The Effect of Boundary Selection on the Stability and Folding of the Third Fibronectin Type III Domain from Human Tenascin†

Stefan J. Hamill,‡ Alison E. Meekhof,‡ and Jane Clarke*,§

MRC Centre for Protein Engineering and Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, U.K.

Received January 22, 1998; Revised Manuscript Received April 1, 1998

ABSTRACT: Correct selection of domain boundaries is critical for structural analysis of single domains from multimodular proteins. Folding and stability studies of the third fibronectin type III domain from human tenascin ($TNfn3_{1-90}$) have shown that it is moderately stable ($\Delta G^{H_2O}_{D-N} \sim 5$ kcal mol^{-1}) and folds with two-state kinetics. In an attempt to stabilize the protein, five domains were constructed with different combinations of extensions to the N- and C-termini. Thermal denaturation studies show that a specific two amino acid (Gly-Leu) extension at the C-terminus is primarily responsible for a significant increase in stability. The $\Delta\Delta G^{H_2O}_{D-N}$ of the Gly-Leu extension ($TNfn3_{1-92}$) is 2.7 ± 0.3 kcal mol^{-1}. Refolding kinetics do not differ significantly, but unfolding is slowed 40-fold. Mutation of leucine 92 to alanine does not affect stability, indicating that the stability of the extension does not come from the packing of the leucine side chain. Hydrogen exchange data suggest that the extension adds new hydrogen bonds and strengthens existing hydrogen bonds in the C-terminal interaction with the A−B and E−F loops. Removal of a very small number of hydrogen bonds substantially increases the unfolding rate, a phenomenon which may be important in stress-relaxation of FNIII-containing muscle proteins such as titin. These experiments demonstrate the importance of a small number of additional long-range interactions in the overall formation of a compact independently folding β-sheet module.

Many protein folds are very common in nature (*1*). Perhaps the best example is the immunoglobulin superfamily (IgSF)[1] fold (*2*−*4*), which includes proteins with a simple sandwich of two β-sheets such as the immunoglobulin (Ig) and fibronectin type III (FNIII) families. IgSF domains are often found in multimodular proteins as a 'string' of consecutive domains (*5*). An important question is the following: What factors precisely define a single module within these multimodular assemblies? Also, what are the boundaries and why? Which interactions are important at the domain boundaries and how do they affect the stability, folding kinetics, and dynamics of the isolated module?

To identify domain boundaries, several methods can be employed. Sequence alignment of homologous domains, in combination with a structural database, is a common and powerful method. The gene sequence may reveal intron−exon boundaries that often naturally delimit domains as they are shuffled throughout evolution (*6*). Limited proteolytic cleavage of a multimodular protein followed by N-terminal sequencing or mass spectroscopy of the resulting fragments (*7*), or crystallization trials of differing domain lengths (*8*), can also identify complete domains. Previous studies of isolated Ig modules from the giant IgSF assembly titin have highlighted the importance of selecting correct boundaries in multimodular IgSF assemblies (*9*, *10*).

We have previously characterized the folding and stability of the third FNIII domain from human tenascin ($TNfn3_{1-90}$) and shown it to fold via a simple two-state process to a native state with a $\Delta G^{H_2O}_{D-N}$ of 5.3 ± 0.5 kcal mol^{-1} (*11*). The domain is 90 amino acids long; the boundaries were chosen in reference to the published crystal structure (*12*). Here we investigate the domain boundaries of TNfn3 further and show that the domain is too short for optimal stability. We examine in detail the domain extended by two residues at the C-terminus ($TNfn3_{1-92}$) and characterize the effect of this extension upon the equilibrium stability, folding, and hydrogen exchange kinetics. In combination with an NMR dynamics and residue-specific titration study to be reported elsewhere (A. E. Meekhof, S. J. Hamill, V. L. Arcus, J. Clarke, and S. M. V. Freund, unpublished results), the results indicate that, although the extension is very small, it has a number of important long-range structural effects which alter many biophysical characteristics of the domain.

EXPERIMENTAL PROCEDURES

Experimental Section

Materials. Deuterated chemicals, imidazole, ^{2}HCl, and ^2H_2O, and [^{15}N]ammonium chloride were obtained from Aldrich Chemicals. Other chemicals and reagents were

† S.J.H. is funded by the Medical Research Council; A.E.M. holds a British Marshall Scholarship. J.C. has a Wellcome career development fellowship.

* To whom correspondence should be addressed. Email: jc162@cam.ac.uk. Telephone: +44 1223 336356. Fax: +44 1223 336445.

‡ MRC Centre for Protein Engineering.

§ Department of Chemistry, University of Cambridge.

[1] Abbreviations: FNIII, fibronectin type III; TNfn3, third FNIII repeat of tenascin; Ig, immunoglobulin; IgSF, immunoglobulin superfamily; ΔG_{D-N}, free energy of unfolding; *m*, dependence of the free energy of unfolding on [denaturant]; HSQC, heteronuclear single-quantum coherence; FNfnX, Xth FNIII domain of human fibronectin; NOE, nuclear Overhauser effect.

S0006-2960(98)00165-2 CCC: $15.00
Published on Web 05/13/1998

obtained from BDH, Fisons, or Sigma. Restriction enzymes and T4 DNA ligase were obtained from New England Biolabs, and Pfu polymerase was from Stratagene.

Construction of Extended Domains. The five extended TNfn3 domains were constructed in two stages in an iterative fashion. The following naturally occurring amino acids were added to residues 1−90 [originally designated as residues 802−891 in the crystal coordinates (*12*)]; ext C2, two residue (Gly-Leu) extension at the C-terminus; ext C4, four residue (Gly-Leu-Asp-Ala) extension to the C-terminus; N4, four residue (Val-Thr-Thr-Thr) extension to the N-terminus; C2N4, combination of the C2 and N4 extensions; C4N4, combination of the C4 and N4 extensions. Extension was performed by PCR using primers which encode the extra sequence and also add *Bam*HI and *Eco*RI restriction enzyme sites to allow cloning into the vector mini-pRSET-A (*11*). DNA sequencing of all the extended genes was performed in both directions.

Site-Directed Mutagenesis. The mutation L92A was created using the Stratagene Quikchange mutagenesis kit. A silent *Bsa*HI restriction site was added near the mutation site to identify mutants, which were then sequenced in both directions.

Protein Expression and Purification. Escherichia coli strain BLR DE3 was transformed with the plasmid pTEN1, which is the TNfn3 gene cloned into a His-tag vector and grown as previously described (*11*). To prepare ^{15}N-labeled samples, cells were grown in minimal media (M9) with ^{15}N-labeled ammonium chloride as the sole nitrogen source. The protein was extensively dialyzed against deionized water and stored in liquid nitrogen. It should be noted that purification using this His-tag method adds two residues, Gly-Ser, to the N-terminus.

Urea Denaturation. Unfolding was monitored by the increase in fluorescence at 357 nm upon unfolding (excitation 280 nm). Protein was mixed to a final concentration of 1 μM in urea solutions in 50 mM buffer (sodium acetate buffer, pH 5.0, 35 mM sodium acetate, 15 mM acetic acid). The solutions were left to equilibrate at 20 or 25 °C for at least 1 h before measurements were recorded in thermostated cuvettes. The fluorescence data were analyzed as described previously (*11*).

