Protein Folding as Flow across a Network of Folding−Unfolding Pathways. 2. The "In-Water" Case

Dmitry N. Ivankov and Alexei V. Finkelstein*

Laboratory of Protein Physics, Institute of Protein Research of the Russian Academy of Sciences, 4 Institutskaya str., Pushchino, Moscow Region 142290, Russia

Received: December 26, 2009

This paper, second in the series, extends our analysis of networks of protein folding−unfolding pathways in vitro from the equilibrium, analytically the simplest, to "in-water" conditions, i.e., to the case when the native state of a protein is, as a rule, much more stable than the unfolded one. Protein folding rates and folding nuclei determined for such "physiological" conditions are of special biological and medical interest because of their relevance to misfolding diseases and other folding-related disorders. Given the native protein structures and their experimental in-water stabilities, the previously developed general theory (see paper 1 of this series) is applied here to compute the in-water folding and unfolding rates and outline the folding nuclei. Agreement between these calculations and experiment appears to be reasonably good.

Introduction

This paper continues our theoretical analysis of protein folding and unfolding in vitro. The preceding paper of the series[1] considered these phenomena at midtransition where the native and denatured states of a protein have equal stabilities and where folding and unfolding reveal their most universal features[2−6] that can be computed in the simplest way.[1−7] However, biochemists are mostly interested in folding under "normal", i.e., "in-water" conditions[8,9] with no "nonphysiological" denaturant involved.

In this paper, we use the term "normal" (or "in-water") conditions with two meanings: (a) the absence of denaturant for denaturant-induced transitions; (b) 25 °C for temperature-induced transitions. The native state of a protein is, as a rule, sufficiently more stable than the denatured one under such normal conditions, which increases complexity of a theoretical analysis of protein folding,[10−16] but now this analysis describes a situation that much better reflects the real events in the cell.

Similar to paper 1 of this series,[1] this study regards protein folding and unfolding as flows through a network of protein folding−unfolding pathways, though in conditions of different stabilities of the native and denatured states of a protein. Special attention is paid to in-water conditions, i.e., to the case when the native state of a protein is, as a rule, much more stable than the unfolded one.

Materials and Methods

Protein Data Set: Structures and Folding−Unfolding Kinetics. All proteins studied in paper 1[1] are considered here. Their folding and unfolding rates, and also some other relevant data about their structure and folding−unfolding kinetics, are taken from the KineticDB database[17] and presented in Table S1 of the Supporting Information. For proteins with multistate folding−unfolding kinetics, we consider the experimental folding and unfolding rate constants (k_f, k_u) that belong to the slowest phases. The values of these rate constants in normal conditions were found from logarithmic extrapolations, as described in the footnotes to Table S1.

Theory. A general theory of computations of the protein folding and unfolding rates for single-domain proteins has been developed in the preceding paper.[1] There, such computations were applied to the simplest case of folding and unfolding at midtransition, where the native and the unfolded states of a protein have equal stability. Here, the same theory is applied to a broader range of conditions, with special attention paid to normal conditions when the native state of a protein is, as a rule, much more stable than the unfolded one. Thus, almost all computational details remain as given in the preceding paper:[1] for the model of a network of folding−unfolding pathways, see "Description of the Model" in ref 1; chain division into links is described in the appropriate section of ref 1; the folding and unfolding rate constants are calculated according to eqs 9 and 10, respectively, given in ref 1; residues' ϕ-values are calculated according to eq 29 of ref 1; free energies G of all chain microstates are calculated according to eq 32 of ref 1; the rates of elementary transitions are calculated according to eq 36 of ref 1.

The only difference between this and the preceding study concerns the strength of interatomic contacts: here, to simulate protein folding and unfolding in conditions different from midtransition, this strength is varied over a wide range.

