Proc. Natl. Acad. Sci. USA
Vol. 96, pp. 477–480, January 1999
Biophysics

Kinetic evidence for an on-pathway intermediate in the folding of cytochrome *c*

YAWEN BAI*

Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Building 37, Room 4A-01, Bethesda, MD 20892

Communicated by S. Walter Englander, University of Pennsylvania School of Medicine, Philadelphia, PA, November 20, 1998 (received for review September 26, 1998)

ABSTRACT **An early folding event of cytochrome *c* populates a helix-containing intermediate (I_{NC}) because of a pH-dependent misligation between the heme iron and nonnative ligands in the unfolded state (U). For folding to proceed, the nonnative ligation error must first be corrected. It is not known whether I is on-pathway, with folding to the native state (N) as in $U \leftrightarrow I_{NC} \leftrightarrow N$, or whether the I must first move back through the U and then fold to the N through some alternative path ($I_{NC} \leftrightarrow U \leftrightarrow N$). By means of a kinetic test, it is shown here that the cytochrome *c* I does not first unfold to U. The method used provides an experimental criterion for rejecting the off-pathway $I \leftrightarrow U \leftrightarrow N$ option.**

Perhaps the most elusive question in protein folding is whether experimentally observed intermediates (I) are on- or off-pathway (1–4). During the last 10 years, a number of protein-folding I have been characterized in some detail (5–10), but it has not been possible to demonstrate convincingly whether they are on-pathway, defined as unfolded state (U) $\leftrightarrow$ I $\leftrightarrow$ native state (N) or off-pathway I $\leftrightarrow$ U $\leftrightarrow$ N (2, 11).

The folding and unfolding process of cytochrome *c* (Cyt *c*) has been studied in detail (5, 12–18). Cyt *c* folding is heterogeneous and sensitive to pH. At pH 6 and higher, a large fraction of the molecules fold slowly and populate an I state in which the amino and carboxyl helices are folded (I_{NC}) (5). In contrast, at pH 5.0 and lower, most of the molecules fold quickly without populating any I state (14, 15). At elevated pH, a nonnative ligation between the heme iron and a peripheral histidine side chain in its neutral form erects a barrier that slows folding and causes I_{NC} to accumulate (14–17). It is not known whether I_{NC} is a dead-end, off-pathway I that has to unfold first and then refold through an alternative pathway ($I_{NC} \leftrightarrow U \leftrightarrow N$) or whether it represents an on-pathway step directly to the N ($U \leftrightarrow I_{NC} \leftrightarrow N$).

This paper describes a kinetic test that can be helpful under favorable conditions. If it can be shown that the rate for I $\rightarrow$ N is greater than that for I $\rightarrow$ U, then the off-pathway model can be rejected. The I_{NC} of Cyt *c* meets this criterion.

MATERIALS AND METHODS

Horse-heart Cyt *c* (type IV) was purchased from Sigma, treated with ferricyanide to oxidize the residual ferrous form, and further purified on a G-25 Sephadex column (Amersham Pharmacia). Ultrapure guanidinium chloride (GdmCl) was from Sigma. GdmCl concentrations were determined from a refractive index (19). ^2H_2O (D_2O; 99.8% D) was from Cambridge Isotope Laboratories (Cambridge, MA). Deuterated GdmCl was prepared by evaporating the solvent D_2O in 6 M GdmCl solution three times with a rotary evaporator.

Stopped-flow fluorescence and absorbance experiments were performed with a Biologic SFM-4 (Grenoble, France). The fluorescence excitation wavelength was 280 nm, and the emission was collected through a 320-nm cut-off filter. For both folding and unfolding experiments, three syringes (numbered 2, 3, and 4) were used to keep the final protein concentration constant while the concentration of denaturant was varied. Refolding experiments were initiated by mixing protein solution in 4 M GdmCl (in syringe 4) with 0.1 M NaOAc buffer solution (in syringe 2) and 6.0 M GdmCl solution (in syringe 3). Each shot contained 0.1 ml of protein solution from syringe 4 and a 0.9-ml mixture from syringe 2 and syringe 3 with an appropriate ratio of the two solutions to yield different concentrations of GdmCl in the final solution. All solutions contained 0.1 M NaOAc at pH 6.2. The final protein concentration in the stopped-flow fluorescence experiments was $\approx$30 μM. Final data were obtained by averaging two shots. A 0.2-cm path-length cuvette (FC-20) was used. Mixing dead time was $\approx$8 ms. Kinetic-unfolding experiments were performed similarly, except that the solution in syringe 4 did not contain GdmCl. In the refolding experiments in D_2O at 30°C and pD 7.0, syringe 4 contained a Cyt *c* in 4 M GdmCl, 0.5 M KCl, and 0.1 M KH_2PO_4. Syringe 3 and syringe 2 contained 0.5 M KCl and 0.1 M KH_2PO_4. The refolding experiment was initiated by mixing 0.175 ml of solution in syringe 4 with 0.4 ml of solution in syringe 3 and 0.425 ml of solution in syringe 2. The dead time was $\approx$3 ms. Stopped-flow absorbance monitored at a 695-nm wavelength was initiated by mixing protein solution in 4.0 M GdmCl with 0.1 M NaOAc buffer solution to yield a final solution of 0.7 M GdmCl and 26 μM protein. A 1.0-cm path-length cuvette (TC-100/15) was used in this experiment.