Thermal Unfolding Monitored by CD. All CD measurements were performed using a JASCO J-720 instrument, using protein at 5 μM in 50 mM sodium acetate buffer, pH 5.0 (35 mM sodium acetate, 15 mM acetic acid), in a 2 cm path length cuvette. Thermal denaturation was monitored at 230 nm. The temperature was controlled by a Neslab RTE-110 water bath, with a rate of temperature increase of 50 °C/h. CD data were analyzed as described previously (*11*).

Refolding. Folding was monitored by changes in fluorescence as described previously (*11*). Data collected from at least four experiments were averaged. Unfolding traces were fitted to a single-exponential function. Refolding traces were fitted to a double exponential function, and the faster rate was taken as the observed refolding rate constant as described (*11*); rate constants determined from regions where the fast rate merges with the slower proline-limited phase were discarded. A term was included to account for base line instability.

Unfolding. To determine the unfolding rate constant accurately over a range of urea concentrations, the temperature dependence of the unfolding rates was studied at several different urea concentrations over a temperature range of 30−64 °C. These data were then analyzed by extrapolating plots of ln k_u versus temperature to 25 °C. By extrapolating over the full range of urea concentrations, an accurate value in the absence of urea, $k_u^{H_2O}$, could be determined.

NMR H/^{2}H Exchange. ^{1}H and ^{15}N chemical shifts of TNfn3$_{1-92}$ backbone amide groups were obtained through a series of triple resonance experiments at pH 5.0. A ^{1}H,^{15}N HSQC collected at pH 7.0 was sufficient to confirm peak assignments at the higher pH used for exchange experiments.

For NMR exchange experiments, the exchange buffer was 100 mM deuterated imidazole, p^{2}H 7.0 [where p^{2}H = p^{2}H$_{read}$ + 0.4 (*13*)], containing 0.05% sodium azide. Fifteen milligrams of lyophilized uniformly ^{15}N-labeled protein was dissolved in the exchange buffer, (final concentration ∼3 mM), centrifuged, and transferred to an NMR tube. Acquisition of gradient-selected sensitivity-enhanced ^{1}H−^{15}N HSQC spectra was started immediately on a Bruker AMX500 spectrometer, with an average dead-time of 5 min. Spectra were acquired at regular intervals with decreasing frequency for several weeks. NMR samples were kept at 25 °C throughout the study. The final p^{2}H did not vary by more than 0.05 p^{2}H unit. The spectra were processed as described previously (*14*).

Analysis

ϕ-Value Analysis. As described fully elsewhere (*15, 16*) a ϕ-value for folding, ϕ_f, can be determined, from analysis of mutant proteins, for a two-state system from eq 1:

$$\phi_f = \frac{\Delta\Delta G_{D-\ddagger}}{\Delta\Delta G_{D-N}} \quad (1)$$

where $\Delta\Delta G_{D-\ddagger}$ is the change in the free energy difference between the denatured (D) and transition (‡) states for folding upon mutation. $\Delta\Delta G_{D-\ddagger} = -RT(k_f/k'_f)$, and k_f and k'_f are refolding rate constants, at 0 M denaturant, for wild-type and mutant proteins, respectively, R is the gas constant, and T is the temperature in degrees kelvin. A ϕ_f-value of 1 indicates that the interaction in question is fully formed in the transition state for folding. A ϕ_f-value of 0 means that the interaction is not formed in the transition state (*17*). The ϕ value for unfolding, ϕ_u, where $\phi_u = \Delta\Delta G_{\ddagger-N}/\Delta\Delta G_{D-N}$, is calculated in exactly the same fashion, substituting the unfolding rate constants, and can be used to derive ϕ_f from unfolding data by $\phi_f = 1 - \phi_u$.

NMR H/^{2}H Exchange Analysis. At sites in the protein where the structure is not directly affected by the extension, $\Delta\Delta G_{ex}$, the change in free energy for exchange upon mutation is determined using eq 2 (*14*):

$$\Delta\Delta G_{ex} = -RT \ln \frac{k_{ex}}{k'_{ex}} \quad (2)$$

where k_{ex} is the rate constant for exchange of the proton in TNfn3$_{1-90}$ and k'_{ex} is the exchange rate constant in the extended protein. In this analysis, we make the assumption that, under these relatively benign conditions (25 °C, pH 7.0),

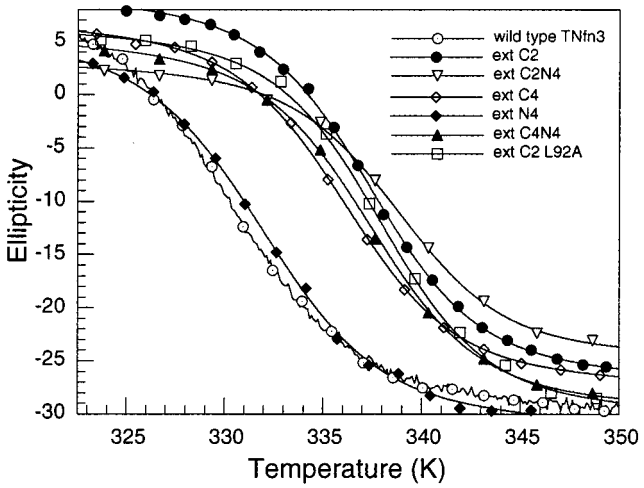

FIGURE 1: Thermal denaturation of $TNfn3_{1-90}$ and extended domains monitored by CD. Thermal denaturation monitored by ellipticity (arbitrary units) at 230 nm. A typical raw trace is shown for the $TNfn3_{1-90}$ transition; for the other proteins, the curve fit is displayed.

protons in TNfn3 exchange by the commonly observed EX2 mechanism in which the exchange rate is proportional to the equilibrium between the exchange-competent and incompetent forms (*18*−*22*). A simple test of the exchange mechanism is to compare exchange rates at two different pH values (*20, 23*). The stability of TNfn3 varies significantly over the pH range 5−7 (*11*) so this test cannot be applied. However, the calculated intrinsic exchange rates for tenascin (*24*) are lower than the value for $k_f^{H_2O}$ obtained by refolding experiments, and the $\Delta\Delta G_{ex}$ value that we obtain from exchange closely matches that obtained from urea denaturation (see Results). It is therefore reasonable to assume an EX2 mechanism.

RESULTS

Thermal Denaturation. To determine precisely the correct domain boundaries, a series of five different extended domains were constructed by PCR: ext C2 (two residue extension to the C-terminus), ext C4, ext N4, ext C2N4, and ext C4N4 (four residue extension to both the C- and N-terminus). Each extension added the amino acids in the neighboring tenascin domains as listed in the Swiss-Prot database (*25*).

To assess relative stability, thermal denaturation was monitored by CD. The far-UV CD spectra of all the extended domains are identical to the wild type spectrum, and denaturation was monitored at 230 nm where the maximal difference between folded and thermally denatured protein is observed (Figure 1, Table 1). Extension of $TNfn3_{1-90}$ can significantly increase the T_m. Stabilization appears to be specific to a two-residue extension at the C-terminus. This raises the T_m by 7.3 °C whereas extending the N-terminus by four residues increases the T_m by only 1.7 °C. Addition of a further two residues at the C-terminus actually lowers the T_m again. Based on these results, ext C2 ($TNfn3_{1-92}$) was studied in more detail to determine the basis for the significant effect on stability of a small extension.

