UC Santa Barbara
UC Santa Barbara Previously Published Works

Title

A comparison of the folding kinetics of a small, artificially selected DNA aptamer with those of equivalently simple naturally occurring proteins

Permalink

Journal

ISSN

Authors

Lawrence, Camille
Vallée-Bélisle, Alexis
Pfeil, Shawn H
et al.

Publication Date

DOI

Peer reviewed

A comparison of the folding kinetics of a small, artificially selected DNA aptamer with those of equivalently simple naturally occurring proteins

Camille Lawrence,[1] Alexis Vallée-Bélisle,[2,3] Shawn H. Pfeil,[4,5] Derek de Mornay,[2] Everett A. Lipman,[1,4] and Kevin W. Plaxco[1,2]*

[1]Interdepartmental Program in Biomolecular Science and Engineering, University of California, Santa Barbara, California 93106
[2]Department of Chemistry and Biochemistry, University of California, Santa Barbara, California 93106
[3]Laboratory of Biosensors and Nanomachines, Département de Chimie, Université de Montréal, Québec, Canada
[4]Department of Physics, University of California, Santa Barbara, California 93106
[5]Department of Physics, West Chester University of Pennsylvania, West Chester, Pennsylvania 19383

Abstract: The folding of larger proteins generally differs from the folding of similarly large nucleic acids in the number and stability of the intermediates involved. To date, however, no similar comparison has been made between the folding of smaller proteins, which typically fold without well-populated intermediates, and the folding of small, simple nucleic acids. In response, in this study, we compare the folding of a 38-base DNA aptamer with the folding of a set of equivalently simple proteins. We find that, as is true for the large majority of simple, single domain proteins, the aptamer folds through a concerted, millisecond-scale process lacking well-populated intermediates. Perhaps surprisingly, the observed folding rate falls within error of a previously described relationship between the folding kinetics of single-domain proteins and their native state topology. Likewise, similarly to single-domain proteins, the aptamer exhibits a relatively low urea-derived Tanford β, suggesting that its folding transition state is modestly ordered. In contrast to this, however, and in contrast to the behavior of proteins, ϕ-value analysis suggests that the aptamer's folding transition state is highly ordered, a discrepancy that presumably reflects the markedly more important role that secondary structure formation plays in the folding of nucleic acids. This difference notwithstanding, the similarities that we observe between the two-state folding of single-domain proteins and the two-state folding of this similarly simple DNA presumably reflect properties that are universal in the folding of all sufficiently cooperative heteropolymers irrespective of their chemical details.

Keywords: contact order; topomer search; tanford beta (β); phi (φ); single molecule Förster resonance energy transfer (FRET); chevron; ionic strength

Introduction

Since Jackson and Fersht reported the first "type specimen" in 1991 (Ref. 1), dozens of small, single domain proteins have been described that fold via a highly concerted, approximately two-state process lacking well-populated intermediate states. The relative simplicity of such folding has allowed for meaningful, quantitative comparisons of folding kinetics
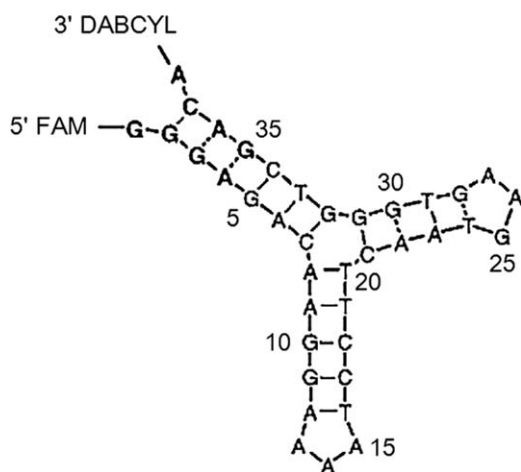
Figure 1. In this study, we have characterized the folding of a DNA aptamer, a DNA molecule selected *in vitro* to bind to a specific target molecule and that adopts a reasonably complex, protein-like structure. Specifically, we have characterized the folding kinetics and thermodynamics of the 38-base cocaine-binding aptamer of Stojanovic, which is thought to adopt the three-way junction structure shown.[13,14] With 185 freely rotating bonds, the number of conformations accessible to the unfolded aptamer is comparable to those of an unfolded, 94-residue protein (186 rotatable bonds) suggesting that comparison of the aptamer's folding kinetics with those of equivalently simple single-domain proteins may prove informative. Of note, what, if any, tertiary contacts the aptamer forms in the absence of cocaine binding is unknown; NMR-based studies of its structure have not identified any base–base contacts other than those that would be seen given the predicted secondary structure.[14]

of this otherwise diverse set of small proteins, and the discovery of a number of general principles. Their folding rates, for example, are strongly correlated with simple, scalar measures of their native state topology[2] (reviewed in Ref. 3). Their folding transition states also bury significant amounts of hydrophobic surface area (between 40 and 96% of that buried in the native state; e.g., Ref. 4) and yet their transition states contain relatively little native-like side-chain packing as defined by ϕ values[5]; the vast majority of reported ϕ values, for example, fall below 0.3, suggesting that the folding transition states of two-state proteins are relatively unstructured (see, e.g., Fig. 1 in Ref. 6).

In contrast to the folding of single domain proteins, the folding kinetics of most of the nucleic acids characterized to date are highly heterogeneous. That is, other than a few studies of the folding of simple secondary structures (such as hairpins, e.g., Ref. 7; and G-quadruplexes, e.g., Ref. 8) or more complex, if still simple topologies (pseudoknots, e.g., Ref. 9), the folding kinetics of the large majority of nucleic acid described in the literature involve highly stable on- and even off-pathway intermediates (see Refs. 10 and 11 for relevant discussions). This lack of well characterized, topologically-complex, yet still

"two-state" nucleic acids has rendered it impossible to determine whether the general rules uncovered via studies of the two-state folding of the simplest proteins also hold for the folding of other simple polymers (although see Ref. 12 for an attempt at reconciliation). In response we describe in this study a small, yet reasonably topologically complex DNA aptamer that is analogous to the large majority of single domain proteins in that it folds in a concerted, approximately two-state process. Detailed comparison of the folding kinetics of this simple, single domain DNA with those of similarly simple, single-domain proteins can thus inform on those properties of folding that are universal and are not dependent on the detailed chemistry of the polymer involved.

Results

As our test bed we have used a short DNA aptamer selected *in vitro* for its ability to bind cocaine and reported[13,14] to adopt a three-way junction structure (Fig. 1). And although, at 38 bases, the aptamer is significantly shorter than all but the shortest single domain proteins studied to date (such as Refs. 15–17), the complexity of the conformational search associated with its folding nevertheless approaches that of more typical single domain proteins. Nucleic acids, for example, contain five rather than two freely rotating bonds per monomer (one of the six bonds in the backbone of nucleic acids is in the pentose ring and therefore rotationally fixed), and thus the Levinthal-esque difficulty of the search for the native conformation[18] of a 38-base oligomer (containing 185 freely rotating bonds in its backbone) is similar to that of a 94-residue protein (with 186 freely rotating bonds). With 830 non-hydrogen atoms the size of the aptamer is likewise similar to the number of non-hydrogen atoms found in a 105-residue protein of average amino acid composition.

