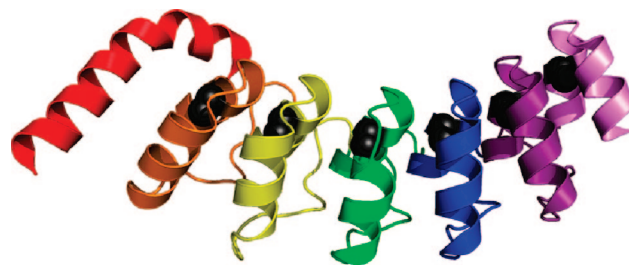


**Figure 1.** Ribbon representation of the 2.0 Å crystal structure of the Notch ankyrin domain.[35] The N-terminal repeat, which is partly disordered, is colored red; repeats two through seven are colored orange through lavender. The conserved alanine in each repeat is represented in CPK. This figure was generated using MacPymol.[36]

determine how the distribution of structure in the folding pathway and transition state ensemble changes by this alteration of the energy landscape, we have measured the effects of conserved alanine to glycine substitutions in each repeat of the C-terminally stabilized Notch ankyrin domain consensus fusions. The results support the existence of specific folding pathways, provide a measure of the degree to which the transition state ensemble can be rerouted, and demonstrate a direct connection between the equilibrium energy landscape and kinetic folding pathways, suggesting that folding pathways are determined by local thermodynamic stabilities.

## Results

Ankyrin repeats are a class of repeat proteins composed of multiple tandem structural elements that are 33 residues in length. Ankyrin repeats contain two antiparallel α-helices that are connected by a short loop; adjacent repeats are connected to each other through an extended loop with a tight β-turn. Adjacent repeats are packed extensively against their nearest-neighbors to form an extended linear array. The Notch ankyrin domain has seven ankyrin sequence repeats, six of which are structured (Figure 1).[35]

**Thermodynamic Consequences of Alanine to Glycine Substitutions.** The tandem architecture of repeat proteins allows substitutions to be made to residues that are in structurally similar environments in different parts of the chain. We have used a series of substitutions of a conserved alanine (Figure 1) with a glycine in each repeat to probe the limits of structure in the transition state ensembles of our C-terminally stabilized constructs.[1,37] Because each alanine contacts residues within the same repeat, this substitution series allows us to interrogate the extent of structure formation within each repeat.

The effect of each substitution on stability in different backgrounds (Nank1–5, Nank1–5C$_1$, and Nank1–5C$_2$) was determined by equilibrium CD- and fluorescence-monitored urea denaturation. Nank1–5C$_1$ and Nank1–5C$_2$ indicate C-terminal fusions of one and two consensus repeats to repeats 1–5 of the Notch ankyrin domain. The construct Nank1–5 consists only of the five N-terminal repeats and is used along with the full-length seven repeat Notch ankyrin domain (Nank1–7Δ) as a reference point for landscape redistribution. With the exception of the first repeat, all glycine substitutions are destabilizing

(21) Makarov, D. E.; Plaxco, K. W. *Protein Sci.* **2003**, *12*, 17–26.
(22) Koga, N.; Takada, S. *J. Mol. Biol.* **2001**, *313*, 171–80.
(23) Clementi, C.; Jennings, P. A.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5871–6.
(24) Clementi, C.; Nymeyer, H.; Onuchic, J. N. *J. Mol. Biol.* **2000**, *298*, 937–53.
(25) Weinkam, P.; Zong, C.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 12401–6.
(26) Wallin, S.; Chan, H. S. *Protein Sci.* **2005**, *14*, 1643–60.
(27) Merlo, C.; Dill, K. A.; Weikl, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 10171–5.
(28) Mello, C. C.; Barrick, D. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 14102–7.
(29) Tang, K. S.; Fersht, A. R.; Itzhaki, L. S. *Structure* **2003**, *11*, 67–73.
(30) Lowe, A. R.; Itzhaki, L. S. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 2679–84.
(31) Mello, C. C.; Bradley, C. M.; Tripp, K. W.; Barrick, D. *J. Mol. Biol.* **2005**, *352*, 266–81.
(32) Bradley, C. M.; Barrick, D. *J. Mol. Biol.* **2005**, *352*, 253–65.
(33) Tripp, K. W.; Barrick, D. *J. Mol. Biol.* **2007**, *365*, 1187–200.