Stability of $TNfn3_{1-92}$. The stability of $TNfn3_{1-92}$ was determined using urea denaturation (Table 2), and a value for $\Delta G_{D-N}^{H_2O}$ was calculated using an averaged *m*-value, $\langle m \rangle$, from several equilibrium unfolding curves (*26*). There is an increase in stability at pH 5.0 and 20 °C of 2.7 ± 0.2 kcal mol^{-1}. As for $TNfn3_{1-90}$, the $TNfn3_{1-92}$ equilibrium unfolding data fit to equations describing two-state equilibrium unfolding.

Table 1: Thermal Stability of $TNfn3_{1-90}$ and Extended Domains[a]

protein	mean T_m (°C)	ΔT_m (°C)
wild type ($TNfn3_{1-90}$)	57.1 ± 0.1	−
ext C2 ($TNfn3_{1-92}$)	64.4 ± 0.1	+7.3 ± 0.1
ext C4	63.0 ± 0.1	+5.9 ± 0.1
ext N4	58.8 ± 0.1	+1.7 ± 0.1
ext C2N4	65.7 ± 0.1	+8.6 ± 0.1
ext C4N4	64.2 ± 0.2	+7.1 ± 0.2
$TNfn3_{1-92}$-L92A*	64.7 ± 0.1	+7.6 ± 0.1

[a] All T_m values are the mean of three determinations except for the protein marked with an asterisk which was performed once. Errors reported are standard errors of the mean except for * is the standard error of the single curve fit.

Table 2: Stability of $TNfn3_{1-90}$, $TNfn3_{1-92}$, and $TNfn3_{1-92}$-L92A Determined from Urea Denaturation[a]

protein/ conditions	$[urea]_{50\%}$ (M)	$\Delta G_{D-N}^{H_2O}$ [b] (kcal mol^{-1})	$\Delta\Delta G_{D-N}^{H_2O}$ (kcal mol^{-1})
$TNfn3_{1-90}$[c] (20 °C), pH 5.0	3.78 ± 0.04	4.76 ± 0.11	−
$TNfn3_{1-92}$ (20 °C), pH 5.0	5.91 ± 0.04	7.44 ± 0.14	2.68 ± 0.18
$TNfn3_{1-90}$ (25 °C), pH 7.0		3.32 ± 0.17[d]	
$TNfn3_{1-92}$ (25 °C), pH 7.0	4.52 ± 0.05	5.69 ± 0.13	2.37 ± 0.21
$TNfn3_{1-92}$ (25 °C), pH 5.0	5.31 ± 0.12	6.68 ± 0.18	
$TNfn3_{1-92}$-L92A (25 °C), pH 5.0	5.28 ± 0.03	6.65 ± 0.13	−0.03 ± 0.13

[a] Standard errors of the data are given. [b] An average *m*-value, $\langle m \rangle$, of 1.26 ± 0.05 kcal mol^{-1} M^{-1} [calculated from 14 different experiments (range 1.0−1.6)] has been used to calculate $\Delta G_{D-N}^{H_2O}$ values using the equation: $\Delta G_{D-N}^{H_2O} = \langle m \rangle [urea]_{50\%}$ (*47*). [c] Data from (*11*) recalculated using $\langle m \rangle$ = 1.26 ± 0.05. [d] Value derived from extrapolation of DSC data (*11*), ΔH_{vh} = 63.24 kcal mol^{-1} and ΔC_p = 1.19 kcal mol^{-1} K^{-1}.

Folding Kinetics. The folding and unfolding kinetics of $TNfn3_{1-92}$ were initially studied at 20 °C at pH 5.0. In comparison to $TNfn3_{1-90}$ (*11*), the folding rate of the stabilized domain changes very little, with a folding rate constant in 0 M urea of 6.0 s^{-1}. In contrast, there is a significant reduction in the unfolding rate such that the protein takes over an hour to unfold completely in 7.5 M urea at 20 °C. The temperature for folding studies was raised to 25 °C to speed unfolding to a manageable rate. It was still difficult to study unfolding at 25 °C due to the high $[urea]_{50\%}$ (5.3 M at 25 °C)—the unfolding amplitude only reaches 100% at 8 M urea. Thus, the temperature dependence of unfolding was analyzed over the temperature range 30−64 °C in regions outside the transition zone at each temperature. Unfolding rate constants at 25 °C for eight different urea concentrations were determined by extrapolation from the higher temperature data (Figure 2). These data fit well with the unfolding data collected at 25 °C above 8 M urea and improve extrapolation to 0 M urea to obtain $k_u^{H_2O}$.

Mutation of Leucine 92 to Alanine. The stability of the mutation $TNfn3_{1-92}$-L92A was analyzed by thermal and urea denaturation to determine whether the hydrophobic packing of the Leu-92 side chain contributes to the observed stability increase (Figure 1, Tables 1 and 2). The $[urea]_{50\%}$ and the

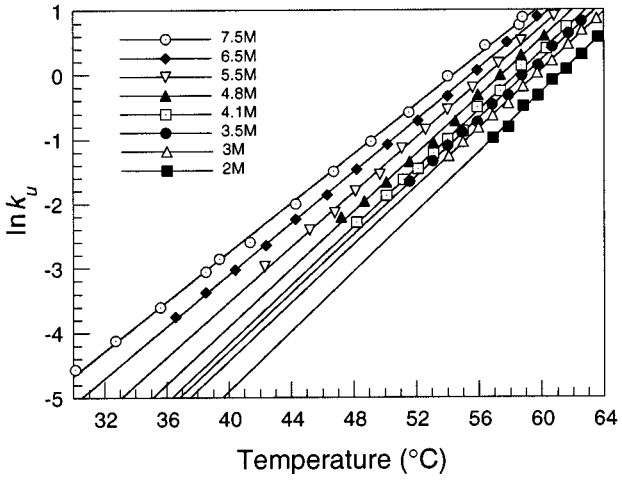

FIGURE 2: Temperature dependence of unfolding of $TNfn3_{1-92}$. Plot of the natural logarithm of the unfolding rate constant against temperature over a range of urea concentrations and temperatures. The unfolding rate constant at 25 °C for each urea concentration was determined from extrapolation.

T_m of L92A are the same, within error, as for wild-type $TNfn3_{1-92}$.

Hydrogen Exchange Kinetics. At pH 7.0, 37 $TNfn3_{1-90}$ and 34 $TNfn3_{1-92}$ amide protons exchange with solvent deuterons within the 5 min deadtime of the exchange experiment. Rate constants for exchange for the remaining 48 and 53 amide protons could be determined for $TNfn3_{1-90}$ and $TNfn3_{1-92}$, respectively (Table 3). The pattern of protection observed is typical for a β-sheet protein. The amides of the three internal β-sheets, B, C, and F, are almost all protected, and there is a binary pattern of protection to nonprotection in the four edge strands. There is less protection of loop amides; some notable exceptions include side chain hydrogen bonding to the β bulge in strand G and the tyrosine corner H-bond in the E−F loop. In general, the exchange rate constants mirror the increased stability of $TNfn3_{1-92}$, which exchanges more slowly than $TNfn3_{1-90}$. Certain amide protons which disappear within the deadtime of the $TNfn3_{1-90}$ exchange experiment are observed in the $TNfn3_{1-92}$ experiment, and exchange rates are determined. Notable inclusions in this group are residues 90 and 91 at the C-terminus and residues 16 and 65 from the A−B and E−F loops which contact the C-terminus. For residues observable in both forms, the change in free energy for exchange for each residue, $\Delta\Delta G_{ex}$, was calculated (Table 3 and Figure 3).

