Aromatic Residues Engineered into the β-Turn Nucleation Site of Ubiquitin Lead to a Complex Folding Landscape, Non-Native Side-Chain Interactions, and Kinetic Traps[†]

Anita M. Rea,[‡] Emma R. Simpson,[‡,§] Jill K. Meldrum,[‡] Huw E. L. Williams, and Mark S. Searle*

Centre for Biomolecular Sciences, School of Chemistry, University Park, Nottingham NG7 2RD, U.K.

Received July 15, 2008; Revised Manuscript Received September 18, 2008

ABSTRACT: The fast folding of small proteins is likely to be the product of evolutionary pressures that balance the search for native-like contacts in the transition state with the minimum number of stable non-native interactions that could lead to partially folded states prone to aggregation and amyloid formation. We have investigated the effects of non-native interactions on the folding landscape of yeast ubiquitin by introducing aromatic substitutions into the β-turn region of the N-terminal β-hairpin, using both the native G-bulged type I turn sequence (TXTGK) as well as an engineered 2:2 XNGK type I′ turn sequence. The N-terminal β-hairpin is a recognized folding nucleation site in ubiquitin. The folding kinetics for *wt*-Ub (TLTGK) and the type I′ turn mutant (TNGK) reveal only a weakly populated intermediate, however, substitution with X = Phe or Trp in either context results in a high propensity to form a stable compact intermediate where the initial U→I collapse is visible as a distinct kinetic phase. The introduction of Trp into either of the two host turn sequences results in either complex multiphase kinetics with the possibility of parallel folding pathways, or formation of a highly compact I-state stabilized by non-native interactions that must unfold before refolding. Sequence substitutions with aromatic residues within a localized β-turn capable of forming non-native hydrophobic contacts in both the native state and partially folded states has the undesirable consequence that folding is frustrated by the formation of stable compact intermediates that evolutionary pressures at the sequence level may have largely eliminated.

A significant number of small proteins fold via an apparent two-state mechanism in which the native properties appear to rapidly accumulate in a concerted kinetic process of collapse and stabilization of secondary and tertiary interactions (*1*). Nonlinear refolding or unfolding kinetics are signatures for deviations from the two-state model and point to either very rapid early folding events, the accumulation of intermediates, or the transient population of high energy states whose formation is driven by large-scale changes in topology of the polypeptide chain associated with significant fractional changes in surface burial (*2*−*10*). The folding rates of many small proteins correlate with various indicators of folding topology that measure separation along the peptide chain of contacting residues (contact order) (*11, 12*). This continuum of folding mechanisms, from highly populated intermediates on a three-state pathway to metastable transiently populated species on an apparent two-state pathway, is underpinned by the relative energies and placement of barriers (early or late, rate-limiting, or otherwise) and whether intermediates lie on-pathway or are kinetic traps requiring error-repair (productive or obstructive).

The nature of the unfolded state remains a largely uncharacterized component in understanding the protein folding problem. Evidence suggests that the initial conformational search may be driven in some contexts by extensive long-range interactions between hydrophobic residues which appear to persist even under strongly denaturing conditions (*13*). The folding of small proteins on a smooth energy landscape is likely to be dominated by the search for native-like contacts, with fast folding facilitated by avoiding the bottlenecks inherent in forming stable non-native interactions (*14*−*17*). In contrast, the opposite is suggested to be the case for large slow-folding multidomain proteins to ensure that they reach a functional folded state (*13, 18*). In these cases, non-native interactions may have a role in sequestering residue side chains to prevent protein aggregation and the development of amyloid structures (*19*−*21*).

The characterization of folding pathways is critically dependent on suitable spectroscopic probes (typically tryptophan) that are able to resolve kinetic processes linked to folding transitions (typically U→I and I→N in three-state pathways) and populations of intermediate states. This distinction becomes problematic when the barrier heights for the two transitions converge, resulting in rate constants that are difficult to separate, or when key transitions are spectroscopically 'silent'. Ubiquitin has proved a paradigm for folding studies (*22*−*28*). The F45W substitution has been used as a fluorescent probe for kinetic studies of ubiquitin over a period of greater than a decade (*24*). However, this

[†] We thank the EPSRC and BBSRC of the UK, and the University of Nottingham, for funding.

* Corresponding author. E-mail: mark.searle@nottingham.ac.uk, tel: (+44) 115 951 3567, fax: (+44) 115 846 6059.

[‡] A.M.R., E.R.S., and J.K.M. contributed equally to this work.

[§] Current address: Lehrstuhl Biotechnologie, Department Chemie, Technische Universität München, Lichtenbergstrasse 4, 85747 Garching, Germany.

10.1021/bi801330r CCC: $40.75
Published on Web 11/08/2008

mutant is an example of where the U→I and I→N transitions appear to be so poorly resolved that they merge to give an apparent U→N transition and a V-shaped chevron characteristic of an apparent two-state folding mechanism. One possible resolution of these difficulties lies in the use of different Trp probes at structurally distinct locations on the protein surface to enhance the visibility of different folding transitions and probe both local and global changes in the conformation of the polypeptide chain as it folds. The work of Vallee-Belisle and Michnick has described such a strategy in which Trp probes with different structural sensitivities were able to distinguish between the U→I and I→N transitions (*27*). These ubiquitin mutants, with enhanced fluorescence of the I-state, resulted in the population of a stable intermediate that was compact, misfolded, and on the folding pathway. Although the various transitions appeared to be ill-resolved for *wt*-Ub containing the F45W reporter, a series of helix mutants with incremental stabilization of the native state through changes to the structural propensity of the main helix (largely through Ala substitutions at solvent exposed sites) resulted in the I-state also becoming more significantly populated with both the U→I and I→N transitions better-resolved (*28*). The poor visibility of the I-state in the folding of *wt*-Ub F45W mutant is most likely associated with the fact that it is only weakly populated. Moreover, accelerated folding kinetics correlated with the increase in helical propensity such as to suggest that the highly populated I-state is productive for folding rather than a misfolded on-pathway kinetic trap. The differences highlighted by theses two approaches suggest that the nature of the substitutions that are introduced (large hydrophobic Trp residues versus stabilized secondary structure) can have a significant impact on the folding landscape by contributing to the stability of the misfolded I-state, making it structurally distinct from intermediates on the folding landscape of wild-type protein.

We have investigated the effects of non-native interactions on the folding landscape of yeast ubiquitin by introducing aromatic substitutions into the β-turn region of the N-terminal β-hairpin, which is a recognized folding nucleation site in ubiquitin (*29*−*32*). We have used ubiquitin for kinetic studies with either the native G-bulged type I turn sequence (T<u>X</u>TGK) or an engineered 2:2 <u>X</u>NGK type I′ turn sequence with aromatic substitution at position X. We have shown that the change in stability of ubiquitin correlates with the statistical preference for different <u>X</u>NGK type I′ turn sequences, with X = Val highly preferred (*33, 34*). The folding kinetics for *wt*-Ub (TLTGK) and the type I′ turn mutant (TNGK) reveal only a weakly populated intermediate, however, substitution with X = Phe or Trp in either context results in a significant increase in stability of the native state and a high propensity to form a stable compact intermediate where U→I and I→N kinetic phases are well resolved. In particular, the introduction of Trp into either of the two host turn sequences results in either complex multiphase kinetics involving parallel folding pathways, or formation of a highly compact I-state stabilized by non-native interactions that must partially unfold in the transition state before refolding to the native structure. Sequence substitutions with aromatic residues within a β-turn implicated in folding nucleation have the undesirable consequence that stable non-native interactions affect the dynamics of the polypeptide chain (*14, 37*) such as to frustrate folding through the formation of stable compact intermediates that evolutionary pressures at the sequence level may have largely eliminated.

MATERIALS AND METHODS

Mutagenesis and Protein Expression. A pKK223-3 plasmid construct containing the wild-type ubiquitin or F45W yeast ubiquitin gene was used as a template for mutagenesis reactions using the QuikChange site-directed mutagenesis kit and as previously described. Proteins were expressed in *Escherichia coli* strain BL21(DE3) under the control of the IPTG-inducible *tac* promoter (*26, 28, 29*).

NMR Analysis of Ubiquitin Mutants. The ubiquitin mutant WN with aromatic substitution Thr7→Trp7 was analyzed by NMR at 600 MHz on a Bruker Avance600 spectrometer with triple resonance inverse-detection probe. Homonuclear ^{1}H 2D NMR (TOCSY and NOESY) data were collected at 298 K on ~1 mM protein samples at pH 5.0 at a number of mixing times (100 and 250 ms for NOESY; 70 ms spin-locking sequence for TOCSY) to aid assignment. A high proportion of backbone NH, CαH, and side-chain ^{1}H resonances were assigned using the CCPNMR software (*36*). Comparisons of Hα chemical shift deviations from random coil values showed gross similarities with those for the native (F45W) protein. The most significant chemical shift perturbations could be directly correlated with structural changes localized to the region of the turn modifications. Chemical shifts, ring current effects, and NOEs were analyzed to identify specific side-chain interactions relevant to the aromatic turn substitutions.

Kinetic Experiments. Kinetic unfolding and refolding measurements were performed using an Applied Photophysics Pi-star 180 spectrophotometer, with temperature regulated using a Neslab RTE-300 circulating programmable water bath. Refolding experiments were performed by 1:10 dilution of unfolded protein (15 μM in 5 M GdmCl) into buffered solutions of different GdmCl concentrations yielding a final protein concentration of 1.36 μM. Experiments at 10-fold lower dilution were performed to assess the possible contribution of protein aggregation on refolding rates, but no differences in rate constant were apparent. Similarly, unfolding experiments also employed a 1:10 dilution and in all cases 25 mM acetate buffer pH 5.0 was used as standard. Kinetic measurements for both unfolding and refolding reactions were averaged 4−6 times at each GdmCl concentration and the GdmCl concentration was determined using a refractometer (*26, 28, 29*).