The kinetic refolding and unfolding curves from stopped-flow fluorescence experiments were fitted with three exponential phases and a single exponential phase, respectively, by using a simplex algorithm (BIOKINE from Biologic). The kinetic refolding curve from the absorbance experiment at a 695-nm wavelength was fitted with two exponential phases. Equilibrium unfolding data were obtained from the baselines of the kinetic-unfolding traces. The unfolding free energy ΔG_{NU} was assumed to be a linear function of GdmCl concentration, i.e., $\Delta G_{NU}([\text{GdmCl}]) = \Delta G_{NU}(H_2O) - m_{NU}[\text{GdmCl}]$ (19). The parameters in the above equation were obtained by fitting the equilibrium unfolding data to the equation of Santoro and Bolen (20). ΔG_{NU} measured in this way does not include a contribution from the population of nonnative peptide conformations of prolines in U. Therefore, its value is slightly larger ($\approx$0.5 kcal/mol) than that obtained from a slow equilibrium unfolding experiment (21). The GdmCl dependence of unfolding rate constants near the transition region was assumed to be linear on a logarithmic scale, i.e., $\log k_{NI} = \log$

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

PNAS is available online at www.pnas.org.

Abbreviations: Cyt *c*, cytochrome *c*; GdmCl, guanidinium chloride; I, intermediate; N, native state; U, unfolded state.

*To whom reprint requests should be addressed. e-mail: yawen@helix.nih.gov.

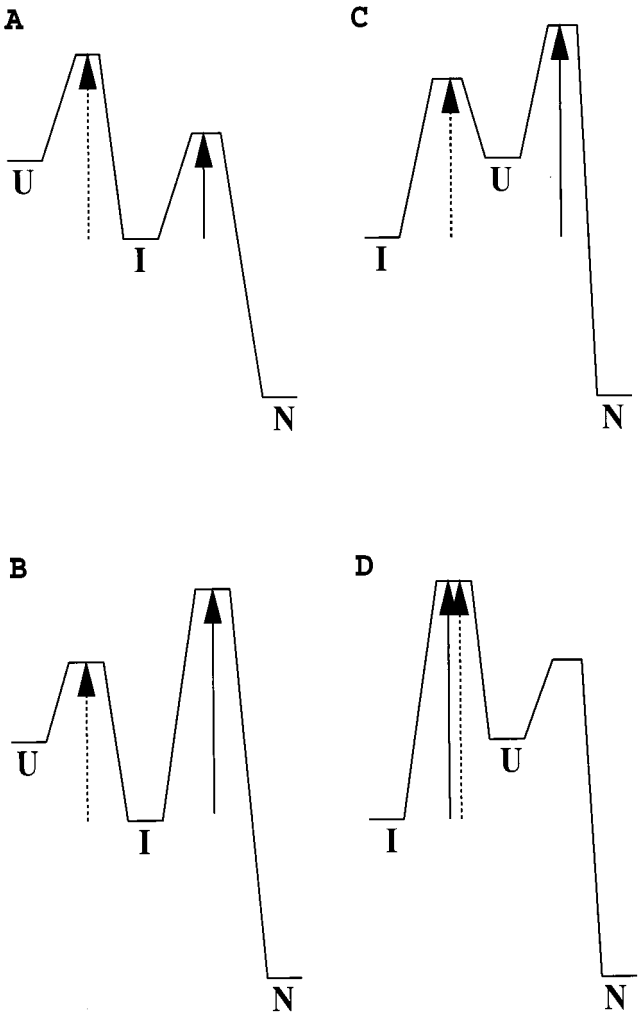

FIG. 1. Illustration of the kinetic difference between an on-pathway and an off-pathway reaction for an apparent three-state system. The effective kinetic barrier from I to N is indicated by the solid arrow and from I to U by the dashed arrow. (*A* and *B*) On-pathway reaction. (*C* and *D*) Off-pathway reaction. Only in *A* is $k_{IN} > k_{IU}$. This characteristic allows one to identify an on-pathway I. This is a sufficient but not a necessary condition. Even though $k_{IN} < k_{IU}$, I can still be on-pathway as in *B*. (*C*) The refolding I can be populated quickly, and a lag phase may be generated before the formation of N. In this case, U is not observable during the folding from I to N even though I must pass through U. Therefore, the appearance of a lag phase of I before formation of N does not necessarily indicate an on-pathway I (33). (*D*) Refolding may seem to be two-state, although a stable off-pathway I exists. Therefore, apparent two-state folding does not exclude the possibility that an equilibrium I may be observed. Similarly apparent two-state folding, as in *A*, can conceal the presence of an on-pathway I.