As is true for the large majority of proteins of similar size and complexity,[19] the equilibrium folding of the aptamer is a concerted process well approximated as two-state (Fig. 2). For example, an equilibrium urea "melt" of a fluorophore- and quencher-modified aptamer is well fitted by the standard model used for the study of proteins, which assumes two-state behavior and a linear relationship between folding free energy and denaturant concentration. Specifically, our data are well fit (Fig. 2; $R^2 = 0.997$) to a two-state model.

$$F_{obs} = \frac{F_f + F_u\, e^{-\Delta G_0 - m_{eq}[\text{urea}]}}{1 + e^{-\Delta G_0 - m_{eq}[\text{urea}]}} \qquad (1)$$

where ΔG_0 is the folding free energy in the absence of denaturant, m_{eq} is a measure of the extent to which the denaturant modulates this free energy, and F_f, F_u, and F_{obs} are the fluorescence of the
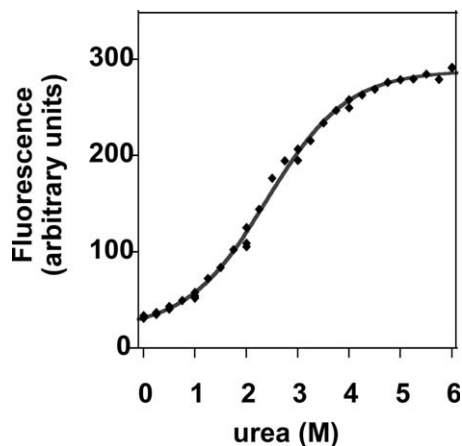
Figure 2. The urea-induced equilibrium unfolding of the aptamer is well approximated ($R^2 = 0.997$) as a two-state process lacking well-populated intermediate states. A fit to the two-state model [Eq. (1)] is shown (solid line).

folded state, the unfolded state, and the equilibrium (observed) mixture at any given urea concentration, respectively. Given the precision of this fit, which produces estimates of the folding free energy and m-value of 7.52 ± 0.40 kJ/mol and 3.13 ± 0.15 kJ/mol.M, respectively, it appears that there are no well-populated intermediates in the equilibrium unfolding of the aptamer.

To probe the extent to which the aptamer's equilibrium folding is two-state in more detail we have studied its folding at the single molecule level. Specifically, we have used single-molecule Förster resonance energy transfer (FRET), which provides a measure of the distance between two appropriate dyes, to monitor the equilibrium unfolding of the aptamer in urea. As expected, we observe only two-states: a relatively low-transfer efficiency, unfolded state, and the higher-transfer efficiency (i.e., more compact) folded state (Fig. 3). As we shift from high to low denaturant, the relative populations of these two-states shift in concert without the population of any states of intermediate transfer efficiency (i.e., intermediate size), offering further evidence that the equilibrium folding of the aptamer is well approximated as two-state and that the unfolded state of the aptamer remains fully unfolded even at low denaturant. This said, the transfer efficiency observed for the unfolded state does decrease slightly with increasing denaturant. Similar denaturant dependence has been seen for energy transfer across the unfolded states of effectively all of the appropriately characterized single-domain proteins described to date; it may or may not reflect contraction of the unfolded state at low denaturant (see discussion in Ref. 20).

As is also true for the large majority of proteins of similar size and complexity,[19] the aptamer's folding/unfolding kinetics are likewise well approximated as a simple, two-state process lacking well-populated

intermediate states. First, the aptamer's folding kinetics are well fitted to a single-exponential decay [Fig. 4(A)], producing fit residuals similar in magnitude ($<2\%$ of total amplitude) to those seen in non-folding controls performed under the same conditions. Second, a plot of the log of observed relaxation rate as a function of urea concentration – a so-called chevron curve – is well fit by the linear free energy relationship [Fig. 4(b); $R^2 = 0.979$] expected for a process lacking well-populated intermediates[1]:

$$\ln(k_{obs}) = \ln\left[e^{(\ln k_f + m_f [D]/RT)} + e^{(\ln k_u + m_u [D]/RT)}\right] \quad (2)$$

where k_{obs} is the experimentally observed relaxation rate, k_f and k_u are the folding and unfolding rates
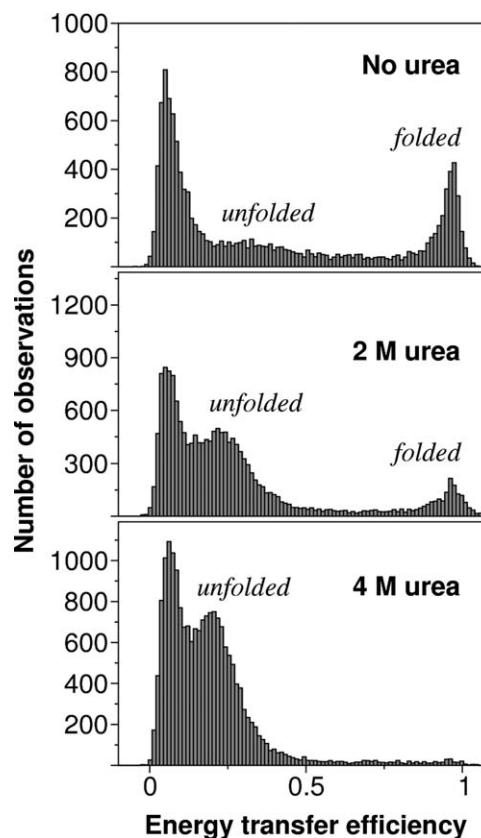

Figure 3. Single-molecule Förster resonance energy transfer (FRET) studies likewise support the claimed two-state equilibrium unfolding of the aptamer. As expected, in addition to a large peak near zero transfer efficiency arising due to mislabeled or photobleached molecules lacking an acceptor dye, we observe only two-states: a relatively low-transfer-efficiency, unfolded state, and the higher-transfer-efficiency (i.e., more compact) folded state. As we shift from high to low denaturant, the relative populations of these two states shift in concert without the population of any states of intermediate transfer efficiency (i.e., intermediate size). This said, the transfer efficiency observed for the unfolded state does decrease slightly with increasing denaturant, which may, or may not, reflect contraction of the unfolded state at low denaturant (see discussion in Ref. 20).
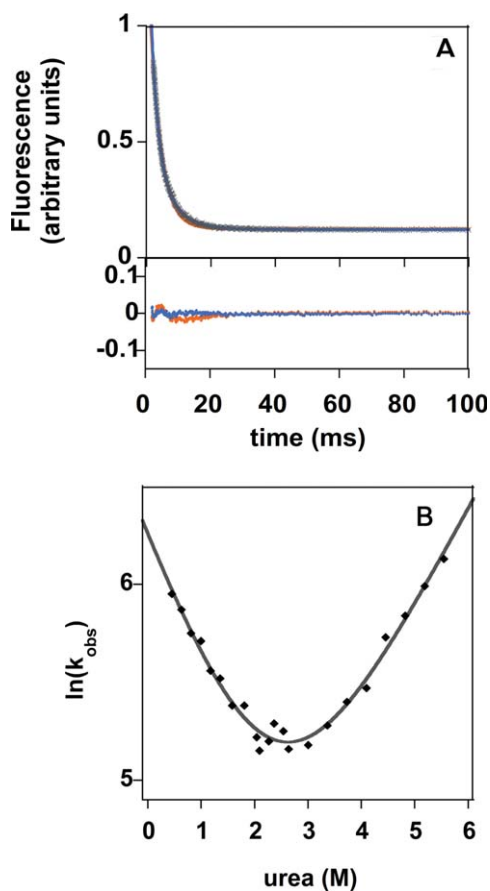
Figure 4. As is true for the majority of similarly simple, single-domain proteins,[19] the aptamer's folding kinetics are well approximated as a concerted, two-state process lacking well-populated intermediates. (**A**) For example, as shown here a representative folding trace (gray data points) is well fitted as single exponential ($R^2 = 0.997$; blue line) and produces residuals (blue) of less than 2% of the total amplitude. A double-exponential fit (red residuals) is not statistically significantly improved. (**B**) Likewise, a chevron plot [a plot of ln(relaxation rate) versus denaturant concentration] exhibits linear arms and produces equilibrium folding free energy and m_{eq} value estimates, 7.07 ± 0.28 kJ/mol and 3.02 ± 0.25 kJ/mol.M, respectively, within error of the 7.52 ± 0.40 kJ/mol and 3.13 ± 0.15 kJ/mol.M estimates obtained from the equilibrium melt.[1] These data were collected in 100 mM NaCl, corresponding to an ionic strength of 105 mM (when the buffer contribution is included).