The values of $\Delta\Delta G_{ex}$ fall into three groups (*14*, *27*). Those where the $\Delta\Delta G_{ex}$ is very close to zero exchange by local unfolding motions (13 amides); these local unfolding reactions are unaffected by mutation. Those where the $\Delta\Delta G_{ex}$ is close to the full change in the free energy of unfolding, $\Delta\Delta G^{H_2O}_{D-N}$, exchange by global unfolding of the protein (26 amides). In the third group, the $\Delta\Delta G_{ex}$ is intermediate in value, and residues exchange by a mixture of local "breathing" and global unfolding (8 amides). From the 26 values of $\Delta\Delta G_{ex}$ obtained from the globally exchanging residues, a mean value of $\Delta\Delta G_{ex}$ was calculated as 2.51 ± 0.03 kcal mol^{-1} (range 2.0−2.7). This is close to the $\Delta\Delta G^{H_2O}_{D-N}$ value of 2.37 ± 0.21, obtained at pH 7.0 (Table 2).

DISCUSSION

The thermal stabilization observed in TNfn3 is specific for extension of the C-terminus by two residues. N-terminal extensions have little effect, and further extension at the C-terminus is unfavorable (Table 1). Measurements by urea denaturation and hydrogen exchange combine to give a $\Delta\Delta G^{H_2O}_{D-N}$ value for $TNfn3_{1-92}$ of 2.4−2.7 kcal mol^{-1}. We use several techniques to assess the reasons for the increased stability.

Comparison of $TNfn3_{1-92}$ with L92A Mutant. The question of whether the extension is stabilizing due to addition of hydrophobic packing can be addressed by replacing the leucine added at position 92 with a smaller side chain (alanine). The mutant $TNfn3_{1-92}$-L92A has an identical $\Delta\Delta G^{H_2O}_{D-N}$ and T_m to $TNfn3_{1-92}$ (Tables 1 and 2). This indicates that the leucine side chain does not pack significantly and that any extra stability due to hydrophobic packing is derived principally from the packing of the backbone atoms.

Hydrogen Exchange. Hydrogen exchange enables us to distinguish not only global effects, but also local effects of the extension on stability. The C-terminus is involved in important interactions with the E−F loop region within the G-F-C-C′ β-sheet (sheet 1) and is also near the A−B loop region in the A-B-E sheet (sheet 2, see Figures 4 and 5). Residues adjacent, in the tertiary structure, to the terminus in $TNfn3_{1-90}$ have hydrogen exchange rates that are slowed significantly on extension (Table 3). These include residues 16 and 17 in the A−B turn and residues 65 and 66 in the E−F loop. Residue 16 shows no protection in $TNfn3_{1-90}$ but is significantly protected in the extended form. [This could, however, reflect exchange broadening of the HSQC cross-peak in the shorter form, due to conformational exchange (A. E. Meekhof and S. M. V. Freund, unpublished data)]. Residue 17 has a $\Delta\Delta G_{ex}$ consistent with a globally exchanging residue. As the amide proton is protected by a main chain−side chain hydrogen bond, this would be an unlikely explanation for the significant slowing of exchange on mutation. The results imply a stabilization of the A−B turn by the extension. The NH of residue 65 exchanges very rapidly in the short domain, but slows very significantly in the longer form. The rapid exchange rate of this proton in $TNfn3_{1-90}$ makes it highly unlikely that the high $\Delta\Delta G_{ex}$ is due to effects on global unfolding; rather, this is a local effect of the extension. Residue 65 has no hydrogen bonding partner in $TNfn3_{1-90}$. To assess whether new hydrogen bonds are created, use can be made of the several crystal structures of FNIII domains available.

Comparison of TNfn3 with Other FNIII Domains. In choosing the correct domain boundaries, much use can be made of structural alignments. For instance, in the case of one of the shorter titin IgSF domains, it is clear from structural alignments that a hydrophobic core residue was deleted in the original construct (*9*, *28*, *29*). It is interesting, therefore, to compare the structure of TNfn3 with other FNIII domains. A structural alignment with 10 FNIII domains shows that the equivalent residue to TNfn3 position 92 never packs into the core. The equivalent stabilities of $TNfn3_{1-92}$ and the L92A mutant, and the lack of NOEs to Leu-92 side chain protons (A. E. Meekhof and S. M. V. Freund, unpublished data), substantiate this.

Table 3: Rate Constants for Hydrogen Exchange of Amide Protons in $TNfn3_{1-90}$ and $TNfn3_{1-92}$ at p^{2}H 7.0, 25 °C

