SI Appendix

Cloning, Protein Production and Purification. The designed proteins CTPRa proteins were cloned, expressed and purified as previously described (1). The sequence of the A-B helices is AEAWYNLGNAYYKQGDYDEAIEYYQKALELDPRS, with the C-terminating solvating helix being AEAKQNLGNAKQKQG.

Equipment and General Procedures. All equilibrium and kinetic experiments, unless otherwise stated, were carried out at 10 °C in 50 mM phosphate pH 7.0.

Equilibrium experiments - urea and guanidinium chloride denaturations

Preparation of samples: A stock solution of urea (8-10 M) or GdnHCl (8 M) was diluted to obtain a large range of denaturant concentrations using a Hamilton Microlab dispenser; 100 µl of a stock solution of CTPRan protein (ca 18-36 µM) containing 450 mM Phosphate (pH 7.0) was added to each denaturant sample (800 µl). This gave a final buffer concentration of 50 mM phosphate pH 7.0 and a protein concentration of between 1-4 µM .The protein/denaturant solutions were pre-equilibrated at 10 °C for at least one hour.

Spectroscopic Measurements: All measurements were performed in thermostatted cuvette holders at 10 °C ± 0.1 °C and without DTT (the CTPRa proteins do not contain cysteine residues). Fluorescence measurements were made on a Varian Cary Eclipse Fluorescence Spectrophotometer. The excitation wavelength was 280 nm, and band passes for excitation and emission typically 2.5-10 nm. The Fluorescence of the CTPRa proteins was measured at the λ_{max} for the native state (335 nm). Ellipticity at 222 nm was measured using a Jasco J-715 Spectropolarimeter. Protein concentrations were typically 2-3 µM.

Equilibrium data analysis

Two state folding model: Each entire denaturation data set for each CTPRan protein monitored by either fluorescence or CD was fitted to equation (1) using

the non-linear regression analysis program *Kaleidagraph* (version 4.0 Synergy Software, PCS Inc.):

$$\lambda_{obs} = \frac{(\alpha_N + \beta_N[\mathrm{D}]) + ((\alpha_D + \beta_D[\mathrm{D}])\exp((m_{D-N}[\mathrm{D}] - [\mathrm{D}]_{50\%})))/RT}{1 + \exp(m_{D-N}[\mathrm{D}] - [\mathrm{D}]_{50\%})/RT} \quad (1)$$

where λ_{obs} is the observed signal (fluorescence or CD), α_N and α_D are the intercepts, and β_N and β_D are the slopes of the baselines at the low (N) and high (D) denaturant concentrations, $[\mathrm{D}]_{50\%}$ is the midpoint of unfolding, [D] is the concentration of denaturant and m_{D-N} is a constant that is proportional to the increase in degree of exposure of the protein on denaturation.

Equation (1) is based on a two-state model of denaturation where only the native and the denatured states are populated, and assumes that λ_{obs} of the native state, λ_N, and the denatured state, λ_D, are linearly dependent on the denaturant concentration ($\lambda_N = \alpha_N + \beta_N[\mathrm{D}]$, $\lambda_D = \alpha_D + \beta_D[\mathrm{D}]$); for a detailed derivation see (2). Values for $[\mathrm{D}]_{50\%}$ and m_{D-N} are obtained with their standard errors. $\Delta G^{\mathrm{H_2O}}_{\mathrm{D-N}}$, the free energy of unfolding in water, can also be calculated using equation (2):

$$\Delta G^{\mathrm{H_2O}}_{\mathrm{D-N}} = m_{D-N}[\mathrm{D}]_{50\%} \quad (2)$$

Ising model (1, 3, 4): The partition function for a of *N*-helix of CTPRan protein was taken as:

$$Z_N = \exp\{(N+1)J\}\exp(-NH)\left\{\left[(1-g_-)g_+^{N+1} - (1-g_+)g_-^{N+1}\right]/(g_+ - g_-)\right\} \quad (3)$$

where *N* is the number of helices within a CTPRan protein, *J* is the interaction between lattice sites (α-helices), $g_\pm = \exp(H)\left(\cosh H \pm \sqrt{\sinh^2 + \exp(-4J)}\right)$ and H is the internal energy of a single lattice site (each α-helix).

GraphPad Prism (using a Marquardt algorithm non-linear least-mean-squares fitting routine) was used to globally fit each series of CTPRan proteins equilibrium data simultaneously (either CD or Fluorescence). This was achieved by numerically calculating the magnetization, $m = d\log Z/dH$, and thus the fraction folded, $f = (1+m)/2$, as a function of *J* and $H = \frac{1}{2} m_1([\mathrm{D}] - [\mathrm{D}]_C)$. Where m_1 is denaturant dependence of a single α-helix in the protein, [D] is the concentration of denaturant and $[\mathrm{D}]_c$ is the value at which $H = 0$. In order to numerically calculate the magnetization a step size of 0.001 M was used. The global fitting produced values for *J*, m_1 and $[\mathrm{D}]_c$ that were the same for all CTPRan proteins.