Connecting the Strength of Interatomic Contacts with Protein Stability. According to eq 38 of paper 1,[1] the midpoint of transition between the two stable states, N* and U*, is determined from the equation

$$\left(\frac{\Delta G}{RT}\right)_{mt} = \left(\frac{\varepsilon}{RT}\right)_{mt} \times [\nu_{N^*} - \nu_{U^*}] - [S_{N^*} - S_{U^*}]/R = 0 \quad (1)$$

Here, $\Delta G = G_{N^*} - G_{U^*}$ is the free-energy difference between the native (N*) and denatured (U*) states of a protein at midtransition, ε is the average free energy of one interatomic contact in this protein, T is the temperature (both ε and T correspond to the midtransition point, "mt"), R is the gas

* Corresponding author. Tel/fax: +7(495)514 0218. E-mail: afinkel@vega.protres.ru.

10.1021/jp912187w
Published on Web 05/24/2010

constant, ν_{N^*} and ν_{U^*} are the number of interatomic contacts in the N* and U* states, respectively, and S_{N^*} and S_{U^*} are conformational entropies of the N* and U* states. In the above expression, the action of solvent is taken into account implicitly, i.e., as a factor that influences the value of ε.

If ε or T changes a little, so that,

$$(\varepsilon/RT)_{mt} \rightarrow (\varepsilon/RT)_{mt} + [d(\varepsilon/RT)]$$

$$\frac{\Delta G}{RT} = \left[\frac{\varepsilon}{RT} - \left(\frac{\varepsilon}{RT}\right)_{mt}\right] \times [\nu_{N^*} - \nu_{U^*}] \quad (2)$$

Note that possible shifts of the N* and/or U* states with a small change of ε/RT do not result in a changed $\Delta G/RT$ since $G_{N^*} = \varepsilon_{mt}\nu_{N^*} - T_{mt}S_{N^*}$ and $G_{U^*} = \varepsilon_{mt}\nu_{U^*} - T_{mt}S_{U^*}$ are the local free-energy minima with respect to the shifts of N* and U*, respectively. Thus, with a slight deviation from the midtransition conditions (as far as it leads to slightly changed ν_{N^*} and ν_{U^*} only), the contact strength can be calculated as

$$\frac{\varepsilon}{RT} = \left(\frac{\varepsilon}{RT}\right)_{mt} + \frac{1}{\nu_{N^*} - \nu_{U^*}} \times \frac{\Delta G}{RT} \quad (3)$$

The $(\varepsilon/RT)_{mt}$ and $\nu_{N^*} - \nu_{U^*}$ values are defined under "Determination of the Midtransition Point" of paper 1.[1]

The dependence of ΔG on solvent conditions can be determined from either equilibrium or kinetic measurements.[18,19]

In two-state kinetics (generally typical of midtransition conditions[2,9]), ΔG is connected with the rate constants of folding (k_f) and unfolding (k_u) by a well-known[18] relationship $\Delta G/RT = -\ln K$, where

$$K = k_f/k_u \quad (4)$$

is the folding−unfolding equilibrium constant (which equals 1 at midtransition). Thus, with small deviations from the midtransition point (until ν_{N^*} or ν_{U^*} is unchanged),

$$\frac{\varepsilon}{RT} = \left(\frac{\varepsilon}{RT}\right)_{mt} - \frac{\ln K}{\nu_{N^*} - \nu_{U^*}} \quad (5)$$

Having linear dependencies of $\ln k_f$, $\ln k_u$, and $\ln K$ on the denaturant concentration M_D around the midtransition point $M_{D,mt}$ (which is used as a reference point), one can evaluate the strength of an interatomic contact in a given protein at various denaturant concentrations:

$$\frac{\varepsilon}{RT} \approx \left(\frac{\varepsilon}{RT}\right)_{mt} - \frac{[\partial(\ln K)/\partial M_D]_{mt}}{\nu_{N^*} - \nu_{U^*}}(M_D - M_{D,mt}) \quad (6)$$

Specifically, under in-water conditions with $M_D = 0$,

$$\left(\frac{\varepsilon}{RT}\right)_w \approx \left(\frac{\varepsilon}{RT}\right)_{mt} - \frac{\ln K^w_{Extr}}{\nu_{N^*} - \nu_{U^*}} \quad (7)$$

where $[\partial(\ln K)/\partial M_D]_{mt}(0 - M_{D,mt}) = \ln K^w_{Extr}$ as a result of linear extrapolation from the midtransition zone.