residue	H-bond partner[a]	location of H-bond (donor-receptor)	$TNfn3_{1-90}$ (min^{-1})	$TNfn3_{1-92}$ (min^{-1})	$\Delta\Delta G_{ex}$[b] (kcal mol^{-1})
6	23 CO	β_A-β_B	4.33×10^{-1}	4.89×10^{-1}	0.07
8			1.22×10^{-1}	1.64×10^{-1}	0.17
9	21 CO	β_A-β_B	1.72×10^{-2}	5.80×10^{-4}	−2.01
11	19 CO	β_A-β_B	1.45×10^{-2}	2.63×10^{-4}	−2.38
14	17 CO	β_A-β_B	1.18×10^{-1}	6.45×10^{-2}	−0.36
16	Thr-14 $O_{\gamma 1}$	$turn_{A-B}$-side $chain_{\beta A}$		7.43×10^{-3}	
17	Thr-14 $O_{\gamma 1}$	β_B-side $chain_{\beta A}$	3.02×10^{-2}	4.65×10^{-4}	−2.47
18	59 CO	β_B-β_E	2.35×10^{-2}	2.98×10^{-4}	−2.59
19	11 CO	β_B-β_A	1.94×10^{-1}	3.25×10^{-1}	0.31
20	57 CO	β_B-β_E	1.12×10^{-2}	1.24×10^{-4}	−2.67
21	9 CO	β_B-β_A	2.61×10^{-2}	3.15×10^{-4}	−2.62
22	55 CO	β_B-β_E	2.20×10^{-2}	3.15×10^{-4}	−2.52
23	6 CO	β_B-β_A	2.59×10^{-2}	4.67×10^{-3}	−1.01
30	75 CO	$loop_{B-C}$-β_F	4.11×10^{-2}	2.97×10^{-2}	−0.19
31			3.04×10^{-2}	1.26×10^{-2}	−0.52
32	50 CO	β_C-$\beta_{C'}$	1.33×10^{-2}	2.95×10^{-3}	−0.89
33	73 CO	β_C-β_F	1.39×10^{-2}	1.58×10^{-4}	−2.65
34	48 CO	β_C-$\beta_{C'}$	1.44×10^{-2}	1.63×10^{-4}	−2.66
35	71 CO	β_C-β_F	2.31×10^{-2}	2.85×10^{-4}	−2.61
36	46 CO	β_C-$\beta_{C'}$	2.18×10^{-2}	5.06×10^{-4}	−2.23
37	69 CO	β_C-β_F	2.74×10^{-2}	5.07×10^{-4}	−2.36
38	Asp-44 $O_{\delta 2}$	β_C-side $chain_{C-C' loop}$	2.35×10^{-1}	1.15×10^{-1}	−0.42
39	67 CO	$loop_{C-C'}$-β_F	8.08×10^{-2}	5.24×10^{-2}	−0.26
41	38 CO	$loop_{C-C'}$-β_C	5.29×10^{-2}	4.25×10^{-2}	−0.13
46	36 CO	$\beta_{C'}$-β_C	1.63×10^{-2}	2.12×10^{-4}	−2.57
48	34 CO	$\beta_{C'}$-β_C	1.62×10^{-2}	2.41×10^{-4}	−2.50
50	32 CO	$\beta_{C'}$-β_C	2.40×10^{-2}	4.47×10^{-4}	−2.36
53			2.63×10^{-1}	4.89×10^{-1}	0.37
56			5.31×10^{-2}	3.26×10^{-2}	−0.29
57	20 CO	β_E-β_B	1.80×10^{-2}	3.31×10^{-4}	−2.37
58				1.62×10^{-1}	
59	18 CO	β_E-β_B	2.05×10^{-2}	2.60×10^{-4}	−2.59
62	16 CO	$loop_{E-F}$-$turn_{A-B}$	2.56×10^{-2}	3.80×10^{-4}	−2.50
63	Tyr-68 OH	tyrosine corner[c]	1.07×10^{-1}	4.78×10^{-2}	−0.48
65	91 CO[d]	$loop_{E-F}$-extension		4.93×10^{-3}	
66	Thr-90 $O_{\gamma 1}$	β_F-side $chain_{\beta G}$	2.03×10^{-2}	2.15×10^{-4}	−2.70
67			3.17×10^{-1}	2.02×10^{-1}	−0.27
68	88 CO	β_F-β_G	1.62×10^{-2}	2.10×10^{-3}	−1.21
69	37 CO	β_F-β_C	1.97×10^{-2}	2.46×10^{-4}	−2.60
70	86 CO	β_F-β_G	1.19×10^{-2}	1.31×10^{-4}	−2.68
71	35 CO	β_F-β_C	2.42×10^{-2}	3.70×10^{-4}	−2.48
72	84 CO	β_F-β_G	1.50×10^{-2}	1.34×10^{-3}	−1.43
73	33 CO	β_F-β_C	1.70×10^{-2}	2.29×10^{-4}	−2.55
74	Ser-81 O_γ	β_F-side $chain_{\beta G-bulge}$	3.11×10^{-2}	4.91×10^{-4}	−2.46
75	31 CO	β_F-β_C	3.15×10^{-2}	5.83×10^{-3}	−1.00
76	79 CO	β_F-β_G	3.87×10^{-2}	1.17×10^{-2}	−0.71
79	76 CO	β_G-β_F	1.31×10^{-1}	9.36×10^{-2}	−0.20
81	74 CO	$\beta_{G-bulge}$-β_F	1.97×10^{-1}	1.39×10^{-1}	−0.21
84	72 CO	β_G-β_F	2.32×10^{-2}	4.39×10^{-3}	−0.99
86	70 CO	β_G-β_F	2.46×10^{-2}	3.42×10^{-4}	−2.53
88	68 CO	β_G-β_F	2.74×10^{-2}	3.69×10^{-4}	−2.55
90	66 CO	β_G-β_F		1.01×10^{-3}	
91	14 CO[d]	extension-$turn_{A-B}$		1.07×10^{-1}	

[a] Hydrogen bond receptor identified from crystal structure (*12*). [b] $\Delta\Delta G_{ex}$ was determined using eq 2. [c] The tyrosine corner is located in $loop_{E-F}$. [d] For residues 65 and 91, the putative H-bond receptor is identified (see Discussion).

The most reliable predictions of the effects of extension on hydrogen bonding come from examining structures of the linker region between individual domains. The structure of 4 tandem FNIII domains 7 to 10 from human fibronectin (FNfn7−10) contains information on 3 such linkers (*30*). An alignment of the $TNfn3_{1-90}$ structure (*12*) with these domains indicates that the C-terminal regions of FNfn8 and FNfn9 are particularly similar (Figure 5). In both of these domains, the amide proton of the residue equivalent to position 90 in TNfn3 makes a hydrogen bond to the carbonyl oxygen of the equivalent residue to position 66. In addition position 91 makes two hydrogen bonds, to position 14 in the A−B loop and position 65 in the E−F loop. The hydrogen exchange data are consistent with the hydrogen bond from 90 to 66 being significantly strengthened and also the addition of the two new hydrogen bonds from position 91 in $TNfn3_{1-92}$. New long-range NOE cross-peaks from the Asp-65 amide to side chain protons of Thr-90 and Leu-92 and also from the Gly 91 amide to the Asp-15 side chain protons (A. E. Meekhof and S. M. V. Freund, unpublished data) are exactly those expected from this interpretation.

In summary, the stabilization effect of the two residue extension seems to result from an increased and strengthened network of hydrogen bonds in the vicinity of the C-terminus. Hydrogen bonds from sheet-1 to sheet-2 such as that from 91 to 14 may be particularly important in stabilizing these edge strands. Sequence alignments of FNIII domains show clearly that the conserved threonine at position 90 is the last residue to pack into the FNIII core. However, close examination of the four domain structure of fibronectin

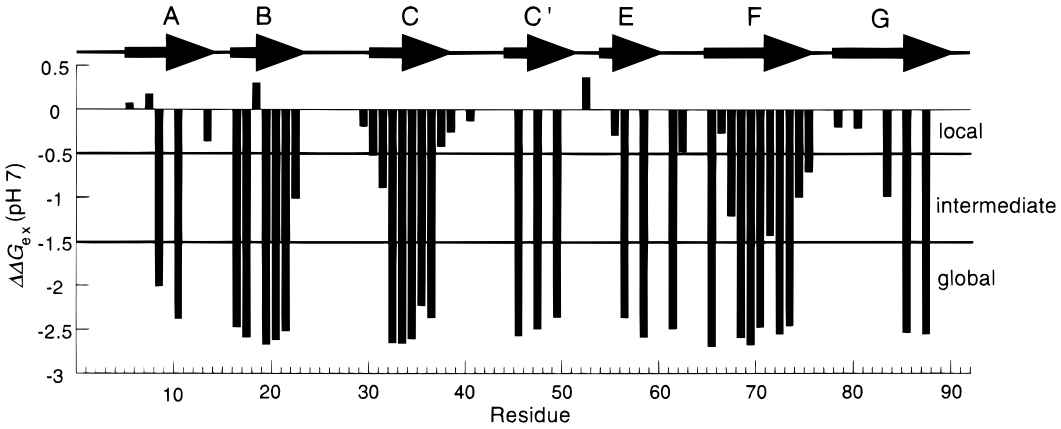

FIGURE 3: Hydrogen exchange: values of $\Delta\Delta G_{ex}$. Histogram showing $\Delta\Delta G_{ex}$ values calculated for each protected amide from hydrogen exchange rate constants of $TNfn3_{1-90}$ and $TNfn3_{1-92}$ at p^2H 7.0. The seven β-strands are indicated at the top. Areas of local, intermediate, and globally exchanging residues are indicated on the right.

