Probing the protein-folding mechanism using denaturant and temperature effects on rate constants

Emily J. Guinn[a,b,1,2], Wayne S. Kontur[c,2], Oleg V. Tsodikov[d,2], Irina Shkel[a,b], and M. Thomas Record, Jr.[a,b,3]

Departments of [a]Chemistry, [b]Biochemistry, and [c]Bacteriology, University of Wisconsin, Madison, WI 53706; and [d]Department of Pharmaceutical Sciences, University of Kentucky, Lexington, KY 40536

Edited* by Robert L. Baldwin, Stanford University, Stanford, CA, and approved August 13, 2013 (received for review June 24, 2013)

Protein folding has been extensively studied, but many questions remain regarding the mechanism. Characterizing early unstable intermediates and the high–free-energy transition state (TS) will help answer some of these. Here, we use effects of denaturants (urea, guanidinium chloride) and temperature on folding and unfolding rate constants and the overall equilibrium constant as probes of surface area changes in protein folding. We interpret denaturant kinetic *m*-values and activation heat capacity changes for 13 proteins to determine amounts of hydrocarbon and amide surface buried in folding to and from TS, and for complete folding. Predicted accessible surface area changes for complete folding agree in most cases with structurally determined values. We find that TS is advanced (50–90% of overall surface burial) and that the surface buried is disproportionately amide, demonstrating extensive formation of secondary structure in early intermediates. Models of possible pre-TS intermediates with all elements of the native secondary structure, created for several of these proteins, bury less amide and hydrocarbon surface than predicted for TS. Therefore, we propose that TS generally has both the native secondary structure and sufficient organization of other regions of the backbone to nucleate subsequent (post-TS) formation of tertiary interactions. The approach developed here provides proof of concept for the use of denaturants and other solutes as probes of amount and composition of the surface buried in coupled folding and other large conformational changes in TS and intermediates in protein processes.

kinetics | protein denaturation | stability | native fold | collapse

Determination of the mechanism of protein folding [unfolded (U) → folded (F)] is a long-standing goal of biophysical research. Folding of a single domain globular protein is a very highly cooperative (thermodynamically two-state) process. From analysis of folding kinetic data for CI2 and barnase, Fersht et al. (1) concluded that proteins fold through a high–free-energy transition state (TS) with partially formed elements of native structure. Recently, Barrick and Sosnick (2) concluded that an ensemble of unstable, rapidly reversible intermediates form early in the folding mechanism, and that the most advanced and unstable of these undergoes a rate-determining conformational change with transition state TS. This transit step, slower than the reverse direction of previous rapidly reversible steps, is followed by rapid propagation of folding. Most proposals for the initial intermediates invoke unstable regions of α-helix and/or β-sheet; these are thought to coalesce and/or rearrange to form TS. Kay and coworkers (3) characterized a marginally stable early intermediate of Fyn SH3 domain and found that interactions of amide groups formed earlier in the folding pathway than interactions of methyl groups, indicating that 2° structure formed before the 3° fold. Recent hydrogen exchange pulse-labeling experiments analyzed with mass spectrometry by Englander, Marqusee, and collaborators (4) indicate that folding of RNase H occurs through a defined set of intermediates where each step of folding adds more native-like elements of 2° structure. Late intermediates in folding that contain most elements of native structure except side-chain close packing have been observed (5).

Mutational analyses of many proteins indicate that TS for protein folding are quite advanced (2). Here, we develop a different approach to characterize TS for noncovalent self-assembly processes like protein folding using solute effects and heat capacity changes as physical probes yielding amounts of hydrocarbon and amide surface buried in folding to and from TS, and apply it to characterize TS and infer properties of intermediates before TS for folding of 13 proteins. In addition to probing the character of TS in the protein-folding mechanism in a unique way, this research serves as a demonstration of the use of solutes as probes of conformational changes and interface formation in the steps of any protein process.

The vast majority of the accessible surface area (ASA) buried in folding an extended polypeptide into a globular protein is either hydrocarbon (H) (~70%) or amide (A) (~20%), with an average ratio $\Delta ASA_H/\Delta ASA_A \sim 3.5$ for complete folding (6). [Structural values of $\Delta ASA_H/\Delta ASA_A$ range from 2.7 to 3.9 for 12 of the proteins analyzed here; only one (FKBP) has a ratio outside this range (~5).] Different folding mechanisms lead to very different predictions of the $\Delta ASA_H/\Delta ASA_A$ ratio for folding from an extended chain to TS. Formation of isolated elements of 2° structure (particularly α-helix) buries proportionately less hydrocarbon surface and proportionately more amide surface than does overall folding. For example, $\Delta ASA_H/\Delta ASA_A$ is 0.6 for formation of an $(AEAAKA)_n$ α-helix (6) and 2.0 for formation of isolated α-helical elements present in phage 434 cro repressor (calculation described below), whereas for β-hairpin [Protein Data

Significance

Analysis of effects of denaturants and temperature on folding and unfolding rate constants of 13 globular proteins yields the amount and composition of the surface buried in folding to and from the high–free-energy transition state (TS), and thereby provides information about folding mechanisms and early unstable folding intermediates. All 13 proteins preferentially bury amide surface in folding to TS; amounts of amide and hydrocarbon surface buried in folding to TS generally exceed those buried in forming the native secondary structure. From this, we conclude that most native secondary structure forms in early unstable folding intermediates and that conversion of these intermediates to TS involves nucleation of tertiary interactions, allowing preferential burial of hydrocarbon surface as TS folds.

Author contributions: E.J.G., W.S.K., O.V.T., and M.T.R. designed research; E.J.G., W.S.K., and O.V.T. performed research; E.J.G., W.S.K., O.V.T., I.S., and M.T.R. analyzed data; and E.J.G. and M.T.R. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

See Commentary on page 16704.

[1]Present address: California Institute for Quantitative Biosciences, University of California, Berkeley, CA 94720.

[2]E.J.G., W.S.K., and O.V.T. contributed equally to this work.

[3]To whom correspondence should be addressed. E-mail: mtrecord@wisc.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1311948110/-/DCSupplemental.