under "standard" conditions (here these are the absence of denaturant), $[D]$ is denaturant concentration, and m_f and m_u are the kinetic "m-values" and denote the extent to which denaturant modulates the stability of the folding transition state. Third, the folding thermodynamics derived from this chevron are effectively indistinguishable from those obtained at equilibrium, suggesting that the folding process lacks any thermodynamically stable intermediate states. That is, the unfolding free energy and m_{eq}-value derived from these kinetic data, 7.07 ± 0.28 kJ/mol and 3.02 ± 0.25 kJ/mol.M, respec-

tively, are within error of the values obtained from the equilibrium urea melt data described above (Fig. 2), suggesting that any structure that may (or may not) be formed in the dead-time of the kinetic experiments (e.g., if the change in transfer efficiency seen for the unfolded state at low denaturant – Fig. 3 – represents a small degree of unfolded state compaction) is thermodynamically inconsequential.

Extrapolation of the chevron plot to zero urea predicts a folding rate for the aptamer of 494 ± 22 s^{-1} in the absence of denaturant. And although comparison of the structure of an oligonucleotide with that of a protein is, of course, subject to potentially significant qualifications, this value is, perhaps surprisingly, close to the folding rate that would be predicted for a protein of comparable topological complexity. To see this we note that the putative structure of the aptamer (see Fig. 1, which was adopted from Ref. 13; also see Refs. 14, 21) includes 15 sequence-distant (more than five bases, or 25 freely rotating bonds, apart) base-pair contacts, giving a long-range-order (number of sequence-distant
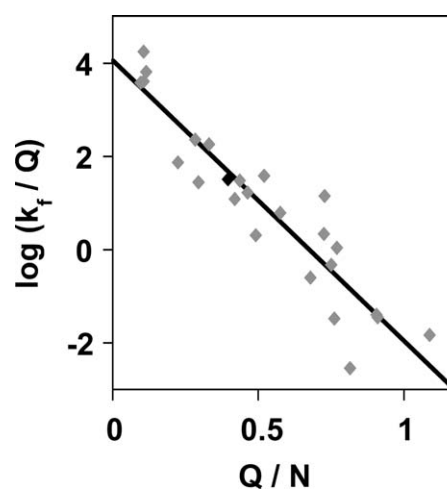

Figure 5. The folding rates of simple, single domain proteins (light gray) are reasonably well correlated with several simple measures of their native-state topology. Shown, for example, is a plot of (log) folding rates versus long-range order,[22] defined as the number of sequence-distant native contacts (Q = number of native contacts separated by 12 or more residues, corresponding to >24 freely-rotating bonds) divided by the total polymer length, N, (data adopted from Makarov and Plaxco[3]). The observed folding rate of the aptamer appears to obey this same relationship (where sequence-distant contacts in the aptamer are defined as >5 bases, corresponding to >25 freely rotating bonds). This suggests that topology's role in defining (two-state) folding rates may not be unique to proteins. Note, the aptamer folding rate shown was determined at 105 mM ionic strength, which is about a third higher than the ∼75 mM ionic strength at which protein folding rates are perhaps most often measured[4]; the folding rate of the aptamer, however, is only mildly dependent on ionic strength, varying only ∼2-fold (Fig. 7) over the 20–100 mM range of ionic strengths typically used in studies of protein folding.
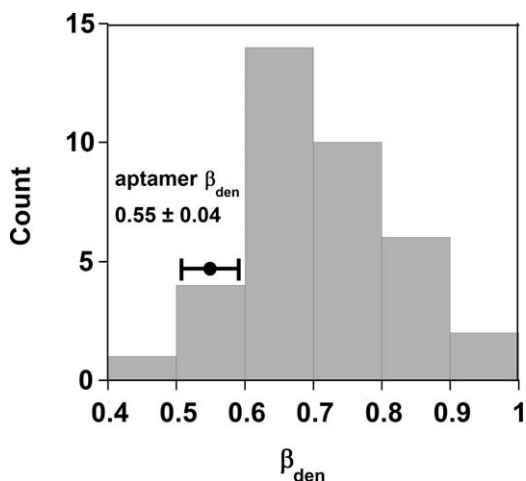
Figure 6. The aptamer's Tanford β_{den}, a metric describing the degree of consolidation of hydrophobic surface area in the folding transition state, lies nearer the bottom of a set of 37 well-characterized two-state proteins adopted from two previously described data sets.[4,51] Since this value is known to anti-correlate with topological complexity,[2] the relatively low value observed for the aptamer may arise due to its relatively low topological complexity. The protein data were collected using either urea or guanidinium chloride (GuHCl) as denaturant. It should be noted that SH3 domains in this dataset are overrepresented; seven of these 37 proteins are SH3 domains, the β_{den} values of which range from 0.64 to 0.89.

contacts per residue[22]) of 0.39. Inserting this value into the known rate/long-range-order relationship for single domain proteins (using a cutoff of 12 residues or 24 freely rotating bonds) predicts a folding rate of 720 s^{-1} for the aptamer, which is within a factor of 1.5 of the experimentally observed value (Fig. 5). The topology-estimated folding rate of the aptamer is thus quite close to the observed value relative to the more than million-fold range of two-state protein folding rates. This said, some approximations and assumptions were, obviously, necessary to make this comparison. First, the tertiary structure of the aptamer is not known with certainty (indeed, the aptamer may not form any tertiary structure; i.e., a lack of sequence-distant NOEs in the aptamer's NMR spectra suggests that the secondary structure plot shown in Fig. 1 may well serve as a complete description of the biopolymer's structure[14]). Second, unfolded DNA may be more prone than unfolded proteins to retain residual structure, which would presumably reduce the effective number of rotatable bonds and thus increase the uncertainty in the placement of the aptamer on the folding rate/long-range-order plot. Finally, the assumed equivalence between one freely rotating bond in the nucleic acid backbone and one freely rotating bond in the polypeptide backbone is likewise speculative. Given these concerns, the study of more (and better structurally characterized) aptamers is clearly necessary before we can confirm

this surprising, if perhaps theoretically expected (see discussion), correlation.