(34) Tripp, K. W.; Barrick, D. *J. Mol. Biol.* **2004**, *344*, 169–78.
(35) Zweifel, M. E.; Leahy, D. J.; Hughson, F. M.; Barrick, D. *Protein Sci.* **2003**, *12*, 2622–2632.
(36) Bradley, C. M.; Barrick, D. *J. Mol. Biol.* **2002**, *324*, 373–386.
(37) McCallister, E. L.; Alm, E.; Baker, D. *Nat. Struct. Biol.* **2000**, *7*, 669–73.
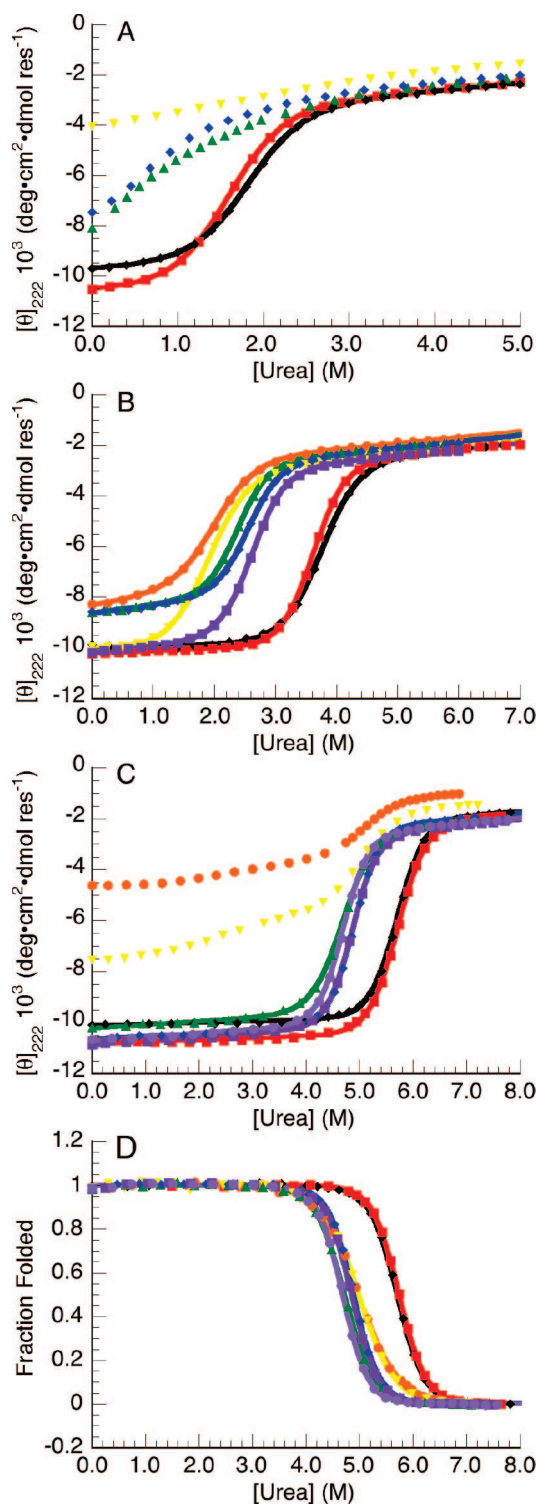
**Figure 2.** Effects of conserved alanine to glycine substitutions on equilibrium stability. CD-monitored urea-induced unfolding transitions of constructs that contain alanine to glycine subsitutions in Nank1–5 (A), Nank1–5C$_1$ (B), and Nank1–5C$_2$ (C). Fluorescence-monitored transitions of substitutions in Nank1–5C$_2$ (D). The position of each substitution is indicated as follows: first repeat, red; second repeat, orange; third repeat, yellow; fourth repeat, green; fifth repeat, blue; first designed, purple; second designed repeat, light purple.

(Figure 2). Moreover, the effect of each glycine substitution is similar in the three different backgrounds (Table 1, Supporting Information Figure 1), suggesting that the structure of the naturally occurring repeats are not strongly perturbed by

C-terminal consensus stabilization. The most destabilizing substitutions are in the second and third repeat (supplemental Figure 1). The substitutions made in the fourth, fifth, and consensus repeats are less destabilizing: $\Delta\Delta G°$ values in these repeats are less than half that for substitutions in the second and third repeats. The substitutions in the first repeat have little to no effect on the unfolding free energy, consistent with it being partly disordered.[38] The variation in $\Delta\Delta G°$ among repeats two through seven (Figure S1) is surprisingly large, given the similar structures of each repeat. This variation, which must be a result of the high sequence variation among repeats, underscores the fact that very subtle structural differences can lead to large variations in the free energy of folding.

Because the unfolding free energy of Nank1–5 is approximately 3.2 kcal/mol, the substitutions made in the second and third repeat, which are the most destabilizing, result in polypeptides that are unfolded under native conditions. Indeed, for Nank1–5 (AG3), no clear transition is observed, and it is likely that the observed signal represents only the denatured baseline. (AG3) indicates that an alanine to glycine substitution has been made in the third repeat of Nank1–5. This is consistent with decreased CD in both the far- and near-UV regions. For the less destabilizing substitutions in the fourth and fifth repeat, only a partial unfolding transition is observed in the Nank1–5 background (Figure 2A).

In Nank1–5C$_2$, the most C-terminally stabilized construct, the substitutions in the second and third repeats [Nank1–5C$_2$ (AG2) and Nank1–5C$_2$ (AG3)] result in an equilibrium unfolding intermediate separated from the native and denatured states by two distinct far-UV CD unfolding transitions (Figure 2C). The transition at low urea concentrations is broad, whereas that at higher concentrations is steep. In contrast, fluorescence-monitored unfolding curves of Nank1–5C$_2$ (AG2) and Nank1–5C$_2$ (AG3) show a single unfolding transition at the same urea concentration as the second transition seen by CD (Figure 2D). The thermodynamic parameters obtained from fits to the fluorescence-monitored unfolding transitions of Nank1–5C$_2$ (AG2) and Nank1–5C$_2$ (AG3) indicate a 5 kcal/mol decrease in the $\Delta G°_{u,H_2O}$ (corresponding to the unfolding of the intermediate) and a 0.6 kcal/mol·M decrease in the $m$-value, consistent with a smaller cooperative unit (Table 1).