Analysis of Kinetic Data. The kinetic data were analyzed using a multiexponential fitting procedure (typically three refolding phases and one unfolding phase were observed) with the quality of the fit determined from an analysis of the residuals (*26, 28, 37*). Chevron plots were constructed by plotting the natural log of the observed rate constants against denaturant concentration (*38*). Kinetic data which exhibited deviations from linearity in the denaturant-dependence of the refolding rates (rollover effects) were fitted to both on-pathway (U↔I↔N) and off-pathway (I↔U↔N) three-state models. The observed rate constants and amplitudes were analyzed globally using IGOR (Wavemetrics) by simultaneously fitting the normalized amplitudes of the refolding and unfolding processes to determine the micro-

Table 1: Kinetic Parameters (obtained from the on-pathway global fitting of data collected at 298 K and pH 5.0 in 25 mM acetate buffer) for the Ubiquitin β-Turn Mutants (all rate constants in s^{-1} and all *m*-values in kJ mol^{-1} M^{-1})

variant	k_{UI} (m_{UI})	k_{IU} (m_{IU})[f]	k_{IN} ($m_{I\text{-}TS}$)	k_{NI} ($m_{N\text{-}TS}$)	m_{kin}[d]	ΔG_{UI}[b] (kJ mol^{-1})	ΔG_{UN}[c] (kJ mol^{-1})	F_I[a]	β_{TS1}[e]	β_I[e]	β_{TS2}[e]
TN	180 ± 28.36 (-4.87 ± 0.27)	9 ± 0.3 (3.07 ± 0.2)	33.5 ± 9.6 (-0.39 ± 0.38)	0.013 ± 0.003 (2.34 ± 0.12)	10.7	$\sim -7 \pm 2$	-26.4 ± 1.7	0.92	0.45	0.74	0.78
VN	1031 ± 108 (-4.70 ± 0.29)	2.0 ± 0.2 (3.12 ± 0.2)	33.3 ± 7.5 (-0.31 ± 0.35)	0.033 ± 0.008 (2.00 ± 0.12)	10.1	-15.5 ± 0.7	-32.6 ± 1.9	0.89	0.46	0.77	0.80
FN	919 ± 51 (-5.05 ± 0.07)	0.06 ± 0.0006 (3.10 ± 0.2)	3.6 ± 0.2 (-0.24 ± 0.08)	0.001 ± 0.0002 (2.67 ± 0.07)	11.1	-23.9 ± 1.1	-44.2 ± 2.8	0.86	0.45	074	0.76
WN^{AA}	259 ± 52 (-5.06 ± 0.44)	0.10 ± 0.03 (3.15 ± 0.2)	4.0 ± 0.6 (-0.29 ± 0.31)	0.00034 ± 0.0001 (3.11 ± 0.12)	11.6	-19.5 ± 0.8	-42.7 ± 2.3	0.88	0.43	0.71	0.73
L8W	110 ± 9 (-6.5 ± 0.15)	0.06 ± 0.02 (3.1 ± 0.2)	2.3 ± 0.3 ($+1.7 \pm 0.3$)	0.0012 ± 0.001 (2.21 ± 0.17)	10.1	-17.7 ± 0.9	-35.4 ± 2.8	0.90	0.64	0.95	0.78

[a] Relative fluorescence of the intermediate where a value of 1 indicates fluorescence of the intermediate is identical to that of the native state and a value of 0 indicates similarity to the denatured state. [b] $\Delta G_{UI} = -RT \ln(K_{UI})$ [c] $\Delta G_{UN} = -RT \ln(K_{UI}(k_{IN}/k_{NI}))$; both ΔG_{UI} and ΔG_{UN} are measured in kJ/mol^{-1}. [d] $m_{kin} = m_{IU} - m_{UI} + m_{NI} - m_{IN}$. [e] β-values were calculated as follows: $\beta_{TS1} = -m_{UI}/m_{kin}$; $\beta_I = (m_{IU} - m_{UI})/m_{kin}$ and $\beta_{TS2} = (m_{kin} - m_{NI})/m_{kin.}$ [f] m_{IU} values in many cases were poorly defined by the data and were constrained by the global fit.

scopic rate constants, as previously described (*28*). Errors represent those calculated directly from the IGOR global fitting procedure. In all of the data sets, the fast chevron for the U↔I transition was poorly defined by a limited number of k_{IU} values in the unfolding arm. Global fitting was guided by the overall best fit between equilibrium *m*-values and total kinetic *m*-values typically in the range -10 to -12 kJ mol^{-1} M^{-1}. Although a variation in m_{IU} between -2 and -4 kJ mol^{-1} M^{-1} produced a similar quality of fit to the chevron and amplitude data, values of -3 ± 0.2 kJ mol^{-1} M^{-1} gave better agreement with equilibrium *m*-values. On this basis, m_{IU} was fixed within this range during the global analysis and values for buried surface areas β_{TS1}, β_I, and β_{TS2} were calculated accordingly (see Table 1). We further discriminated between on-pathway and off-pathway mechanisms using the approach described for cytochrome *c* based on the observed relative rate constants (*39*). Under conditions that favor the native state, only the on-pathway model allows the observed rate constant for the unfolding of the intermediate k_{IU} to be slower than the microscopic rate constant for the I→N transition. Consistent with this model, we observe that $k_{IU} < k_{IN}$ in all well-resolved cases, supporting the on-pathway model.

RESULTS

Effects of Stabilizing β-Turn Sequences on Folding Kinetics. The 2:2 type I′ turn (TNGK) is readily accommodated in place of the native G-bulged 3:5 type I turn (TLTGK) of ubiquitin with a small stabilizing effect of only ~1 kJ mol^{-1} (*33*). We analyzed the refolding and unfolding kinetics in detail making particular use of the amplitude data to resolve the various kinetic phases. The refolding data for the TN mutant (<2 M GdmCl) resolve into three phases with rate constants λ_1, λ_2, and λ_3, and amplitudes A_1, A_2, and A_3. The slowest phase ($\lambda_3 = 0.3$ s^{-1}) has a low amplitude ($A_3 < 0.1$) which is invariant to denaturant concentration and suggestive of a slow isomerization-limited folding process (*26*). The rate constants for the two faster phases ($\lambda_1 \sim 60$ s^{-1} and $\lambda_2 \sim 20$ s^{-1} at ~1 M GdmCl) have relative amplitudes that are quite similar (Figure 2a and 2b). A_1 decreases at higher denaturant concentrations, and A_2 increases and dominates the refolding data. The observation of a low amplitude fast-phase (λ_1) is consistent with the rapid population of a weakly stabilized I-state (U→I) under conditions that strongly favor refolding; however, the I-state is not significantly populated at denaturant concentrations >1 M GdmCl. Only a few data points at the beginning of the fluorescence decay define λ_1 which, when combined with the rapid decrease in amplitude with [GdmCl], makes the fitted rate constants subject to large uncertainty. We constructed a chevron plot of all of the refolding and unfolding data, as shown in Figure 2a for the TN mutant and see obvious similarities with the kinetic data for *wt*-Ub with the native TLTGK type I G-bulged turn (*28*), reflecting the fact that the two proteins are isoenergetic. A global analysis of the rate and amplitude data was fitted to an on-pathway three-state model, however, a detailed determination of all kinetic parameters was limited by the poorly defined refolding and unfolding kinetics for the U↔I transition arising from the scatter in the data for λ_1 (see Materials and Methods). The crossover in the amplitudes for A_1 and A_2 demonstrate the formation of a weakly populated I-state that we can detect directly as a distinct kinetic phase.

Substitution of Thr with Val at the -B1 position leads to the statistically preferred type I′ turn (VNGK) (*34*) which is reflected in a stabilizing effect of ~5 kJ mol^{-1} (*33*). We observe the same three kinetic refolding phases for both the TN and the VN mutant; however, the rate and amplitude profiles are shifted to higher denaturant concentrations for VN, enabling us to more readily resolve the populated I-state. The amplitude of the fastest phase in the refolding data for VN dominates at low denaturant concentrations, with the crossover between A_1 and A_2 now occurring at around ~2 M GdmCl (compared to ~0.8 M GdmCl for TN) (Figure 2c and 2d). The small effect of the turn substitution on the stability of the I-state results in a more apparent rollover effect in the denaturant dependence of λ_2, reflecting the build-up in the population of the I-state. Although the refolding component of λ_1 (U→I) is more clearly resolved, enabling us to determine m_{UI} with greater certainty, we are unable to resolve any data points for the I→U transition such that extrapolated k_{IU} and m_{IU} parameters can only be estimated from the global analysis taking into consideration the equilibrium *m*-value for the U→N transition (see Table 1).

Engineered Non-Native Interactions Stabilize the I-State. The highly co-operative folding that is observed for many small naturally occurring proteins is likely the result of selectional pressures that minimize the possibility of hydrophobic contacts that could trap the folding polypeptide chain

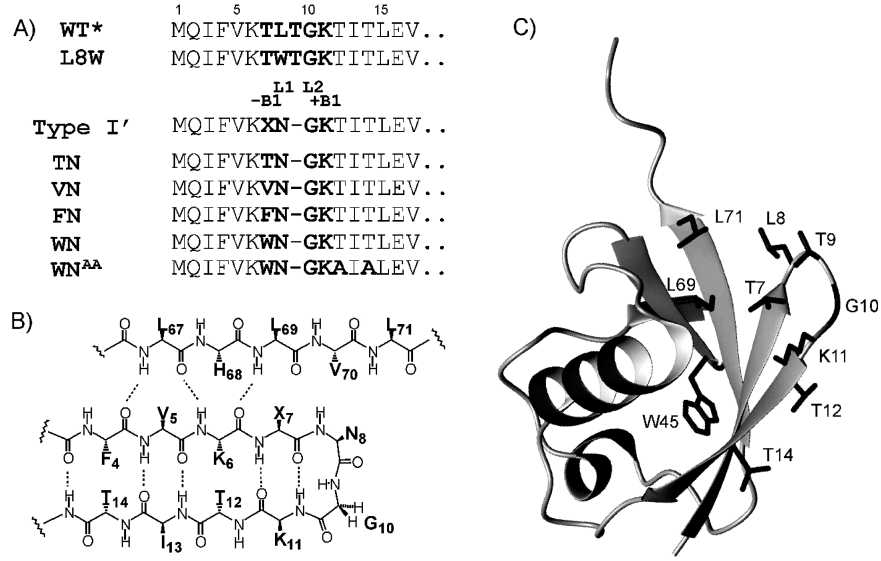


FIGURE 1: (A) N-Terminal 17-residue sequence of yeast wild-type ubiquitin involved in β-hairpin formation with a G-bulged type I β-turn sequence (sequence of the L8W mutant also shown). The native turn sequence was mutated to XNGK (type I′ β-turn) with several substitution considered at position X (T, V, F, and W), giving the TN, VN, FN, and WN mutants. The introduction of a T12A/T14A double mutation in the WN sequence is shown as WN^{AA}. (B) Schematic representation of the alignment of the β-hairpin sequence with the C-terminal β-strand that brings the side chain of residue 7 (X_7) into close proximity with Leu69 and Leu71 which are on the same face of the β-sheet. (C) Ribbon structure of *wt*-ubiquitin showing residue side chains involved in substitutions or that form side-chain contacts with the native β-turn (*44, 50*).