Similarly, for temperature-induced (at $T \neq T_{mt}$) transitions in water,

$$\frac{\varepsilon}{RT} \approx \left(\frac{\varepsilon}{RT}\right)_{mt} - \frac{[\partial(\ln K)/\partial(1/T)]_{mt}}{\nu_{N^*} - \nu_{U^*}}\left(\frac{1}{T} - \frac{1}{T_{mt}}\right) \quad (8)$$

and, at $T = 298$ K (or 25 °C),

$$\left(\frac{\varepsilon}{RT}\right)_{w,298K} \approx \left(\frac{\varepsilon}{RT}\right)_{mt} - \frac{\ln K^w_{Extr}}{\nu_{N^*} - \nu_{U^*}} \quad (9)$$

where $\ln K^w_{Extr} = [\partial(\ln K)/\partial(1/T)]_{mt}(1/298K - 1/T_{mt})$.

Results and Discussion

To investigate protein folding that occurs beyond midtransition, we have to apply our theory[1] to ε/RT values that deviate from the equilibrium $(\varepsilon/RT)_{mt}$ points given by eq 1; special attention should be paid to in-water $(\varepsilon/RT)_w$ given by eqs 7 and 9.

A deviation of ε/RT from $(\varepsilon/RT)_{mt}$ does not change the microstates composing the network of folding−unfolding pathways, but changes free energies of these microstates and, consequently, the rates of transitions between them. Also, such deviations can change the microstates corresponding to the free-energy minima, including demolition of the free-energy barrier responsible for "all-or-none" transitions between the folded and unfolded states of the protein.

Checking the Existence of an "All-or-None" Transition. Similarly to the previous paper, we begin with checking the protein folding profiles, i.e., projections of the protein free-energy landscapes onto the coordinate describing the number of links having achieved their native states. Three characteristic examples are given in the Figure 1, where we compare midtransition free-energy profiles with those for normal conditions. As for midtransition, we did not find any profile without the free-energy barrier. The comparison gives the following variants of condition-dependent profile behavior:

1. Two microstates, completely unfolded U and completely folded N, that are most stable at midtransition, remain so under normal conditions as well (case of protein L, Figure 1A,B). This is true for 44 proteins (## 1, 3, 5, 16, 17−20, 23, 25, 26, 32, 35, 36, 38, 40, 44−49, 51, 54, 56−64, 66−70, 74, 76−79, 81). It should be noted that the members of this group α-helix and β-hairpin (## 1, 3) in normal conditions (298 K in water) preserve the same stability.
2. Two microstates, completely unfolded U and "nearly completely" folded N*, that are most stable at midtransition, remain so under normal conditions as well (case of ubiquitin, Figure 1C,D). This is true for 32 proteins (## 2, 4, 6−11, 13, 14, 21, 22, 24, 27−31, 34, 37, 39, 42, 43, 52, 53, 55, 65, 71−73, 75, 82).
3. Protein folding at midtransition involves the completely unfolded microstate U and the nearly *co*mpletely folded microstate N*, whereas under in-water conditions the microstate N* turns either into the native microstate N (for proteins, ## 15, 33, 41, 50) or into a novel nearly folded microstate N** (this occurs only in # 80; see Figure 1E,F).

It should be noted that the most stable denatured microstate found by our calculations is always a completely unfolded microstate, even for in-water conditions, whereas in experiment the folding sometimes proceeds from a partly folded denatured state.[20] At the same time, the most stable folded microstate found by our calculations often contains some "disordered" parts,