Bank (PDB) ID code 2EVQ] formation $\Delta ASA_H/\Delta ASA_A \sim 2.7$. For hydrophobic collapse mechanisms, the ratio $\Delta ASA_H/\Delta ASA_A$ for folding to TS is presumably as large or larger than the value for complete folding.

Denaturant Probes of Changes in Amide and Hydrocarbon Surface Area

Recently, the interactions of key denaturants, osmolytes, and Hofmeister salts with the functional groups of proteins and nucleic acid have been determined, allowing these solutes to be used as thermodynamic and kinetic–mechanistic probes of the amount and the composition of the surface buried or exposed in the steps of a noncovalent protein or nucleic acid process like helix formation, folding, or binding (6–14). For protein folding, urea and guanidinium chloride (GuHCl) thermodynamic *m*-values (derivatives of the standard free energy of folding with respect to denaturant concentration) are determined almost entirely by the preferential interactions of these denaturants with the hydrocarbon and amide surface buried in folding (6):

$$m\text{-value} \approx \alpha_H \Delta ASA_H + \alpha_A \Delta ASA_A, \quad [1]$$

where α_H and α_A (Table 1) are interaction potentials quantifying the preferential interaction of urea or GuHCl per unit area with these two types of surface, obtained from analysis of thermodynamic data for the preferential interactions of urea and GuHCl with model compounds (6, 10). For GuHCl, α_A is 20-fold larger in magnitude than α_H (Table 1) (10). For overall protein folding (U → F), therefore, $\alpha_A \Delta ASA_A$ is more than fivefold larger in magnitude than $\alpha_H \Delta ASA_H$, so GuHCl folding *m*-values are primarily determined by ΔASA_A. For urea, α_A is 3.7-fold larger than α_H (Table 1) (6), and therefore contributions from ΔASA_A and ΔASA_H to urea folding thermodynamic *m*-values are similar. For cases in which folding and unfolding kinetics are single-exponential (2, 14), Eq. **1** is also applicable to interpret kinetic *m*-values (obtained from derivatives of the logarithms of the folding and unfolding rate constants with respect to denaturant concentration), in which case ΔASA_H and ΔASA_A are the changes in water-accessible hydrocarbon and amide surface area in folding to or from TS (U → TS; TS → F).

Heat Capacity Probes of Changes in Hydrocarbon (and Amide) Surface Area

Neither denaturant is optimal as a probe of changes in hydrocarbon ASA (ΔASA_H), but the heat capacity change is. The overall (thermodynamic) heat capacity change for folding (ΔC_p°) is determined to a good approximation by the amounts of hydrocarbon and amide surface buried in folding (15–17):

$$\Delta C_p^\circ \approx \varepsilon_H \Delta ASA_H + \varepsilon_A \Delta ASA_A, \quad [2]$$

where ε_H and ε_A are intrinsic values of ΔC_p° per unit area for these two types of surface, obtained from analysis of hydrocarbon and amide transfer thermodynamic data (Table 1). Whereas ε_H is only 2.4-fold larger in magnitude than ε_A, the contribution of $\varepsilon_H \Delta ASA_H$ to the overall ΔC_p° of protein folding is ∼9-fold larger in magnitude than $\varepsilon_A \Delta ASA_A$. Eq. **2** (like Eq. **1**) is also applicable to interpret activation heat capacity data obtained from the second derivatives of the unfolding and folding rate constants with respect to temperature.

Table 1. α-Values for urea and GuHCl (Eq. 1) and ε-values (Eq. 2) for the composition of the amide and hydrocarbon surfaces buried in protein folding (*SI Materials*)

Property	Amide	Hydrocarbon
GuHCl α-value, cal·mol^{-1}·m^{-1}·Å^{-2}	−2.2 ± 0.2	0.11 ± 0.05
Urea α-value, cal·mol^{-1}·m^{-1}·Å^{-2}	−0.42 ± 0.01	−0.11 ± 0.03
ΔCp ε-value, cal·mol^{-1}·K^{-1}·Å^{-2}	−0.14 ± 0.04	0.34 ± 0.02

Results and Discussion

Quantifying Amounts of Amide and Hydrocarbon Surface Buried in Folding to and from TS. The kinetics of folding and unfolding of 13 single-domain globular proteins with known structures and ranging in size from 56 to 164 residues have been determined as a function of denaturant concentration and temperature (18–32). GuHCl or urea kinetic *m*-values and activation heat capacity changes for folding and unfolding are summarized in Table S1. Amounts of amide and hydrocarbon surface buried in folding to and from TS (i.e., U → TS and TS → F) and for overall folding (U → F), predicted from activation heat capacity change and denaturant kinetic *m*-values using Eq. **S1**, are reported in Table 2.

Thermodynamic *m*-values and ΔC_p° values for overall folding (U → F) are determined from differences between kinetic/activation quantities for folding and unfolding. To validate the use of kinetic *m*-values and activation ΔC_p° values as structural probes to determine amounts of amide and hydrocarbon ASA buried in folding to and from TS, we compared values of ΔASA_H and ΔASA_A for U → F predicted from the thermodynamic *m*-value and ΔC_p° with those calculated from structural data assuming the extended polypeptide model of the denatured state (6). Fig. 1 and Table 2 demonstrate the very good agreement obtained between predicted and structural values of ΔASA_H and ΔASA_A for most of the proteins in the dataset. Individual (Fig. 1) and summed (Fig. S1) values of ΔASA_H and ΔASA_A predicted from *m*-values and activation heat capacities agree within experimental uncertainty with those calculated from structural information for 9 of 13 proteins. (See *SI Materials* for discussion of uncertainties in ΔASA and of proteins deviating from the line in Fig. 1.) Fig. S1 shows that, in seven cases, the experimentally determined value of ΔASA_H plus ΔASA_A is equal to or slightly larger in magnitude than that calculated from the structural data, validating the use of the extended model of the unfolded chain for these calculations.

