The N-Terminal Domain of the Ribosomal Protein L9 Folds Through a Diffuse and Delocalized Transition State

Satoshi Sato[1, +], Jae-Hyun Cho[2, +], Ivan Peran[3], Rengin G. Soydaner-Azeloglu[4], and Daniel P. Raleigh[3, 5, 6, *]

[(1)] Nomadic Bioscience Co., Ltd.
Okayama Research Park Incubation Center
5303 Haga Kita-ku Okayama City Okayama 701-1221, Japan

[(2)] Department of Biochemistry and Biophysics
Texas A&M University
300 Olsen Boulevard
College Station, TX 77843-2128, USA

[(3)] Department of Chemistry,
Stony Brook University
Stony Brook, NY 11794-3400, USA

[(4)] Department of Biochemistry and Molecular Pharmacology
New York University School of Medicine
522 First Avenue, New York, New York 10016, USA

[(5)] Graduate Program in Biochemistry and Structural Biology,
Stony Brook University
Stony Brook, NY 11794, USA

[(6)] Laufer Center for Physical and Quantitative Biology
Stony Brook University
Stony Brook, NY 11794, USA

(+) These authors contributed jointly to this work

Author to whom correspondence should be addressed: daniel.raleigh@stonybrook.edu; 631 632 9547

Running Title: Folding of NTL9

Key Words: Protein Folding, Protein Stability, Phi-value, Unfolded State, Ribosomal Protein L9

Abstract: The N-terminal domain of L9, (NTL9), is a 56 residue mixed α–β protein that lacks disulfides, does not bind cofactors and folds reversibly. NTL9 has been widely used as model system for studies of protein folding and for investigations of the unfolded state. The role of sidechain interactions in the folding of NTL9 is probed by mutational analysis. ϕ-values, the ratio of the change in the log of the folding rate upon mutation to the change in the log of the equilibrium constant for folding, are reported for 25 point mutations and 15 double mutants. All ϕ-values are small with an average over all sites probed of only 0.19, with a largest value of 0.4. The effect of modulating unfolded state interactions is studied by measuring ϕ-values in second site mutants and under solvent conditions that perturb unfolded state energetics. Neither of these alterations affects the distribution of ϕ-values. The results, combined with earlier studies that probe the role of hydrogen bond formation in folding and the burial of surface area [Sato, S., and Raleigh, D. P. (2007), *J. Mol. Biol. 370*, 349-355] reveal that the transition state for folding contains extensive backbone structure, buries a significant fraction of hydrophobic surface area, but lacks well developed sidechain-sidechain interactions. The folding transition state for NTL9 does not contain a specific "nucleus" consisting of a few key residues; rather it involves extensive backbone hydrogen bonding and partially formed structure delocalized over almost the entire domain.

Introduction

The ribosomal protein L9 consists of two globular domains connected by a long α-helix.[1] The N-terminal domain (NTL9) contains one of the simplest examples of a common super secondary structural motif known as the split β–α–β motif (Figure-1). It lacks disulfides, does not bind metals or co-factors, and folds in a reversible two-state fashion over a wide range of conditions. The domain has been widely used as a model system for experimental and computational studies of protein folding, and for studies of the properties of the unfolded state ensemble, but comparably little is known about the development of sidechain-sidechain interactions during its folding.[2-29] The effect of mutating six hydrophobic core residues has been examined and the role of four surface sites has been probed, but the role of other positions in the folding of the domain is unknown.[6, 7] The unfolded ensemble of NTL9 populated under conditions which favors folding contains electrostatic interactions, hydrophobic clusters involving both native and nonnative interactions, and significant residual helical structure, in at least the C-terminal α-helix.[6, 8, 9, 27-29] Some of these interactions are energetically significant and the effects of modulating them on the transition state for folding is not understood.[6, 9, 27, 29]

The role of sidechains in protein folding is traditionally probed using mutational analysis, the so-called "phi-value" (ϕ-value) approach.[30-32] ϕ–values are the ratio of the change in the log of the folding rate to the change in the log of the equilibrium constant upon mutation and are generally thought to provide a picture of the development of sidechain interactions during folding, although some have questioned their information content.[33, 34] A ϕ-value of one indicates that the effect of the mutation upon the free energy difference between the unfolded ensemble and the transition state for folding is equal to the effect upon equilibrium stability and is interpreted to mean that the interaction being probed is fully developed in the transition state. Conversely, a ϕ-value of zero

indicates that the mutation has an effect upon equilibrium stability, but not upon the folding rate, and is interpreted to indicate that the interaction under investigation is no more developed in the transition state for folding than it is in the unfolded ensemble. Fractional ϕ-values are interpreted to represent interactions that are partially formed in the folding transition state.[32] The interpretation of ϕ-values is more complicated if the mutation impacts the energetics of the unfolded ensemble, or if nonnative interactions are being probed or if folding involves parallel microscopic processes.[35, 36] For example, a small ϕ–value can result when a native interaction is only weakly formed in the transition state for folding or when a nonnative interaction is present in the transition state, but is only weakly perturbed relative to the unfolded state.[9] Nevertheless, ϕ-values can provide important information about the transition state for folding and offer experimental tests to benchmark simulations or to use as constraints in restrained molecular dynamics simulations which seek to provide a structural interpretation of the folding transition state.

The folding of NTL9 has previously been examined as a function of temperature, pH and added denaturant. What is missing is a detailed picture of the development of sidechain interactions in the folding transition state and an understanding of how altering unfolded state energetics impacts these interactions. Here we report the results of an extensive mutational analysis of the folding of NTL9. In total 25 single mutants and 15 double mutants were examined. This data, in combination with prior studies, provides a detailed picture of the transition state for the folding of the domain.

Results

Choice of mutations.

The NTL9 fold is comprised of a three stranded antiparallel β-sheet sandwiched between two α-helices. Each helix packs against one face of the β-sheet (Figure-1). The C-terminal helix in the domain is part of the long helical segment which links the N and C-terminal domains of the intact protein and has been shown to adopt partial helical structure in the unfolded state.[8, 37] The folded domain contains two Gly residues which adopt positive values of the backbone dihedral angle ϕ, a conformation which is energetically unfavorable for an L-amino acid. NTL9 is a basic protein, consistent with its function as a ribosomal protein, containing 12 Lys or Arg residues and six Asp/Glu residues (Figure-1). All of the Lys and Arg residues are on the surface of the protein and the charged portion of their sidechains are exposed to solvent, and none are involved in salt bridges. There is one native state salt bridge involving the N-terminus and the sidechain of Asp-23.[3] Prior studies have identified nonnative electrostatic interactions in the unfolded state which involve K12.[6, 27] Paramagnetic relaxation enhancement (PRE) NMR studies of a significantly destabilized double mutant, V3A I4A, have revealed the presence of long range contacts in the unfolded state under conditions where the wildtype domain is folded and studies of protein fragments have shown that the C-terminal helix of NTL9 partially folds in isolation and is populated in the unfolded ensemble.[8, 37]