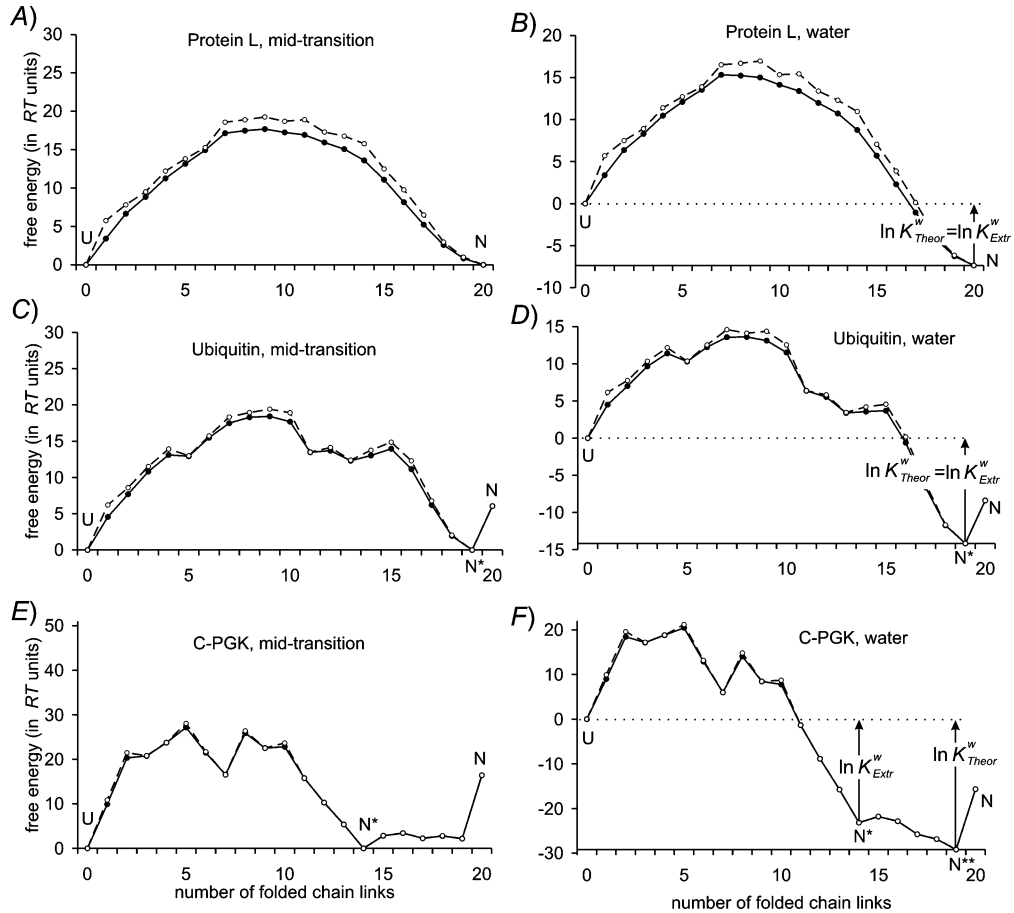

Figure 1. Comparison of midtransition and normal free-energy profiles of folding−unfolding barriers for protein L (A, B), ubiquitin (C, D), and C-PGK (E, F). In all cases, the polypeptide chain is divided into 20 links; the number of folded links ν plays the role of abscissa in all these plots. The lowest free-energy value for the microstate having the given number of folded links is presented by open circles and a dashed line; the free energy for the ensemble of all microstates having the given number of folded links is presented by solid circles (where the open circles do not shadow them) and a black line. (A, C, E) Profiles for midtransition. (B, D, F) Profiles for normal (in-water) conditions.

which are not observed experimentally; this artifact arises from crude energy estimates used in this work.

In-Water Folding Rates. Now, having ascertained that all proteins have free-energy barriers under in-water conditions, we can proceed to calculation of the in-water protein folding rates.

Table S1 gives both experimentally measured (column 6) and theoretically calculated (column 9) logarithms of the in-water folding rate constants ($\ln k_f^w$). The correlation between these is reasonable (see Figure 2A−C), better for two-state proteins and worse for multistate proteins. It is noteworthy that the calculations underestimate rate constants for slow-folding proteins, specifically for large proteins (like midtransition rate constants), and especially for large multistate proteins.