To determine how advanced TS is on the folding pathway from U and F, the fraction θ of the overall denaturant *m*-value or heat capacity change that occurs in formation of TS has been calculated. As noted by Plaxco et al. (33) for a subset of these proteins, the fraction $\theta_C^{U\to TS}$ of the overall heat capacity change that occurs in folding to TS of these 13 proteins is less than the corresponding fraction $\theta_m^{U\to TS}$ of the thermodynamic denaturant *m*-value for folding to TS (Table S2). Because burial of hydrocarbon surface largely determines the heat capacity change, and burial of amide surface is equally or more important for determining denaturant *m*-values as discussed above, the ranking $\theta_m^{U\to TS} > \theta_C^{U\to TS}$ indicates that amide ASA is preferentially buried in folding to TS.

Values of the fraction $\theta_{ASA}^{U\to TS}$ of total (hydrocarbon plus amide) ASA buried in folding (Eq. **S2**) ranges from 0.53 to 0.94, with 9 of the 13 proteins in the range 0.53–0.70 (Fig. 2 and Table S2). For all 13 proteins, more than one-half of the total (amide plus hydrocarbon) surface buried in folding is buried in folding to TS; the average for the dataset is 65 ± 10%. Mutational analyses and simulations have previously concluded that TS of many proteins is advanced as discussed below. For all 13 proteins, values of $\theta_{ASA}^{U\to TS}$ are smaller than $\theta_m^{U\to TS}$ and greater than $\theta_C^{U\to TS}$, as expected.

Several metrics can be used to quantify the extent to which amide or hydrocarbon ASA is preferentially buried in folding to TS and from TS to the globular protein. The simplest, applicable to each protein individually, compares hydrocarbon:amide ΔASA ratios r, defined as follows:

$$r = \Delta ASA_H/\Delta ASA_A, \quad [3]$$

Values of r for U → TS, TS → F, and U → F are listed in Table S3. For all 10 proteins where values of r for TS → F can be accurately determined, Table S3 demonstrates that $r_{U\to TS} <$

Table 2. Predicted values of amide and hydrocarbon ΔASA for U → TS, TS → F, and U → F; comparison with structural values for U → F

Protein	PDB ID code	Predicted U → TS ΔASA		Predicted TS → F ΔASA		Predicted U → F ΔASA		U → F ΔASA from structure*	
		Hydrocarbon, Å^2	Amide, Å^2	Hydrocarbon, Å^2	Amide, Å^2	Hydrocarbon, Å^2	Amide, Å^2	Hydrocarbon, Å^2	Amide, Å^2
NTL9	1CQU (51)	−1,282 ± 339	−471 ± 86	−958 ± 335	−328 ± 61	−2,241 ± 485	−799 ± 128	−2,726†	−729†
Bc-Csp	1C9O (52)	−1,800 ± 122	−801 ± 106	−908 ± 45	−133 ± 27	−2,708 ± 154	−934 ± 129	−3,436	−1213
Bs-Csp	1NMG (53)	−2,292 ± 238	−923 ± 238	−207 ± 266	−2 ± 71	−2,499 ± 358	−926 ± 256	−3,425	−1,054
FynSH3	1NYF (54)	−1,534 ± 94	−577 ± 75	−1,362 ± 66	−306 ± 48	−2,905 ± 152	−883 ± 122	−2,726	−691
434 Cro	2CRO (55)	−2,673 ± 469	−1,278 ± 333	−840 ± 506	−113 ± 147	−3,514 ± 698	−1,391 ± 400	−3,705†	−1,106†
PLB1	1HZ6 (56)	−1,231 ± 115	−703 ± 84	−686 ± 101	−308 ± 39	−1,916 ± 180	−1,011 ± 122	−3,259	−1,220
CI2	2CI2 (57)	−1,672 ± 95	−562 ± 75	−1,575 ± 93	−397 ± 59	−3,248 ± 174	−959 ± 134	−3,259	−834
HPr	1POH (58)	−2,629 ± 186	−890 ± 125	−2,370 ± 346	−542 ± 88	−4,999 ± 428	−1,431 ± 207	−4,656	−1,525
ACP	1APS (59)	−2,472 ± 256	−1,647 ± 344	−2,638 ± 166	94 ± 174	−5,110 ± 326	−1,554 ± 439	−5,391	−1,652
ACP-1	2VH7 (60)	−2,647 ± 346	−2,215 ± 455	−2,786 ± 266	−409 ± 207	−5,434 ± 491	−2,624 ± 608	−5,765	−1,818
RPS6	1RIS (61)	−3,026 ± 315	−909 ± 127	−3,065 ± 431	−494 ± 96	−6,091 ± 564	−1,402 ± 219	−5,796	−1,561
FKBP	1FKD (62)	−2,760 ± 271	−1,916 ± 396	−3,243 ± 253	−447 ± 242	−6,003 ± 437	−2,364 ± 588	−6,529†	−1,295†
T4L	1B6I (63)	−5,534 ± 570	−1,796 ± 255	−1,841 ± 189	−542 ± 79	−7,375 ± 643	−2,338 ± 328	−9,806	−2,606

*ΔASA calculated for an extended model of the denatured state.
†ΔASA values from Guinn et al. (6).

$r_{TS\to F}$, showing that folding to the TS preferentially buries amide ASA, whereas folding from the TS to the globular protein preferentially buries hydrocarbon ASA.

Because overall ratios $r_{U\to F}$ are different for different proteins, we normalize $r_{U\to TS}$ and $r_{TS\to F}$ by the thermodynamically determined $r_{U\to F}$ and define these ratios as Ω values as follows:

$$\Omega_{U\to TS} = r_{U\to TS}/r_{U\to F}; \quad \Omega_{TS\to F} = r_{TS\to F}/r_{U\to F}, \qquad [4]$$

For all proteins $\Omega_{U\to TS} \le 1$ (13 examples) and $\Omega_{TS\to F} \ge 1$ (10 examples), confirming the preferential burial of amide ASA in U → TS and of hydrocarbon ASA in TS → F (Fig. 3).