The equilibrium curves could also be globally fitted, in the same manner as above, without converting the observed signal to fraction folded by using equation 4:

$$\lambda_{obs} = (\alpha_N + \beta_N[\mathrm{D}])f + (\alpha_N + \beta_N[\mathrm{D}])(1-f) \quad (4)$$

where λ_{obs} is the observed signal (fluorescence or CD), α_N and α_D are the intercepts, and β_N and β_D are the slopes of the baselines at the low (N) and high (D) denaturant concentrations, [D] is the concentration of denaturant and $f = f(H,J)$ is the fraction of the protein that is folded according to the Ising Model. Here, *J*, m_1 and $[\mathrm{D}]_c$ were globally fitted to obtain values that were the same for all CTPRan proteins. Whereas, α_N, α_D, β_N and β_D were not globally fit, but were specific for each CTPRan protein's equilibrium unfolding curve. Equation (4) assumes that λ_{obs} of each proteins native state, λ_N, and each proteins denatured state, λ_D, are linearly dependent on the denaturant concentration ($\lambda_N = \alpha_N + \beta_N[\mathrm{D}]$, $\lambda_D = \alpha_D + \beta_D[\mathrm{D}]$).

Free energies for folding were then calculated using equation (5) (3):

$$\Delta G_{0\to \mathrm{j}} = -nRTH + 4RTJ \quad (5)$$

where, $\Delta G_{0\to j}$ is the free energy of folding in water for a CTPRa protein with j α-helices, *n* is the number of α-helices in each protein, *R* is the gas constant, *T* is 283 K, *H* is the internal energy of a single lattice site (α-helices) and is calculated from $H = \frac{1}{2} m_1([D] - [D]_C)$ and *J* is the interaction between the lattice site (α-helices).

Kinetic Experiments

Kinetic experiments were performed on an Applied Photophysics π*-180 Stopped-Flow Spectrometer. Refolding and Unfolding were monitored by fluorescence using an excitation wavelength of 280 nm and monitoring the emission through a cut-off filter of 300 nm. All experiments were carried out at 10 °C ± 0.1 °C and without DTT (the CTPRa proteins do not contain cysteine residues).

Unfolding Studies: Unfolding was performed by [denaturant]-jump experiments in the following manner. Unfolding was initiated by diluting 1 volume of an aqueous protein solution (approx. 22 μM CTPRa protein in 50 mM Phosphate, pH 7.0) into 10 volumes of concentrated GuHCl (containing 50 mM Phosphate, pH 7.0) such that the final concentrations of GuHCl were between 1.5 and 5.5 M (depending on the protein being unfolded).

Refolding Studies: Refolding was initiated by either [denaturant]- or pH-jump experiments (2). pH-jumps were employed to measure the rate of folding in the absence of denaturant. Each type of experiment produced the same rate constants when refolding into identical denaturant concentrations, confirming that both methods were probing the same folding pathway.

[denaturant]-jump: CTPRa proteins (approx. 22 μM) were initially unfolded in 6.0 M GuHCl that contained 50 mM Phosphate, pH 7.0. Refolding was then initiated by rapid mixing of 1 volume of denatured protein into 10 volumes of a refolding buffer containing 50 mM Phosphate, pH 7.0, and appropriate amounts of denaturant to give final concentrations between 0.54 – 3.2 M GuHCl (depending on which CTPRa protein was under investigation).

pH-jump (alkali): CTPRa proteins (approx. 4 μM) were initially denatured by changing the pH to 12.4 (by addition of NaOH to give a final concentration of 25 mM NaOH). Each protein was then refolded by rapid mixing (1:1) with a refolding buffer that contained 100 mM Phosphate, pH 6.42 to produce a final solution of 50 mM Phosphate, pH 7. The pH- jump experiments were performed in the absence and presence of low concentrations of GuHCl (0 M to 1 M).

Protein Concentration Dependence: To test for protein concentration dependence of refolding rates, refolding kinetics at GuHCl concentrations of 0 M (using alkali jumps), 0.54 M and 1 M (using [denaturant] jumps) were performed at protein concentrations of between 0.1 – 10 μM.

Confirmation of a populated intermediate state by 1-anilinonaphthalene-8-sulfonate binding: Refolding experiments using [denaturant]-jump were performed at two GuHCl concentrations for each CTPRa protein – (i) at 0.54 M, which was fully within the rollover section of each CTPRa protein and (ii) at a concentration of GuHCl that was outside the rollover section of each CTPRa protein. In each case 11 to 22 μM of CTPRa protein was initially unfolded in 6.0 M GuHCl that contained 50 mM Phosphate, pH 7.0. Refolding was then initiated by rapid mixing of 1 volume of denatured protein into 10 volumes of a refolding buffer containing 50 mM Phosphate, pH 7.0, 200 μM ANS, and appropriate amounts of denaturant to give the final GuHCl concentration required. The experiments were performed on an Applied Photophysics π*-180 Stopped-Flow Spectrometer an excitation wavelength of 370 nm and monitoring the emission through a cut-off filter of 420 nm. All experiments were carried out at 10 °C ± 0.1 °C. Control experiments were performed that showed that ANS did not bind to the native or denatured CTPRa proteins. Thus the change is due to binding to the Intermediate, followed by release upon folding of the protein.