Prior work examined the effect of mutating six sites in the hydrophobic core of NTL9; M1, V3, I4, L6, V21, and L30. ϕ-values were measured for a range of substitutions at each site, including variants prepared using non-genetically coded amino acids, and in all cases the observed ϕ-values were 0.4 or smaller. In the present study we focused on first 50 residues of the domain since the C-terminal portion of the C-terminal helix is known to be frayed in the native state in solution and analysis of C-terminal truncation mutants has shown that helical structure in this region is no more developed in the transition state for folding than it is in the unfolded state.[38] We

targeted 21 additional sites scattered throughout the structure; in total 25 single mutants and 15 double mutants were examined, including mutations deliberately designed to modulate unfolded state energetics. The sites probed include an additional 14 hydrophobic residues, the two glycines with positive ϕ-angles and 4 charged surface sites as well as the N-capping residue which helps to stabilize the C-terminal α-helix. The present study, in combination with earlier work, probes all of the hydrophobic residues in NTL9 and all of the residues whose sidechains are more than 75% buried in the native state (Figure-1). Hydrophobic residues were replaced by smaller hydrophobic residues. Acid residues were replaced by the corresponding amides, Gln or Asn and by Ala. Lysine residues were replaced by Met as this is the closest isosteric replacement for Lys which removes the polar ε-amino group. More drastic mutants were made if the Lys to Met replacements did not generate a large enough change in stability to allow reliable ϕ-value analysis. The C-terminal α-helix contains a classic N-capping motif where Thr-40 acts as the N-capping residue. Thr-40 was replaced by a Ser, which possesses the ability to act as an N-cap hydrogen bond acceptor and by Ala. There are no conservative substitutions for glycine residues with positive values of the backbone dihedral angle ϕ amongst the 20 genetically coded amino acids. In the present case we used an Ala replacement as Ala contains the smallest sidechain. A set of 15 double mutants were prepared which involved mutation of hydrophobic core residues in a K12M background. The K12M mutant reduces unfolded state interactions, thus the double mutants allow us to evaluate the effects of altering unfolded state energetics upon ϕ-values. This has proven very difficult to do in other studies since the role of specific residues in unfolded state interactions is difficult to define under conditions which favor the native state.

All of the mutants exhibit two state folding.

The stability of the various mutants was determined by CD monitored urea or guanidine HCl denaturation. Folding and unfolding rates were determined by measuring first order relaxation rates as a function of final denaturant concentration and fitting to the standard expressions for two-state folding. Plots of the natural log of the first order rate constant vs. denaturant concentration, so-called "chevron plots", displayed the characteristic V-shape expected for two-state folding and no evidence of deviation from linearity in the plot (rollover) was observed at low denaturant concentration (supporting figure-1). The stabilities calculated from the kinetic measurements are in good agreement with stabilities measured in equilibrium denaturation experiments. The data is summarized in table-1.

Specific side chain interactions are not well developed in the transition state for folding.

ϕ-values determined for mutations which exhibit a small change in stability can be unreliable since the uncertainty in $\Delta\Delta G^o$ usually becomes larger in percentage terms for small differences in folding free energy.[39, 40] In the present case, we restricted ourselves to mutations that lead to a value of $\Delta\Delta G^o = 0.5$ kcal mol^{-1}, or larger. Table-1 includes the measured folding and unfolding rates and calculated ϕ-values. A few of the mutants were too stable to accurately measure the unfolding branch of the chevron plot and for those k_u could not be reliably determined. Data reported in earlier studies is included for completeness.

The ϕ-values for the hydrophobic core mutants are small, the largest value among the hydrophobic site mutants is 0.4, while the average value for this subset of mutations is 0.19. The ϕ-values were also small for the charged sidechains and polar residue (Thr-40) with a largest value of 0.27 and an average of 0.17. It is more difficult to probe the role of a glycine in folding since introduction of a sidechain may lead to new interactions in the folded state or introduce native state backbone stain, and since a Gly to X replacement will alter the energetics of the unfolded state by

reducing its configurational entropy. The least perturbing substitution is an Ala. We prepared Gly to Ala mutants for G24 and G34. Both of these sites adopt conformations in the folded state with a positive backbone dihedral angle ϕ. The substitution at position 24 had no effect on the folding rate, while replacement of G34 with Ala lead to an approximately twofold slower folding rate and a modest effect upon stability. The data indicates that Gly residues are not required at these sites, and, within the limitations outlined above, suggest that development of interactions at these sites is not required in the transition state for folding.

Several of the Lys mutants, K12M, K14M, and K15M, in the Lys rich loop lead to variants which fold faster. The origin of this effect is not clear for all of the sites, but the effect has been investigated in detail for postion-12 and it has been shown that alteration of unfolded state energetics plays a significant role.[6, 9, 27, 28] The enhancement in folding rates for the K14M and K15M mutants are more modest than for the K12M mutant. A histogram of the measured ϕ-values, including previously reported values, is displayed in figure-2.

ϕ–values are a dimension-less ratio and most attention is focused on the interpretation of the subset, if any, with large values. However, this can sometimes be misleading because calculation of the ϕ–value involves dividing by the change in equilibrium unfolding free energy, $\Delta\Delta G^{o}$, thus a large ϕ-value could result from a small energetic perturbation and even though ϕ is large, the net contribution that the site makes to stabilizing the transition state could be small. Consequently we also examined the effect on the natural log of k_f by plotting RT $\ln(k_f^{wildtype} / k_f^{mutant})$ (Figure-3). The histogram largely parallels the plot of ϕ-values vs. sequence.

The effect of modulating unfolded state energetics on NTL9 ϕ-values.

The ϕ-value is formally defined as

$$\phi = \frac{\Delta\Delta G^{\circ\ddagger}}{\Delta\Delta G^{\circ}} \quad (1)$$

Where $\Delta\Delta G^{\circ\ddagger}$ is the difference in stability of the transition state upon mutation and $\Delta\Delta G^{\circ}$ is the difference in equilibrium stability. Equation-1 can be written as,

$$\phi = \frac{\left(G^{\circ\ddagger}_{wt} - G^{\circ unfolded}_{wt}\right) - \left(G^{\circ\ddagger}_{mutant} - G^{\circ unfolded}_{mutant}\right)}{\left(G^{\circ folded}_{wt} - G^{\circ unfolded}_{wt}\right) - \left(G^{\circ folded}_{mutant} - G^{\circ unfolded}_{mutant}\right)} \quad (2)$$

$G^{\circ\, folded}$ is the free energy of the folded state, $G^{\circ\, unfolded}$ is the free energy of the unfolded state, and $G^{\circ\ddagger}$ is the free energy of the transition state. The ϕ-value has a simple quasi-structural interpretation if the mutation does not alter the energetics of the unfolded state. In this case equation-2 reduces to equation-3.

$$\phi = \frac{\left(G^{\circ\ddagger}_{wt} - G^{\circ\ddagger}_{mutant}\right)}{\left(G^{\circ folded}_{wt} - G^{\circ folded}_{mutant}\right)} \quad (3)$$

The parameter is just the ratio of the change in transition state free energy relative to the change in the free energy of the native state and the ϕ-value represents the extent to which the interaction being probe is formed in the transition state relative to its full development in the native state. The interpretation is less straightforward if the mutation alters the energetics of the unfolded state.[35]

The unfolded state interactions which are most relevant for folding are those that are formed in the unfolded state that is populated under folding conditions, i.e. in buffer where the native state is stable. Normally the free energy balance of folding and high cooperativity of the folding process make it extremely difficult to define unfolded state interactions under these conditions. However, prior work has shown that there are energetically significant interactions in the unfolded state of NTL9 under native conditions and has demonstrated that they can be modulated by mutation and by the addition of salt.[27-29] Consequently, we re-examined a subset of 9 key hydrophobic core residues in high salt. The stability and folding kinetics were determined

in 750 mM NaCl and the ϕ-values calculated (Table-2). The ϕ-values are slightly smaller, but the overall pattern is persevered. The fact that the ϕ-values are all still small argues that the small ϕ-values observed under "normal conditions" are not an artifact caused by the presence of significant unfolded state interactions and can modulate hydrophobic interactions by the Hofmeister effect. High NaCl is not a selective perturbant since it also modulates native state electrostatic interactions. Thus, we used a second, independent, method to alter unfolded state interactions. We have previously demonstrated that K12 participates in unfolded state interactions and these can be significantly weakened by a K12M mutation.[27, 28] We exploited this effect to reexamine the NTL9 ϕ-values in a mutant background in which wildtype unfolded state interactions are modulated. We measured the stability and refolding for 15 hydrophobic core mutants in the K12M background. The ϕ-values measured in the K12M background are all small and are well correlated with the values measured in the wildtype NTL9 background (Table-3, Figure-4), providing further evidence that the modest ϕ-values observed in the wildtype background are not due to just to modulation of unfolded state interactions.