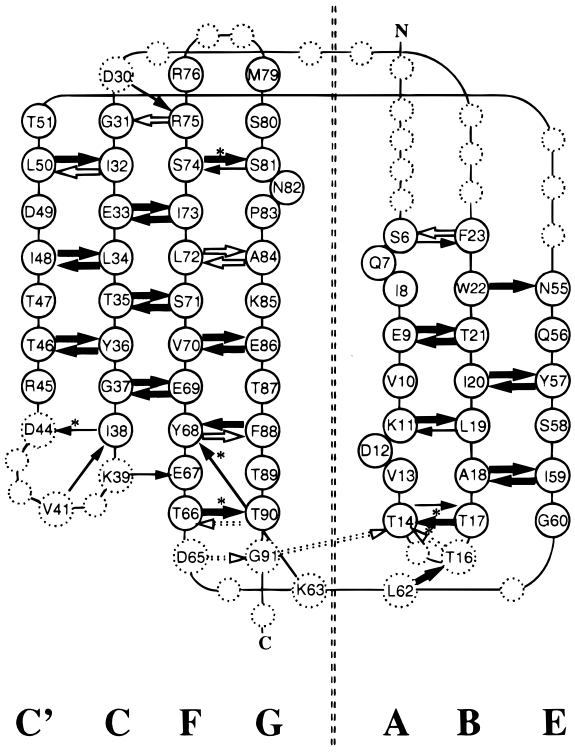

FIGURE 4: Hydrogen exchange: secondary structure. Schematic showing the secondary arrangement of the seven β-strands in TNfn3. The two sheets are separated by the dashed line, and the strands are labeled in the conventional FNIII manner underneath. The amino acid and sequence number are shown in bigger circles for all residues that are protected from exchange in one or both of the two forms, both residues in the β-sheet (closed circles) and in residues the loops and turns (dashed circles). Arrows connect the protected amides to the hydrogen bond acceptor (those marked * have side chain acceptors). The three subclasses of exchange are indicated by different arrows. Locally exchanging residues are indicated by thin arrows; intermediate exchange by thick hollow arrows; and the globally exchanging residues by thick bold arrows. Dotted arrows represent amide protons which show no protection in $TNfn3_{1-90}$ but are protected in $TNfn3_{1-92}$.

indicates that separation of individual FNIII domains may hinge on residue 92 (*30*). Inclusion of this residue in the preceding domain allows the maintenance of the hydrogen bonding network surrounding the C-terminus, whereas it should also be included in the next domain where the side chain is often buried (Figure 5).

Another factor which may influence TNfn3 stability may be the location of the charged terminal carboxylate group in an unfavorable buried environment (see Figure 5). Extension by two residues would move this group to the solvent-exposed surface. Experiments examining this possibility will be published elsewhere (Meekhof et al., unpublished results).

Kinetics. Stopped-flow kinetic studies show that whereas the folding rate of TNfn3 is little affected by extension, the unfolding rate is slowed considerably (Figure 6). This contrasts previous experimental evidence and theoretical studies which suggest that intrinsic refolding rates of homologous FnIII domains will strongly parallel their stabilities (*31*, *32*). The plot of ln k_u against [urea] is significantly curved, and it is evident that the unfolding amplitude does not reach 100% until unfolded in 8 M urea. Thus, extrapolation of unfolding data to 0 M urea, to obtain $k_u^{H_2O}$ is problematic. This was overcome by combining the data collected at 25 °C with those extrapolated to 25 °C from higher temperatures outside the transition region (Figure 2). Figure 6 illustrates the combined data plot of ln k_{obs} against [urea] which allows a value for $k_u^{H_2O}$ of 7.2 ($\pm$0.3) $\times$ 10^{-5} s^{-1} to be determined (Table 4).

It is important to analyze kinetic data for evidence for accumulation of intermediates, which could be stabilized by extension and made detectable. Several tests were employed, which are described in detail previously (*11*, *33*), which show that $TNfn3_{1-92}$ folding kinetics are still described by a two-state model. First, there is no significant deviation from linearity observed in the refolding region of Figure 6. Two remaining tests are to derive values for $\Delta G_{D-N}^{H_2O}$ and m_{D-N} from kinetic data (Table 4). The equilibrium constant derived from kinetics, $K_k = k_f^{H_2O}/k_u^{H_2O}$, adjusted to account for the proline isomerization rate [which constitutes 6% of the total refolding amplitude extrapolated to 0 M urea; see Jackson and Fersht (1991)], gives a kinetic $\Delta G_{D-N}^{H_2O}$ value of 6.68 $\pm$ 0.04 kcal mol^{-1}. The *m*-value derived from kinetics, $m_k = RT(m_{k_f} + m_{k_u})$, is 1.23 $\pm$ 0.03 kcal mol^{-1} M^{-1}. These values compare well with values derived from equilibrium experiments (6.68 $\pm$ 0.18 kcal mol^{-1} and 1.26 $\pm$ 0.05 kcal mol^{-1} M^{-1}). From folding data presented here and previously (*11*), a ϕ_f-value for the truncation of TNfn3 by two residues of 0.2 is calculated from both the folding and unfolding data. This indicates, as might be expected for the terminus of a protein domain, that the extended region