Unfolding Rates Extrapolated to In-Water Conditions. In experiment, in-water conditions do not allow direct measuring of the unfolding rates. Instead, these are obtained by extrapolation from the midtransition region. To make a direct comparison of our calculations with experiment, we herein compare the experimentally measured and calculated unfolding rates extrapolated from midtransition to in-water conditions.

Table S1 gives both experimentally measured (column 7) and theoretically calculated (column 10) logarithms of the unfolding rate constants extrapolated to in-water conditions ($\ln k_u^w$). The correlation between these is very good (Figure 2D−F). Again, it is better for two-state proteins and worse for multistate proteins. However, there is an underestimation of the unfolding rate constants by the calculations, which is again higher for slow-folding proteins, specifically for large proteins, and especially for large multistate proteins.

As seen, the correlation improves from in-water folding (0.70) to midtransition folding−unfolding (0.79) and (further) to in-water unfolding (0.87). In part, this can result from the effect of folding intermediates observed experimentally in low-denaturant conditions (note that both the folding and unfolding rate constants calculated for two-state proteins have a better correlation with experiment than those for multistate proteins, see Figure 2). Importantly, however, that a higher correlation with experiment demonstrated by calculations of the in-water unfolding rates may be explained by the fact that the free-energy barrier looks higher when watched in the direction of unfolding; this means that during unfolding the absolute error in its height estimate gives a smaller error in its relative estimate, as compared with that during folding of the protein.

Chevron Plots. Experimental investigation of protein folding kinetics yields a "chevron plot", a dependence of logarithm of the apparent protein relaxation rate constant k_{app} on denaturant concentration;[21] for two-state folding, $k_{app} = k_f + k_u$.

It is of interest to compare the calculated and experimental chevron plots. As an illustration, both the calculated and experimental chevrons are given in Figure 3. Because our model[1] simulates the two-state folding, we calculated $\ln k_{app} = \ln(k_f + k_u)$ for each protein as a function of denaturant concentration, M_D. The ε/RT values describing the strength of an interatomic contact at various denaturant concentrations were calculated