Discussion

The data presented here, combined with earlier, more limited studies, shows that the ϕ-values for NTL9 are all small. There is no evidence for the formation of a nucleus involving a few key residues, rather sidechain-sidechain interactions are weakly developed in the transition state for folding. Combining the present mutational analysis with more global probes of folding provides a unified view of the folding of this domain.

The transition state for the folding of NTL9 lies closer to the native state than to the unfolded state as judged by classic order parameters such as the Tanford β-parameter, β_T, which

is a rate-equilibrium free energy relationship that utilizes the denaturant dependence of the log of the folding rate and the log of the equilibrium constant to place the transition state. The closer the value is to one the closer the transition state is to the native state. The parameter is between 0.65 and 0.7 for NTL9, a value which is similar to that measured for a number of other single domain proteins.[2] The mutations studied here do not significantly change the value of β_T (supporting figure-2). Temperature dependent studies of folding rates and folding equilibria can be combined to report on the burial of hydrophobic surface area in the transition state for folding via a comparison of the apparent heat capacity change between the unfolded ensemble and the transition state relative to the change in equilibrium heat capacity. Such studies indicate that approximately 2/3 of the hydrophobic surface area buried during folding is achieved in the transition state for folding of NTL9.[2] Conversely, pH dependent studies have shown that native electrostatic interactions involving surface residues are weakly developed in the transition state for folding of NTL9, a view which is confirmed by analysis of the effect of mutating the acidic residues on the folding rate.[4, 6] Isotope effect experiments that utilized proteins with deuterated and protonated backbone amides have shown that that significant hydrogen bonded structure is formed in the NTL9 folding transition state.[5] Peptide fragment studies and studies of a destabilized mutant have revealed that there is considerable 2^o structure in the unfolded state of the domain, particularly in the C-terminal helix.[8, 37] However, studies with helix truncation mutants reveal that destabilizing the C-terminal helix does not impact the folding rate, indicating that the helix is significantly formed in the unfolded state and the transition state for folding, but consolidation of helical structure does not take place until after crossing the dominant free energy barrier for folding.[38] The small ϕ-value associated with mutation of the N-capping residue, Thr-40, is consistent with this hypothesis.

The global picture which emerges from these studies is that the protein folds from an unfolded ensemble which contains considerable structure via a transition state that is relatively "native like" in the sense that there is significant burial of surface area and extensive hydrogen bonded structure, but which has weakly developed surface interactions. The mutational analysis presented here shows that specific sidechain interactions are not developed. Thus the transition state for the folding of NTL9 bears significant similarity to the classic molten globule state that has been studied at equilibrium. The molten globule state is traditionally defined as being compact, having a large proportion of its hydrophobic surface area buried, containing native like secondary structure and a largely native-like topology, but lacking specific sidechain-sidechain tertiary interactions. NTL9 is a fairly small domain and small proteins usually do not populate partially folded states at equilibrium such as the molten globule state. In the case of NTL9 a molten globule like state appears be a saddle point on the free energy landscape, i.e. the transition state for folding rather than a minima. The small size of NTL9 likely makes it difficult to bury enough surface area to compensate for the desolvation of the protein backbone and lower the free energy of the state enough that it can be populated at equilibrium.[41] The transition state for NTL9 resembles the dry molten globule transition state ensemble.[41, 42] The picture which emerges of NTL9's transition state may be fairly general for small single domain proteins and is consistent with the observation that the average ϕ-value over a large number of proteins is low and with the well-known dependence of folding rate on contact order, and with studies that have shown significant hydrogen bonded structure in the transition state.[34, 43-45]

Materials and Methods

The N-terminal domain of L9 was expressed and purified as previously described.[46] CD-monitored denaturation experiments were performed on an AVIV 62A DS spectrophotometer and an Applied Photophysics Chirascan spectrophotometer both equipped with a temperature control unit as previously described.[7] The concentration of denaturant was determined by measuring the refractive index. Denaturation curves were fit to equation-4,

$$\theta_{222} = \frac{a_n + b_n[denaturant] + (a_d + b_d[denaturant])exp(-\frac{\Delta G°([denaturant])}{RT})}{1 + exp(-\frac{\Delta G°([denaturant])}{RT})} \quad (4)$$

where θ_{222} is the ellipticity in millidegrees, T is the temperature in Kelvin, and R is the gas constant in kcal mol^{-1} K^{-1}. a_n is the intercept and b_n is the slope of the curve in the pre-transition region. a_d is the intercept and b_d is the slope of the curve in the post-transition region. $\Delta G°([denaturant])$ is the change in free energy upon unfolding in kcal mol^{-1} at a particular concentration of denaturant, and is given by equation-5.

$$\Delta G°([denaturant]) = \Delta G°(H_2O) - m[denaturant] \quad (5)$$

$\Delta G°(H_2O)$ is the change in free energy upon unfolding in the absence of denaturant and m is a parameter that is related to the change in solvent accessible surface area upon unfolding with units of kcal mol^{-1} M^{-1}.

Fluorescence detected stopped-flow experiments were performed on an Applied Biophysics SX18MV stopped-flow instrument. The fluorescence signal of the single tyrosine at residue 25 was followed. Folding and unfolding reactions were initiated by diluting the protein sample 11-fold. The trace for each folding or unfolding experiment was fit to a single-exponential function to determine the observed rate constant, k_{obs}, at different concentrations of denaturant. The natural log of k_{obs} vs. denaturant concentration was plotted and fit to equation-6,

$$\ln(k_{obs}) = \ln\left(k_f^{H_2O} exp(-m_f[denaturant]) + k_u^{H_2O} exp(m_u[denaturant])\right) \quad (6)$$

where $k_f^{H_2O}$ and $k_u^{H_2O}$ are the folding and unfolding rates in the absence of denaturant, and m_f and m_u describe how the log of the folding and unfolding rates depends on the concentration of denaturant.

ASSOCIATED CONTENT

Supporting Information. Chevron plots for the NTL9 mutants and a plot of the Tanford β-parameter for all NTL9 mutants. This material is available free of charge via the Internet at http://pubs.acs.org.

Author Contributions

This manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

(+) These authors contributed jointly to this work

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

This work was supported by NSF grant MCB-1330259 to D.P.R.

References

[1] Hoffman, D. W., Davies, C., Gerchman, S. E., Kycia, J. H., Porter, S. J., White, S. W., and Ramakrishnan, V. (1994) Crystal structure of prokaryotic ribosomal protein L9: a bi-lobed RNA-binding protein. *EMBO J. 13*, 205-212.

[2] Kuhlman, B., Luisi, D. L., Evans, P. A., and Raleigh, D. P. (1998) Global analysis of the effects of temperature and denaturant on the folding and unfolding kinetics of the N-terminal domain of the protein L9. *J. Mol. Biol. 284*, 1661-1670.

[3] Luisi, D. L., Snow, C. D., Lin, J. J., Hendsch, Z. S., Tidor, B., and Raleigh, D. P. (2003) Surface salt bridges, double-mutant cycles, and protein stability: an experimental and computational analysis of the interaction of the Asp 23 side chain with the N-terminus of the N-terminal domain of the ribosomal protein l9. *Biochemistry 42*, 7050-7060.