**Kinetic Consequences of Alanine to Glycine Substitutions.** To determine the effect that each substitution has on the refolding and unfolding kinetics as the energy landscape is shifted by addition of C-terminally stabilizing consensus repeats, we measured the kinetics of refolding and unfolding of the alanine to glycine substitutions in Nank1–5, Nank1–5C$_1$, and Nank1–5C$_2$ using stopped-flow fluorescence. Previous studies have shown that Nank1–5, Nank1–5C$_1$, and Nank1–5C$_2$ show two kinetic refolding phases: a dominant fast phase and a second, minor phase that is limited by prolyl isomerization, consistent with Nank1–7Δ (the parent construct).[33] With the exception of the highly destabilized Nank1–5 (AG4), all of the alanine to glycine substitutions also show two refolding phases.

However, in contrast to the parent construct, Nank1–5, Nank1–5C$_1$, and Nank1–5C$_2$ show a single unfolding phase,[33] as do the glycine substitutions in those backgrounds. For the parent construct, a second observed unfolding phase results from an on-pathway kinetic intermediate; the observation of two unfolding phases and their urea dependencies allows a kinetic three-state mechanism to be fitted to the rate constants and

(38) Nauli, S.; Kuhlman, B.; Baker, D. *Nat. Struct. Biol.* **2001**, *8*, 602–5.

***Table 1.*** Urea-Induced Unfolding Parameters for Variants That Contain Alanine to Glycine Substitutions in Nank1−5, Nank1−5C$_1$, and Nank1−5C$_2$

| | $\Delta G^{\circ}_{H_2O}$(kcal·mol$^{-1}$) | | $m$-value (kcal·mol$^{-1}$·M$^{-1}$) | | $C_m$ (M) | |
|---|---|---|---|---|---|---|
| | CD | fluorescence | CD | fluorescence | CD | fluorescence |
| **Nank1−5** | 3.21 ± 0.09 | 3.29 ± 0.03 | 1.78 ± 0.03 | 1.84 ± 0.01 | 1.80 ± 0.09 | 1.78 ± 0.02 |
| Nank1−5 (AG1) | 3.06 ± 0.06 | 3.17 ± 0.13 | 1.83 ± 0.02 | 1.89 ± 0.05 | 1.67 ± 0.02 | 1.68 ± 0.03 |
| | | | | | | |
| **Nank1−5C$_1$** | 7.31 ± 0.17 | 7.50 ± 0.09 | 1.98 ± 0.10 | 2.05 ± 0.03 | 3.67 ± 0.04 | 3.65 ± 0.02 |
| Nank1−5C$_1$ (AG1) | 8.21 ± 0.06 | 8.16 ± 0.24 | 2.25 ± 0.01 | 2.22 ± 0.07 | 3.65 ± 0.01 | 3.67 ± 0.01 |
| Nank1−5C$_1$ (AG2) | 3.72 ± 0.09 | 3.88 ± 0.09 | 1.77 ± 0.04 | 1.85 ± 0.05 | 2.10 ± 0.07 | 2.10 ± 0.07 |
| Nank1−5C$_1$ (AG3) | 3.61 ± 0.13 | 4.00 ± 0.16 | 1.84 ± 0.03 | 2.01 ± 0.07 | 1.96 ± 0.04 | 1.99 ± 0.01 |
| Nank1−5C$_1$ (AG4) | 5.52 ± 0.04 | 5.69 ± 0.11 | 2.32 ± 0.01 | 2.41 ± 0.04 | 2.38 ± 0.01 | 2.36 ± 0.01 |
| Nank1−5C$_1$ (AG5) | 5.71 ± 0.09 | 5.77 ± 0.07 | 2.22 ± 0.03 | 2.26 ± 0.02 | 2.58 ± 0.01 | 2.56 ± 0.01 |
| Nank1−5C$_1$ (AGC1) | 5.68 ± 0.06 | 5.66 ± 0.08 | 2.18 ± 0.01 | 2.17 ± 0.02 | 2.61 ± 0.01 | 2.61 ± 0.02 |
| | | | | | | |
| **Nank1−5C$_2$** | 13.0 ± 0.03 | 13.25 ± 0.40 | 2.28 ± 0.01 | 2.30 ± 0.07 | 5.70 ± 0.01 | 5.70 ± 0.01 |
| Nank1−5C$_2$ (AG1) | 12.8 ± 0.12 | 12.62 ± 0.09 | 2.21 ± 0.02 | 2.17 ± 0.01 | 5.77 ± 0.02 | 5.82 ± 0.01 |
| Nank1−5C$_2$ (AG2) | NA$^a$ | (8.98, 7.95) | NA$^a$ | (1.78, 1.59) | NA$^a$ | (5.06, 4.98) |
| Nank1−5C$_2$ (AG3) | NA$^a$ | 8.12 ± 0.22 | NA$^a$ | 1.64 ± 0.04 | NA$^a$ | 4.97 ± 0.01 |
| Nank1−5C$_2$ (AG4) | 10.1 ± 0.05 | 9.59 ± 0.10 | 2.14 ± 0.01 | 2.06 ± 0.03 | 4.72 ± 0.02 | 4.66 ± 0.02 |
| Nank1−5C$_2$ (AG5) | 11.9 ± 0.07 | 11.46 ± 0.28 | 2.44 ± 0.02 | 2.35 ± 0.05 | 4.89 ± 0.01 | 4.88 ± 0.01 |
| Nank1−5C$_2$ (AGC1) | 11.8 ± 0.08 | 11.65 ± 0.13 | 2.45 ± 0.02 | 2.41 ± 0.03 | 4.85 ± 0.01 | 4.84 ± 0.01 |
| Nank1−5C$_2$ (AGC2) | 11.7 ± 0.04 | 11.61 ± 0.14 | 2.50 ± 0.01 | 2.48 ± 0.03 | 4.69 ± 0.01 | 4.68 ± 0.01 |