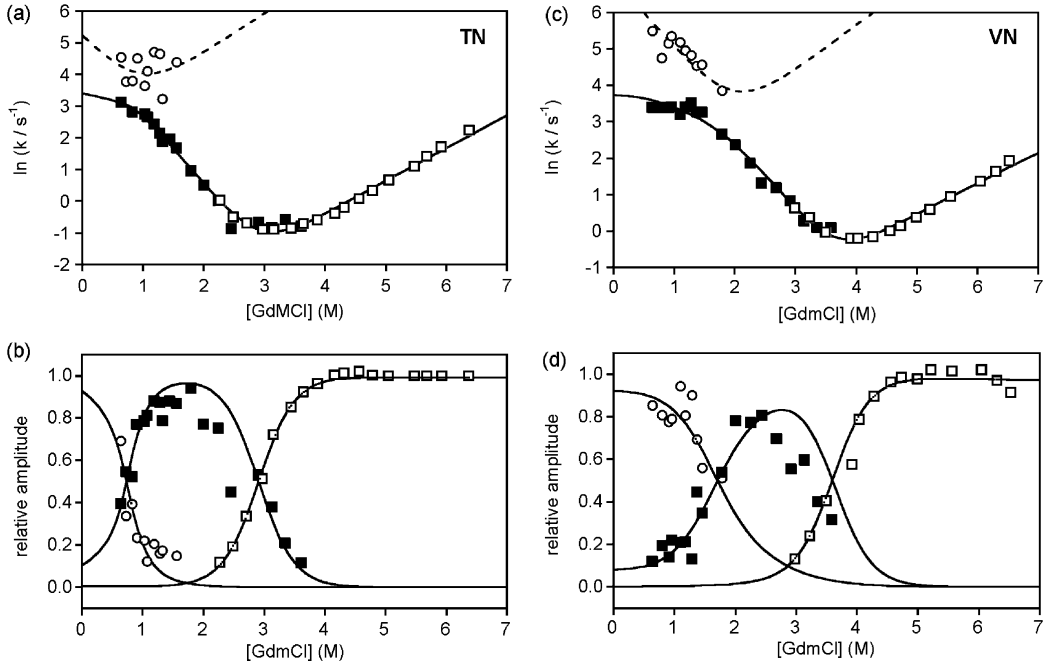


FIGURE 2: Denaturant-dependence of the refolding and unfolding rate constants for the TN turn mutant at 298 K (a), and the corresponding amplitude variation (b) of the two major refolding phases and unfolding phase. The amplitude of the fastest phase (open circles) shows that the intermediate is only weakly populated at low denaturant concentrations, resulting in a small rollover in the data for the I→N transition (black squares) in the chevron plot in part a. The amplitudes of the two phases (b) crossover at <1 M GdmCl as the I-state is further destabilized at increasing concentrations of denaturant. Analogous data for the VN mutant are shown in parts c and d. The amplitude of the fastest phase (open circles) shows that the intermediate is now significantly populated at low denaturant concentrations, resulting in a well-defined rollover in the data for the I→N transition (black squares) in the chevron plot in part c. The amplitude data is dominated by the initial collapse of U→I, with slow formation of the N-state. The phases crossover at ~ 2 M GdmCl as the I-state is destabilized. The lines of best-fit represent the global analysis of all of the data for each mutant using a three-state on-pathway model, with kinetic parameters presented in Table 1.

in non-native energy minima. This has been demonstrated in a number of de novo protein design studies using computational approaches or phage display libraries (*37*). To examine this hypothesis we have inserted both Phe and Trp at position X within our XNGK engineered turn sequence. Although FNGK and WNGK are statistically uncommon type I′ turn sequences (*34*), both aromatic residues enhance protein stability by ≥ 10 kJ mol^{-1} (*33*). We examined the origin of this stability enhancement using NMR. Homonuclear ^1H NMR analysis (TOCSY and NOESY experiments) enabled us to assign a significant proportion of backbone NH, CαH, and side-chain resonances of residues

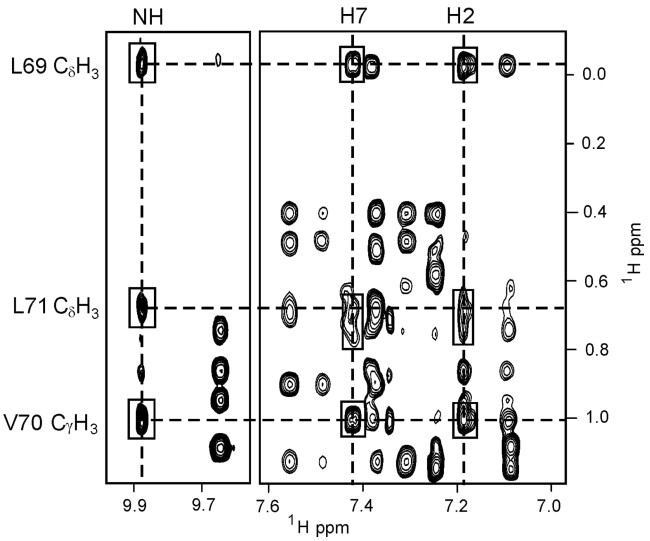


FIGURE 3: NMR analysis of side-chain contacts of Trp7 in the WN β-turn mutant. A portion of the 600 MHz 2D NOESY spectrum is shown at 298 K and pH 5.0, highlighting NOEs involving the aromatic side chain of Trp7 and methyl resonances of residues on the C-terminal β-strand. The large ring current perturbation to the $C\delta H_3$ of Leu69 ($\Delta\delta = 0.9$ ppm) is consistent with close side-chain contacts with the indole ring. The ^{1}H resonances of the indole ring protons of Trp7 (H2, H7, and NH, at 7.18, 7.42, and 9.87 ppm, respectively) are identified by vertical dashed lines and the relevant NOE crosspeaks highlighted in boxes are for Leu69 $C\delta H_3$ (−0.03 ppm), Leu71 $C\delta H_3$ (0.68 and 0.75 ppm), and Val70 $C\gamma H_3$ (1.01 ppm).

in close contact with the β-turn residues. As evident from Figure 1c, in the case of the WN mutant, substitution of Trp at position Thr7 is predicted to introduce non-native aromatic side-chain contacts with the indole ring, particularly with Leu69 but also the adjacent Val70 and Leu71 at the end of the C-terminal β-strand. A clear manifestation of these interactions was large ring-current-induced chemical shift perturbations to the $C\delta H_3$ and $C\gamma H$ resonances of Leu69, which appear at −0.03 and 0.92 ppm, respectively ($\Delta\delta$ = 0.93 and 0.72 ppm from random coil values). The $C\delta H_3$ resonances of Leu71 shift by a more modest amount (0.19 and 0.26 ppm). We are readily able to assign a string of medium/strong NOEs from the ^{1}H resonances of the indole side chain of Trp7 (H2, H7, and NH) to the perturbed side chains of Leu69, Val70, and Leu71 (Figure 3). These NOEs confirm the close contacts between the aromatic β-turn residue and adjacent residues in the C-terminal β-strand, providing a rationale for the significant increase in protein stability associated with these aromatic substitutions. Analogous ring current effects and interactions are also apparent with the Phe7 side chain of the FN mutants, as previously described (*33*).

Kinetic studies of the FN mutant resolve the same three refolding phases described for the TN and VN mutants. However, the I-state now appears to be significantly more stable. The amplitude changes for λ_1 and λ_2 show that the fastest kinetic refolding phase for the U→I transition dominates at low denaturant concentrations with the crossover between A_1 and A_2 occurring at ~2.5 M GdmCl (Figure 4). The chevron analysis shows that the slope of λ_1 for the initial U→I collapse is well-defined, and the high stability of the I-state clearly manifested as a pronounced rollover in the denaturant-dependence of λ_2 for the I→N transition. The variation of λ_2 with denaturant concentration (m_{IN}) is effectively zero at <2 M GdmCl, demonstrating that the I-state and the rate-limiting transition state are both highly collapsed and bury comparable amounts of hydrophobic surface area. We concluded that the rate-limiting step from I→N is associated largely with a structural rearrangement in the transition state rather than further surface area burial. Global analysis enabled us to define m_{IU} on the basis that the sum of the kinetic *m*-values equals the equilibrium *m*-value. On the basis of the quality of the global fit to an on-pathway model (see Materials and Methods (*39*)), and from systematic variation of the fitting parameters to assess effects on errors, we determined a full parameter set from which we were able to estimate m_{UI}, m_{IU}, m_{IN}, and m_{NI} and the corresponding β-values, $\beta_I \sim 0.74$ and $\beta_{TS} \sim 0.76$. To complete the analysis we repeated the global fitting of the rate and amplitude data for TN and VN using initial estimates for m_{IU} and m_{UI} derived from the FNGK data. We were able to produce satisfactory fits and global m_{UN} values that were consistent within the family of mutants (see Table 1).