[4] Luisi, D. L., and Raleigh, D. P. (2000) pH-dependent interactions and the stability and folding kinetics of the N-terminal domain of L9. Electrostatic interactions are only weakly formed in the transition state for folding. *J. Mol. Biol. 299*, 1091-1100.

[5] Sato, S., and Raleigh, D. P. (2007) Kinetic isotope effects reveal the presence of significant secondary structure in the transition state for the folding of the N-terminal domain of L9. *J. Mol. Biol. 370*, 349-355.

[6] Cho, J. H., and Raleigh, D. P. (2006) Electrostatic interactions in the denatured state and in the transition state for protein folding: effects of denatured state interactions on the analysis of transition state structure. *J. Mol. Biol. 359*, 1437-1446.

[7] Anil, B., Sato, S., Cho, J. H., and Raleigh, D. P. (2005) Fine structure analysis of a protein folding transition state; distinguishing between hydrophobic stabilization and specific packing. *J. Mol. Biol. 354*, 693-705.

[8] Meng, W., Luan, B., Lyle, N., Pappu, R. V., and Raleigh, D. P. (2013) The denatured state ensemble contains significant local and long-range structure under native conditions: analysis of the N-terminal domain of ribosomal protein L9. *Biochemistry 52*, 2662-2671.

[9] Cho, J. H., Meng, W., Sato, S., Kim, E. Y., Schindelin, H., and Raleigh, D. P. (2014) Energetically significant networks of coupled interactions within an unfolded protein. *Proc. Natl. Acad. Sci. U. S. A. 111*, 12079-12084.

[10] Chen, W., Shi, C., and Shen, J. (2015) Nascent beta-hairpin formation of a natively unfolded peptide reveals the role of hydrophobic contacts. *Biophys. J. 109*, 630-638.

[11] Dai, W., Sengupta, A. M., and Levy, R. M. (2015) First passage times, lifetimes, and relaxation times of unfolded proteins. *Phys. Rev. Lett. 115*, 048101.

[12] Chen, T., Song, J., and Chan, H. S. (2015) Theoretical perspectives on nonnative interactions and intrinsic disorder in protein folding and binding. *Curr. Opin. Struct. Biol. 30*, 32-42.

[13] Kundrotas, P. J., and Karshikoff, A. (2002) Modeling of denatured state for calculation of the electrostatic contribution to protein stability. *Protein Sci. 11*, 1681-1686.

[14] Zhou, H. X. (2002) Residual electrostatic effects in the unfolded state of the N-terminal domain of L9 can be attributed to nonspecific nonlocal charge-charge interactions. *Biochemistry 41*, 6533-6538.

[15] Baiz, C. R., and Tokmakoff, A. (2015) Structural disorder of folded proteins: isotope-edited 2D IR spectroscopy and Markov state modeling. *Biophys. J. 108*, 1747-1757.

[16] Beauchamp, K. A., McGibbon, R., Lin, Y. S., and Pande, V. S. (2012) Simple few-state models reveal hidden complexity in protein folding. *Proc. Natl. Acad. Sci. U. S. A. 109*, 17807-17813.

[17] Lindorff-Larsen, K., Piana, S., Dror, R. O., and Shaw, D. E. (2011) How fast-folding proteins fold. *Science 334*, 517-520.

[18] Luan, B., Lyle, N., Pappu, R. V., and Raleigh, D. P. (2014) Denatured state ensembles with the same radii of gyration can form significantly different long-range contacts. *Biochemistry 53*, 39-47.

[19] Wallace, J. A., and Shen, J. K. (2011) Continuous constant pH molecular dynamics in explicit solvent with pH-based replica exchange. *J. Chem. Theory Comput. 7*, 2617-2629.

[20] Contessoto, V. G., de Oliveira, V. M., de Carvalho, S. J., Oliveira, L. C., and Leite, V. B. (2016) NTL9 folding at constant pH: the importance of electrostatic interaction and pH dependence. *J. Chem. Theory Comput. 12*, 3270-3277.

[21] Shen, J. K. (2010) Uncovering specific electrostatic interactions in the denatured states of proteins. *Biophys. J. 99*, 924-932.

[22] Bowman, G. R., and Pande, V. S. (2010) Protein folded states are kinetic hubs. *Proc. Natl. Acad. Sci. U. S. A. 107*, 10890-10895.

[23] Shen, J. K. (2010) A method to determine residue-specific unfolded-state pKa values from analysis of stability changes in single mutant cycles. *J. Am. Chem. Soc. 132*, 7258-7259.

[24] Azia, A., and Levy, Y. (2009) Nonnative electrostatic interactions can modulate protein folding: molecular dynamics with a grain of salt. *J. Mol. Biol. 393*, 527-542.

[25] Song, B., Cho, J. H., and Raleigh, D. P. (2007) Ionic-strength-dependent effects in protein folding: analysis of rate equilibrium free-energy relationships and their interpretation. *Biochemistry 46*, 14206-14214.

[26] Anil, B., Craig-Schapiro, R., and Raleigh, D. P. (2006) Design of a hyperstable protein by rational consideration of unfolded state interactions. *J. Am. Chem. Soc. 128*, 3144-3145.

[27] Cho, J. H., and Raleigh, D. P. (2005) Mutational analysis demonstrates that specific electrostatic interactions can play a key role in the denatured state ensemble of proteins. *J. Mol. Biol. 353*, 174-185.

[28] Cho, J. H., Sato, S., and Raleigh, D. P. (2004) Thermodynamics and kinetics of non-native interactions in protein folding: a single point mutant significantly stabilizes the N-terminal domain of L9 by modulating non-native interactions in the denatured state. *J. Mol. Biol. 338*, 827-837.

[29] Kuhlman, B., Luisi, D. L., Young, P., and Raleigh, D. P. (1999) pKa values and the pH dependent stability of the N-terminal domain of L9 as probes of electrostatic interactions in the denatured state. Differentiation between local and nonlocal interactions. *Biochemistry 38*, 4896-4903.

[30] Fersht, A. R., and Sato, S. (2004) Phi-value analysis and the nature of protein-folding transition states. *Proc. Natl. Acad. Sci. U. S. A. 101*, 7976-7981.

[31] Goldenberg, D. P., Frieden, R. W., Haack, J. A., and Morrison, T. B. (1989) Mutational analysis of a protein-folding pathway. *Nature 338*, 127-132.

[32] Bulaj, G., and Goldenberg, D. P. (2001) Phi-values for BPTI folding intermediates and implications for transition state analysis. *Nat. Struct. Biol. 8*, 326-330.

[33] Matouschek, A., Kellis, J. T., Jr., Serrano, L., and Fersht, A. R. (1989) Mapping the transition state and pathway of protein folding by protein engineering. *Nature 340*, 122-126.

[34] Naganathan, A. N., and Munoz, V. (2010) Insights into protein folding mechanisms from large scale analysis of mutational effects. *Proc. Natl. Acad. Sci. U. S. A. 107*, 8611-8616.

[35] Cho, J. H., and Raleigh, D. P. (2006) Denatured state effects and the origin of nonclassical phi values in protein folding. *J. Am. Chem. Soc. 128*, 16492-16493.

[36] Ozkan, S. B., Bahar, I., and Dill, K. A. (2001) Transition states and the meaning of Phi-values in protein folding kinetics. *Nat. Struct. Biol. 8*, 765-769.

[37] Luisi, D. L., Wu, W. J., and Raleigh, D. P. (1999) Conformational analysis of a set of peptides corresponding to the entire primary sequence of the N-terminal domain of the ribosomal protein L9: evidence for stable native-like secondary structure in the unfolded state. *J. Mol. Biol. 287*, 395-407.