$^a$ Parameters were estimated by fitting urea-induced unfolding transitions using nonlinear least-squares as described.[38] Uncertainties are standard errors on the mean of at least three titrations. Unfolding parameters determined using CD could not be obtained for Nank1−5C$_2$ (AG2) and Nank1−5C$_2$ (AG3) due to the existence of a folding intermediate. Conditions: 25 mM Tris·HCl (pH 8.0), 150 mM NaCl, 20 °C.

amplitudes for unfolding and refolding, providing tight constraints on the underlying kinetic constants.[31] Although the urea dependencies of the refolding and unfolding rate constants (Figure 3) are nonlinear for the glycine variants, suggesting a complex kinetic mechanism, for most variants this nonlinearity is not sufficient to adequately constrain the underlying kinetic constants in a three-state fit. Thus, we have simply used a nonlinear model (eq 2, Materials and Methods) to fit the urea dependencies of the refolding and unfolding rate constants of the glycine variants. Although this model is comparatively phenomenological, it provides an accurate estimate for the rate of folding in the absence of denaturant, allowing the effect of glycine substitution on the rate-limiting step in folding to be determined.

The glycine substitutions in Nank1−5C$_1$ have varied effects on the refolding and unfolding rate constants. Substitutions in the first, second, third, and fourth repeat have little effect on the refolding rate constants, although at urea concentrations near the $C_m$, the refolding rate constant of Nank1−5C$_1$ (AG4) is smaller than that of Nank1−5C$_1$ (Figure 3). Instead, the destabilization seen in the equilibrium data seems to arise from increases in the unfolding rates. In contrast, the substitutions in the two C-terminal repeats (the fifth and consensus repeat) decrease the refolding rate constants at all urea concentrations, corresponding to decreases in $k_{f,H_2O}$ of 10- and 6-fold for Nank1−5C$_1$ (AG5) and Nank1−5C$^1$ (AGC1), respectively (Table 2). AGC1 and AGC2 refer to substitutions in the first and second C-terminal consensus repeat, respectively.

As with the glycine substitutions in Nank1−5C$_1$, only the C-terminal substitutions in Nank1−5C$_2$ have an effect on the refolding rate constants. None of the substitutions in the N-terminal or central repeats of Nank1−5C$_2$ (repeats one through five) significantly affect the refolding rate constants but instead increase the unfolding rate constants (Figure 3). In contrast, substitutions in the two C-terminal consensus repeats have a large effect on the refolding rate constants at all urea concentrations. Values for $k_{f,H_2O}$ are decreased by 20- and 3-fold for Nank1−5C$_2$ (AGC1) and Nank1−5C$_2$ (AGC2), respectively.

Since the cost of the alanine to glycine substitutions in repeats two and three exceeds the global stability of Nank1−5 (3.2 kcal/mol), we have only been able to examine the kinetic consequences of glycine substitutions in the first, fourth, and fifth repeat. The refolding and unfolding rates of Nank1−5 (AG1) are nearly identical to those measured for Nank1−5 at all urea concentrations (Figure 3). In contrast, the refolding rate constants of the variants that contain substitutions in the fourth and fifth repeat are smaller than the refolding rate constants of Nank1−5 at all urea concentrations. The extrapolated refolding rate constants in water ($k_{f,H_2O}$; 0 M urea) are decreased by a factor of 10 and 5, respectively (Table 2), for Nank1−5 (AG4) and Nank1−5 (AG5). In addition, the unfolding rate constants of Nank1−5 (AG4) and Nank1−5 (AG5) are larger than the unfolding rate constants of Nank1−5.

## Discussion

The tolerance of the Notch ankyrin domain to deletion of multiple repeats has allowed the equilibrium energy landscape to be mapped at single-repeat resolution.[28] We have used the same analysis to compare how the folding energy landscape of Nank1−5C$_2$ is altered from that of the Nank1−7Δ.[33] Although Nank1−7Δ and Nank1−5C$_2$ contain the same number of ankyrin repeats, the unfolding free energy of Nank1−5C$_2$ is 13 kcal/mol, which is over 5 kcal/mol more stable than Nank1−7Δ (Figure 4A and 4D). Because this enhancement results from stabilization on the end of this linear domain, it results in a major perturbation in the energy landscape of Nank1−5C$_2$ (compare Figure 4A and 4D). The partially folded forms that include the two C-terminal consensus repeats are much more stable than other partially folded forms of the same repeat length, whereas in Nank1−7Δ the same C-terminal conformations are less stable than those including repeats three through five. This shift suggests that if the kinetic pathways for refolding are thermodynamically selected, the folding pathway of Nank1−5C$_2$ should be C-terminally rerouted.