The formation and decay of the populated I-state of the FN mutant was further characterized using sequential mixing stopped-flow experiments to first refold the protein and then rapidly jump to unfolding conditions (*40*−*42*). The Trp45 probe undergoes changes in fluorescence intensity corresponding to both the U↔I and I↔N transitions (folding or unfolding) that allow the formation and decay of the I-state to be observed directly. The sequential mixing experiment enabled us to initially refold the FN mutant by dilution of protein in 7 M GdmCl (25 mM acetate buffer at pH 5.0, 298 K) into the same buffer, giving a final denaturant concentration of 1 M GdmCl. A subsequent unfolding step was initiated by rapidly jumping the protein back into unfolding buffer (6.1 M GdmCl). The aging time (τ) between the two mixing steps was varied between 41 ms and 10 s and the unfolding kinetics measured along with the build-up in the amplitudes of these various phases as a function of the aging time. When short aging times were used (τ < 200 ms), we were able to resolve both a fast and a slow unfolding phase with rate constants of 33 ± 3 s^{-1} and 0.5 ± 0.04 s^{-1}. The rate constant for the major slow unfolding phase is completely consistent with that for the N→I transition observed in single-jump experiments (0.9 ± 0.3 s^{-1}) under similar unfolding conditions. However, the faster minor phase, which we attribute to the I→U transition, is not observed during single-jump unfolding.

As the aging time in the sequential mixing experiment increases (τ > 200 ms), the amplitude of the faster phase decreases considerably; however, the slow major phase is unaffected (Figure 4c). Time-dependent changes in the amplitudes of these phases reflect rates of formation of native-like molecules in the first refolding part of the sequential mixing experiment before they are induced to unfold in the final denaturation step. The amplitudes of the two unfolding kinetic phases were plotted versus aging time to show that the fastest phase, representing formation of the I-state, rapidly increases and then decays as I converts to N. The decay rate (k_{IN} = 3.9 ± 0.3 s^{-1}) is in good agreement with the value (3.3 ± 0.4 s^{-1}) derived directly from the single-jump experiment. Although the build-up rate is poorly defined because of the long dead-time of the experiment (~40 ms) and could not be measured directly, we were

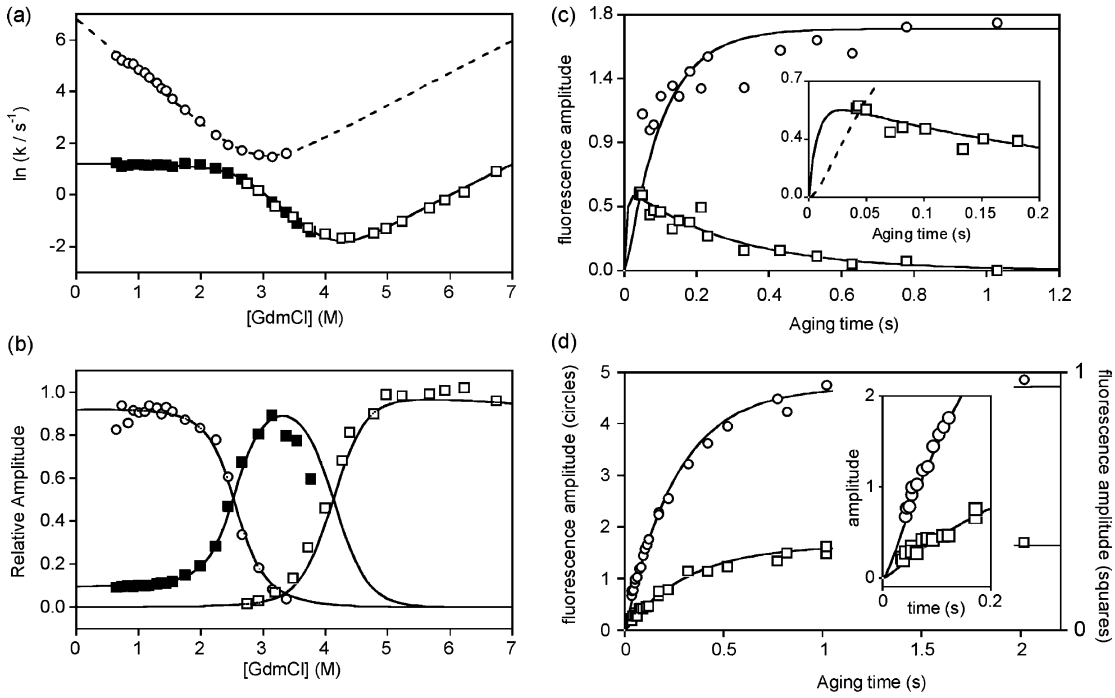


FIGURE 4: Denaturant-dependence of the refolding and unfolding rate constants for the FN turn mutant at 298 K (a), and the corresponding amplitude variation of the two major refolding phases and unfolding phase (b). The amplitude of the fastest phase (open circles) shows that the intermediate is highly populated at low denaturant concentrations, resulting in an initially flat rate profile for the I→N transition ($m_{IN} \sim 0$; black squares) in the chevron plot in part a, indicating that the I-state is highly compact. Interrupted refolding data is shown in part c using the sequential mixing approach in which the protein is initially refolding in 1 M GdmCl at 298 K followed by a jump back to unfolding buffer (6.1 M GdmCl) after a variable aging time of between 41 ms and 10 s. The time-dependent variation in the amplitude of the two unfolding phases provides a measure of the rate of formation of the native state in the first refolding step. The observation of a fast phase whose amplitude first increases rapidly and decays is clear evidence for the population of an intermediate state with rate constants consistent with those derived directly from single-jump stopped-flow experiments. In part d, sequential mixing experiments were performed that allowed an initial unfolding step followed by a second refolding step with the time-dependence of the amplitude of the two refolding phases illustrated. The rates fit well with a double exponential reflecting the formation of the unfolded state with k_{IU} contributing as a lag-phase to the build up curves (d). The long deadtime of the sequential mixing experiment and the scatter in the data prevented the lag phase from being determined directly with any accuracy; however, we were readily able to model the data with an estimated $k_{IU} > 100$ s^{-1} (see inset).

readily able to fit this data with the same rate constant ($k_{UI} \sim 130$ s^{-1}) measured directly for the U→I conversion in single-jump experiments (Figure 4a). The behavior of the amplitude of the slower major phase in the sequential mixing experiment should also reflect the same rate constants (k_{UI} and k_{IN}) for formation of the native state along the pathway U to I to N. The initial rapid U to I conversion, and subsequent population of the I-state, was expected to result in a lag in the increase in amplitude for the formation of the N-state that reflects the U→I transition (k_{UI}). Again, the long dead-time of the experiment and scatter in the amplitude data at short aging times precluded direct measurement of the kinetics of the lag phase; however, we were readily able to model the build-up in amplitude in terms of the population and decay of the I-state in a three-state on-pathway mechanism (as shown in Figure 4c).

Sequential mixing experiments were subsequently repeated by interchanging the two steps by first performing an unfolding reaction by dilution into 6.67 M GdmCl, followed by a rapid final refolding step in 1.4 M GdmCl after a variable aging time (*40*−*42*). Consistent with the above model, in the final refolding step we initially observe two refolding phases (U→I and I→N transitions), followed by a third minor phase which appears at longer aging times and is consistent with a slow isomerization process which is allowed to take place in the unfolded state. At the shortest aging time (41 ms), the two major refolding phases are described by rate constants of 74 ± 7 s^{-1} and 7 ± 4 s^{-1} which are (within the errors of the measurements) consistent with values for k_{UI} and k_{IN} measured in single-jump experiments (65 ± 4 and 4 ± 2 s^{-1}, respectively). The time-dependent increase in amplitude of these two phases should correlate with the rates of formation of the unfolded state (k_{NI} and k_{IU}) in the first step of the experiment, before being induced to refold in the final dilution step (Figure 4d). The amplitudes of the two phases fit well to a double exponential where k_{NI} is well-defined in both cases (3.9 ± 0.2 and 3.2 ± 0.2 s^{-1}, respectively), in good agreement with $k_{NI} = 3 \pm 0.3$ s^{-1} derived from single jump-experiments. However, k_{IU} is expected to contribute as a fast lag phase which under the conditions of the first unfolding step of this sequential mixing experiment we estimate to be $k_{IU} > 100$ s^{-1} (see Figure 4a). Again, the fast lag phase is poorly defined as a consequence of the long dead-time in the second step of the double-jump; however, we were able to model the data with a lag phase $k_{IU} > 100$ s^{-1} (Figure 4d). In conclusion, sequential mixing experiments that interrupt both the unfolding and refolding of the FN mutant under different conditions appear to fully account for the different populations of the N-state, I-state, and U-state within the context of a three-state folding mechanism with an on-pathway intermediate.