[38] Luisi, D. L., Kuhlman, B., Sideras, K., Evans, P. A., and Raleigh, D. P. (1999) Effects of varying the local propensity to form secondary structure on the stability and folding kinetics of a rapid folding mixed alpha/beta protein: characterization of a truncation mutant of the N-terminal domain of the ribosomal protein L9. *J. Mol. Biol. 289*, 167-174.

[39] Sanchez, I. E., and Kiefhaber, T. (2003) Origin of unusual phi-values in protein folding: evidence against specific nucleation sites. *J. Mol. Biol. 334*, 1077-1085.

[40] de los Rios, M. A., Muralidhara, B. K., Wildes, D., Sosnick, T. R., Marqusee, S., Wittung-Stafshede, P., Plaxco, K. W., and Ruczinski, I. (2006) On the precision of experimentally determined protein folding rates and phi-values. *Protein Sci.15*, 553-563.

[41] Shakhnovich, E. I., and Finkelstein, A. V. (1989) Theory of cooperative transitions in protein molecules. I. Why denaturation of globular protein is a first-order phase transition. *Biopolymers 28*, 1667-1680.

[42] Baldwin, R. L., and Rose, G. D. (2013) Molten globules, entropy-driven conformational change and protein folding. *Curr. Opin. Struct. Biol. 23*, 4-10.

[43] Plaxco, K. W., Simons, K. T., and Baker, D. (1998) Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol. 277*, 985-994.

[44] Pandit, A. D., Jha, A., Freed, K. F., and Sosnick, T. R. (2006) Small proteins fold through transition states with native-like topologies. *J. Mol. Biol. 361*, 755-770.

[45] Baxa, M. C., Freed, K. F., and Sosnick, T. R. (2008) Quantifying the structural requirements of the folding transition state of protein A and other systems. *J. Mol. Biol. 381*, 1362-1381.

[46] Sengupta, R., Pantel, A., Cheng, X., Shkel, I., Peran, I., Stenzoski, N., Raleigh, D. P., and Record, M. T., Jr. (2016) Positioning the intracellular salt potassium glutamate in the hofmeister series by chemical unfolding studies of NTL9. *Biochemistry 55*, 2251-2259.

[47] Tsodikov, O. V., Record, M. T., Jr., and Sergeev, Y. V. (2002) Novel computer program for fast exact calculation of accessible and molecular surface areas and average surface curvature. *J. Comput. Chem. 23*, 600-609.

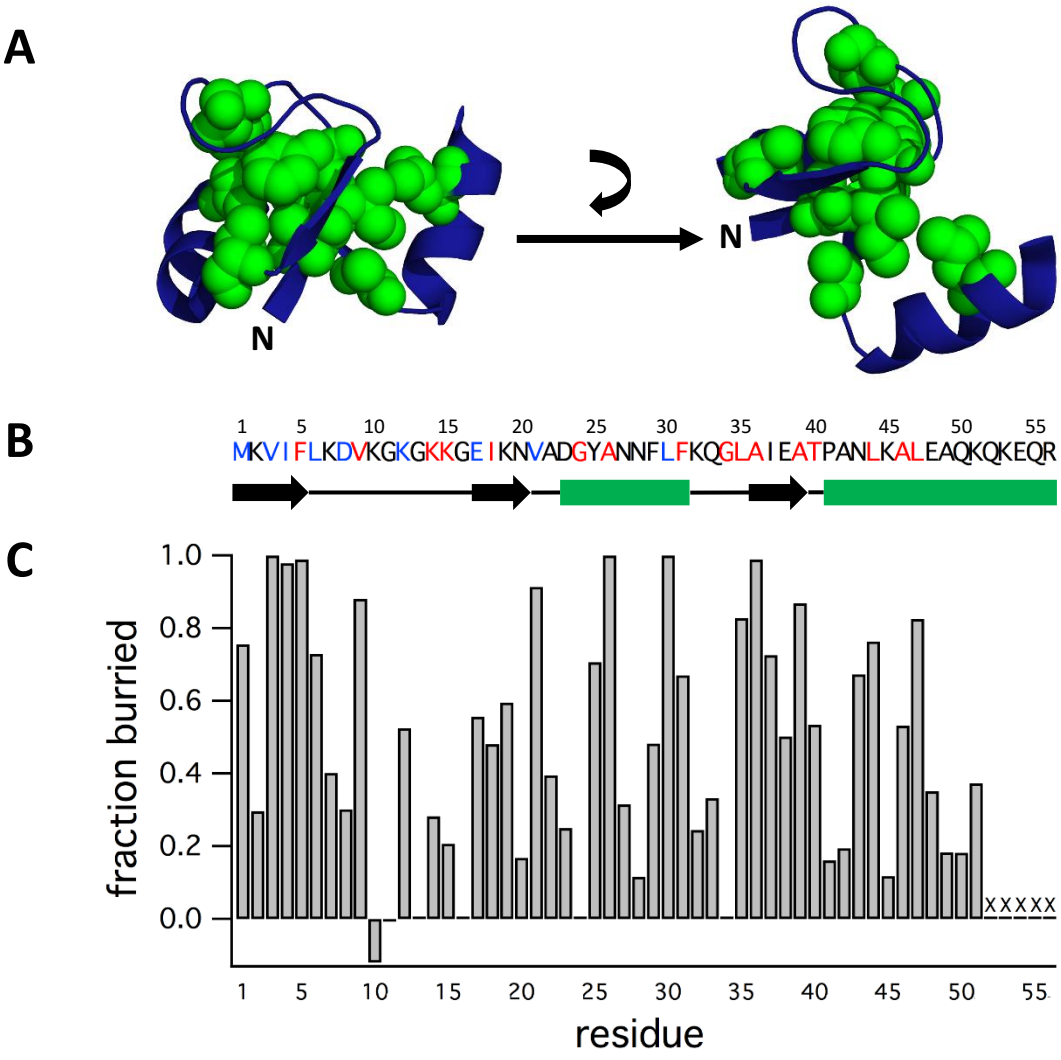


Figure 1. (A) Two views of a ribbon diagram of NTL9 (PDB code: 2HBB). The structures are rotated 90° with respect to each other. The N-terminus is labeled. Residues with sidechains that are more than 80% buried are shown in green space-filling representation. These are V3, I4, F5, V9, V21, A26, L30, L35, A36, A39, and L47. (B) The primary sequence of NTL9 is shown with a diagram of the secondary structure. β-strands are indicated by black arrows and α-helices by green rectangles. The residues colored red represent the sites studied here. Folding kinetics have been previously reported for the residues colored in blue. (C) Sidechain burial for NTL9. The fraction of the sidechain buried was calculated using the program SurfRacer.[47] An extended β-

strand was used to construct the maximally extended model of the unfolded state. The crystal structure of NTL9 was used for the folded state (PDB code: 2HBB). Residues 52-56 are disordered in the crystal structure and were not included in the calculation. These residues are indicated by an X in the plot.