Urea chevron plots also provide a means of probing the extent to which hydrophobic surface is buried in the folding transition state. Specifically, the ratio of slope of the folding arm of the chevron to the difference in the slopes of the folding and unfolding arms is thought to reflect the fraction of the total hydrophobic surface area buried upon folding that is buried in the folding transition state[1] (we note that this may be an oversimplified interpretation of β_{den}; urea may also disrupt hydrogen bonding: e.g., backbone hydrogen bonding in proteins and base pairing in nucleic acids). The value so obtained, $\beta_{den} = m_f/(m_f - m_u)$, is 0.55 ± 0.04. Although this falls toward the low end of the range of values seen for similarly small, single-domain proteins (Fig. 6), it has previously been shown that, like folding rates, β_{den} is correlated with topological complexity[2] and thus, given the aptamer's relatively sequence-local structure, its relatively low β_{den} would appear to represent yet another similarity with proteins.

Aptamers, like all nucleic acids, are more highly charged than is typical for proteins, suggesting that studies of the ionic strength dependence of the aptamer's folding may highlight differences between the two. To characterize this we modulated the ionic strength by varying the concentration of sodium chloride, under the argument that sodium and chloride are monovalent (minimizing the possibility of specific binding effects) and fall in the middle of their respective Hofmeister series (minimizing chaotropic or kosmotropic effects). To access low ionic strengths we performed these experiments in 10 mM tris at pH 8.1, conditions in which the buffer contributes only $\sim$5 mM to the overall ionic strength. To measure both the folding and unfolding of the aptamer we conducted these experiments in 4 M urea.

To employ salt to probe the role that electrostatics play in folding we must first establish an equation that describes the manner in which the presence of salt alters the stability of the folding transition state. For studies using the denaturants urea or guanidine (e.g., Fig. 2), a linear free energy relationship is usually used [i.e., the transition state energy, and thus log(relaxation rate) is linear in denaturant concentration: Eq. (2)]. Unfortunately, however, the correct functionality for $\Delta G(I)$, the free energy as a function of ionic strength and/or ion concentration, I, is not clear, with some authors (including us[23]) claiming Debye-Huckel-type square-root of ionic strength dependence, others claiming that free energy (or T_m) goes with the log of salt concentration (e.g., Refs. 24–26), and still others claiming more complex behaviors.[27] Moreover, given that the Debye-Huckel square root dependence arises due to the assumption that ions are point charges,[28] and
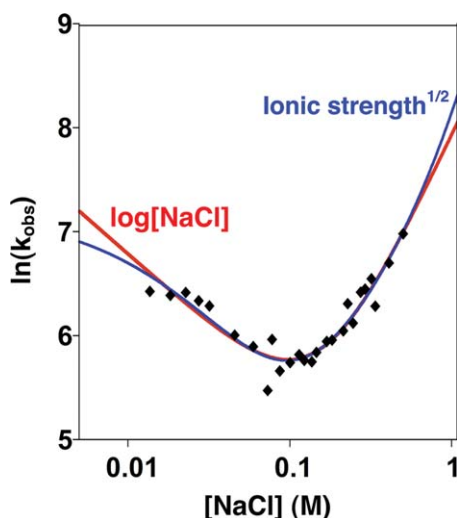
Figure 7. Due to the high charge of the aptamer's backbone it would appear likely that electrostatics would play an important role in its folding. To explore this we constructed a chevron plot (analogous to the denaturant chevron presented in Fig. 2) varying salt concentration. We fit this chevron under two assumptions: that the free energy of the folding transition state is linear in log[NaCl] or in square-root of ionic strength (see text for rationale). The experiment was conducted in 4 M urea, conditions under which both folding and unfolding can be accessed. Raw data for this figure can be found in the supplementary material.

that the logarithmic dependence of free energy on ionic strength is a consequence of the model in which a line of charges interacts with a point charge,[29] it is possible that neither model accurately captures electrostatics in a complex, highly charged structure such as an aptamer.

In the absence of a single, compelling model for $\Delta G(I)$, we present in this study, the kinetic parameters for both: a Debye-Huckel-type square-root-ionic strength dependence,

$$\ln\left(k_{obs}\right)=\ln\left[e^{(\ln k_f + m_f^{DH}\sqrt{I})}+e^{(\ln k_u + m_u^{DH}\sqrt{I})}\right] \quad (3)$$

and a logarithmic salt concentration dependence,

$$\ln(k_{obs})=\ln\left[e^{(\ln k_f + m_f^{\log}\log[NaCl])}+e^{(\ln k_u + m_u^{\log}\log[NaCl])}\right] \quad (4)$$

where k_f and k_u are the folding and unfolding rates under standard conditions (here in either zero ionic strength for Eq. (3) or in 1 M NaCl for Eq. (4), and, by analogy to m_f and m_u, m_f^{DH} and m_u^{DH} and $m_f^{\log}$ and $m_u^{\log}$ describe the effects of ionic strength or salt concentration on the folding rate or unfolding rate, respectively. Measuring the folding and unfolding of our aptamer over a range of salt concentrations we find that both relationships fit observed kinetics reasonably well [$R^2 = 0.946$ and 0.937 for Eq. (3) and (4), respectively; Fig. 7], rendering it difficult to

determine which model better describes the aptamer's behavior.

Using the linear in log-salt strength and linear in square-root-ionic-strength models we can, by analogy to the urea chevron analysis performed above, define a NaCl-derived Tanford β indicative of the degree to which native-like ionic interactions are formed in the transition state. The resultant $\beta_{NaCl}^{\log}$ (i.e., using the linear in log-salt model) is 0.71 ± 0.15, suggesting that the phosphate–sodium electrostatic interactions in the folding transition state are as well or better formed than are hydrophobic interactions (from above, $\beta_{den} = 0.55 \pm 0.04$). The linear in square-root-ionic-strength value, β_{NaCl}^{sqrt} is, however, just 0.27 ± 0.05, suggesting that the electrostatic interactions in the transition state are less well formed than the hydrophobic interactions. Given that electrostatic interactions are likely longer range than are hydrophobic interactions this interpretation of the effects of salt would appear less physically sound.

In contrast to β_{den}, which suggests that the aptamer's folding transition state is only modestly consolidated, ϕ-value analysis[5] suggests that many, if not the large majority, of the aptamer's nucleobases are in well-structured, native-like environments in the transition state. Φ-values are obtained by comparing the folding kinetics and thermodynamics of a substituted variant sequence to a reference sequence (for a protein this is typically the wild type sequence; in this study, we have used the parent aptamer sequence). Specifically, ϕ is the ratio of the extent to which the substitution (de)stabilizes the folding transition state ($\Delta\Delta G_{U-\ddagger}$; that is, the difference between the substituted protein compared to the wild type in the change in free energy between the unfolded and transition state) to the extent to which it (de)stabilizes the native state ($\Delta\Delta G_f$; or, the difference between the substituted protein and wild type in folding free energy):

$$\frac{\Delta\Delta G_{u-\ddagger}}{\Delta\Delta G_f} = \frac{\Delta\ln(k_f)}{\Delta\ln(k_f) - \Delta\ln(k_u)} \quad (5)$$

Although the interpretation of ϕ can be complicated by issues such as residual structure in the unfolded state (see, e.g., discussion in Ref. 30), the most commonly held view is that ϕ reflects the extent to which the mutated side chain participates in native-like structure in the folding transition state.