**Structure of the Transition State Ensemble of Nank1−5, Nank1−5C$_1$, and Nank1−5C$_2$.** The structure of the transition state ensembles in refolding of the C-terminally stabilized constructs
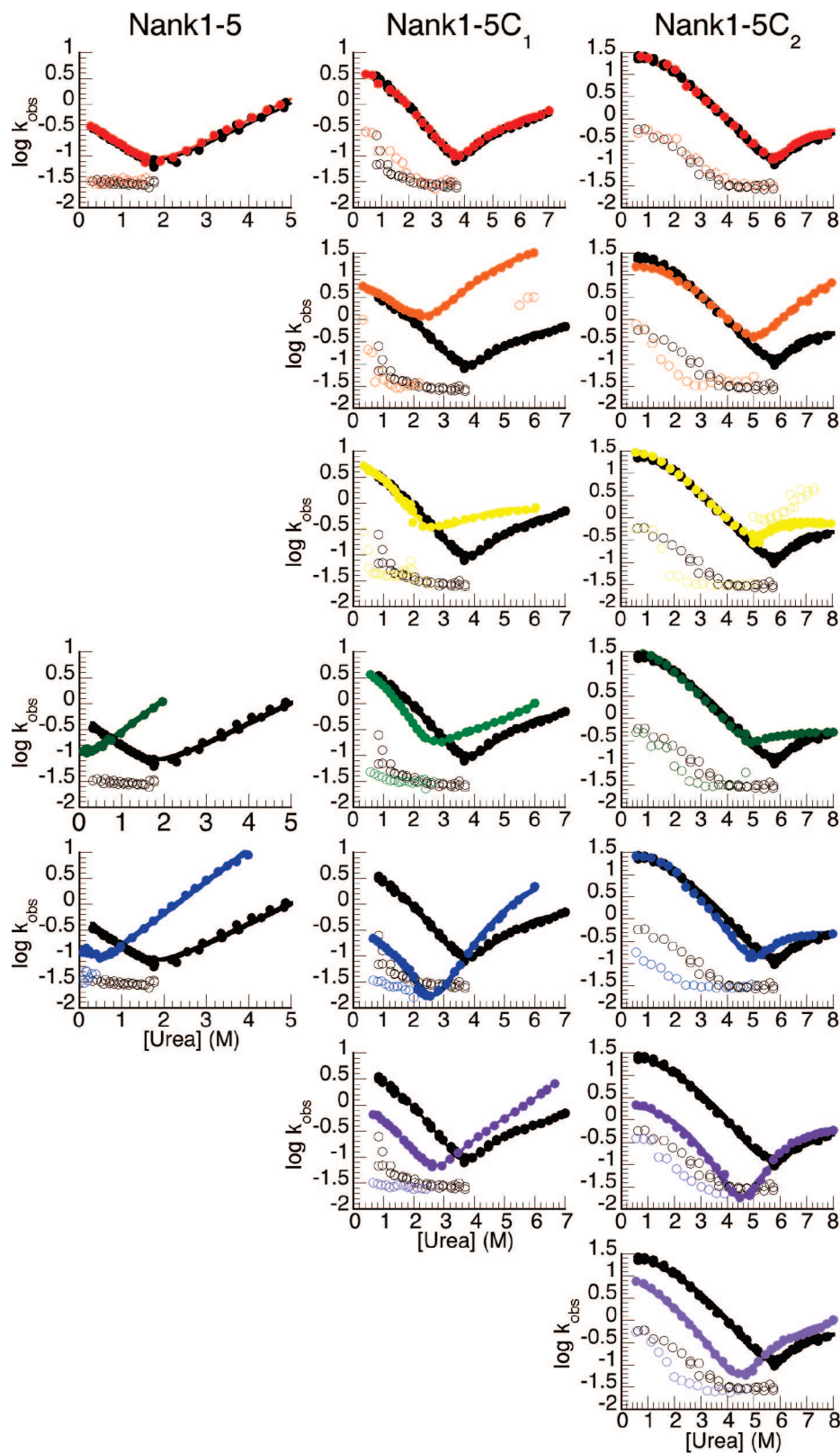
**Figure 3.** Effects of conserved alanine to glycine substitutions on folding kinetics. Chevron plots (log $k_{obs}$ vs urea concentraion), monitored by tryptophan fluorescence, alanine to glycine variants. Rate constants of constructs that contain alanine to glycine substitutions in Nank1–5 (left column), Nank1–5C$_1$ (center column), and Nank1–5C$_2$ (right column). Filled circles represent rate constants for the major refolding and unfolding phases; open circles represent rate constants for the minor refolding and unfolding phases. Each panel shows a single alanine to glycine substitution (colored as in Figure 2), compared to the parent construct (black), starting in the first repeat (top) and progressing toward the consensus repeats (bottom, see numbering at right).

***Table 2.*** Fitted Unfolding and Refolding Kinetic Parameters for Variants That Contain Alanine to Glycine Substitutions in Nank1−5, Nank1−5C$_1$, and Nank1−5C$_2$[a]