Complex Multiphase Kinetics in the Folding of the WN Mutant. We extended the study to consider the β-turn substitution with X = Trp (WNGK type I′ turn). The

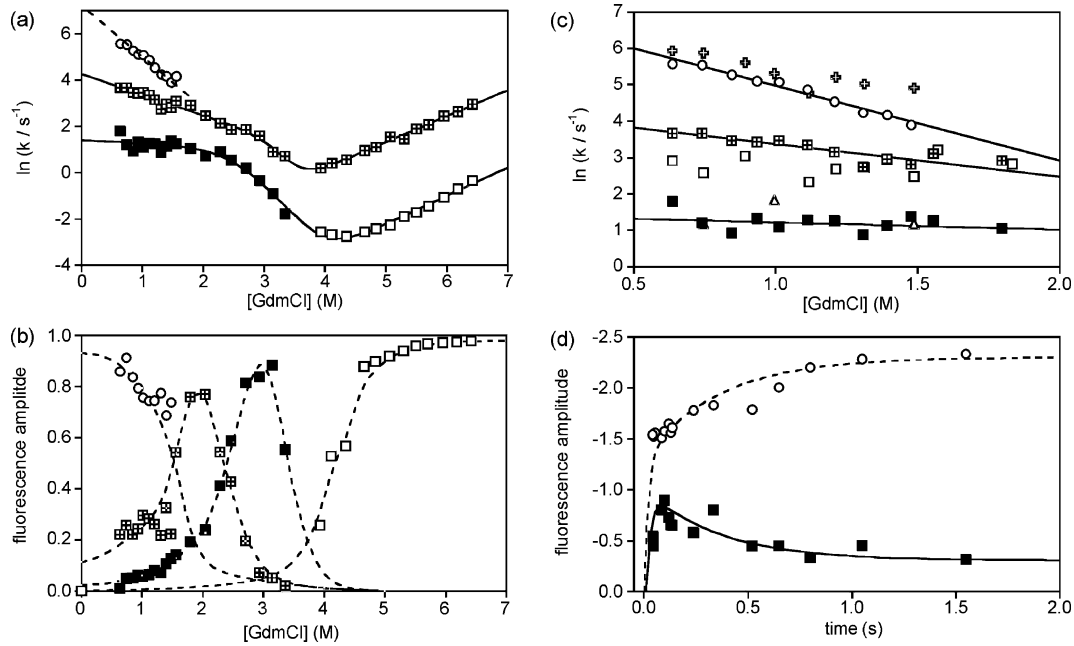


FIGURE 5: Chevron plots constructed from the refolding and unfolding rate profiles for WN at 298 K showing the resolution of three kinetic phases (a), two of which form chevrons with curvature in the refolding data. The fourth slow phase, which is of low amplitude and shows little variation with denaturant concentration, is assigned to an isomerization process and has been excluded from the plot for clarity. The amplitude changes for the three phases, together with that for the major unfolding phase, are represented in part b. Curves plotted through the data are to guide the eye only. The amplitude changes are consistent with the formation and decay of a number of significantly populated species. (c) Dependence of the three major refolding phases of WN on protein concentration over a 10-fold range (1.0 and 0.1 μM). Refolding data at denaturant concentrations below 2 M GdmCl were collected. The data points corresponding to open circles, crossed squares, and black squares correspond exactly with the data shown in part a at 1.0 μM protein concentration. The corresponding data at the lower protein concentration are represented by crosses, open squares and open triangles, respectively. Although the signal-to-noise ratio for the data collected at the lowest protein concentration was low, the same refolding phases and rate constants could be measured, indicating no obvious dependence of rates on possible aggregation phenomena. (d) Interrupted-refolding data for WN showing the variation in amplitude of the two unfolding phases that reflect the rate of formation of the native state in the initial refolding step of the sequential mixing experiment. The amplitude of the faster phase shows rapid formation (>45 s^{-1}) and decay (~3.5 s^{-1}) of an intermediate state (black squares), with the rates in reasonable agreement with the single-jump data corresponding to $k_{UI} \sim 50$ s^{-1} and $k_{IN} \sim 3.3$ s^{-1} (slower of the two chevrons) under similar experimental conditions. The second unfolding phase, which is both slower and higher in amplitude (open circles), fits to a double exponential consisting of a rapid initial increase in amplitude during the dead-time of the experiment and a second slower phase (rate constants of ~30 s^{-1} and 2.8 s^{-1}, respectively) that indicate that molecules reach the N-state via two routes with different rates. These rates are consistent with those observed for the two parallel chevrons in the single-jump experiments (Figure 4a).

introduction of the bulky indole ring resulted in kinetics with additional refolding and unfolding phases that were not apparent in any of the other mutants so far described. The refolding data are consistent with four kinetic phases, the slowest of which is assigned to an isomerization-limited process. Two of the three remaining phases are quite similar to the data for the FN mutant. The observation of three kinetic phases (excluding the very slowest isomerization phase) is consistent with the population of at least four species. However, in contrast to FN, we now also observe two well-resolved unfolding phases in the data for WN. The kinetic data and the associated amplitude changes for each phase as a function of denaturant concentration resulted in the construction of the chevron plots shown in Figure 5.

We also considered the possibility that the introduction of a large hydrophobic residue could lead to complexities in the folding kinetics as a consequence of transient aggregation effects in the unfolded and partially folded states (*35*). We carried out parallel refolding studies at a 10-fold lower protein concentration (0.1 μM), representing the limit of fluorescence sensitivity. Although we could not resolve the slowest low-amplitude isomerization phase, all three faster phases could be identified, albeit with slightly larger scatter in the data (Figure 5c). We concluded that the three refolding phases are not dependent on protein aggregation phenomena.

An overlay of the WN data with the FN chevrons reveals some similarities between the two data sets (Figure 6). For clarity, the slow isomerization phase which shows little variation with denaturant concentration has been omitted. The fastest phase, identified as the U→I transition in the FN data has very similar rate constants and *m*-value in the WN data set, consistent with a similar rapid initial collapse to a compact intermediate. The slower of the two parallel chevrons is also characteristic of the I→N transition in the FN data set with a pronounced rollover effect at low denaturant concentration (<2 M GdmCl). Thus we appear to observe U→I and I→N transitions that are common to both proteins. In addition, the WN data reveal a second faster chevron which also demonstrates pronounced rollover effects, although there is some merging of data points with those for the U→I transition at around 2 M GdmCl (see Figure 5a). An initial analysis in which the chevron plots were fitted independently indicated that the sum of the *m*-values over all three kinetic profiles is significantly higher than described for the VN and FN mutants. This would appear to preclude a simple sequential four-state folding model (Scheme 1a), but suggest one involving either parallel folding routes to

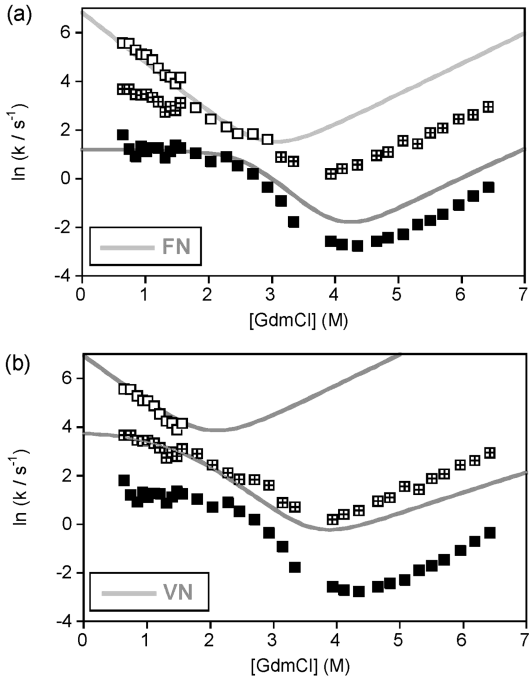
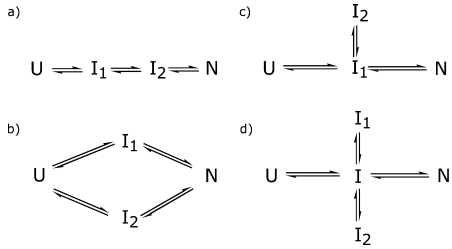


FIGURE 6: Overlay of the data points for WN taken from Figure 4a with the three-state chevron plots for FN (a) and VN (b) shown as gray lines. The fastest phase for WN (open squares) corresponds closely with the U→I chevrons for both FN and VN [upper V-shaped chevron in parts a and b, although there is overlap of data points with the faster of the two chevrons], suggesting a common initial U→I collapse among the family of mutants. The two curved chevrons seen in the WN data (black squares and crossed squares) suggest two parallel folding pathways and are well described by a combination of the I→N chevrons observed in the FN (a) and VN (b) data. In the two plots, the assignment of some data points to the fast phase (open squares) has been changed to emphasize the correlation with the overlaid chevron for the U→I transitions for FN (a) and VN (b).

Scheme 1



the native state or the population of species caught in an off-pathway kinetic traps (Scheme 1b−d). The initially weak denaturant dependence of these two chevrons indicates that they both involve collapsed states with subsequent transitions not involving significant changes in surface area.

We also overlaid the kinetic data for WN with those for VN and observed some similarities. The fastest phase for the initial U→I collapse again concurs with that of the two aromatic mutants; however, the main I→N chevron for VN is much more similar to the faster of the two chevrons of WN (Figure 6b). Thus, the chevrons constructed from the kinetic data for WN appear to be reasonably well represented by a sum of the data sets for the VN and FN mutants, representing the possibility of two parallel folding pathways with distinct intermediates and different rates of production of the native state, reflecting different barrier heights influenced by non-native interactions requiring different degrees of error-repair in the transition state. The intermediate formed on the folding pathway of VN converts to the native state more rapidly (I→N) but also unfolds more rapidly (N→I) than is the case for the FN mutant, indicating that the higher rating limiting barrier for folding and unfolding of the aromatic mutant correlates with the requirement for the destabilization of non-native contacts in the transition state.

We investigated the parallel pathway model in more detail using interrupted refolding sequential mixing experiments. Parallel pathways should result in molecules reaching the N-state with very different rates. The unfolded WN mutant in 6.5 M GdmCl was first refolded into 1.5 M GdmCl and then a second unfolding event induced by jumping into 6.1 M GdmCl after a variable aging time (between 40 ms and 5 s). The final unfolding kinetics revealed two unfolding phases with rate constants of 31.5 ± 13.5 s^{-1} and 0.42 ± 0.01 s^{-1} and amplitudes of 15% and 85%, respectively, of the total fluorescence change. The former is less well defined on account of the lower amplitude of the signal and shows further variation with the aging time of the sequential mixing experiment. The build-up in the amplitudes of these unfolding phases provides a measure of the rate at which molecules are reaching the native state in the first refolding step prior to being unfolded. Thus, a plot of the variation in amplitude with aging time reveals a build-up and decay of the signal associated with the faster of the two phases, which is characteristic of the rapid population of one or more I-states with a slower conversion to the N-state, as clearly observed for the FN mutant in Figure 4c. The increase in the amplitude of the slower, more highly populated phase is initially very rapid with many molecules reaching the native state essentially within the dead-time of the experiment (~40 ms). We were unable to fit the data to a single exponential process but to a double exponential with two rate constants differing by 1 order of magnitude. The faster process is not well defined due to the long dead-time, but we estimate ~30 s^{-1}; however, the slower process has a rate constant of 2.8 s^{-1}. Thus, molecules appear to be reaching the native state via two processes. The corresponding refolding rate constants measured in single-jump experiments under the initial refolding conditions of the sequential mixing experiment (1.5 M GdmCl) suggest rates of interconversion for I→N for these parallel processes of 22 s^{-1} and 3.1 s^{-1}. Thus, the double exponential growth in the amplitude of the faster unfolding phase in the interrupted refolding experiment appears to be consistent with molecules reaching the native state via two parallel routes via distinct intermediate states I_1 and I_2 (Scheme 1b).