Interpreting Amide ΔASA_A Values for Folding to TS; Comparison with Results of Other Methods. Values of ΔASA_A for folding to TS determined from the analysis of kinetic *m*-values and activation heat capacities (Table 2) allow us to estimate the amount of secondary structure formed in folding to TS for each protein. Formation of $(AEAAKA)_n$ α-helix (6) and β-hairpin (PDB ID code 2EVQ) buries ∼22 Å^2 and ∼12 Å^2 of backbone amide ASA per residue, respectively. Calculations of amide ASA (34) buried in formation of individual 2° structural elements of Bs-Csp, HPr, and phage 434 Cro protein (434 Cro) agree well with these values (Table S4). From the number of residues in α-helices or β-sheets in these folded proteins and the above per-residue values, we estimate the contribution to ΔASA_A from burial of the chain backbone in forming all of the native 2° structure in folding from U to F (Table S5) and compare with values of ΔASA_A determined from the experimental data for folding from U to TS (Table 2). For 11 of the 13 proteins investigated, the amount of amide ASA buried in folding from U to TS is either greater than (seven proteins) or approximately equal to (four proteins) the amount of backbone amide ASA buried in forming all of the native 2° structure (Table S5). Only for RPS6 and T4L is the amount of amide ASA buried in folding from U to TS less than the amount of backbone amide ASA buried in forming all of the native 2° structure (Table S5). This comparison indicates most simply that the 2° structural elements of these folded proteins are already present in TS. The additional amide ASA buried in seven of the TS investigated may result from reductions in side-chain amide ASA in 2° structure formation or interactions of 2° structures, or from organization of residues that are not involved in 2° structures formed in TS.

Analysis of All-2° and More Advanced Structural Models of Folding Intermediates and TS. As a structural test of the interpretation that the native 2° structure is present in TS, models of putative

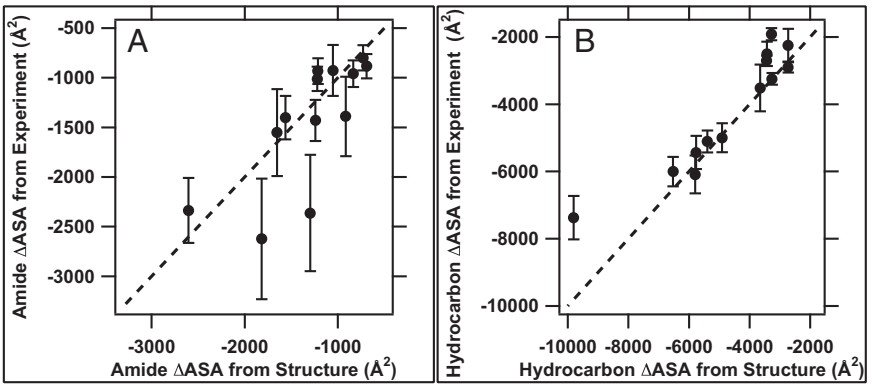

Fig. 1. Comparison of amide (*A*) and hydrocarbon (*B*) ΔASA values for U → F from analysis of experimental data (using Eqs. **1** and **2**) with values calculated from structural data for F and an extended-chain model for U (Table 2). The lines represent equality of experimental and structural values. Fig. S1 compares experimental and structural values for the sum of amide and hydrocarbon ΔASA for U → F.

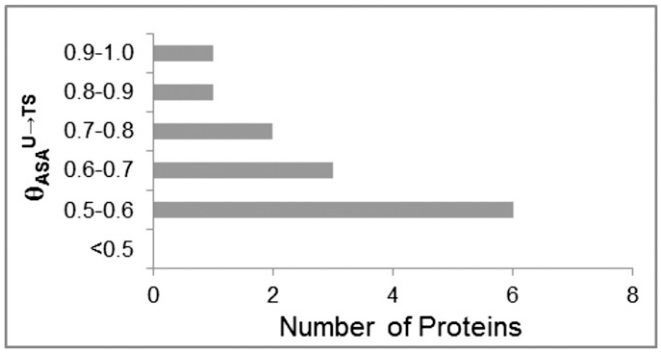

Fig. 2. Distribution of $\theta_{ASA}^{U\rightarrow TS}$ values quantifying the fraction of the total hydrocarbon plus amide ΔASA that is buried in folding to TS for 13 proteins (Table S2).

folding intermediates (designated I_{ss}), which preserve the native *2°* structure (abbreviated ss) but eliminate as much as possible of the native *3°* structure, were made for proteins with α-helical (434 cro), β-sheet (Bs-Csp), and mixed (HPr) secondary structures. In these models, the native elements of α-helix and β-sheet structure are connected to the extent possible by extended chain segments like those in the model of U. Values of amide and hydrocarbon ΔASA for folding of U to form these proposed I_{ss} intermediates are summarized in Table S4, together with θ and r values calculated from these ΔASA values.

Table S4 reveals that I_{ss} models for all three proteins examined are significantly less advanced than TS in terms of the sum of amide and hydrocarbon ASA buried ($\theta_{ASA}^{U\rightarrow Iss} << \theta^{U\rightarrow TS}$). For all-α 434 Cro and all-β Bs-Csp, the amount of amide surface buried in forming all of the elements of native 2° structure is only 50–60% of that buried in folding to TS, whereas for HPr our I_{ss} model buries almost 100% of the amount of amide surface buried in folding to TS. To create more advanced models that may approximate the TS, we added native-like φ and ψ angles in the regions connecting 2° structure elements but avoided creating the contacts between these elements present in the native state. For all three proteins, we found more advanced models for which ΔASA^{TS}, $\theta_{ASA}^{U\rightarrow TS}$, and r values (Table S4) are more comparable to those predicted from ΔC_p and *m*-values and so designate these as TS. These structures for 434 Cro and Bs-Csp are shown on the activation free-energy diagrams of Fig. 4 using θ_{ASA} values for the fraction of total ASA buried in folding to the structure as the position on the reaction coordinate.