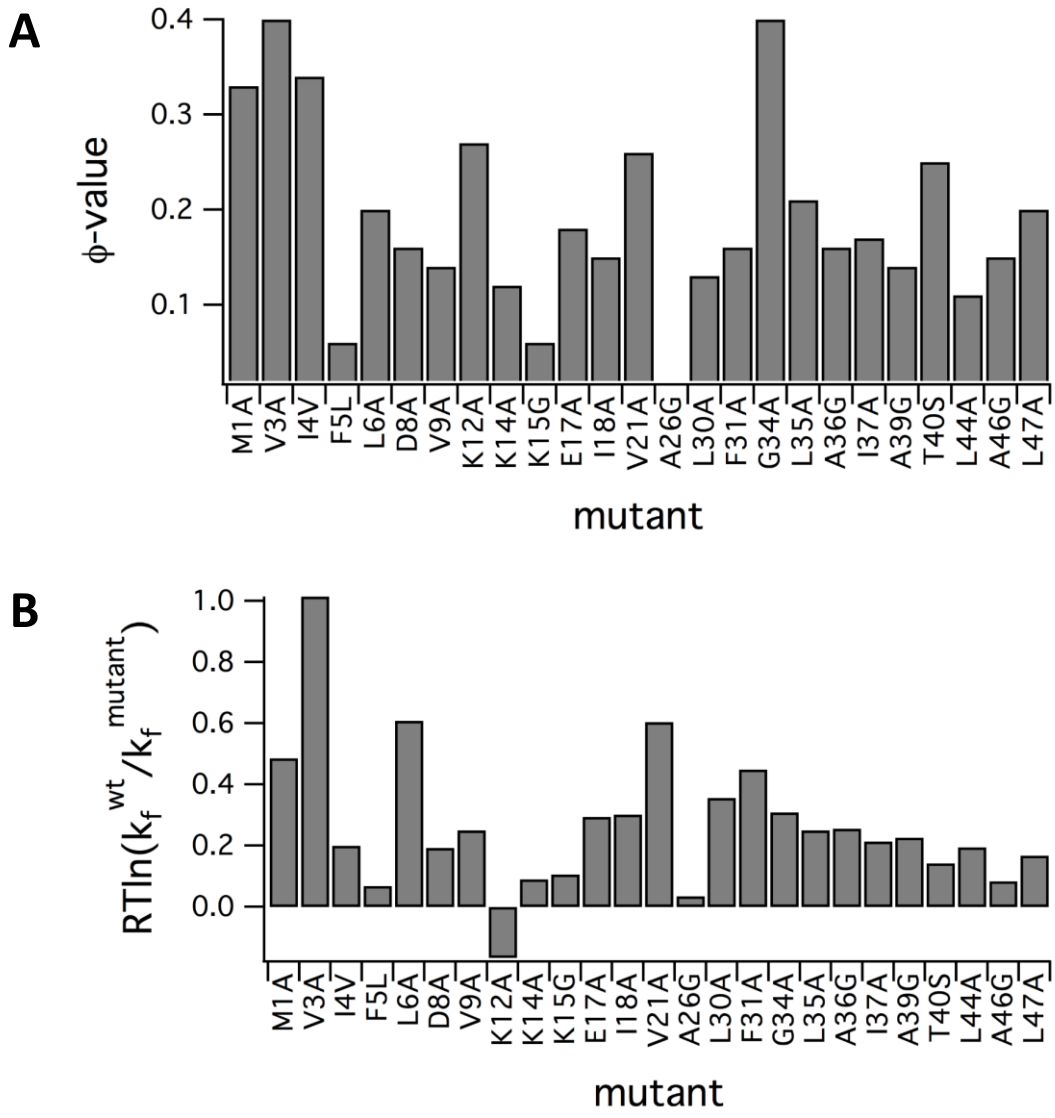


Figure 2. (A) ϕ-values for residues in NTL9 are modest and indicate that specific interactions between sidechains are not strongly formed in the transition state. ϕ-values were determined using mutants that differed in stability from wild-type NTL9 by 0.5 kcal mol^{-1} or greater. (B) Histogram of $\Delta\Delta G^{\ddagger}$ [RT ln(k_f^{WT}/k_f^{mutant})] vs. residue number. The trend in the plot is similar to that in the ϕ-value vs. residue plot. The plots include data collected in this study and data for M1, V3, I4, L6, D8, K12, E17, V21 and L30 which were previously reported.[6, 7]

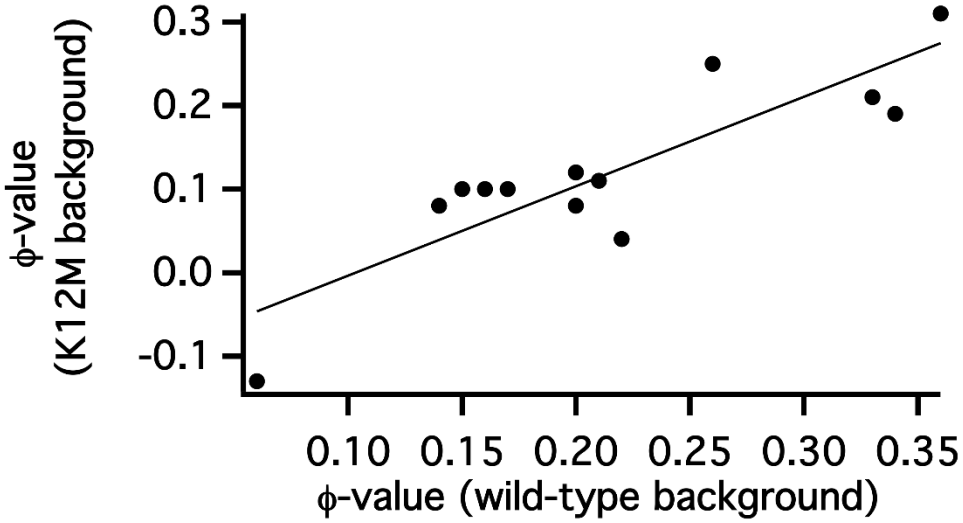


Figure 3. There is a correlation between ϕ-values measured in the wildtype background and the K12M background indicating that the trend of the values across the sequence is not significantly influenced by interactions in the unfolded state that involve K12. The r^2 value is 0.751 and p = 0.000126.

Table 1: Thermodynamic and Kinetic Parameters and ϕ-Values for NTL9 Mutants [a]