It would seem to be a plausible assumption that all of our mutants fold along a pathway which is predetermined by the foldon structure of ubiquitin (*43*). In the case of WN, the chances of a misfolding error that blocks productive folding are compounded by the introduction of the large hydrophobic indole side chain. These error repair-type mechanisms have recently been reported to successfully describe the complex folding kinetics of lysozyme (see Scheme 1c and 1d). The high-pH folding of lysozyme, originally described by Beri and Kiefhaber (*42*), shows four kinetic species which are fitted to a 'diamond' model (Scheme 1b) involving multiple independent parallel routes to the native state, including direct conversion of U to N

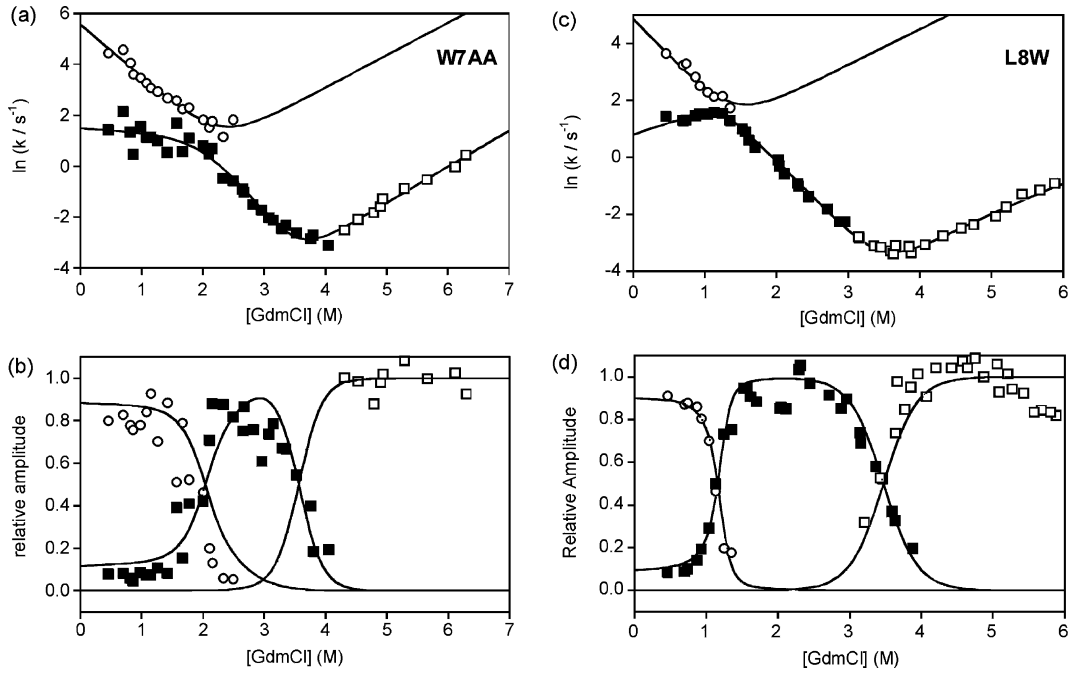


FIGURE 7: Denaturant-dependence of the refolding and unfolding rate constants at 298 K for the WN turn mutant with the T12A/T14A double substitution (WN^{AA}) within the N-terminal β-hairpin sequence (a), and the corresponding amplitude variation of the two major refolding phases and unfolding phase (b). The amplitude of the fastest phase (open circles) shows that the intermediate is highly populated at low denaturant concentrations resulting in an initially flat rate profile for the I→N transition (black squares) in the chevron plot in (a), indicating that the I-state is highly compact. The additional unfolding phase observed for WN (see Figure 4a) is no longer visible. The data fit well to a three-state on-pathway model (see data in Table 1). Global analysis of the rate and amplitude data is shown by the lines of best fit in parts a and b. The plots in parts c and d illustrate similar data for the L8W mutant (as also described in ref *30*). The initial positive slope for the I→N transition at low denaturant concentrations (black squares) shows that the I-state must partially unfold ($m_{IN} < 0$) to reach the transition state indicating a highly compact misfolded state stabilized by non-native interactions. The global fitting of the rate and amplitude data (as shown in parts c and d) is consistent with this model.

(not shown). The alternative predetermined pathway model with optional errors is suggested by Krishna and Englander (*43*) to fit equally well and with fewer kinetic parameters (Scheme 1c and 1d). In Scheme 1d, an on-pathway intermediate (I) generates the two different error-dependent off-pathway kinetic traps I_1 and I_2. The interconversion between I_1 and I_2 results in I not being significantly populated, resulting in four kinetic species whose fractional populations increase and/or decay with folding time. At present, our analysis appears to rule out the sequential four-state model (Scheme 1a), with some evidence to support the possibility of parallel pathways to the native state (Scheme 1b). However, we cannot exclude the possibility of error-dependent off-pathway traps analogous to those described by Krishna and Englander (*43*) (Scheme 1c and 1d). The most attractive model for the folding heterogeneity of WN builds upon a hybrid version of the three-state model described and characterized in detail for the VN and FN mutants where all use the same folding nucleation site encoded within the initial U→I transition (*43*) but populate intermediates of different stability with different numbers of non-native contacts and transition state barriers that reflect repair of these errors.

Point Mutations to Probe the Folding Complexity. The complex folding kinetics and the stability of populated intermediates were further examined by making a number of additional destabilizing substitutions. We probed whether the complex folding kinetics for the WN mutant could be simplified as a result of the destabilization of the N-terminal β-hairpin. We introduced the double mutation T12A/T14A on the solvent exposed β-strand (Figure 1c). In the X-ray structure of ubiquitin (*44*), the side chains of T12 and T14 form contacts with residues (F4 and K6) in the complementary antiparallel β-strand of the N-terminal hairpin (Figure 1c). These substitutions resulted in an overall reduction in protein stability of ~5 kJ mol^{-1} and a simplification of the folding kinetics of the WN mutant such that we were no longer able to resolve two unfolding phases (Figure 7). The variation in amplitude of the two major refolding phases were more consistent with the kinetic data for the FN mutant and fit globally to a three-state on-pathway folding model with similar properties of the I-state to FN (see Table 1). Thus, destabilizing the β-hairpin template of the WN mutant appears to favor the slower folding of the two parallel pathways.

In agreement with our structural model, a significant loss of stability is associated with the L69A substitution in *wt*-Ub, VN, and particularly in FN and WN. Earlier refolding studies with *wt*-Ub using a two-state approximation reported a ϕ-value <0.2 for the L69A substitution (*30*), indicating that the interactions with the C-terminal β-strand of the native β-turn sequence are consolidated late in the folding process and are not significantly structured in the transition state. In the case of VN, we see the same three-state folding kinetics for the L69A mutant as described in Figure 2 but a reduction in refolding rates (U→I and I→N) and acceleration of the unfolding rate (N→I), which suggest that all of the folding barriers are partially perturbed, with estimated ϕ-values in all cases <0.5 (Figure 8a). The initial collapse from U→I and conversion of I→N are decelerated, indicating that long-range contacts with the C-terminal β-strand are

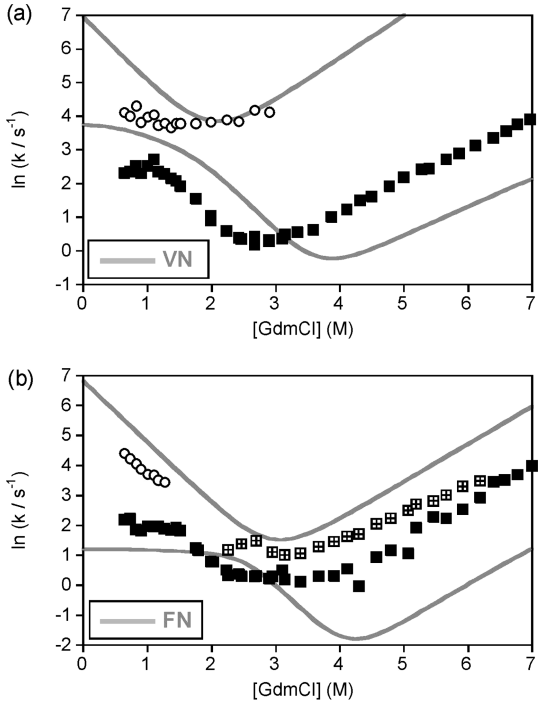


FIGURE 8: Kinetic data for the L69A mutants of VN (a) and FN (b). The rate data are overlaid with the chevron plots for FN and VN shown as gray lines. The truncation of the hydrophobic side chain enhances the rate of unfolding of VN but slows both the U→I and I→N transitions, suggesting contributions of L69 to the global collapse of the polypeptide chain. In the case of L69A-FN, the initial fast U→I phase (open circles) shows the same deceleration; however, the I→N transition appears to have been accelerated by the L69A substitution, indicative that some non-native interactions in the I-state involving the Phe substitution are destabilized by the L69A mutation. A second low amplitude unfolding phase for L69A-FN is also evident.

destabilized, suggesting that global collapse of the polypeptide chain drives the initial barrier crossing event and formation of the I-state. Since we are considering a side-chain truncation, the assumption is that the mutation destabilizes intermediates and transition states relative to the unfolded state through loss of a subset of the native interactions evident in the fully folded state. This would be expected to decelerate the U→I and I→N transitions as native contacts are lost. This appears to be the case with the kinetic data for the L69A-VN mutant.