To perform ϕ-value analysis on the aptamer we have used abasic substitutions in which the nucleobase on one sugar has been ablated, leaving behind its ribose-phosphate backbone. Specifically, we have characterized the folding kinetics of eight abasic mutants (Fig. 8 and Table I). Two of these, ab17 and ab28, produce ϕ of 0.80 ± 0.07 and 0.76 ± 0.04,
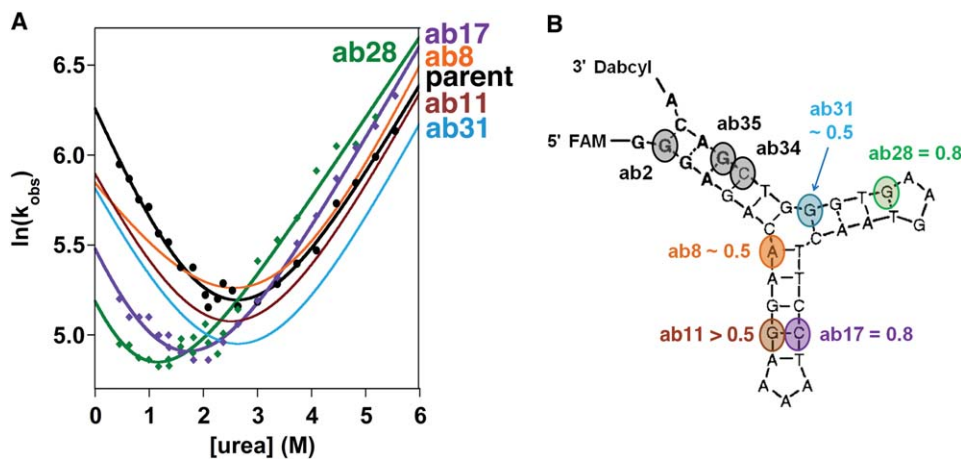
Figure 8. Φ value analysis suggests that the folding transition state of the aptamer may be rather well structured. **(A)** To probe this, we characterized eight abasic substitution variants. Unfortunately, however, only two of these eight, those at positions 17 and 28, produce changes in folding free energy that fall within the narrow window necessary to perform accurate φ-value analysis (see body of text). For both of these the observed φ value falls above 0.75. Although we cannot accurately estimate φ for the remaining three variants, we can nevertheless state that the φ-value for ab11 falls statistically significantly above 0.5. Data points collected are only shown for ab17, ab28, and the parent aptamer to minimize cluttering; the remaining data are presented in the SI (Supporting Information Fig. S1 and SI_phichev.xlsx). **(B)** Abasic sites mapped onto the proposed aptamer secondary structure. Those in gray failed to fold at all. For ab17 and ab28, φ are calculated directly from Eq. (4); for ab11, simple rearrangement of the φ relationship allows us to demonstrate that, although the exact φ-value cannot be determined with precision, it can be shown with good statistical significance to be greater than 0.5.

respectively. Unfortunately, however, because φ is the *ratio* of *differences* between experimental observables, this analytic approach tends to suffer from large errors unless the mutation produces a fairly large change in folding free energy.[31] Given that the folding free energy of the parent aptamer is just -7.5 kJ/mol, achieving an appropriately large denominator whilst still allowing the aptamer to fold is problematic. Problematic enough that three of the eight abasic variants we have characterized, ab2, ab34, and ab35, do not fold at all, precluding φ–value analysis. In contrast, the change in free energy associated with the remaining three substitutions are all too low to accurately constrain φ. This said, however, inspection of Eq. (5) shows that, if $-\Delta\ln(k_f) \geq \Delta\ln(k_u)$, φ must be ≥ 0.5. This relationship provides an alternate means of characterizing the transition state that, while less quantitative than formal φ-value analysis, has the advantage of not placing an often poorly constrained experimental value ($\Delta\Delta G_f$ is the difference in differences between four experimental measurements and thus often subject to significant error) in

the denominator. Using this approach we find that, in addition to the mutants ab17 and ab28, for which φ >0.75, the φ value observed for the abasic variant at position 11 is *also* above 0.5. (For positions 8 and 31 the values of $\Delta\ln(k_f)$ and $\Delta\ln(k_u)$ are within error and thus no such analysis is possible.)

It appears that, in terms of φ-defined structure, the folding transition state of our aptamer differs significantly from those of single-domain proteins. First, φ-value analysis suggests that the aptamer's folding transition state is *more* structured than would be predicted from the β_{den} analysis described above. This contrasts strongly with the case for single-domain proteins, for which β_{den} is invariably higher than the mean observed φ-value. The φ-values of the aptamer also differ from those of proteins in an absolute sense; whereas both of the well-defined φ values we have measured fall above 0.75, the mean φ value observed for a data set of 424 characterized substitutions in 15 single-domain proteins is just 0.28. Indeed, only 6% of all reportedly statistically significant φ values (i.e., those for which the reported error bars do

Table I. *Kinetic and Thermodynamic Parameters for the Parent Aptamer and Various Abasic Variants*

|	$\ln(k_f)$	$\ln(k_u)$	ΔG_u kJ/mol	Φ	$\Delta\Delta G_{u\text{-}\ddagger}$ kJ/mol	$\Delta\Delta G_u$ kJ/mol	$\Delta\ln(k_f)$	$\Delta\ln(k_u)$
parent	6.11 ± 0.03	3.23 ± 0.08	7.06 ± 0.12	–	–	–	–	–
ab8	5.95 ± 0.03	3.35 ± 0.08	6.38 ± 0.12	0.58 ± 0.25	0.40 ± 0.10	0.69 ± 0.26	-0.16 ± 0.04	0.12 ± 0.11
ab11	5.84 ± 0.03	3.18 ± 0.08	6.52 ± 0.12	1.20 ± 0.59	0.65 ± 0.10	0.54 ± 0.26	-0.27 ± 0.04	-0.05 ± 0.11
ab17	5.20 ± 0.03	3.46 ± 0.07	4.27 ± 0.13	0.80 ± 0.07	2.22 ± 0.11	2.79 ± 0.28	-0.91 ± 0.04	0.23 ± 0.11
ab28	4.81 ± 0.05	3.65 ± 0.06	2.86 ± 0.16	0.76 ± 0.04	3.18 ± 0.14	4.21 ± 0.31	-1.30 ± 0.06	0.42 ± 0.10
ab31	5.78 ± 0.03	3.00 ± 0.08	6.83 ± 0.12	3.33 ± 3.61	0.79 ± 0.10	0.24 ± 0.26	-0.33 ± 0.04	-0.23 ± 0.11
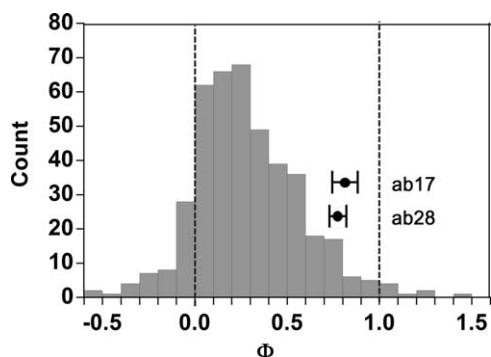
Figure 9. The aptamer's folding transition state would appear to contain significantly more φ-value defined structure than is typical for similarly simple, single-domain proteins. For example, although the two sites in the aptamer with precisely measurable φ-values both fall above 0.75, only 6% of the φ in a set of 424 single-site φ-values taken from the literature on 15 two-state proteins are similarly large (data set from Ref. 51).