Comparison with Other Models of TS. TS for several of the proteins studied here have been investigated previously by other methods. Relaxation dispersion NMR spectroscopy showed that a marginally stable intermediate in Fyn-SH3 (1NYF) folding contains the β-structure formed by the central portion of the polypeptide chain, but lacks the 2° structural elements nearer the N and C termini. The kinetic *m*-value/activation heat capacity analysis presented here predicts that ∼56% of the ΔASA of folding occurs in forming the high–free-energy Fyn-SH3 TS and that the amount of amide surface buried in forming TS is ∼1.5-fold greater than the backbone amide surface buried in forming the elements of Fyn-SH3 2° structure. Our results indicate that TS may be quite close on the reaction coordinate to the intermediate characterized by Kay and coworkers. Mutant ψ-value analysis, where metal ion binding sites are introduced to probe for contacts in TS, indicates that TS for protein L (1HZ6) contains the entire (28-residue) β-sheet network (35) (and some nonnative structure) but little if any of the native (21-residue) α-helical structure. Simulations and Φ-value analysis, where the effect of Ala substitutions on TS stability are observed, indicate that TS for chymotrypsin inhibitor (2CI2) contains the 14 residues of α-helix but not the 18 residues of β-sheet of the native protein (36). The kinetic *m*-value/activation heat capacity analysis presented here predicts that the amide ASA buried in forming TS is comparable to the amount of backbone amide surface predicted to be buried in forming all of the 2° structure of both these proteins (Table S5). Possibly there is extensive burial of side-chain amide surface in the TS of these proteins, or possibly the mutation analyses may underestimate the amount of 2° structure in TS. Additional experiments and structural analysis (like that performed here) will be needed to distinguish between these and other explanations.

Structural models of the proposed TS ensemble for six of the proteins investigated here have been predicted from unfolding or folding simulations. These included two small all-β cold shock proteins [Bc-Csp (37), Bs-Csp (38)] and four larger mixed α/β proteins [chymotrypsin inhibitor (39, 40), HPr (41), acylphosphatase (42), and ribosomal protein S6 (43)]. Properties like radius of gyration or numbers of native and nonnative contacts are reported for U, TS, and F states, but these characteristics are not directly related to ΔASA_H and ΔASA_A. Qualitatively, these simulated TS ensembles are quite advanced on the folding pathway, in agreement with the experimentally determined results in Fig. 2 and Table S2. However, simulations in many cases predict a relatively concerted formation (or disruption) of *2°* and *3°* structure in folding (or unfolding) to TS, whereas analysis of the experimental data indicates that most if not all of the 2° is formed before TS in folding. ASA calculations on predicted TS and intermediate complexes from simulations will permit more detailed comparisons.

Implications for the Protein-Folding Mechanism. The amide and hydrocarbon ASA results and structural analysis presented here provide strong experimental support for a general mechanism of protein folding in which most if not all of the native elements of secondary structure form before TS, and begin to coalesce into more native-like structures in TS, as previously proposed (2, 4, 44–46). A rapidly equilibrating mixture of largely unfolded chains with various subsets of these elements of 2° structure comprises the ensemble of early folding intermediates. Nucleating each structural element is unfavorable and propagation in the observed range of element lengths is not sufficiently favorable to make them stable so formation of each is characterized by an equilibrium constant $K_i < 1$; the overall equilibrium constant for forming a species with all these elements present (the putative I_{ss}) is the product of these K_i ($K_{Iss} = \Pi K_i$). Although I_{ss} should be very unstable, ASA analysis of I_{ss} and possible TS species for 434 Cro and Bs-Csp described above indicates that I_{ss}

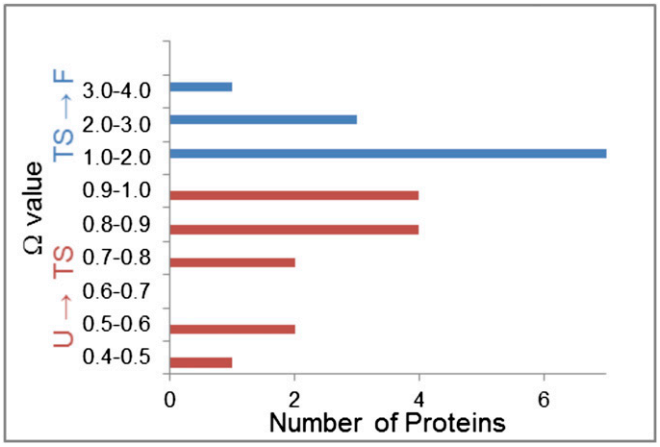

Fig. 3. Distribution of Ω-values quantifying preferential burial of amide surface in folding to TS ($\Omega_{U\rightarrow TS} \leq 1$) and preferential burial of hydrocarbon surface in folding from TS to F ($\Omega_{TS\rightarrow F} \geq 1$) (Table S3).