In contrast, the L69A substitution within FN appears to make the kinetics more complex. We observe significant changes in the rates for the folding transitions but now also resolve two unfolding phases both of which are significantly faster than that of FN (Figure 8b). The effect of the L69A substitution on the rate of initial collapse parallels that observed for VN, suggesting that the U→I transition is driven by the same global collapse and burial of hydrophobic side chains and is decelerated by removal of some specific native contacts. In contrast, the initial I→N transition at low denaturant concentrations is accelerated by the L69A substitution in FN, suggesting that L69A destabilizes non-native interactions in the transition state that would otherwise increase the size of the rate-limiting barrier. The kinetic differences appear to correlate with the Val→Phe substitution within the β-turn. The emergence of two parallel unfolding routes suggests that the partial loss of non-native interactions opens up the possibility of folding through two structurally different intermediates represented by the scenarios described above for VN and FN. Thus, the L69A mutation appears to destabilize the slower of the two possible folding pathways such that a detectable number of molecules now follow the alternative faster route that is not evident for FN but is visible for WN. Given the complexities of the data, only broad conclusions may be drawn that point to non-native aromatic interactions playing an important role in stabilizing different intermediates and in resolving parallel pathways or trapped intermediates.

We also considered whether the complex kinetics observed for the WN mutant could be reproduced in the *wt*-Ub sequence within the context of the native G-bulged type I turn by introducing the L8W mutation in to *wt*-Ub (TLTGK→TWTGK) (*28*). This resulted in a highly fluorescent protein that similarly underwent a large change in fluorescence intensity upon folding. We observed single exponential unfolding kinetics for L8W and double exponential refolding kinetics which we were readily able to fit to a three-state model with a highly populated on-pathway I-state of comparable stability to that observed for the FN and WN mutants. We also observed a pronounced rollover in the refolding data for the I→N transition which has an initially positive slope for m_{IN} in contrast to the denaturant-independent initial rate profile for k_{IN} observed for FN and WN ($m_{IN} \leq 0$) (Figure 7b). This suggested that the intermediate was required to partially unfold to reach the rate-limiting transition state, consistent with a highly compact and misfolded species (*27, 28*). We do not, however, observe the additional refolding and unfolding phases for L8W that are indicative of the more complex kinetics identified for WN. However, the introduction of a bulky hydrophobic aromatic residue within the β-turn irrespective of the turn sequence has the same impact in stabilizing a highly collapsed intermediate on the folding pathway.

Evolutionary Pressures on Amino Sequences of Small Proteins. We examined the frequency of occurrence of aromatic residues in small proteins to provide some insights into the evolutionary pressures at the primary sequence level that facilitate a smooth folding landscape that is not beset by aggregation-prone partially folded structures that may be the consequence of formation of stable non-native contacts during folding. We examined the frequency of occurrence of aromatic residues as a function of protein sequence length for a high resolution set of 650 nonhomologous structures ranging from 33 to 3200 residues (Figure 9). We analyzed the data in groups, representing structures of up to 60 residues, all structures up to 99 residues, sequences of between 100−199, 200−300, 300−500, and 500−3200 residues. The shortest sequences (<100 residues) are most likely to have the smoothest folding energy landscape and indeed appear to have the lowest mean percentage of Phe and Trp residues. While the trend is interesting, the standard deviation from the mean is large in all cases.

DISCUSSION

Evolutionary Pressures on Avoiding Non-Native Interactions in Proteins. An interesting property of small single domain proteins is their general ability to fold rapidly on a

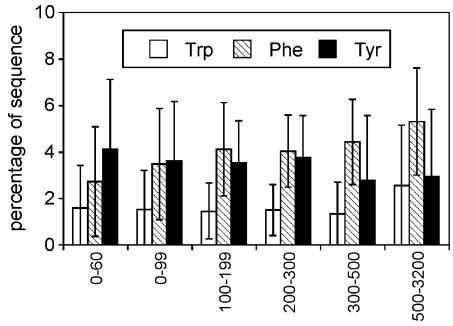


FIGURE 9: Statistical analysis of the occurrence of aromatic residues in proteins of different sequence length (0−60 residues, 0−99, 100−199, 200−300, 300−500, and 500−3200). The percentage of the sequence corresponding to Trp, Phe, and Tyr versus sequence length is shown as a histogram. The variation versus sequence length demonstrates that Phe is disfavored in small proteins with a small hydrophobic core, but is more abundant than Tyr in much larger globular protein structures. Although a general trend is apparent, there are large standard deviations from the mean, represented by the error bars. Data were derived from 650 high-resolution nonhomologous structures. The total number of data points within each group of sequences was in all cases >50.

smooth energy landscape that largely avoids stable compact or misfolded non-native states (*13*−*17, 35*). It has been suggested that this is the consequence of evolutionary pressures at the level of the primary sequence to minimize the necessity for proofreading and correction mechanisms that are more common in protecting larger multidomain proteins from misfolding and aggregation (*13, 35*). The evolutionary disadvantage for proteins that form stable long-lived non-native structures, or have regions of the sequence with high intrinsic secondary structure propensities, is their tendency to nucleate amyloid formation through the aggregation of exposed hydrophobic surface area (*18*−*21*). Intriguingly, molecular chaperones appear to have evolved to suppress the accumulation of misfolded states for larger slow-folding multidomain structures, whereas small single domain proteins appear to have achieved smooth co-operative folding landscapes through evolutionary pressure at the level of the amino acid sequence (*13, 35*). The latter has proved difficult to reproduce in completely non-natural designed proteins. This scenario also argues for the gate-keeping role of some residues that define the specificity of the native state but which may contribute thermodynamic instability to the folded structure through the burial of specific polar side-chain interactions in the hydrophobic core (*35, 45*).

Solvent-exposed protein β-strands provide an excellent template for promoting intermolecular aggregation effects that plagued early studies of model peptide systems designed to form autonomously stable β-sheets (antiparallel hairpins). Charge-rich sequences partially alleviated aggregation phenomena in simple systems (*46*), while negative design features have been identified that inhibit nonspecific aggregation between native proteins across β-sheet interfaces (*47*). Eliminating non-native β-strand interactions through intramolecular association is more problematic because any pair of aligned strands are typically stabilized by the same sort of interactions whether the pairing is native or non-native. In the current context, we have demonstrated that aromatic side chains engineered into the N-terminal β-strand of ubiquitin form stabilizing non-native contacts with the C-terminal strand. These interactions occur between side chains as far apart in the ubiquitin sequence as it is possible to engineer (Phe7/Trp7 ↔ Leu69/Leu71; see Figure 1b). Estimates of contact orders in proteins predict that those that fold fastest have a higher proportion of native contacts between residues that are close together in the primary amino acid sequence (*11, 12, 35*). However, the opposite structural topology appears most likely to lead to a smooth folding landscape that is not beset by formation of partially folded structures stabilized by native-like contacts between residues close together in sequence. In small proteins, the N- and C-terminal parts of the structure are more commonly found in contact than would be expected (*48*), arguing for evolutionary pressures on small protein structures (including *wt*-Ub) to optimize the folding landscape for a co-operative folding process that avoids populated partially folded states that arise from too many local interactions. This may also argue for minimizing the number of aromatic side chains in small proteins to reduce the possibilities of forming these stable non-native contacts. We examined the frequency of occurrence of aromatic residues as a function of protein sequence length for a high resolution set of nonhomologous structures (Figure 9). The data showed this general trend, although the statistical variations were large.

In studies of the unfolded state of lysozyme, extensive clusters of hydrophobic residues appear to persist even under strongly denaturing conditions (*13*). Although the clusters are widely dispersed and use distinctly different parts of the sequence, a single Trp→Gly mutation was shown to simultaneously disrupt all of these clusters, indicating that they are intimately linked and involve extensive long-range interactions. The conclusions suggest that in larger proteins with multiple domains, non-native interactions may be an important determinant of the folding process by burying from solvent, at an early stage, residues that will have a tendency to lead to protein aggregation and the development of amyloid structures. Thus, although the folding of small proteins on a smooth energy landscape may be dominated by the search for native-like contacts and have evolved to avoid the bottle-necks inherent in forming stable non-native interactions, the opposite is suggested to be the case for large slow-folding multidomain proteins to ensure that they reach a functional folded state.

In the current context, engineering aromatic substitutions either within a native β-turn sequence of *wt*-Ub, or an engineered variant with a type I′ β-turn, enhances protein stability through non-native contacts which we were able to identify in the native state by NMR. However, the consequence of these mutations for promoting local non-native interactions as the polypeptide chain folds are less readily identified structurally but are manifested in the kinetic barriers and folding pathways. The effects on stability affect the population of kinetically distinct stable intermediate states that we are able to characterize. We have focused our studies on one small element of structure, namely the β-turn sequence of a key β-hairpin folding nucleation site, and show that the folding landscape of a small protein such as ubiquitin becomes significantly more rugged and complex as non-native interactions are introduced that lead to populated intermediates that frustrate folding (*14, 37, 43, 49*). Destabilizing mutations partially simplify the folding mechanism in some cases but further complicate it in others by favoring one or other putative parallel folding pathways. The initial collapse of the

polypeptide chain (U→I), which is evident in all of the data examined, appears to be an intrinsic property of the sequence rather than specifically linked to the stability of the native state. Evolutionary pressures to avoid stable non-native states in the fast and efficient folding of small single domain proteins suggests that the observed intermediates in our engineered mutants, which are structurally highly compact, are kinetic artifacts of re-engineering the sequence.