not overlap with zero) are greater than 0.75 (Fig. 9). Likewise, although all three of the φ values we can put constraints on are above 0.5, only 21% of the φ values reported for single-domain proteins are similarly high (Fig. 9), and many of these are artifacts (i.e., not statistically significant; see discussion in Ref. 32). Moreover, we are not aware of *any* single domain protein for which more than 33% of the reported φ-values are above 0.5 (i.e., eight of 24 characterized positions in alpha spectrin SH3[33–35]), much less any proteins for which *all* reported φ-values are this high. Thus, as defined by φ, the aptamer's folding transition state appears to be much more highly structured than those of the equivalent single-domain proteins.

Discussion

The 38-base cocaine-binding aptamer of Stojanovic,[13] which is of similar complexity to a ~100 residue protein, folds via a process lacking readily observed, well-populated intermediates. The simplicity of this behavior allows for ready comparison with the apparently two-state folding kinetics of a large set of similarly simple, single-domain proteins. Making this comparison we find that, perhaps surprisingly, the aptamer's folding rate is quite close to the rate that would be predicted for a protein of similar topological complexity. We likewise find that, although the aptamer's Tanford β_{den} falls within the range of values observed for proteins, it falls at the lower end of this range suggesting that, as is often true for topologically simple proteins,[2] the hydrophobic interactions in its folding transition state are relatively poorly consolidated. Finally, the aptamer's folding differs from those of proteins in terms of its φ-value-defined transition state. That is, unlike the case for proteins, where β_{den}, which is a global measure of transition state consolidation, invariably predicts

more structure than is seen by φ-value analysis, for the aptamer this is reversed. Indeed, φ-value analysis suggests that many, if not the large majority, of the aptamer's nucleobases are in far more native-like environments in its folding transition state than is typical for the side chains of single domain proteins.

The late 1990s saw the development of a number of theories of protein folding kinetics, predominantly derived from simulations of the folding of highly simplified "toy models," including lattice polymers and simple, off-lattice Gō polymers. These include the hypothesis of Wolynes, that to ensure rapid folding it is sufficient that nature create a "funnel-shaped," monotonically decreasing "energy landscape" (reviewed in Ref. 36); the hypothesis of Karplus, that it is a necessary and sufficient condition to ensure rapid folding that the native state be a pronounced minimum relative to all other compact states and that folding rates are correlated with the size of the gap (e.g., Ref. 37); and the hypothesis of Thirumalai, that folding *kinetics* are related to the extent to which *equilibrium* collapse and folding are concomitant as the temperature or solvent quality drops from destabilizing to permissive of folding (e.g., Ref. 38).

Given that the above theories were derived from observations of the folding of highly simplified toy models no more (or less) protein-like than DNA-like, we would believe that their predictive value, if any, should also hold for the aptamer we have investigated. This said, the data we have collected here (and indeed, the vast body of data collected from proteins) provide unfortunately few direct tests of these theories (see discussion in Ref. 39). For example, as is true with proteins, there is no experimental method by which we can map the folding energy landscape of an aptamer, thus rendering it impossible to test the "funnel hypothesis" experimentally.[39] Likewise, in the absence of detailed knowledge of the aptamer's energy landscape we cannot estimate the size of its "energy gap," and thus cannot discern whether or not it is larger (or smaller) than those of biopolymers that fold faster (or slower). In contrast, our single-molecule FRET studies (Fig. 3) suggest that the aptamer's collapse and folding are more-or-less concomitant, as is predicted to be true for the most rapidly folding polymers.[38] Previous studies have shown, however, that collapse and folding are concomitant for many, if not all two-state proteins and thus that the extent to which these processes overlap is not, as predicted,[38] correlated with folding rates.[40]

Although the above arguments suggest that it is difficult to use our data (or, again, any experimental data–again, see discussion in Ref. 39) to directly test most current theories of folding kinetics, our data do provide an important, if indirect, test of lattice-

polymer-derived theories of folding kinetics. Specifically, based on the results of lattice polymer simulations multiple authors have argued that the properties necessary to support rapid folding (e.g., a sufficient energy gap, a smooth landscape, sufficiently concerted collapse and folding) are rare among even "foldable" sequences (i.e., sequences that adapt a unique and stable native conformation).[41–44] This prediction, however, is difficult to test using naturally occurring biopolymers, as these *must* fold in a biologically relevant timeframe to provide a selective advantage for their host. Our aptamer, however, was produced via an *in vitro* selection scheme that places no significant pressure on the sequence to fold rapidly (given the timescale of the typical selection, even folding times measured in hours would not produce a selective disadvantage [M. Stojanovic, Pers. Comm.]). The rapid folding of our aptamer thus speaks against an important, if indirect, prediction of many of the dominant existing theories of protein folding kinetics (for further discussion of this issue and its exploration using proteins see Ref. 45).

In contrast to the above arguments, our results may be consistent with the topomer search model of two-state folding. This model argues that, because the formation of local structural elements is rapid, the folding of any sufficiently cooperative polymer (i.e., polymers lacking well populated kinetic traps) will be limited by the entropic cost of finding the correct overall topology (the so-called "topomer search" process[3,46,47]). Moreover, this theory says nothing about the nature of the polymer itself; that is, the theory argues that, irrespective of the chemical makeup of the polymer the entropy of the topomer search will dominate the folding barrier for any sufficiently cooperative folding reaction. Given this, the correlation between the observed and the topology-predicted folding rates of this highly cooperative aptamer is perhaps gratifyingly close (Fig. 5), albeit with the obvious caveat that this speculation is based on a single observation that uses an, at best, low-resolution predicted structure and that relies on several assumptions regarding how to map measures of protein topological complexity onto an oligonucleotide. Given this, we look forward to watching how this story evolves as more data are reported regarding the folding kinetics of two-state nucleic acids.

Of course, the topomer search model is not without potential significant limitations. The model, for example, completely ignores the contribution of specific side chain interactions in the folding transition state. This said, the model does predict the experimentally observed folding rates of proteins with reasonable accuracy, suggesting that the impact of this omission is not significant. That is, although it is clear from φ-value analysis that side-chain interac-

tions form in the folding transition state, that the topomer search model ignores them and yet still predicts relative rates with good accuracy suggests that the φ-value-defined structure in the transition state plays only a minor role in explaining why some proteins fold more rapidly than others (see arguments in Gillespie and Plaxco[39]). Consistent with this, the φ values observed for two-state proteins are, as noted above, generally quite low (Fig. 9), arguing, perhaps, that the structure they reflect may not impact folding rates significantly (see discussion in Ref. 39). We note, too, that an alternative metric of residue-level transition state structure, Ψ-value analysis,[48] predicts that, for many proteins, the folding transition state involves the formation of a more native-like topology than may be implied by φ-value analysis.[49] Conversely, however, the limited φ-value analysis we have performed on our aptamer suggests that specific, native-like interactions are a common feature of its folding transition state. This suggests that the relative contribution of the topomer search process and the formation of specific, native-like interactions are better balanced in the folding of the aptamer than is the case for two-state proteins, suggesting in turn that, despite the correlation reported here (Fig. 5), the topomer search model may prove a less complete model of two-state DNA folding than it has of two-state protein folding.