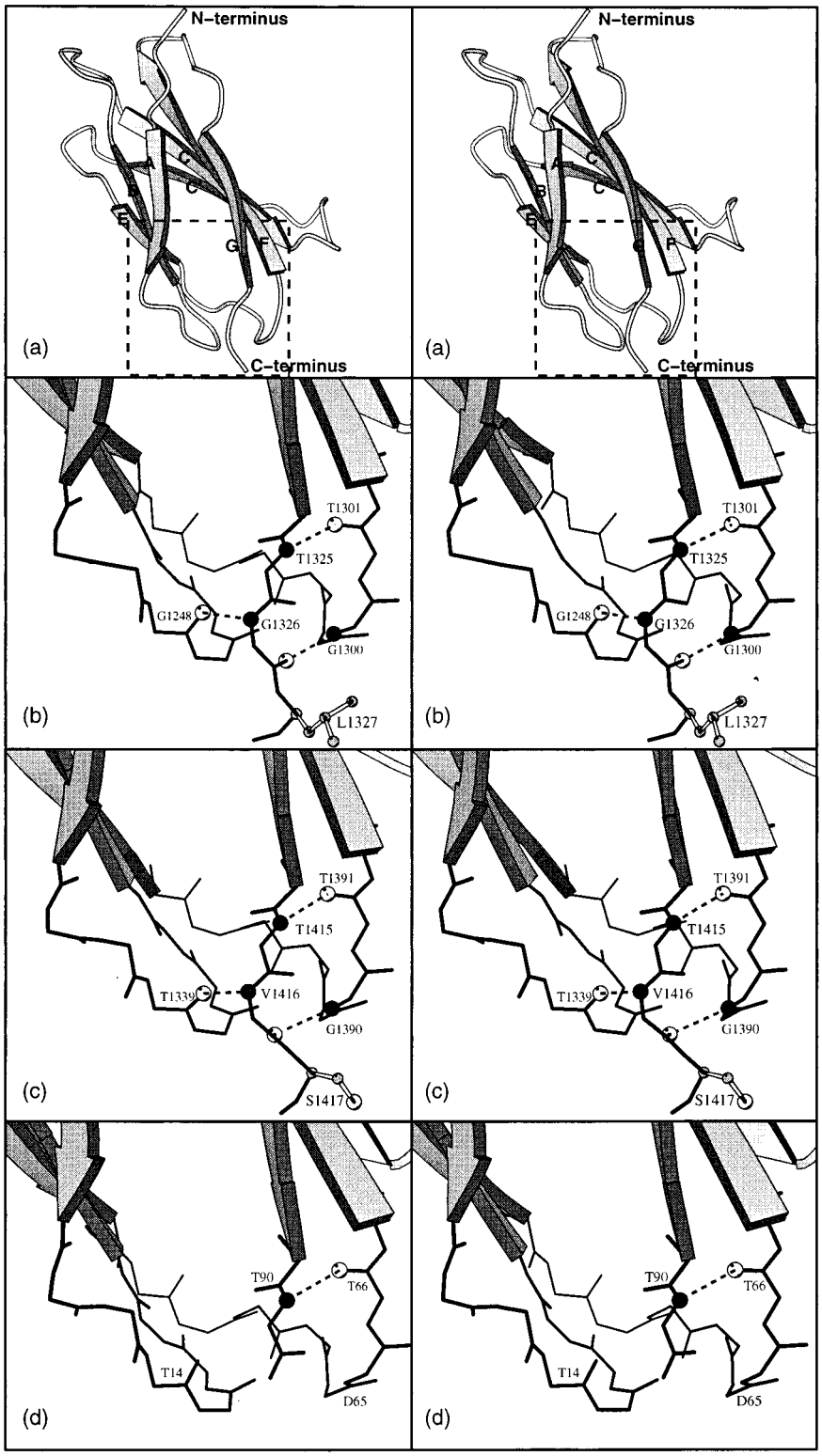

FIGURE 5: Comparison of the C-terminal regions of truncated $TNfn3_{1-90}$ to FNIII modules from fibronectin. (a) Ribbon representation in stereo of the eighth FNIII domain of human fibronectin (FNfn8); the seven β-strands are labeled in the conventional manner, and the boxed region indicates the C-terminal area shown (in the same orientation) in the remaining panels. (b) Stereo representation of the C-terminal region of FNfn8. (c) Stereo diagram of the C-terminal region of FNfn9. (d) Stereo diagram of the C-terminal region of $TNfn3_{1-90}$ (*12*). The stereo diagrams show the three FNIII domains in the same orientation with the A−B loop to the left, the C-terminal strand down the center, and the E−F loop crossing behind. Hydrogen bonds are shown as dots between oxygen (white spheres) and nitrogen (dark gray spheres). Residues involved in hydrogen bonds and their equivalents in $TNfn3_{1-90}$ are labeled. The side chains of the equivalent residues to position 92 in $TNfn3_{1-92}$ are shown as ball-and-stick in (b) and (c); carbon atoms are represented as small light gray spheres. These structures suggest that the side chain of residue 92 in $TNfn3_{1-90}$ points out of the core and that the added backbone makes the equivalent hydrogen bonds between the C-terminal strand and the A−B and E−F loops which are shown in (b) and (c). The FNIII domains are those from the four-domain structure of fibronectin (*30*). This diagram was produced using Molscript (*48*).

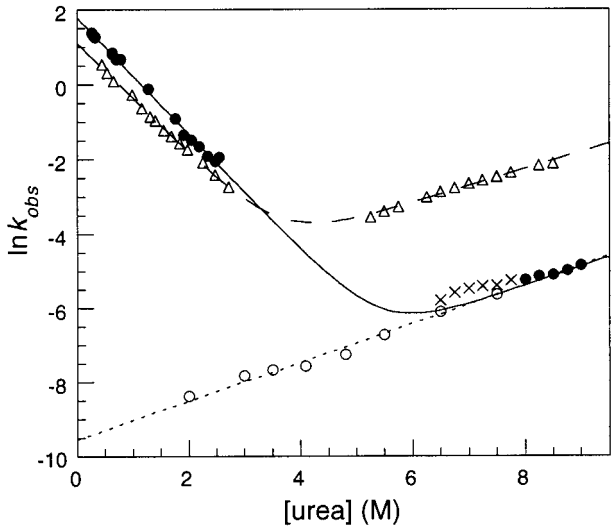

FIGURE 6: Extrapolated kinetics of folding. Urea dependence of the natural logarithm of the folding and unfolding rate constants of $TNfn3_{1-90}$ and $TNfn3_{1-92}$. Open triangles (△) show the data for $TNfn3_{1-90}$ [data taken from *(11)*]. The filled circles (●) indicate the rate constants (s^{-1}) determined by stopped-flow folding and unfolding of $TNfn3_{1-92}$ at 25 °C; open circles (○) indicate unfolding rate constants extrapolated from higher temperatures (Figure 2); and crosses (×) indicate values where the $TNfn3_{1-92}$ unfolding rate constant is at less than 100% of the full unfolding amplitude. The extension of TNfn3 has no significant effect on refolding but slows unfolding considerably.

Table 4: Folding Kinetics of $TNfn3_{1-90}$ and $TNfn3_{1-92}$

	$TNfn3_{1-90}$[a]	$TNfn3_{1-92}$
$k_f^{H_2O}$ (s^{-1})	2.90 ± 0.40	6.02 ± 0.22
$k_u^{H_2O}$ (s^{-1})	2.78 (±0.27) × 10^{-3}	7.21 (±0.26) × 10^{-5}
m_{k_f} (M^{-1})	1.41 ± 0.02	1.55 ± 0.03
m_{k_u} (M^{-1})	0.45 ± 0.01	0.52 ± 0.01

[a] Data from (*11*) reanalyzed with a linear fit to the unfolding data.

becomes stable after the major transition state for folding. The ratio $m_{D-\ddagger}/m_{D-N}$ [where $m_{D-\ddagger} = RTm_{k_f}$ (Table 4), and is the *m*-value on going from the denatured state D to the transition state; and m_{D-N} is the equilibrium *m*-value for unfolding (Table 2)] measures the relative changes in exposure to urea between the transition and natives states and is thus an index of the position of the transition state on the reaction coordinate (*34*). The value of $m_{D-\ddagger}/m_{D-N}$ for both $TNfn3_{1-90}$ and $TNfn3_{1-92}$ is 0.7.

In summary, the broad features of TNfn3 folding are maintained in the extended form, and the effects of extension fit well to the two-state model proposed for its folding. The extended region is not energetically important in the transition state for folding. The energy and position of the folding transition state along the reaction coordinate, relative to the denatured state, are unchanged. Thus, the only significant effect of the increased native state stability is a considerable slowing of the intrinsic unfolding rate constant. FNIII domains are found in many muscle proteins such as titin and twitchin. These results indicate that removal of a small number of terminal hydrogen bonds can dramatically increase the rate of unfolding. This may have implications for the mechanism of unfolding of β-sandwich domains which have been proposed to occur during stress-relaxation in muscle fibers (*35*−*39*).

Domain Boundary Selection Strategies. Rational domain boundary selection can be crucial for structural studies (*9*). In the case of modules such as TNfn3, the selection of a single domain for study completely removes interactions with other modules within the multimodular assembly. Truncation can have unexpected consequences, which may or may not be present in the complete assembly, such as introducing substantial mobility, or simply compromising domain stability. TNfn3 is reasonably stable and can fold and crystallize if the domain boundaries selected are nonoptimal (*12*), but less stable protein domains with similar truncations may not fold at all. Structural alignments make it possible to make reasonable predictions of domain boundaries. Our results make it clear, however, that other factors, more difficult to predict, may be important.

A strategy much like the one presented can be used to correctly identify boundaries, using those predicted as a starting point. Directed or random 'dirty' PCR can produce many different domain lengths. Subsequent expression tests can then be used as an approximate test of stability to select clones to be assayed for thermal stability. This strategy can proceed iteratively to identify the precise boundaries (*9*). Alternatively, several extra residues could be included for safe measure after the predicted boundaries at each end. For detailed structural analyses, initial characterization of each domain boundary is worthwhile. $TNfn3_{1-92}$ is an ideal model protein for use in studying the folding pathway of an FNIII domain by protein engineering. The increased stability should facilitate a mutagenic study of folding and stability.