REFERENCES

1. Jackson, S. E. (1998) How do small single-domain proteins fold? *Folding Des. 3*, R81–R91.
2. Whittaker, S. B. M., Spence, G. R., Grossmann, J. G., Radford, S. E., and Moore, G. R. (2007) NMR analysis of the conformational properties of the trapped on-pathway folding intermediate of the bacterial immunity protein Im7. *J. Mol. Biol. 366*, 1001–1015.
3. Latypov, R. F., Cheng, H., Roder, N. A., Zhang, J., and Roder, H. (2006) Structural characterization of an equilibrium unfolding intermediate in cytochrome c. *J. Mol. Biol. 357*, 1009–1025.
4. Sanchez, I. E., and Kiefhaber, T. (2003) Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J. Mol. Biol. 325*, 367–376.
5. Maki, K., Cheng, H., Dolgikh, D. A., and Roder, H. (2007) Kinetics of Staphylococcal Nuclease Studied by Tryptophan Engineering and Rapid Mixing Methods. *J. Mol. Biol. 368*, 244–255.
6. Mayor, U., Guydosh, N. R., Johnson, C. M., Grossmann, J. G., Sato, S., Jas, G. S., Freund, S. M. V., Alonso, D. O. V., Daggett, V., and Fersht, A. R. (2003) The complete folding pathway of a protein from nanoseconds to microseconds. *Nature 421*, 863–867.
7. Roder, H., Maki, K., Latypov, R. F., Cheng, H., Ramachandra Shastry, M. C. (2005) Early events in protein folding explored by rapid mixing methods In *Protein Folding Handbook* (Kiefhaber, T., Buchner, J., Eds.) Part I, Wiley-VCH, Weinheim, Germany.
8. Chiti, F., Taddei, N., Webster, P., Hamada, D., Fiaschi, T., Ramponi, G., and Dobson, C. M. (1999) Acceleration of the folding of acylphosphatase by stabilization of local secondary structure. *Nat. Struct. Biol. 6*, 380–387.
9. Gianni, S., Guydosh, N. R., Khan, F., Caldas, T. D., Mayor, U., White, G. W. N., DeMarco, M. L., Daggett, V., and Fersht, A. R. (2003) Unifying features in protein-folding mechanisms. *Proc. Natl. Acad. Sci. U.S.A. 100*, 13286–13291.
10. Gianni, S., Geierhaas, C. D., Calosci, N., Jemth, P., Vuister, G. W., Travaglini-Allocatelli, C., Vendruscolo, M., and Brunori, M. (2007) A PDZ domain recapitulates a unifying mechanism for protein folding. *Proc. Natl. Acad. Sci. U.S.A. 103*, 128–133.
11. Plaxco, K. W., Simons, K. T., and Baker, D. (1998) Contact order, transition state placement and refolding rates of single domain proteins. *J. Mol. Biol. 277*, 985–994.
12. Chiti, F., Taddei, N., White, P. M., Bucciantini, M., Magherini, F., Stefani, M., and Dobson, C. M. (1999) Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. *Nat. Struct. Biol. 6*, 1005–1009.
13. Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L. J., Dobson, C. M., and Schwalbe, H. (2002) Long-range interactions within a non-native protein. *Science 295*, 1719–1722.
14. Krantz, B. A., Mayne, L., Rumbley, J., Englander, S. W., and Sosnick, T. R. (2002) Fast and slow intermediate accumulation and the initial barrier mechanism in protein folding. *J. Mol. Biol. 324*, 359–371.
15. Fersht, A. R. (1995) Optimisation of rates of protein folding: the nucleation-condensation mechanism and its implications. *Proc. Natl. Acad. Sci. U.S.A. 92*, 10869–10873.
16. Sanchez, I. E., and Kiefhaber, T. (2003) Origin of unusual ϕ-values in protein folding: evidence against specific nucleation sites. *J. Mol. Biol. 334*, 1077–1085.
17. Jacob, J., Krantz, B., Dothager, R. S., Thiyagarajan, P., and Sosnick, T. R. (2004) Early collapse is not an obligate step in protein folding. *J. Mol. Biol. 338*, 369–382.
18. Dobson, C. M. (2003) Protein folding and misfolding. *Nature 426*, 884–890.
19. Stefani, M., and Dobson, C. M. (2003) Protein aggregation and aggregation troxicity: new insights into protein folding, misfolding diseases and biological evolution. *J. Mol. Med. 81*, 678–699.
20. Wright, C. F., Teichmann, S. A., Clarke, J., and Dobson, C. M. (2005) The importance of sequence evolution in the aggregation and evolution of proteins. *Nature 438*, 878–881.
21. Scalley-Kim, M., and Baker, D. (2004) Characterisation of the folding energy landscape of computer generated proteins suggests high folding free energy barriers and co-operativity may be consequences of natural selection. *J. Mol. Biol. 338*, 573–583.
22. Krantz, B. A., and Sosnick, T. R. (2000) Distinguishing between two-state and three-state models for ubiquitin folding. *Biochemistry 39*, 11696–11701.
23. Went, H. M., Benitez-Cardoza, C. G., and Jackson, S. E. (2004) Is an intermediate state populated on the folding pathway of ubiquitin? *FEBS Lett. 567*, 333–338.
24. Khorasanizadeh, S., Peters, I. D., and Roder, H. (1996) Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. *Nat. Struct. Biol. 3*, 193–205.
25. Gladwin, B. F., and Evans, P. A. (1996) Structure of very early protein folding intermediates: new insights through a variant of hydrogen exchange labelling. *Folding Des. 1*, 407–417.
26. Crespo, M. D., Simpson, E. R., and Searle, M. S. (2006) Population of on-pathway intermediates in the folding of ubiquitin. *J. Mol. Biol. 360*, 1053–1066.
27. Vallee-Belisle, A., and Michnick, S. W. (2007) Multiple tryptophan probes reveal that ubiquitin folds via a late misfolded intermediate. *J. Mol. Biol. 374*, 791–805.
28. Rea, A. M., Simpson, E. R., Crespo, M. D., and Searle, M. S. (2008) Helix mutations stabilise a late productive intermediate on the folding pathway of ubiquitin. *Biochemistry 47*, 8225–8236.
29. Platt, G. W., Simpson, S. A., Layfield, R., and Searle, M. S. (2003) Stability and folding kinetics of a ubiquitin mutant with a strong propensity for nonnative beta-hairpin conformation in the unfolded state. *Biochemistry 42*, 13762–13771.
30. Went, H. M., and Jackson, S. E. (2004) Ubiquitin folds through a highly polarised transition state. *Protein Eng. 18*, 239–246.
31. Simpson, E. R., Meldrum, J. K., and Searle, M. S. (2006) Engineering diverse changes in β-turn propensities in the N-terminal β-hairpin of ubiquitin reveals significant effects on stability and kinetics but a robust folding transition state. *Biochemistry 45*, 4220–4230.
32. Jourdan, M., and Searle, M. S. (2000) Co-operative assembly of a native-like ubiquitin structure through peptide fragment complexation: Energetics of peptide association and folding. *Biochemistry 39*, 12355–12364.
33. Simpson, E. R., Meldrum, J. K., Bofill, R., Crespo, M. D., Holmes, E., and Searle, M. S. (2005) Engineering enhanced protein strability through β-turn optimisation: insights for the design of stable peptide β-hairpin systems. *Angew. Chem., Int. Ed. 44*, 4939–4944.
34. Griffiths-Jones, S. R., Maynard, A. J., and Searle, M. S. (1999) Dissecting the stability of a β-hairpin peptide that folds in water: NMR and molecular dynamics analysis of the β-turn and β-strand contributions to folding. *J. Mol. Biol. 292*, 1051–1069.
35. Watters, A. L., Deka, P., Corrent, C., Callender, D., Varani, G., Sosnick, T., and Baker, D. (2007) The highly co-operative folding of small naturally occurring proteins is likely the result of natural selection. *Cell 128*, 613–624.
36. Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J., and Laue, E. D. (2005) The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins, 59*, 687–698.
37. Bofill, R., and Searle, M. S. (2005) Engineering stabilising interactions into a conformationally flexible region of the folding transition state of ubiquitin. *J. Mol. Biol. 353*, 373–384.
38. Bachmann, A., Kiefhaber, T. (2005) Kinetic mechanisms in protein folding In *Protein Folding Handbook* (Kiefhaber, T., Buchner, J.) Part I, Wiley-VCH, Weinheim, Germany.
39. Travaglini-Allocatelli, C., Gianni, S., and Brunori, M. (2004) A common folding mechanism in the cytochrome c family. *Trends Biochem. Sci. 29*, 535–541.
40. Sanchez, I. E., Morillas, M., Zobeley, E., Kiefhaber, T., and Glockshuber, R. (2004) Fast folding of the two-domain semliki forest virus capsid protein explains co-translational proteolytic activity. *J. Mol. Biol. 338*, 159–167.
41. Pappenberger, G., Aygun, H., Engels, J. W., Reimer, U., Fischer, G., and Kiefhaber, T. (2001) Nonprolyl cis peptide bonds in unfolded proteins cause complex folding kinetics. *Nat. Struct. Biol. 8*, 452–458.

42. Bieri, O., Wildegger, G., Bachmann, A., Wagner, C., and Kiefhaber, T. (1999) A salt-induced kinetic intermediate is on a new parallel pathway of lysozyme folding. *Biochemistry 38*, 12460–12470.
43. Krishna, M. M. G., and Englander, S. W. (2007) A unified mechanism for protein folding: predetermined pathways with optional errors. *Protein Sci. 16*, 449–464.
44. Vijaykumar, S., Bugg, C. E., and Cook, W. J. (1987) Structure of Ubiquitin Refined at 1.8 A Resolution. *J. Mol. Biol. 194*, 531–544.
45. Bolon, D. N., and Mayo, S. L. (2001) Polar residues in the protein core of Escherichia coli thioredoxin are important for fold specificity. *Biochemistry 40*, 10047–10053.
46. Searle, M. S., and Ciani, B. (2004) Design of β-sheet systems for understanding the thermodynamics and kinetics of protein folding. *Curr. Opin. Struct. Biol. 14*, 458–464.
47. Richardson, J. S., and Richardson, D. C. (2002) Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc. Natl. Acad. Sci. U.S.A. 99*, 2754–2759.
48. Krishna, M. M. G., and Englander, S. W. (2005) The N-terminal to C-terminal motif in protein folding and function. *Proc. Natl. Acad. Sci. U.S.A. 102*, 1053–1058.
49. Capaldi, A. P., Kleanthous, C., and Radford, S. E. (2002) Im7 folding mechanism: misfolding on a path to the native state. *Nat. Struct. Biol. 9*, 209–216.
50. Koradi, R., Billeter, M., and Wuthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structure. *J. Mol. Graph. 14*, 51–55.

BI801330R