The aptamer examined here exhibits similarities to proteins in its folding behavior, including highly concerted, two-state folding, a transition state that is only modestly consolidated in terms of hydrophobic burial and electrostatics, and, perhaps, the predictability of folding rates based on the topomer search model. That these similarities occur, despite the seemingly significant chemical differences between DNA and proteins, and despite the fact that the aptamer was created artificially and is not the product of natural selection suggests that they speak to properties that are universal in the folding of all sufficiently cooperative heteropolymers.

Materials and Methods

Materials

Urea (ultrapure) was purchased from Affymetrix (Cleveland, OH). Tris base and sodium chloride were purchased from Fisher (Fair Lawn, NJ). All buffers were filtered before use. Fluorescent oligonucleotides were purchased HPLC purified from IBA (Göttingen, Germany) and used as received. For all save our single molecule studies we used the fluorophore 5-Carboxyfluorescein (5-FAM) at the 5′ terminus and the quencher 4-((4-(dimethylamino)phenyl)azo) benzoic acid (DABCYL) on the 3′ terminus of the aptamer. For our single molecule studies we used a

construct with an Alexa fluor 488 at its 5′ terminus and an Alexa fluor 594 on its 3′ terminus.

Equilibrium titrations

Fluorescence was monitored at 520 nm using a Cary Eclipse fluorimeter (Varian). Instrument settings were: 5 nm slit widths, 490 nm excitation. Aptamer concentration was 100 nM. All experiments were performed at 22°C. The buffer used for urea titrations was 10 mM tris, 100 mM NaCl, pH 8.1. Salt titrations were performed in conditions identical to those for the urea titrations except that, instead of holding the NaCl concentration constant at 100 mM and varying the urea concentration, the urea was kept at a constant 4 M and the salt concentration was varied. All samples were equilibrated for 30 min at room temperature before measurement.

Using the nonlinear curve fitting function in Kaleidograph we fit our equilibrium unfolding data to Eq. (1) to obtain thermodynamic parameters from the urea equilibrium titration. We chose a four parameter fit (fixed, non-sloped baselines) as the use of a more complex model with fitted baseline slopes returned estimates of these slopes well within error of zero. Reported errors on parameters from the fit are estimated standard errors.

Single molecule measurements

Single-molecule measurements were performed on a custom built confocal microscope, described previously.[50] In brief, this consists of a confocal fluorescence microscope, with excitation provided by a 488 nm continuous laser. Individual photons, collected from molecules as they diffuse through the collection volume, are split onto a donor and acceptor channel for FRET measurements.

Kinetic measurements

Kinetic traces were captured using an Applied Photophysics SX.18MV stopped flow fluorimeter (Leatherhead, UK). The excitation wavelength was 490 nm, the path length was 10 mm, and a long pass filter of 495 nm was used. The volume ratio was 1:10. Post-mixing aptamer concentration was 180 nM. For folding experiments in urea, the unfolded aptamer in 4–6 M urea (depending on the mutant) was diluted into 10 mM tris, 100 mM NaCl, pH 8.1 buffer, with varying concentrations of urea. For unfolding experiments in urea, the folded aptamer sample before reaction was in 0 M urea.

We measured kinetics as a function of sodium chloride concentration under the same conditions we used for the urea kinetic experiments above (at a constant urea concentration of 4 M urea). However, these were slightly more complicated to capture; because of the limit 1:10 mixing ratio and the 6 order of magnitude span of sodium chloride concentrations, several starting (premixing) sodium chloride concentrations were necessary in the unfolding region of the plot.

We obtained the best fit values and estimated errors for $\ln(k_f)$, $\ln(k_u)$, m_f, m_u, and ϕ for kinetic data obtained over varying urea [Figs. 4(b) and 8(a)] using previously reported methods.[32] We fit the kinetic data obtained over varying salt to Eq. (3) and (4) using Kaleidograph, with the reported errors reflecting estimated standard errors.

Protein kinetic parameter data sets

The rate and long-range-order data set (Fig. 5) was adapted from a previously described data set[3] and used without modification. The tabulated β_{den} values we used (Fig. 6) were taken from two previously described data sets of two-state proteins[4,51]; all proteins in the two data sets were included except muscle acylphosphatase, the β_{den} of which was excluded as being erroneous (F. Chiti, Pers. Comm.). The ϕ-value data set (Fig. 9) we used comes from a set of the 15 two-state proteins that have been the most exhaustively characterized in terms of ϕ-values and was taken from the literature as presented.[51]

References

1. Jackson SE, Fersht AR (1991) Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. Biochemistry 30:10428–10435.
2. Plaxco KW, Simons KT, Baker D (1998) Contact order, transition state placement and the refolding rates of single domain proteins. J Mol Biol 277:985–994.
3. Makarov DE, Plaxco KW (2003) The topomer search model: a simple, quantitative theory of two-state protein folding kinetics. Protein Sci 12:17–26.
4. Maxwell KL, Wildes D, Zarrine-Afsar A, De Los Rios MA, Brown AG, Friel CT, Hedberg L, Horng J-C, Bona D, Miller EJ, Vallée-Bélisle A, Main ERG, Bemporad F, Qiu L, Teilum K, Vu N-D, Edwards AM, Ruczinski I, Poulsen FM, Kragelund BB, Michnick SW, Chiti F, Bai Y, Hagen SJ, Serrano L, Oliveberg M, Raleigh DP, Wittung-Stafshede P, Radford SE, Jackson SE, Sosnick TR, Marqusee S, Davidson AR, Plaxco KW (2005) Protein folding: defining a "standard" set of experimental conditions and a preliminary kinetic data set of two-state proteins. Protein Sci 14:602–616.
5. Matouschek A, Kellis JT, Serrano L, Fersht AR (1989) Mapping the transition state and pathway of protein folding by protein engineering. Nature 340:122–126.
6. De Los Rios MA, Daneshi M, Plaxco KW (2005) Experimental investigation of the frequency and substitution dependence of negative Φ-values in two-state proteins. Biochemistry 44:12160–12167.
7. Bonnet G, Krichevsky O, Libchaber A (1998) Kinetics of conformational fluctuations in DNA hairpin-loops. Proc Natl Acad Sci U S A 95:8602–8606.
8. Gray RD, Chaires JB (2008) Kinetics and mechanism of K+- and Na+-induced folding of models of human telomeric DNA into G-quadruplex structures. Nucleic Acids Res 36:4191–4203.
9. Narayanan R, Velmurugu Y, Kuznetsov SV, Ansari A (2011) Fast folding of RNA pseudoknots initiated by laser. J Am Chem Soc 133:18767–18774.