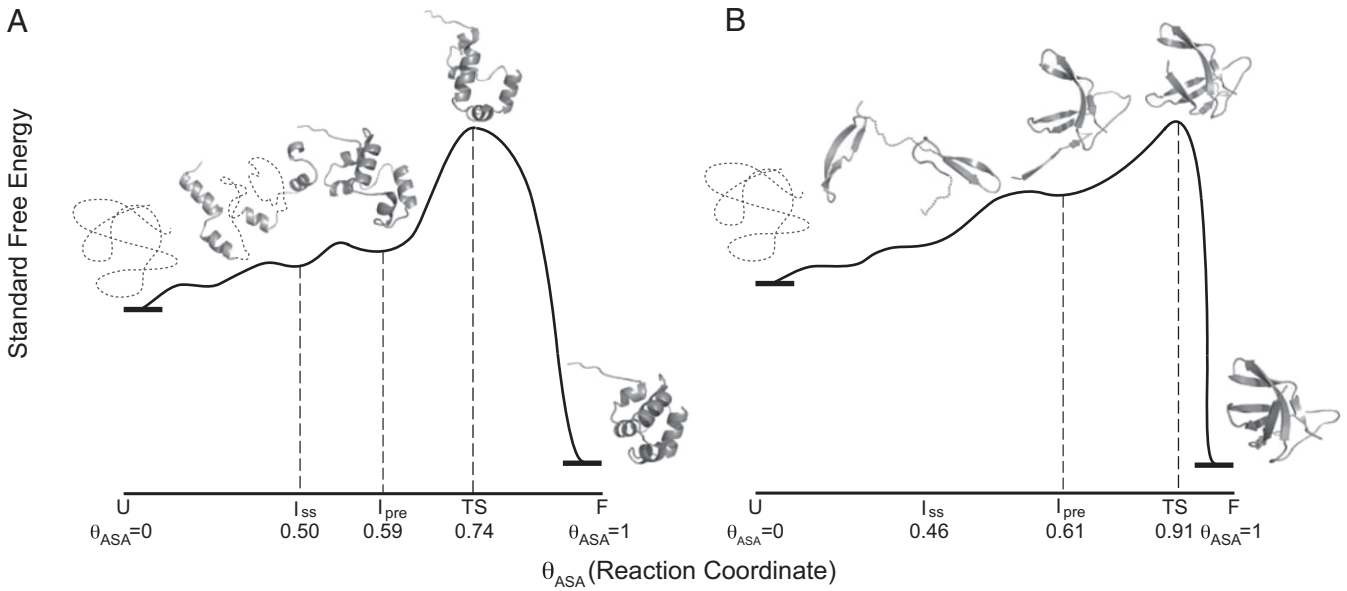

Fig. 4. Proposed reaction progress diagrams for (*A*) 434 cro (all α-helix) and (*B*) Bs-Csp (all β-sheet). Models of the TS and of earlier folding intermediates, one with the native secondary structure (I_{ss}) and a more advanced intermediate in which tertiary structure formation is nucleated [a candidate for I_{pre} (2)], are placed along the reaction coordinate based on their θ_{ASA}-values (Table S5). Free energies are arbitrarily chosen; only the ASA scale is quantitative.

is significantly less advanced than TS in terms of ASA buried, and therefore that I_{ss} may not be the least stable on-pathway folding intermediate before TS (i.e., I_{pre}). Therefore, organization of these elements of 2° structure into native-like tertiary structure begins before the TS.

In the Barrick–Sosnick analysis (2), the folding rate constant $k_f = K_{Ipre}k_{transit}$, where K_{Ipre} is the equilibrium constant for forming I_{pre} from U and $k_{transit}$ is the rate constant for the rate-determining step with TS. We conclude that $K_{Iss} \geq K_{Ipre}$ and that the transit step with rate constant $k_{transit}$ is either (*i*) organization of the backbone (and side chains) of one or more of the regions that are not part of the 2° structure in the most unstable on pathway intermediate, (*ii*) coalescence of one or more elements of structure, or (*iii*) both. TS models like those in Fig. 4 with sufficient backbone organization to nucleate formation of the native 3° structure from the isolated elements of 2° structure agree with the ASA characteristics of TS.

Conclusions

The analysis of kinetic *m*-values and activation heat capacity changes for protein folding and unfolding presented here quantifies the changes in amide and hydrocarbon surface area (ASA) in folding to and from the highest free energy TS; this information was not obtainable by other methods. We propose that changes in amide, hydrocarbon, and total ASA are the natural reaction coordinates of protein folding. From both the kinetic and structural analyses, we conclude that TS has both the elements of 2° structure of the native protein and the organization of other regions of the backbone required to nucleate the formation of 3° structure in subsequent folding from TS. The instability of TS is the result of the instability of the individual elements of 2° structure and of nucleation of 3° structure; subsequent propagation of 3° structure formation is of course highly favorable.

In addition to its contribution to our understanding of the protein-folding mechanism, this research provides proof of concept for the use of solute *m*-values and heat capacity changes to probe mechanisms, determine steps with coupled folding or other large conformational changes, and quantify the amount and composition of the surface buried or exposed in these intermediates or TS, information that is difficult or impossible to obtain by other methods. An example is the mechanism of open complex formation and stabilization in transcription initiation by *Escherichia coli* RNA polymerase (RNAP) at a DNA promoter (47). Here, information about interactions of solutes and noncoulombic (Hofmeister) interactions of salt ions with nucleic acid functional groups is also needed; this has been obtained for urea (11, 12), and research with other solutes and salts is in progress. The small effects of urea and Hofmeister salts on the rate and equilibrium constants for the step involving opening of 13 bp of duplex promoter DNA provide evidence that RNAP opens this region in the active site cleft (14, 48, 49). The large effects of urea, glycine betaine, and different Hofmeister salts on the subsequent step(s) that stabilize the initial open complex provide evidence for large-scale coupled folding and assembly of mobile elements of RNAP to form a clamp on the downstream duplex DNA (14, 48–50). Solutes are also excellent probes of intermediates and TS for RNA folding, because burial of base ASA is disfavored by all solutes whereas burial of backbone ASA is favored by some solutes (12, 13).

ACKNOWLEDGMENTS. We thank the reviewers for their thoughtful consideration of the manuscript and valuable questions and suggestions. We also thank R. L. Baldwin, S. Cavagnero, V. Daggett, K. Dill, W. Englander, S. Marqusee, K. Plaxco, D. Raleigh, G. Rose, and T. Sosnick for their very helpful comments, and thank Rituparna Sengupta for her help creating extended models of proteins. This research was supported by National Institutes of Health Grant GM 47022 (to M.T.R.).

1. Fersht AR, Itzhaki LS, elMasry NF, Matthews JM, Otzen DE (1994) Single versus parallel pathways of protein folding and fractional formation of structure in the transition state. *Proc Natl Acad Sci USA* 91(22):10426–10429.
2. Sosnick TR, Barrick D (2011) The folding of single domain proteins—have we reached a consensus? *Curr Opin Struct Biol* 21(1):12–24.
3. Mittermaier A, Korzhnev DM, Kay LE (2005) Side-chain interactions in the folding pathway of a Fyn SH3 domain mutant studied by relaxation dispersion NMR spectroscopy. *Biochemistry* 44(47):15430–15436.
4. Hu W, et al. (2013) Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry. *Proc Natl Acad Sci USA* 110(19): 7684–7689.
5. Baldwin RL, Rose GD (2013) Molten globules, entropy-driven conformational change and protein folding. *Curr Opin Struct Biol* 23(1):4–10.
6. Guinn EJ, Pegram LM, Capp MW, Pollock MN, Record MT, Jr. (2011) Quantifying why urea is a protein denaturant, whereas glycine betaine is a protein stabilizer. *Proc Natl Acad Sci USA* 108(41):16932–16937.