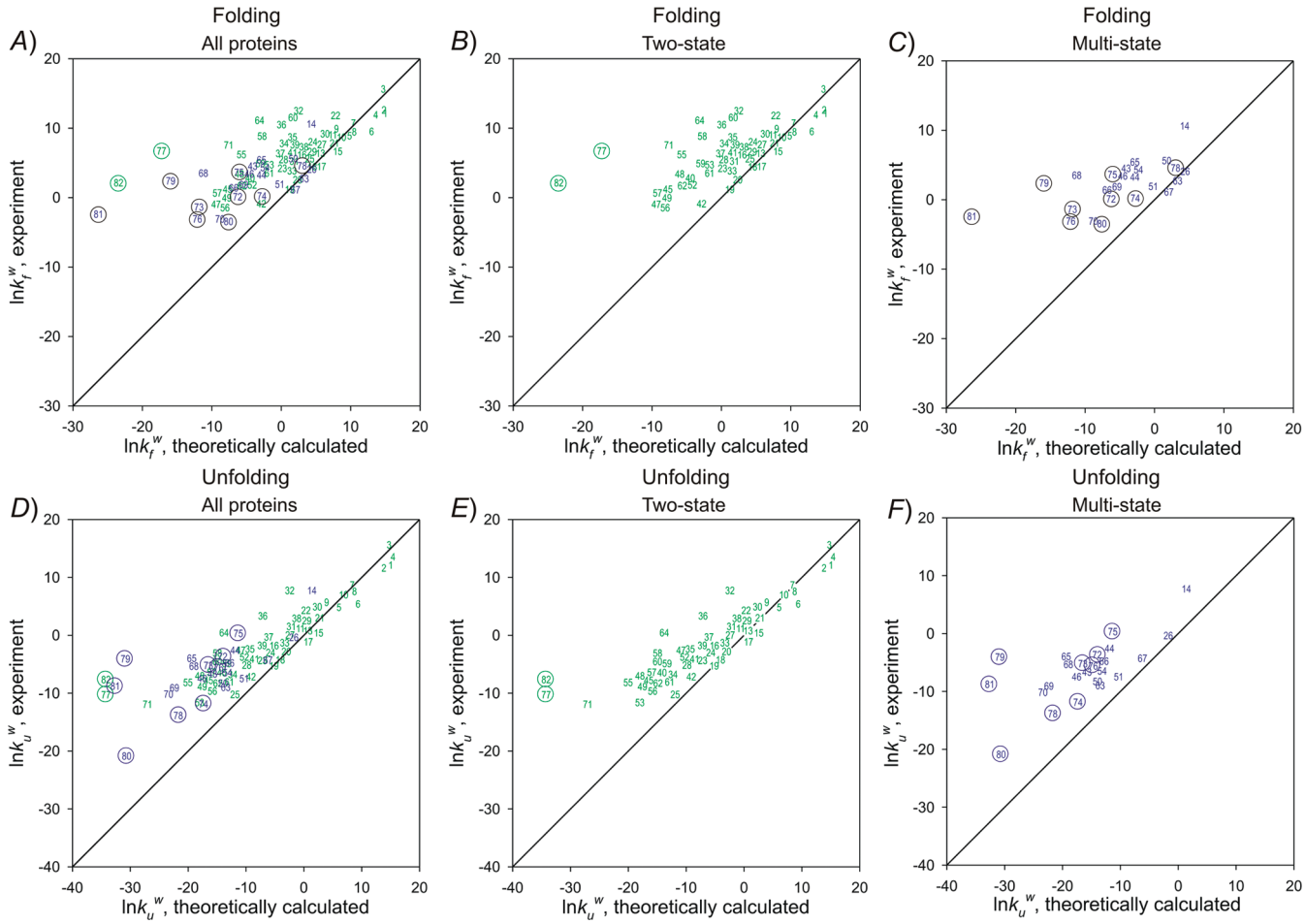

Figure 2. Comparison of calculated (horizontal axis) and known from experiment (vertical axis) logarithms of folding and unfolding rate constants under in-water conditions. (A) Folding rate constants for all proteins; the correlation is 0.70 ± 0.06. (B) For two-state proteins, the correlation is 0.67 ± 0.07 (0.63 ± 0.08 without α-helix and β-hairpin). (C) For multistate proteins, the correlation is 0.53 ± 0.11; (*D*) extrapolated unfolding rate constants for all proteins, the correlation is 0.87 ± 0.03.(*E*) For two-state proteins, the correlation is 0.90 ± 0.03 (0.88 ± 0.03 without α-helix and β-hairpin). (*F*) For multistate proteins, the correlation is 0.65 ± 0.04. After exclusion of proteins with >150 folded residues (encircled, see the text), the correlation becomes 0.73 ± 0.06 for folding of all proteins, 0.73 ± 0.06 for folding of two-state proteins (0.70 ± 0.07 without α-helix and β-hairpin), and 0.38 ± 0.12 for folding of multistate proteins, and 0.91 ± 0.02 for unfolding of all proteins, 0.92 ± 0.02 for unfolding of two-state proteins (0.91 ± 0.02 without α-helix and β-hairpin), and 0.77 ± 0.04 for unfolding of multistate proteins. Proteins are given the numbers used in Table S1.

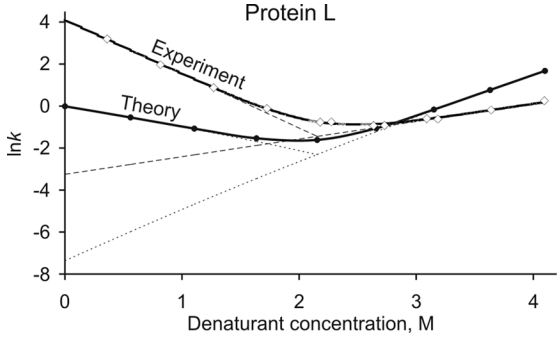

Figure 3. Theoretically calculated ("Theory") and experimentally obtained ("Experiment") chevron plots for protein L. The experimental data are taken from ref 9. The denaturant concentration is given on the horizontal axis. The logarithms of apparent rate constants (experimental and theoretical) are given on the vertical axis. The extrapolated (but unobservable) part of the unfolding chevron arm at low-denaturant concentrations (where ln $K > 0$) is shown as a dashed line. The dotted line shows the extrapolated logarithm of the calculated unfolding rate constant.

according to eqs 6 and 8 and used in the calculations of the k_f and k_u values for various M_D.