10. Bokinsky G, Zhuang X (2005) Single-molecule RNA folding. Accounts Chem Res 38:566–573.
11. Treiber DK, Williamson JR (2001) Beyond kinetic traps in RNA folding. Curr Opin Struct Biol 11:309–314.
12. Sosnick TR, Pan T (2004) Reduced contact order and RNA folding rates. J Mol Biol 342:1359–1365.
13. Stojanovic MN, de Prada P, Landry DW (2001) Aptamer-based folding fluorescent sensor for cocaine. J Am Chem Soc 123:4928–4931.
14. Neves MAD, Reinstein O, Johnson PE (2010) Defining a stem length-dependent binding mechanism for the cocaine-binding aptamer. A combined NMR and calorimetry study. Biochemistry 49:8478–8487.
15. Jäger M, Nguyen H, Crane JC, Kelly JW, Gruebele M (2001) The folding mechanism of a beta-sheet: the WW domain. J Mol Biol 311:373–393.
16. Qiu L, Pabit SA, Roitberg AE, Hagen SJ (2002) Smaller and faster: the 20-residue trp-cage protein folds in 4 micros. J Am Chem Soc 124:12952–12953.
17. Kubelka J, Eaton WA, Hofrichter J (2003) Experimental tests of villin subdomain folding simulations. J Mol Biol 329:625–630.
18. Levinthal C (1969) How to fold graciously. Mossbauer Spectroscopy in biological systems. Proc Univ Illinois Bull 67:22–24.
19. Jackson SE (1998) How do small single-domain proteins fold? Fold Des 3:R81–91.
20. Yoo TY, Meisburger SP, Hinshaw J, Pollack L, Haran G, Sosnick TR, Plaxco K (2012) Small-angle X-ray scattering and single-molecule FRET spectroscopy produce highly divergent views of the low-denaturant unfolded state. J Mol Biol 418:226–236.
21. Cekan P, Jonsson EO, Sigurdsson ST (2009) Folding of the cocaine aptamer studied by EPR and fluorescence spectroscopies using the bifunctional spectroscopic probe Ç. Nucleic Acids Res 37:3990–3995.
22. Miller EJ, Fischer KF, Marqusee S (2002) Experimental evaluation of topological parameters determining protein-folding rates. Proc Natl Acad Sci U S A 99:10359–10363.
23. De Los Rios MA, Plaxco KW (2005) Apparent Debye-Huckel electrostatic effects in the folding of a simple, single domain protein. Biochemistry 44:1243–1250.
24. Dove WF, Davidson N (1962) Cation effects on the denaturation of DNA. J Mol Biol 5:467–478.
25. Bond JP, Anderson CF, Record MT (1994) Conformational transitions of duplex and triplex nucleic acid helices: thermodynamic analysis of effects of salt concentration on stability using preferential interaction coefficients. Biophys J 67:825–836.
26. SantaLucia J (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc Natl Acad Sci U S A 95:1460–1465.
27. Manning GS (1979) Counterion binding in polyelectrolyte theory. Accounts Chem Res 12:443–449.
28. Debye P, Hückel E (1923) The interionic attraction theory of deviations from ideal behavior in solution. Phys Z 24:185–206.
29. Record MT (1967) Electrostatic effects on polynucleotide transitions. I. Behavior at neutral pH. Biopolymers 5:975–992.
30. Raleigh DP, Plaxco KW (2005) The protein folding transition state: what are Φ-values really telling us? Protein Peptide Lett 12:117–122.
31. De Los Rios MA, Muralidhara BK, Wildes D, Sosnick TR, Marqusee S, Wittung-Stafshede P, Plaxco KW, Ruczinski I (2006) On the precision of experimentally determined protein folding rates and Φ-values. Protein Sci 15:553–563.
32. Ruczinski I, Sosnick TR, Plaxco KW (2006) Methods for the accurate estimation of confidence intervals on protein folding Φ-values. Protein Sci 15:2257–2264.
33. Martinez JC, Serrano L (1999) The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. Nat Struct Mol Biol 6:1010–1016.
34. Cobos ES, Filimonov VV, Vega MC, Mateo PL, Serrano L, Martinez JC (2003) A thermodynamic and kinetic analysis of the folding pathway of an SH3 domain entropically stabilised by a redesigned hydrophobic core. J Mol Biol 328:221–233.
35. Fernandez-Escamilla AM, Cheung MS, Vega MC, Wilmanns M, Onuchic JN, Serrano L (2004) Solvation in protein folding analysis: combination of theoretical and experimental approaches. Proc Natl Acad Sci U S A 101:2834–2839.
36. Oliveberg M, Wolynes PG (2005) The experimental survey of protein-folding energy landscapes. Q Rev Biophys 38:245–288.
37. Dinner AR, Abkevich V, Shakhnovich E, Karplus M (1999) Factors that affect the folding ability of proteins. Proteins 35:34–40.
38. Camacho CJ, Thirumalai D (1996) A criterion that determines fast folding of proteins: a model study. Europhys Lett 35:627–632.
39. Gillespie B, Plaxco KW (2004) Using protein folding rates to test protein folding theories. Annu Rev Biochem 73:837–859.
40. Millet IS, Townsley LE, Chiti F, Doniach S, Plaxco KW (2002) Equilibrium collapse and the kinetic "foldability" of proteins. Biochemistry 41:321–325.
41. Shakhnovich E, Farztdinov G, Gutin AM, Karplus M (1991) Protein folding bottlenecks: a lattice Monte Carlo simulation. Phys Rev Lett 67:1665–1668.
42. Sali A, Shakhnovich E, Karplus M (1994) How does a protein fold? Nature 369:248–251.
43. Sali A, Shakhnovich E, Karplus M (1994) Kinetics of protein folding. A lattice model study of the requirements for folding to the native state. J Mol Biol 235:1614–1636.
44. Dinner AR, So SS, Karplus M (1998) Use of quantitative structure-property relationships to predict the folding ability of model proteins. Proteins 33:177–203.
45. Gillespie B, Vu DM, Shah PS, Marshall SA, Dyer RB, Mayo SL, Plaxco KW (2003) NMR and temperature-jump measurements of de novo designed proteins demonstrate rapid folding in the absence of explicit selection for kinetics. J Mol Biol 330:813–819.
46. Debe DA, Goddard WAI (1999) First principles prediction of protein folding rates. J Mol Biol 294:619–625.
47. Makarov DE, Keller CA, Plaxco KW, Metiu H (2002) How the folding rate constant of simple, single-domain proteins depends on the number of native contacts. Proc Natl Acad Sci U S A 99:3535–3539.
48. Krantz BA, Sosnick TR (2001) Engineered metal binding sites map the heterogeneous folding landscape of a coiled coil. Nat Struct Biol 8:1042–1047.
49. Yoo TY, Adhikari A, Xia Z, Huynh T, Freed KF, Zhou R, Sosnick TR (2012) The folding transition state of protein L is extensive with nonnative interactions (and not small and polarized). J Mol Biol 420:220–234.
50. Pfeil SH, Wickersham CE, Hoffmann A, Lipman EA (2009) A microfluidic mixing system for single-molecule measurements. Rev Sci Instrum 80:055105.
51. Lawrence C, Kuge J, Ahmad K, Plaxco KW (2010) Investigation of an anomalously accelerating substitution in the folding of a prototypical two-state protein. J Mol Biol 403:446–458.