| | $k_{f,H_2O}$ (s$^{-1}$) | $m_f$ (M$^{-1}$) | $k_{u,H_2O}$ (s$^{-1}$) | $m_u$ (M$^{-1}$) | $b$ (M$^{-2}$) | $c$ (M$^{-3}$) | $\Phi_f^{H_2O}$ |
|---|---|---|---|---|---|---|---|
| **Nank1–7Δ**[b] | 0.48 | −0.22 | 0.00 | 1.702 | −0.17 | | |
| **Nank1–5** | 0.69 | −0.76 | 0.01 | 0.40 | | | |
| Nank1–5 (AG1) | 0.65 | −0.72 | 0.01 | 0.41 | | | 0[c] |
| Nank1–5 (AG4) | 0.07 | −1.76 | 0.07 | 0.61 | | | 0.60 |
| Nank1–5 (AG5) | 0.13 | −1.52 | 0.04 | 0.62 | | | 0.69 |
| **Nank1–5C$_1$** | 5.92 | −0.24 | $1.68 \times 10^{-5}$ | 1.27 | −0.09 | | |
| Nank1–5C$_1$ (AG1) | 5.92 | −0.20 | $1.78 \times 10^{-5}$ | 1.28 | −0.09 | | 0[c] |
| Nank1–5C$_1$ (AG2) | 6.97 | −0.26 | $3.54 \times 10^{-3}$ | 1.14 | −0.08 | | −0.03 |
| Nank1–5C$_1$ (AG3) | 8.43 | −0.43 | $3.64 \times 10^{-3}$ | 0.86 | −0.08 | | −0.06 |
| Nank1–5C$_1$ (AG4) | 4.53 | −0.01 | $3.70 \times 10^{-4}$ | 1.65 | −0.33 | 0.02 | 0.09 |
| Nank1–5C$_1$ (AG5) | 0.44 | −0.32 | $1.59 \times 10^{-6}$ | 1.72 | −0.12 | | 0.87 |
| Nank1–5C$_1$ (AGC1) | 0.89 | −0.03 | $4.04 \times 10^{-5}$ | 1.62 | −0.22 | 0.02 | 0.6 |
| **Nank1–5C$_2$** | 34.7 | −0.11 | $1.16 \times 10^{-6}$ | 1.25 | −0.07 | | |
| Nank1–5C$_2$ (AG1) | 37.2 | −0.14 | $1.71 \times 10^{-6}$ | 1.21 | 0.07 | | 0[c] |
| Nank1–5C$_2$ (AG2) | 18.4 | −0.02 | $7.01 \times 10^{-7}$ | 1.47 | −0.07 | | 0.08 |
| Nank1–5C$_2$ (AG3) | 35.3 | −0.06 | $1.83 \times 10^{-5}$ | 1.27 | −0.09 | | 0.00 |
| Nank1–5C$_2$ (AG4) | 51.5 | −0.20 | $1.36 \times 10^{-4}$ | 0.98 | −0.07 | | −0.06 |
| Nank1–5C$_2$ (AG5) | 32.3 | −0.03 | $4.77 \times 10^{-7}$ | 1.60 | −0.11 | | 0.02 |
| Nank1–5C$_2$ (AGC1) | 1.91 | 0.10 | $5.02 \times 10^{-10}$ | 2.24 | −0.14 | | 1.05 |
| Nank1–5C$_2$ (AGC2) | 11.1 | −0.21 | $5.70 \times 10^{-7}$ | 1.47 | −0.09 | | 0.40 |

[a] Parameters were estimated by fitting eq 2 to the urea dependencies of the rate constants of the major refolding and unfolding phases. [b] Data taken from ref 22. [c] Φ-Values of zero were assumed for substitutions in the first repeat, because folding rates are not affected by the substitution. Conditions: 25 mM Tris·HCl (pH 8.0), 150 mM NaCl, 20 °C.

can be compared using Φ-values obtained for glycine substitutions in different repeats, which represent the degree to which folding is slowed by destabilizing substitutions. The structure of the transition state for Nank1–5 folding appears to be similar to that for Nank1–7Δ, in that Φ-values in repeats four and five are high but in repeat one is low (Figure 4A, B; Table 1).

In contrast the structure in the transition state for folding of Nank1–5C$_1$ appears to be shifted C-terminally compared to Nank1–5 and Nank1–7Δ (Figure 4C; Table 2). The Φ-values determined for the first through fourth repeats are zero, whereas the Φ-values determined for the fifth and consensus (sixth) repeat are near 1 (0.87 and 0.6, respectively). The distribution of Φ-values in Nank1–5C$_1$ suggests that the transition state ensemble involves structure in the fifth and consensus repeat.

This shift in the transition state structure continues toward the stabilized C-terminus in Nank1–5C$_2$. In this construct, the Φ-values for the five N-terminal repeats are zero, whereas the Φ-values for the two C-terminal consensus (sixth and seventh) repeats are nonzero (1.05 and 0.4, respectively). The distribution of Φ-values in Nank1–5C$_2$ suggests that the structure in the transition state ensemble has shifted away from repeats four and five to the two C-terminal consensus repeats (Figure 4D). These findings are consistent with recent results showing that the transition state structure of myotrophin, a four-ankyrin repeat protein, can be shifted by destabilizing substitutions.[30]

**Alteration of the Transition State Ensemble and the Folding Pathway through C-Terminal Stabilization.** Previous studies have determined the distribution of structure in the transition state ensemble of Nank1–7Δ using analogous alanine to glycine substitutions in each repeat.[1] Whereas no kinetic effect was found for substitutions in the first, sixth, or seventh repeats, a large kinetic effect was found in the third, fourth, and fifth repeats. Thus, the Notch ankyrin domain has a polarized transition state with high Φ-values in the three internal repeats and Φ-values of zero elsewhere. This picture of the transition state ensemble is consistent with the observation that Nank1–5 folds at the same rate as that for Nank1–7Δ (Table 2)[31]

with the observation that repeats four and five remain structured in the transition state during Nank1–5 folding.

In contrast, substitutions in the third and fourth repeats in the C-terminally stabilized constructs have no kinetic effect, producing Φ-values of zero. Furthermore, in Nank1–5C$_2$, where C-terminal stabilization is largest, substitution in repeat five has no effect on refolding kinetics, whereas, in Nank1–7Δ, it defines the C-terminal boundary for the transition state ensemble. Instead, for the consensus-stabilized constructs, the transition state ensemble is shifted toward the C-terminus (Figure 4). This shift matches the new low energy channel through the consensus stabilized landscape (Figure 4C, D), demonstrating that, for the Notch ankyrin domain, the folding pathway selection is under thermodynamic control. This conclusion is consistent with the observation that Nank1–5C$_1$ and Nank1–5C$_2$ fold 1 and 2 orders of magnitude faster than either parent construct (Nank1–7Δ and Nank1–5).[33] Thus, the addition of the terminal consensus repeats not only stabilizes the entire domain but also alters the dominant folding pathway of the domain.