Having examined all 82 proteins listed in Table S1, we observed the linear and nonlinear chevron arms, both for folding and unfolding. However, the found linearity (and nonlinearity) of the chevron arms has poor correlation with experiment (results not shown). Of note, for all studied proteins, both the experimental and calculated nonlinear chevron arms show solely upward curvature.

From the chevron plot of each protein we calculated its transition state coordinate as $\beta_T = (m_f)/(m_f - m_u)$,[22] where $m_f \equiv \partial(\ln k_f)/\partial M_D$ and $m_u \equiv \partial(\ln k_u)/\partial M_D$. Regretfully, there is almost no correlation between the experimental and calculated β_T values.

Folding Nuclei. For 17 proteins listed in ref 6 with experimentally extensively explored folding nuclei, we calculated the folding nuclei under various conditions. The calculations were made, as described in ref 1, for ϵ/RT values spanning from midtransition to in-water conditions. Figure 4 shows correlation between experimental ϕ-values and ϕ-value calculations made for in-water conditions and for midtransition conditions. As seen, in both cases the calculations have similar correlation with experiment. On the average, the correlation coefficients for in-water calculations are slightly worse than those for midtransition, 0.41 ± 0.29 vs 0.45 ± 0.29. This could hardly be expected taking into account that experimental ϕ-values in the majority of experimental data correspond to the

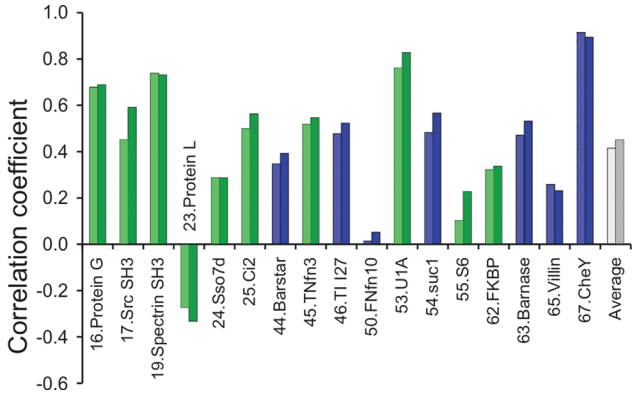

Figure 4. Correlation between experimentally measured ϕ-values listed in ref 6 and those calculated under in-water (left columns, light bars) and midtransition (right columns, dark bars) conditions. Green bars denote ϕ-values of two-state proteins and blue bars are for multistate folding proteins. The average correlation coefficients are given as gray bars on the right. They are virtually the same for two-state and multistate proteins.

in-water conditions. However, the errors in the averaged correlation coefficients are too large to discuss the difference between them.

Conclusion

Thus, the proposed method can be used to calculate the folding and unfolding rates both at midtransition and under in-water conditions. The obtained agreement with experiment is better for two-state proteins than for multistate proteins. For folding rate constants, the predictions are better in the midtransition case than in the case of in-water conditions; for unfolding rate constants, the predictions are better for in-water than for midtransition conditions. Overall, the correlation coefficients are as high as 0.7−0.9, although our calculations used no adjustable parameters at all. Large deviations are observed mostly for large proteins; most probably, they arise from too large a size of chain links, which follows from the limited number of links that we can use in our computations. Folding nuclei computed by the proposed method (again, without any adjustable parameters) are in worse, but still satisfactory, agreement with those outlined by experiment: the correlation coefficients of computed and experimental ϕ-values are, on the average, 0.41 for in-water and 0.45 for midtransition calculations.