	$\Delta G°$ (eq) (kcal mol^{-1})	m_u (kcal mol^{-1} M^{-1})	m_f (kcal mol^{-1} M^{-1})	k_u (s^{-1})	k_f (s^{-1})	ϕ-value
WT	4.30 ± 0.36	0.20 ± 0.09	-0.45 ± 0.06	0.90 ± 0.40	865.0 ± 120	–
M1A [b]	2.84 ± 0.18	0.20 ± 0.09	-0.51 ± 0.06	4.18 ± 1.94	381.0 ± 110	0.33
V3A [b]	1.75 ± 0.28	0.29 ± 0.11	-0.59 ± 0.06	11.0 ± 4.40	156.0 ± 16.0	0.40
I4V [b]	3.71 ± 0.08	0.20 ± 0.09	-0.46 ± 0.06	1.22 ± 0.49	617.0 ± 62.0	0.34
F5L	3.15 ± 0.15	0.22 ± 0.09	-0.49 ± 0.06	3.76 ± 1.20	771.0 ± 95.0	0.06
L6A [b]	1.20 ± 0.33	0.26 ± 0.10	-0.62 ± 0.07	25.7 ± 17.0	310.0 ± 40.0	0.20
K7M	4.51 ± 0.15	0.23 ± 0.01	-0.43 ± 0.01	0.89 ± 0.19	1132.9 ± 35.4	–
D8N	3.92 ± 0.03	0.23 ± 0.01	-0.45 ± 0.01	1.34 ± 0.18	626.5 ± 17.2	–
D8A [c, d]	3.00 ± 0.07	0.25 ± 0.01	-0.46 ± 0.01	3.98 ± 0.37	625 ± 23	0.16
V9A	2.49 ± 0.20	0.27 ± 0.11	-0.46 ± 0.06	9.20 ± 3.92	567.0 ± 66.0	0.14
K10M	4.77 ± 0.09	0.22 ± 0.03	-0.43 ± 0.01	0.67 ± 0.29	825.7 ± 41.9	–
K10E	4.43 ± 0.10	0.18 ± 0.14	-0.46 ± 0.06	1.07 ± 0.92	1059.0 ± 129	–
K12M [e]	6.10 ± 0.30	[f]	-0.44 ± 0.01	[f]	1900 ± 78.3	0.26
K12A	4.91 ± 0.15	[f]	-0.44 ± 0.06	[f]	1146.0 ± 130	0.27
K12G	4.33 ± 0.05	0.17 ± 0.10	-0.43 ± 0.06	1.02 ± 0.44	1170 ± 122	–
K14M	3.96 ± 0.05	0.23 ± 0.01	-0.43 ± 0.01	1.13 ± 0.16	917.0 ± 28.1	–
K14A	3.57 ± 0.20	0.22 ± 0.10	-0.44 ± 0.06	2.13 ± 1.05	743.0 ± 95.0	0.12
K14G	3.00 ± 0.14	0.25 ± 0.11	-0.45 ± 0.06	4.93 ± 2.65	747.0 ± 102	0.07
K15M	3.86 ± 0.09	0.21 ± 0.02	-0.44 ± 0.01	2.21 ± 0.58	1120.7 ± 45.5	–
K15G	2.56 ± 0.18	0.22 ± 0.09	-0.48 ± 0.06	11.3 ± 4.95	724.0 ± 108	0.06
E17Q	3.97 ± 0.05	0.22 ± 0.01	-0.47 ± 0.01	0.82 ± 0.19	569.1 ± 33.7	–
E17A [c, d]	2.61 ± 0.05	-0.23 ± 0.01	-0.51 ± 0.01	6.36 ± 0.35	526 ± 17	0.18
I18V	4.23 ± 0.05	0.21 ± 0.09	-0.45 ± 0.06	0.90 ± 0.38	873.0 ± 100	–
I18A	2.25 ± 0.15	0.19 ± 0.09	-0.52 ± 0.06	12.8 ± 5.20	520.0 ± 88.0	0.15
V21A [b]	1.97 ± 0.25	0.30 ± 0.14	-0.53 ± 0.06	6.83 ± 2.93	312.0 ± 42.0	0.26
A22G	3.98 ± 0.09	0.22 ± 0.09	-0.47 ± 0.06	1.33 ± 0.55	750.0 ± 92.0	–
D23N	3.51 ± 0.02	n.d.	n.d.	n.d.	n.d.	n.d.
D23A [c, d]	4.27 ± 0.14	0.30 ± 0.01	-0.41 ± 0.01	0.36 ± 0.07	486 ± 11	–
G24A [g]	4.32 ± 0.05	0.57 ± 0.01	-0.87 ± 0.02	0.55 ± 0.05	861 ± 47	–
A26G	2.51 ± 0.19	0.23 ± 0.09	-0.54 ± 0.06	12.8 ± 5.04	816.0 ± 105	0.02
L30A [b]	1.54 ± 0.30	0.29 ± 0.13	-0.53 ± 0.06	26.3 ± 15.0	474.0 ± 52.0	0.13
F31A	1.46 ± 0.31	0.23 ± 0.09	-0.44 ± 0.06	18.0 ± 8.50	405.0 ± 44.0	0.16
G34A [h]	3.28 +0.24/-0.17	0.51 ± 0.01	-0.95 ± 0.1	1.94 ± 0.25	489 ± 107	0.40
L35A	3.09 ± 0.15	0.22 ± 0.09	-0.53 ± 0.06	4.04 ± 1.82	567.0 ± 62.0	0.21
A36G	2.75 ± 0.18	0.20 ± 0.09	-0.60 ± 0.07	6.66 ± 2.92	562.0 ± 63.0	0.16
I37V	4.39 ± 0.05	0.20 ± 0.09	-0.46 ± 0.06	0.86 ± 0.39	880.0 ± 108	–
I37A	3.02 ± 0.30	0.17 ± 0.08	-0.51 ± 0.06	3.45 ± 1.19	603.0 ± 77.0	0.17
E38Q	3.62 ± 0.02	n.d.	n.d.	n.d.	n.d.	n.d.
E38A [c, d]	3.89 ± 0.10	0.21 ± 0.01	-0.43 ± 0.01	1.05 ± 0.14	750 ± 17	–
A39G	2.71 ± 0.19	0.22 ± 0.10	-0.55 ± 0.06	6.12 ± 3.19	591.0 ± 72.0	0.14
T40S [c]	3.50 +0.63/-0.31	0.20 ± 0.09	-0.46 ± 0.12	1.85 ± 1.09	680.0 ± 80.0	0.25
A42G	3.97 ± 0.08	0.19 ± 0.08	-0.46 ± 0.06	1.43 ± 0.76	729.0 ± 105	–
L44A	2.57 ± 0.20	0.13 ± 0.04	-0.51 ± 0.06	8.83 ± 4.29	623.0 ± 73.0	0.11
A46G	3.75 ± 0.07	0.19 ± 0.09	-0.47 ± 0.06	2.65 ± 1.02	751.0 ± 82.0	0.15
L47A	3.46 ± 0.09	0.20 ± 0.09	-0.45 ± 0.06	2.02 ± 0.96	652.0 ± 77.0	0.20
E48Q	4.37 ± 0.38	n.d.	n.d.	n.d.	n.d.	–
A49G	4.08 ± 0.25	0.20 ± 0.09	-0.46 ± 0.06	1.20 ± 0.79	824.0 ± 110	–
E54Q	4.56 ± 0.18	n.d.	n.d.	n.d.	n.d.	–

[a] All experiments were performed in 20 mM sodium acetate and 100 mM NaCl at pH 5.5 and 25 °C.
[b] Data were taken from ref. 7.

[c] ΔG° values were obtained from kinetic experiments using urea. For the wildtype protein, ΔG° = 4.07 ± 0.10 kcal mol^{-1}.
[d] Data were taken from ref. 6.
[e] Data were taken from ref. 46.
[f] The protein is so stable that the unfolding kinetic parameters are not defined well.
[g] The experiment was performed using guanidine hydrochloride. For the wildtype protein the measured ΔG° using GndCl as a denaturant was 4.17 ± 0.07 kcal mol^{-1}.
[h] The ΔG° value was obtained from kinetic experiments using guanidine hydrochloride. For the wildtype protein, ΔG° = 4.05 ± 0.09 kcal mol^{-1}.

Table 2. Experiments performed in low and high concentrations of salt yield a similar trend of ϕ-values. Uncertainties in ΔG° values range from ±0.15 to ±0.30 and the uncertainties in ϕ-values range from ±0.03 to ±0.07. Experiments were performed in 20 mM sodium acetate with either 100 mM NaCl or 750 mM NaCl at pH 5.5 and 25 °C.