7. Pegram LM, et al. (2010) Why Hofmeister effects of many salts favor protein folding but not DNA helix formation. *Proc Natl Acad Sci USA* 107(17):7716–7721.
8. Record MT, Jr., Guinn E, Pegram L, Capp M (2013) Introductory lecture: Interpreting and predicting Hofmeister salt ion and solute effects on biopolymer and model processes using the solute partitioning model. *Faraday Discuss* 160:9–44, discussion 103–120.
9. Capp MW, et al. (2009) Interactions of the osmolyte glycine betaine with molecular surfaces in water: Thermodynamics, structural interpretation, and prediction of m-values. *Biochemistry* 48(43):10372–10379.
10. Pegram LM, Record MT, Jr. (2008) Thermodynamic origin of hofmeister ion effects. *J Phys Chem B* 112(31):9428–9436.
11. Guinn EJ, et al. (2013) Quantifying functional group interactions that determine urea effects on nucleic acid helix formation. *J Am Chem Soc* 135(15):5828–5838.
12. Lambert D, Draper DE (2012) Denaturation of RNA secondary and tertiary structure by urea: Simple unfolded state models and free energy parameters account for measured m-values. *Biochemistry* 51(44):9014–9026.
13. Lambert D, Leipply D, Draper DE (2010) The osmolyte TMAO stabilizes native RNA tertiary structures in the absence of Mg^{2+}: Evidence for a large barrier to folding from phosphate dehydration. *J Mol Biol* 404(1):138–157.
14. Kontur WS, Capp MW, Gries TJ, Saecker RM, Record MT, Jr. (2010) Probing DNA binding, DNA opening, and assembly of a downstream clamp/jaw in *Escherichia coli* RNA polymerase-lambdaP(R) promoter complexes using salt and the physiological anion glutamate. *Biochemistry* 49(20):4361–4373.
15. Spolar RS, Record MT, Jr. (1994) Coupling of local folding to site-specific binding of proteins to DNA. *Science* 263(5148):777–784.
16. Livingstone JR, Spolar RS, Record MT, Jr. (1991) Contribution to the thermodynamics of protein folding from the reduction in water-accessible nonpolar surface area. *Biochemistry* 30(17):4237–4244.
17. Spolar RS, Livingstone JR, Record MT, Jr. (1992) Use of liquid hydrocarbon and amide transfer data to estimate contributions to thermodynamic functions of protein folding from the removal of nonpolar and polar surface from water. *Biochemistry* 31(16):3947–3955.
18. Kuhlman B, Luisi DL, Evans PA, Raleigh DP (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(5):1661–1670.
19. Perl D, et al. (2002) Thermodynamics of a diffusional protein folding reaction. *Biophys Chem* 96(2–3):173–190.
20. Schindler T, Schmid FX (1996) Thermodynamic properties of an extremely rapid protein folding reaction. *Biochemistry* 35(51):16833–16842.
21. Plaxco KW, et al. (1998) The folding kinetics and thermodynamics of the Fyn-SH3 domain. *Biochemistry* 37(8):2529–2537.
22. Laurents DV, et al. (2000) Folding kinetics of phage 434 Cro protein. *Biochemistry* 39(45):13963–13973.
23. Scalley ML, Baker D (1997) Protein folding kinetics exhibit an Arrhenius temperature dependence when corrected for the temperature dependence of protein stability. *Proc Natl Acad Sci USA* 94(20):10636–10640.
24. Jackson SE, Fersht AR (1991) Folding of chymotrypsin inhibitor 2. 2. Influence of proline isomerization on the folding kinetics and thermodynamic characterization of the transition state of folding. *Biochemistry* 30(43):10436–10443.
25. Tan YJ, Oliveberg M, Fersht AR (1996) Titration properties and thermodynamics of the transition state for folding: Comparison of two-state and multi-state folding pathways. *J Mol Biol* 264(2):377–389.
26. Van Nuland NAJ, et al. (1998) Slow cooperative folding of a small globular protein HPr. *Biochemistry* 37(2):622–637.
27. Chiti F, et al. (1998) Structural characterization of the transition state for folding of muscle acylphosphatase. *J Mol Biol* 283(4):893–903.
28. Taddei N, et al. (1999) Thermodynamics and kinetics of folding of common-type acylphosphatase: Comparison to the highly homologous muscle isoenzyme. *Biochemistry* 38(7):2135–2142.
29. Main ERG, Fulton KF, Jackson SE (1999) Folding pathway of FKBP12 and characterisation of the transition state. *J Mol Biol* 291(2):429–444.
30. Chen BL, Baase WA, Schellman JA (1989) Low-temperature unfolding of a mutant of phage T4 lysozyme. 2. Kinetic investigations. *Biochemistry* 28(2):691–699.
31. Otzen DE, Kristensen O, Proctor M, Oliveberg M (1999) Structural changes in the transition state of protein folding: Alternative interpretations of curved chevron plots. *Biochemistry* 38(20):6499–6511.
32. Otzen DE, Oliveberg M (2004) Correspondence between anomalous m- and DeltaCp-values in protein folding. *Protein Sci* 13(12):3253–3263.
33. Plaxco KW, Simons KT, Baker D (1998) Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* 277(4):985–994.
34. Tsodikov OV, Record MT, Jr., Sergeev YV (2002) Novel computer program for fast exact calculation of accessible and molecular surface areas and average surface curvature. *J Comput Chem* 23(6):600–609.
35. Yoo TY, et al. (2012) The folding transition state of protein L is extensive with non-native interactions (and not small and polarized). *J Mol Biol* 420(3):220–234.
36. Ladurner AG, Itzhaki LS, Daggett V, Fersht AR (1998) Synergy between simulation and experiment in describing the energy landscape of protein folding. *Proc Natl Acad Sci USA* 95(15):8473–8478.