CONCLUSIONS

Many biologically important proteins are made up of independently folding modules. The stability of these modules is affected by the presence of interactions with adjacent modules (*40*, *41*). To study isolated modules, boundary selection is clearly important (*9*, *10*). Incorrect selection of the C-terminal boundary of TNfn3 by two residues removes several long-range interactions which contribute to the overall stability of the β-sandwich structure. This has local and global effects on stability and alters related biophysical characteristics such as the unfolding rates, hydrogen exchange kinetics, and overall flexibility of the domain. Previous biophysical studies of isolated FNIII domains have focused on domains of equivalent length to $TNfn3_{1-90}$ (*11*, *31*, *41*−*46*). The results presented here imply that truncation can have a very significant effect on the properties of a marginally stable domain whereas intrinsically stable domains may be less affected.

ACKNOWLEDGMENT

We thank Susan Fowler for technical assistance and Ernesto Cota, Jean Chatelier, Chris Johnson, Stefan Freund and Mark Bycroft for very helpful discussions and advice.

REFERENCES

1. Chothia, C. (1992) *Nature 357*, 543−544.
2. Gerstein, M., and Levitt, M. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 11911−11916.
3. Bork, P., and Doolittle, R. F. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 8990−8994.

4. Williams, A. F., and Barclay, A. N. (1988) *Annu. Rev. Immunol. 6*, 381−405.
5. Baron, M., Norman, D. G., and Campbell, I. D. (1991) *Trends Biochem. Sci. 16*, 13−17.
6. Patel, R. S., Odermatt, E., Schwarzbauer, J. E., and Hynes, R. O. (1987) *EMBO J. 6*, 2565−2572.
7. Cohen, S. L. (1996) *Structure 4*, 1013−1016.
8. Dickinson, C. D., Gay, D. A., Parello, J., Ruoslahti, E., and Ely, K. R. (1994) *J. Mol. Biol. 238*, 123−127.
9. Pfuhl, M., Improta, S., Politou, A. S., and Pastore, A. (1997) *J. Mol. Biol. 265*, 242−256.
10. Politou, A. S., Gautel, M., Joseph, C., and Pastore, A. (1994) *FEBS Lett. 352*, 27−31.
11. Clarke, J., Hamill, S. J., and Johnson, C. M. (1997) *J. Mol. Biol. 270*, 771−778.
12. Leahy, D. J., Hendrickson, W. A., Aukhil, I., and Erickson, H. P. (1992) *Science 258*, 987−991.
13. Glasoe, P. F., and Long, F. A. (1960) *J. Phys. Chem. 64*, 188−193.
14. Clarke, J., Hounslow, A. M., Bycroft, M., and Fersht, A. R. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 9837−9841.
15. Jackson, S. E., elMasry, N., and Fersht, A. R. (1993) *Biochemistry 32*, 11270−11278.
16. Itzhaki, L. S., Otzen, D. E., and Fersht, A. R. (1995) *J. Mol. Biol. 254*, 260−288.
17. Fersht, A. R., Matouschek, A., and Serrano, L. (1992) *J. Mol. Biol. 224*, 771−782.
18. Hvidt, A. A., and Nielsen, S. O. (1966) *Adv. Protein Chem. 21*, 287−386.
19. Roder, H., Wagner, G., and Wütrich, K. (1985) *Biochemistry 24*, 7396−7407.
20. Perrett, S., Clarke, J., Hounslow, A. M., and Fersht, A. R. (1995) *Biochemistry 34*, 9288−9298.
21. Bai, Y., Sosnick, T. R., Mayne, L., and Englander, S. W. (1995) *Science 269*, 192−197.
22. Swint-Kruse, L., and Robertson, A. D. (1996) *Biochemistry 35*, 171−180.
23. Skelton, N. J., Kordel, J., Akke, M., and Chazin, W. J. (1992) *J. Mol. Biol. 227*, 1100−1117.
24. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) *Proteins: Struct., Funct., Genet. 17*, 75−86.
25. Nies, D. E., Hemesath, T. J., Kim, J. H., Gulcher, J. R., and Stefansson, K. (1991) *J. Biol. Chem. 266*, 2818−2823.
26. Serrano, L., Kellis, J., Cann, P., Matouschek, A., and Fersht, A. R. (1992) *J. Mol. Biol. 224*, 783−804.
27. Kim, K. S., Fuchs, J. A., and Woodward, C. K. (1993) *Biochemistry 32*, 9600−9608.
28. Pfuhl, M., and Pastore, A. (1995) *Structure 3*, 391−401.
29. Fong, S., Hamill, S. J., Procter, M., Freund, S. M. V., Benian, G. M., Chothia, C., Bycroft, M., and Clarke, J. (1996) *J. Mol. Biol. 264*, 624−639.
30. Leahy, D. J., Aukhil, I., and Erickson, H. P. (1996) *Cell 84*, 155−164.
31. Plaxco, K. W., Spitzfaden, C., Campbell, I. D., and Dobson, C. M. (1997) *J. Mol. Biol. 270*, 763−770.
32. Finkelstein, A. V. (1991) *Proteins: Struct., Funct., Genet. 9*, 23−27.
33. Jackson, S. E., and Fersht, A. R. (1991) *Biochemistry 30*, 10428−10435.
34. Chen, B.-L., Baase, W. A., Nicholson, H., and Schellman, J. A. (1992) *Biochemistry 31*, 1464−1476.
35. Tskhovrebova, L., Trinick, J., Sleep, J. A., and Simmons, R. M. (1997) *Nature 387*, 308−312.
36. Politou, A. S., Thomas, D. J., and Pastore, A. (1995) *Biophys. J. 69*, 2601−2610.
37. Soteriou, A., Clarke, A., Martin, S., and Trinick, J. (1993) *Proc. R. Soc. London B. 254*, 83−86.
38. Erickson, H. P. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 10114−10118.
39. Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J. M., and Gaub, H. E. (1997) *Science 276*, 1109−1112.
40. Litvinovich, S. V., and Ingham, K. C. (1995) *J. Mol. Biol. 248*, 611−626.
41. Spitzfaden, C., Grant, R. P., Mardon, H. J., and Campbell, I. D. (1997) *J. Mol. Biol. 265*, 565−579.
42. Akke, M., Liu, J., Cavanagh, J., Erickson, H. P., and Palmer, A. G. (1998) *Nat. Struct. Biol. 5*, 55−59.
43. Plaxco, K. W., Spitzfaden, C., Campbell, I. D., and Dobson, C. M. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 10703−10706.
44. Carr, P. A., Erickson, H. P., and Palmer, A. G. (1997) *Structure 5*, 949−959.
45. Baron, M., Main, A. L., Driscoll, P. C., Mardon, H. J., Boyd, J., and Campbell, I. D. (1992) *Biochemistry 31*, 2068−2073.
46. Main, A. L., Harvey, T. S., Baron, M., Boyd, J., and Campbell, I. D. (1992) *Cell 71*, 671−678.
47. Pace, C. N. (1986) *Methods Enzymol. 131*, 266−279.
48. Kraulis, P. (1991) *J. Appl. Crystallogr. 24*, 946−950.

BI9801659