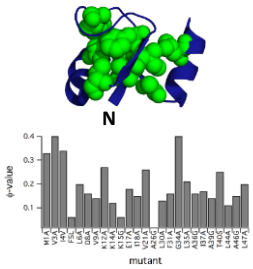
	100 mM NaCl		750 mM NaCl	
	$\Delta G°$ (kcal mol^{-1})	ϕ-value	$\Delta G°$ (kcal mol^{-1})	ϕ-value
WT	4.30		5.02	
M1A	2.84	0.33	3.32	0.23
V3A	1.75	0.40	2.32	0.35
I4A	2.08	0.38	2.40	0.33
V9A	2.49	0.14	2.87	0.06
I18A	2.25	0.15	3.18	0.05
V21A	1.97	0.26	2.71	0.22
L30A	1.54	0.13	2.13	0.01
L35A	3.09	0.21	3.58	0.12
A36G	2.75	0.16	3.24	0.04

Table 3. Thermodynamic and Kinetic Parameters and ϕ-Values for NTL9-K12M Mutants [a]

	$\Delta G°$ (eq) (kcal mol^{-1})	m_u (kcal mol^{-1} M^{-1})	m_f (kcal mol^{-1} M^{-1})	k_u (s^{-1})	k_f (s^{-1})	ϕ-value
K12M	6.10 ± 0.30	[b]	-0.44 ± 0.01	[b]	1900.0 ± 78.3	
M1A/K12M	3.75 ± 0.07	0.13 ± 0.01	-0.46 ± 0.01	2.05 ± 0.16	827.7 ± 9.0	0.21
V3A/K12M	2.36 ± 0.03	0.18 ± 0.01	-0.56 ± 0.01	5.69 ± 0.35	261.0 ± 10.7	0.31
I4A/K12M	2.70 ± 0.05	0.12 ± 0.01	-0.53 ± 0.02	4.68 ± 1.16	670.7 ± 58.3	0.19
F5L/K12M	4.95 ± 0.05	[b]	-0.39 ± 0.01	[b]	2440.6 ± 1.03	-0.13
L6A/K12M	2.48 ± 0.06	0.15 ± 0.01	-0.45 ± 0.01	12.5 ± 1.03	892.9 ± 17.7	0.12
I18A/K12M	2.99 ± 0.08	0.07 ± 0.01	-0.48 ± 0.01	10.5 ± 0.98	1115.5 ± 24.9	0.10
V21A/K12M	2.81 ± 0.04	0.17 ± 0.01	-0.55 ± 0.01	5.33 ± 0.27	559.6 ± 13.1	0.25
A22G/K12M	4.78 ± 0.10	[b]	-0.42 ± 0.01	[b]	1422.2 ± 1.02	0.10
L35A/K12M	3.87 ± 0.15	0.15 ± 0.01	-0.48 ± 0.01	2.07 ± 0.60	1255.7 ± 57.6	0.11
A36G/K12M	3.58 ± 0.10	0.14 ± 0.01	-0.50 ± 0.01	4.12 ± 0.39	1265.9 ± 29.9	0.10
I37A/K12M	4.31 ± 0.30	[b]	-0.42 ± 0.02	[b]	1422.2 ± 28.8	0.10
A39G/K12M	3.46 ± 0.17	0.15 ± 0.01	-0.47 ± 0.01	4.12 ± 0.47	1350.9 ± 30.8	0.08
T40S/K12M	4.30 ± 0.04	[b]	-0.43 ± 0.01	[b]	1685.8 ± 86.4	0.04
A42G/K12M	4.54 ± 0.05	[b]	-0.42 ± 0.01	[b]	1669.0 ± 103	0.09
L47A/K12M	3.75 ± 0.05	[b]	-0.42 ± 0.01	[b]	1422.2 ± 28.7	0.08

[a] All experiments were performed in 20 mM sodium acetate and 100 mM NaCl at pH 5.5 and 25 °C.
[b] The protein is so stable that the unfolding kinetic parameters are not defined well.

Table of Contents Figure:



SUPPORTING INFORMATION FOR

The N-Terminal Domain of the Ribosomal Protein L9 Folds Through a Diffuse and Delocalized Transition State

Satoshi Sato, Jae-Hyun Cho, Ivan Peran and Daniel P. Raleigh

Figure S1. Plots of ln k vs. urea concentration for NTL9 and its mutants. (A) Mutants in the first β strand. (B) Mutants in the first loop and the second β strand. (C) Mutants in the second loop and the first (short) helix. (D) Mutants in the third β strand. (E) Mutants in the second helix. The data for the wild-type protein are included in all graphs.

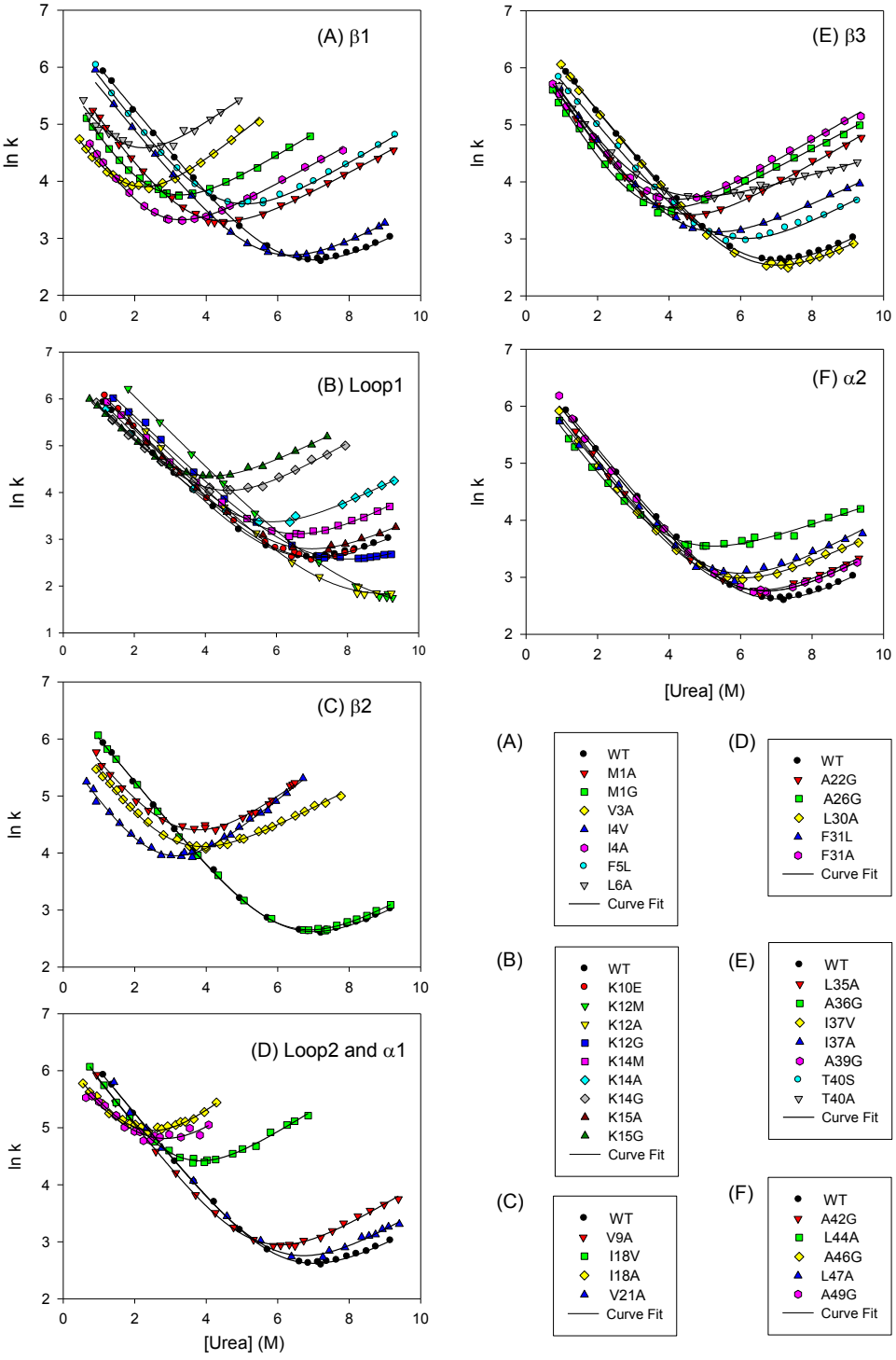


Figure S2. The Tanford β-parameter for wildtype NTL9 and its mutants is similar. The bold line represents the average values for all NTL9 variants which is 0.69, the same as the wildtype value. The dashed line represents 1.5 x standard deviations, respectively. The points corresponding to the wildtype protein and two mutants that are outliers are labeled.