37. Morra G, Hodoscek M, Knapp EW (2003) Unfolding of the cold shock protein studied with biased molecular dynamics. *Proteins* 53(3):597–606.
38. Huang L, Shakhnovich EI (2012) Is there an en route folding intermediate for cold shock proteins? *Protein Sci* 21(5):677–685.
39. Pan YP, Daggett V (2001) Direct comparison of experimental and calculated folding free energies for hydrophobic deletion mutants of chymotrypsin inhibitor 2: Free energy perturbation calculations using transition and denatured states from molecular dynamics simulations of unfolding. *Biochemistry* 40(9):2723–2731.
40. Li AJ, Daggett V (1994) Characterization of the transition state of protein unfolding by use of molecular dynamics: Chymotrypsin inhibitor 2. *Proc Natl Acad Sci USA* 91(22):10430–10434.
41. Suenaga A, Okimoto N, Ebisuzaki T (2002) Refolding molecular dynamics simulations of small- and middle-sized proteins in an explicit solvent. *Mol Simul* 28(4):337–357.
42. Pandit AD, Jha A, Freed KF, Sosnick TR (2006) Small proteins fold through transition states with native-like topologies. *J Mol Biol* 361(4):755–770.
43. Hubner IA, Oliveberg M, Shakhnovich EI (2004) Simulation, experiment, and evolution: Understanding nucleation in protein S6 folding. *Proc Natl Acad Sci USA* 101(22):8354–8359.
44. Kurt N, Cavagnero S (2005) The burial of solvent-accessible surface area is a predictor of polypeptide folding and misfolding as a function of chain elongation. *J Am Chem Soc* 127(45):15690–15691.
45. Gong H, Isom DG, Srinivasan R, Rose GD (2003) Local secondary structure content predicts folding rates for simple, two-state proteins. *J Mol Biol* 327(5):1149–1154.
46. Makarov DE, Plaxco KW (2003) The topomer search model: A simple, quantitative theory of two-state protein folding kinetics. *Protein Sci* 12(1):17–26.
47. Saecker RM, Record MT, Jr., Dehaseth PL (2011) Mechanism of bacterial transcription initiation: RNA polymerase–promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *J Mol Biol* 412(5):754–771.
48. Kontur WS, Saecker RM, Davis CA, Capp MW, Record MT, Jr. (2006) Solute probes of conformational changes in open complex (RPo) formation by *Escherichia coli* RNA polymerase at the lambdaPR promoter: Evidence for unmasking of the active site in the isomerization step and for large-scale coupled folding in the subsequent conversion to RPo. *Biochemistry* 45(7):2161–2177.
49. Kontur WS, Saecker RM, Capp MW, Record MT, Jr. (2008) Late steps in the formation of *E. coli* RNA polymerase-lambda P R promoter open complexes: Characterization of conformational changes by rapid [perturbant] upshift experiments. *J Mol Biol* 376(4):1034–1047.
50. Drennan A, et al. (2012) Key roles of the downstream mobile jaw of *Escherichia coli* RNA polymerase in transcription initiation. *Biochemistry* 51(47):9447–9459.
51. Luisi DL, Kuhlman B, Sideras K, Evans PA, Raleigh DP (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(1):167–174.
52. Mueller U, Perl D, Schmid FX, Heinemann U (2000) Thermal stability and atomic-resolution crystal structure of the *Bacillus caldolyticus* cold shock protein. *J Mol Biol* 297(4):975–988.
53. Schnuchel A, et al. (1993) Structure in solution of the major cold-shock protein from Bacillus subtilis. *Nature* 364(6433):169–171.
54. Morton CJ, et al. (1996) Solution structure and peptide binding of the SH3 domain from human Fyn. *Structure* 4(6):705–714.
55. Mondragón A, Wolberger C, Harrison SC (1989) Structure of phage 434 Cro protein at 2.35 A resolution. *J Mol Biol* 205(1):179–188.
56. O'Neill JW, Kim DE, Baker D, Zhang KY (2001) Structures of the B1 domain of protein L from *Peptostreptococcus magnus* with a tyrosine to tryptophan substitution. *Acta Crystallogr D Biol Crystallogr* 57(Pt 4):480–487.
57. McPhalen CA, James MN (1987) Crystal and molecular structure of the serine proteinase inhibitor CI-2 from barley seeds. *Biochemistry* 26(1):261–269.
58. Jia Z, Quail JW, Waygood EB, Delbaere LT (1993) The 2.0-A resolution structure of *Escherichia coli* histidine-containing phosphocarrier protein HPr. A redetermination. *J Biol Chem* 268(30):22490–22501.
59. Pastore A, Saudek V, Ramponi G, Williams RJ (1992) Three-dimensional structure of acylphosphatase. Refinement and structure analysis. *J Mol Biol* 224(2):427–440.
60. Lam SY, Yeung RC, Yu TH, Sze KH, Wong KB (2011) A rigidifying salt-bridge favors the activity of thermophilic enzyme at high temperatures at the expense of low-temperature activity. *PLoS Biol* 9(3):e1001027.
61. Lindahl M, et al. (1994) Crystal structure of the ribosomal protein S6 from *Thermus thermophilus*. *EMBO J* 13(6):1249–1254.
62. Becker JW, et al. (1993) FK-506-binding protein: Three-dimensional structure of the complex with the antagonist L-685,818. *J Biol Chem* 268(15):11335–11339.
63. Yang G, et al. (2000) Solid-state synthesis and mechanical unfolding of polymers of T4 lysozyme. *Proc Natl Acad Sci USA* 97(1):139–144.
64. Hong J, Capp MW, Saecker RM, Record MT, Jr. (2005) Use of urea and glycine betaine to quantify coupled folding and probe the burial of DNA phosphates in lac repressor-lac operator binding. *Biochemistry* 44(51):16896–16911.