Acknowledgment. The authors thank V. V. Filimonov, O. V. Galzitskaya, and S. O. Garbuzinskiy for helpful discussions and E. V. Serebrova for assistance in manuscript preparation. This work was supported by a grant of the President of Russian Federation (#MK-4894.2009.4), by programs "Molecular and Cellular Biology" of the Russian Academy of Sciences and "Leading Scientific Schools of Russia" (Grant #NSh-2791.2008.4), by a grant from the Federal Agency for Science and Innovation (#02.740.11.0295), by the Russian Foundation for Basic Research (Grant #07-04-00388), by the INTAS (grant #05-1000004-7747), and by an International Research Scholar's Award #55005607 to A.V.F. from the Howard Hughes Medical Institute.

Supporting Information Available: List of studied proteins with their experimental and calculated folding rates and other relevant details, presented as Table S1. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Ivankov, D. N.; Finkelstein, A. V. *J. Phys. Chem. B*, DOI: 10.1021/jp912186z.
(2) Galzitskaya, O. V.; Ivankov, D. N.; Finkelstein, A. V. *FEBS Lett.* **2001**, *489*, 113.
(3) Galzitskaya, O. V.; Finkelstein, A. V. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11299.
(4) Alm, E.; Baker, D. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11305.
(5) Alm, E.; Morozov, A. V.; Kortemme, T.; Baker, D. *J. Mol. Biol.* **2002**, *322*, 463.
(6) Garbuzynskiy, S. O.; Finkelstein, A. V.; Galzitskaya, O. V. *J. Mol. Biol.* **2004**, *336*, 509.
(7) Ivankov, D. N.; Finkelstein, A. V. *Biochemistry* **2001**, *40*, 9957.
(8) Fersht, A. R. *Curr. Opin. Struct. Biol.* **1995**, *5*, 79.
(9) Maxwell, K. L.; Wildes, D.; Zarrine-Afsar, A.; De Los Rios, M. A.; Brown, A. G.; Friel, C. T.; Hedberg, L.; Horng, J. C.; Bona, D.; Miller, E. J.; Vallee-Belisle, A.; Main, E. R.; Bemporad, F.; Qiu, L.; Teilum, K.; Vu, N. D.; Edwards, A. M.; Ruczinski, I.; Poulsen, F. M.; Kragelund, B. B.; Michnick, S. W.; Chiti, F.; Bai, Y.; Hagen, S. J.; Serrano, L.; Oliveberg, M.; Raleigh, D. P.; Wittung-Stafshede, P.; Radford, S. E.; Jackson, S. E.; Sosnick, T. R.; Marqusee, S.; Davidson, A. R.; Plaxco, K. W. *Protein Sci.* **2005**, *14*, 602.
(10) Abkevich, V. I.; Gutin, A. M.; Shakhnovich, E. I. *Biochemistry* **1994**, *33*, 10026.
(11) Finkelstein, A. *UFJ NATO ASI. Les Houches, Session LXXVII, 2002. Slow relaxations and nonequilibrium dynamics in condenced matter* **2003**, 649.
(12) Munoz, V.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11311.
(13) Plaxco, K. W.; Simons, K. T.; Baker, D. *J. Mol. Biol.* **1998**, *277*, 985.
(14) Sali, A.; Shakhnovich, E.; Karplus, M. *J. Mol. Biol.* **1994**, *235*, 1614.
(15) Thirumalai, D. *J. Phys. (Paris)* **1995**, *5*, 1457.
(16) Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 6170.
(17) Bogatyreva, N. S.; Osypov, A. A.; Ivankov, D. N. *Nucleic Acids Res.* **2009**, *37*, D342.
(18) Emanuel, N. M.; Knorre, D. G. *Course of Chemical Kinetics.* Vyschaya shkola: Moscow, 1984.
(19) Tanford, C. *Adv. Protein Chem.* **1970**, *24*, 1.
(20) Myers, J. K.; Oas, T. G. *Nat. Struct. Biol.* **2001**, *8*, 552.
(21) Matouschek, A.; Kellis, J. T., Jr.; Serrano, L.; Fersht, A. R. *Nature* **1989**, *340*, 122.
(22) Jackson, S. E. *Fold. Des.* **1998**, *3*, R81.

JP912187W