## Conclusions

We have taken advantage of the modularity, sequence simplicity, and linear structure of a repeat protein to probe the robustness and determinants of protein folding pathways. By C-terminally stabilizing the Notch ankyrin domain using consensus sequence repeats and interrogating the folding pathway using conservative, structurally homologous substitutions in each repeat, we find that a single well-defined folding pathway is identified in each construct. Moreover, the folding pathway shifts to track the lowest channel through the energy landscape, demonstrating experimentally that folding pathways

(39) Clementi, C.; Garcia, A. E.; Onuchic, J. N. *J. Mol. Biol.* **2003**, *326*, 933–54.
(40) Islam, S. A.; Karplus, M.; Weaver, D. L. *Structure* **2004**, *12*, 1832–45.
(41) Zweifel, M. E.; Barrick, D. *Biochemistry* **2001**, *40*, 14344–14356.
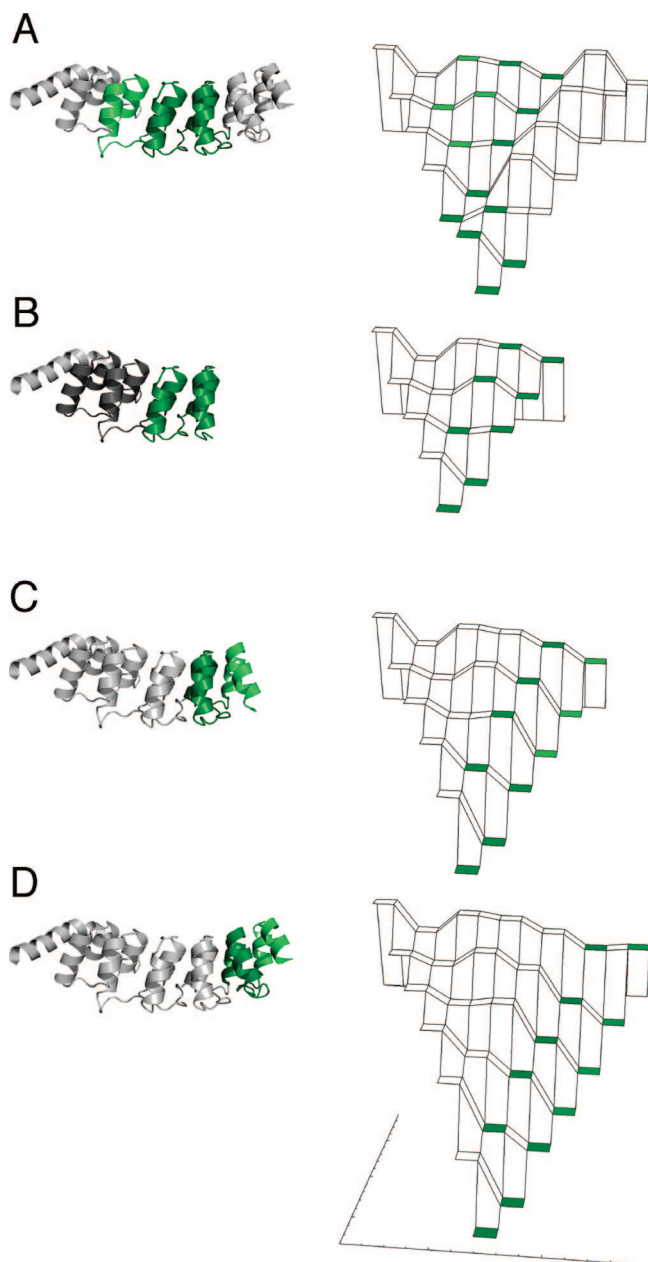(42) Zweifel, M. E.; Barrick, D. *Biochemistry* **2001**, *40*, 14357–14367.

**Figure 4.** The distribution of structure in the transition state ensembles of C-terminally stabilized ankyrin repeat constructs and the shift in the energy landscape from adding consensus repeats. Left: The transition state ensemble of the full length Notch ankyrin domain (A) is compared with that of Nank1–5 (B), Nank1–5C$_1$ (C), and Nank1–5C$_2$ (D). Repeats with high Φ-values are shaded green; those with zero Φ-values are shaded white. Constructs for which Φ-values could not be determined, Nank1–5 (AG2) and Nank1–5 (AG3), are colored gray. All Φ-values are mapped onto the crystal structure of Nank1–7Δ. Ribbon diagrams were generated using MacPymol.[45] Right: Energy landscapes, showing free energy (vertical direction) as a function of the number of folded repeats (back to front for denatured to native state, respectively) and the localization of structure (left to right for N- and C-terminally localized structure, respectively). Landscape tiers are shaded corresponding to Φ-values as in the ribbon diagrams.

are determined by local energetics. Given that the folding pathway in this repeat protein is localized to a subset of the repeats, even though multiple structurally and topologically similar pathways are available, it seems likely that the finding here of selection of specific pathways by local thermodynamics can be generalized to globular proteins. Indeed, in the rotationally symmetric globular proteins G and protein L, the folding

pathway has been shown to correspond to the region of local stability,[37,38] a result that can be captured by both an all-atom Gō model[39] and the sequence dependent components of the diffusion–collision model.[40] These results underscore the importance of developing accurate energy functions (that can accurately reproduce the stability differences among the ankyrin repeats of Notch, for example) for an accurate prediction of folding.