Kinetic Data Fitting

Monitored by intrinsic fluorescence: At each denaturant concentration the rate constant (k_{obs}) was obtained by averaging at least five individual experiments. Both the unfolding and folding kinetics of each CTPRa protein

were found to be monophasic and the data fitted well to a single-exponential process. No slow, proline isomerisation phases were observed in the refolding experiments over a 200 s timescale. It is quite possible that slower phases exist, but are difficult to detect due to instrumental drift.

Monitored by ANS fluorescence: At each denaturant concentration the rate constant (k_{obs}) was obtained by averaging at least fifteen individual experiments. Folding kinetics were found to be biphasic and the data fitted well to a double exponential process. Although the faster phase's amplitude was dependent on protein concentration, the second slower phase's amplitude was not. Further, although the faster phase changed with each CTPRa protein the slower phase did not. This showed that the slower phase was caused by photolysis.

Kinetic Data Analysis

The natural logarithm of the observed rate constants measured as a function of GuHCl concentration for each CTPRa protein were plotted as chevrons graphs. The dependence of ln k_{obs} on [denaturant] for each CTPRa protein was fitted to either a two-state model with a linear dependence (Equation 6) or non-linear dependence (Equation 7) on denaturant:

$$\ln k_{obs} = \ln\left\{k_F^{H_2O}\exp\left(-m_{‡\text{-}F}[D]\right) + k_U^{H_2O}\exp\left(-m_{‡\text{-}U}[D]\right)\right\} \quad (6)$$

$$\ln k_{obs} = \ln\left\{k_F^{H_2O}\exp\left(-m_{‡\text{-}F}[D] + m^*_{‡\text{-}F}[D]^2\right) + k_U^{H_2O}\exp\left(m_{‡\text{-}U}[D] + m^*_{‡\text{-}U}[D]^2\right)\right\} \quad (7)$$

or a sequential three state model where an intermediate is on pathway (Scheme 1). In the sequential three state model, all four microscopic rate constants are defined by the solution of a quadratic equation according to Equations 8 to 10 and assuming a linear dependence on [denaturant]:

$$k_{obs} = \frac{1}{2}\left\{-A_1 \pm \sqrt{A_1^2 - 4A_2}\right\} \quad (8)$$

Where,

$$A_1 = -\left(k_{DI} + k_{ID} + k_{IN} + k_{NI}\right) \quad .. \quad (9)$$

$$A_2 = k_{DI}\left(k_{IN} + k_{NI}\right) + k_{ID}k_{NI} \quad (10)$$

In equations 6 to 10: D or F, I and N or F represent the denatured, intermediate and native states (respectively), $k_F^{H_2O}$ and $k_U^{H_2O}$ are the rate constants of folding and unfolding in water (respectively), $m_{x\text{-}y}$ are constants that can be related to the solvent-accessible area in the activation process of un/folding and k_{xy} is the microscopic rate constant for the conversion of *x* to *y*.

Finally, the refolding arms of CTPRa8 and 10 chevrons were fitted to a minimal dead-end scheme where a compact off pathway intermediate species I equilibrates with the denatured state (Scheme 2 and Equation 11). As is common practice, we assume rapid equilibrium between the intermediate I and the denatured state D to give:

$$k_{obs} = f_D k_F = \frac{D}{1+D} k_F = \frac{k_F}{1+K_I}, .. \qquad (11)$$

where k_{obs} is the measured rate constant, K_I = [I]/[D], and the k_F is the refolding rate constant for the D → N step. This gives the following Equation 12:

$$\ln k_{obs} = \ln \frac{\exp\left\{\ln k_F^{H_2 0} + m_F[D]\right\}}{1+\exp\left\{\ln K_I^{H_2 0} + m_I[D]\right\}} \qquad (12)$$

where m_I and m_F are the linear dependencies of ln K_I and ln k_F on [GuHCl].

References

1. Kajander T, Cortajarena AL, Main ERG, Mochrie SGJ, & Regan L (2005) A new folding paradigm for repeat proteins. *J Am Chem Soc* 127(29):10188-10190.
2. Jackson SE & Fersht AR (1991) Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* 30(43):10428-10435.
3. Kloss E, Courtemanche N, & Barrick D (2008) Repeat-protein folding: New insights into origins of cooperativity, stability, and topology. *Arch Biochem Biophys* 469(1):83-99.
4. Aksel T & Barrick D (2009) Analysis of Repeat-Protein Folding Using Nearest-Neighbor Statistical Mechanical Models. *Methods in Enzymology: Biothermodynamics,Vol 455, Part A* 455:95-125.