## Materials and Methods

**Mutagenesis, Protein Expression, And Purification.** Constructs encoding Nank1–5, Nank1–5C$_1$, and Nank1–5C$_2$ have been described previously.[33,39] The conserved alanine in the first helix of each repeat was substituted with glycine using QuickChange (Stratagene, La Jolla, CA). Constructs that contain individual alanine to glycine substitutions are referred to as Nank1–5 (AG1), Nank1–5 (AG2), etc. depending on whether the single alanine to glycine substitution is in repeat one, two, etc. Proteins were expressed in *E. coli* BL21 (DE3) and purified as described.[33,41]

**Equilibrium Unfolding Experiments.** Urea-induced unfolding was performed using an Aviv 62A DS spectropolarimeter, equipped with a computer-controlled Hamilton Microlab syringe titrator (Hamilton, Reno, NV) as described.[42,36,34] CD measurements were made at 222 nm, monitoring α-helical structure. Tryptophan fluorescence was measured by exciting at 280 nm and recording emission using a perpendicular 320 nm cutoff filter. Samples contained 2–4 μM protein, 25 mM Tris·HCl, 150 mM NaCl, pH 8.0, and were maintained at 20 °C. With the exception of Nank1–5C$_2$ (AG2) and Nank1–5C$_2$ (AG3), equilibrium unfolding was assumed to follow a two-state process.

Thermodynamic parameters were estimated assuming a linear dependence of the unfolding free energy on denaturant ($\Delta G = \Delta G°_{H_2O} - m$[urea]).[40,41] Free energies were related to unfolding transitions by treating the observed spectroscopic signal as a population-weighted average of native and denatured signals, as described.[36,42]

**Kinetic Refolding and Unfolding Experiments.** Refolding and unfolding kinetic measurements were made on an Applied Photophysics SX.18MV-R stopped-flow rapid mixing device (Leatherhead, UK). Fluorescence measurements were detected perpendicular to a 280 nm excitation using a 320 nm cutoff filter. For each variant, refolding was initiated from urea concentration just above the $C_m$, to permit refolding to be measured as close to 0 M urea as possible. For variants made in the stabilized Nank1–5C$_2$ background, unfolding was initiated from moderate urea concentrations, to allow unfolding to be measured at high urea concentrations. Final protein concentrations, following dilution, were between 1.0 and 3.0 μM.

Rates of refolding and unfolding were obtained from fits of the individual progress curves to the following equation:

$$Y_{obs} = Y_{\infty} + \sum_{i=l}^{n} Y_i e^{-k_i t} \qquad (1)$$

where the observed signal ($Y_{obs}$) is a function of the signal at infinite time ($Y_{\infty}$) and exponentials that correspond to individual kinetic phases, each with amplitudes $Y_i$ and rate constants $k_i$, where $n = 1$ for unfolding and $n = 2$ for refolding to account for the additional proline isomerization phase.

Analysis of Chevron plots and Φ-values. The urea dependence of the rate constants for the major refolding and unfolding phases (chevron plot) were fit according to the following equation:

$$\log k_{obs} = \log(k_f + k_u) = \log(k_{f,H_2O}10\{m_f[urea]\} +$$
$$k_{u,H_2O}10\{m_u[urea]\}) + b[urea]^2 + c[urea]^3 \qquad (2)$$

(43) Pace, C. N. *Methods Enzymol.* **1986**, *131*, 266–80.
(44) Santoro, M. M.; Bolen, D. W. *Biochemistry* **1988**, *27*, 8063–8.
(45) DeLano, W. L. *MacPymol*; DeLano Scientific LLC: South San Francisco, 2005.

The log of the refolding and unfolding rate constants are allowed to have a linear, quadratic, and cubic dependence, given by $m_f$ and $m_u$, $b$, and $c$, as required. In general, the nonlinear terms, which are the same for both folding and unfolding, were needed when a large range of urea concentrations was required to fully define the chevron (i.e., Nank1–5C$_2$ and Nank1–5C$_1$).

$\Phi$-values were calculated using the following equation:

$$\Phi_f = \frac{RT\ln\left(k_{f,H_2O}^{mut}/k_{f,H_2O}^{wt}\right)}{\Delta G_{H_2O}^{\circ mut} - \Delta G_{H_2O}^{\circ wt}} \tag{3}$$

where the folding rate constant of variants and the parent constructs were determined from fits of the chevron plots. Equilibrium unfolding free energies were determined from urea-induced denaturations. $\Phi$-Values could not be determined for Nank1–5 (AG2) and Nank1–5 (AG3) because these constructs are not folded.

The $\Phi$-values calculated for Nank1–5C2 (AG2) and Nank1–5C2 (AG3) are from kinetic and thermodynamic data obtained monitoring folding with fluorescence. Although an additional equilibrium unfolding transition is seen by CD at low denaturant concentrations, the structural transition probed in fluorescence is likely to be largely localized to the C-terminus of the molecule, based on the location of the single tryptophan (repeat five, packed against repeat four) and the observation that the both the equilibrium curves and chevrons of these two variants are sensitive to the C-terminal stabilization (compare the folding arm of and minimum in the chevron of Nank1–5C$_2$ (AG2) and Nank1–5C$_2$ (AG3) to Nank1–5 and Nank1–7$\Delta$). As a result, the actual effect of these substitutions on the free energy of unfolding of the native state (repeats two and 3-folded) would actually be larger than that estimated from fluorescence, making these low $\Phi$-values even lower.