$k_{NI}^{H_2O} - m^{\ddagger}_{NI}$[GdmCl] (6). Linear extrapolation was used to obtain k_{NI} at 1.0 M GdmCl.

RESULTS

Fig. 1 illustrates the kinetic picture for on- and off-pathway folding when an I is populated. An off-pathway I must pass through the U to arrive at the N. In this mode, the time it takes for I to reach N cannot be shorter than the time for U to reach N. If the reverse condition holds, i.e., $k_{IN} > k_{IU}$, then the off-pathway option can be rejected. When an I is clearly populated in kinetic folding, the apparent rate constant, k_{IN}, for conversion from I to N (through U in the off-pathway case) normally can be measured directly. Therefore, to make the desired comparison, the key is to obtain the unfolding rate constant from I to U, k_{IU}, under native folding conditions. This paper shows how k_{IU} can be obtained indirectly from several experimental approaches by using the relationship $\Delta G_{IU} = -RT \ln(K_{IU}) = -RT \ln (k_{IU}/k_{UI})$.

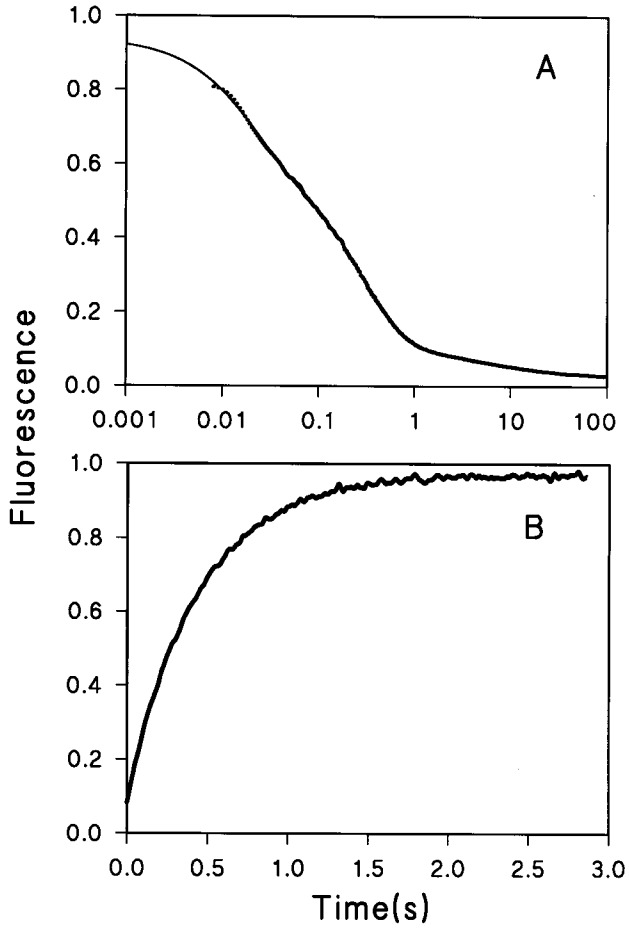

FIG. 2. Kinetic traces of refolding and unfolding. The unit of fluorescence is arbitrary. (*A*) Refolding at 1.0 M GdmCl. A three-exponential fitting yields a residual signal that is 1/70 that of the total amplitude. The apparent relaxation rate constants are $\lambda_1 = 50\ sec^{-1}$, $\lambda_2 = 3.5\ sec^{-1}$, and $\lambda_3 = 0.1\ sec^{-1}$. The solid line is the fitting curve. (*B*) Unfolding at 3.0 M GdmCl. The data fitted with a single exponential have a residual 1/20 that of the total amplitude. The solid line is the fitting curve.

Direct Measurement of k_{IN}. Cyt *c* folding and unfolding experiments were carried out at pH 6.2 and 10°C, as were earlier pulse-labeling experiments (5). Fig. 2 *A* and *B* shows typical kinetic traces for folding in 1.0 M GdmCl and unfolding in 3.0 M GdmCl. The folding trace can be fitted with three exponential phases. Previous experiments under similar conditions showed that the fast phase represents the folding process from U to I_{NC} (5, 22). Also, some small fraction of the protein population becomes N at this same rate (5, 15). The middle phase corresponds to folding from I_{NC} to N, shown specifically by the loss of the I_{NC} and the rise of the absorbance band at 695 nm (Fig. 3; refs. 5 and 22), a specific probe for the Cyt *c* N (23). The very slow phase represents a small fraction of the protein that may be blocked by a prolyl *cis–trans* isomerization process (13). We are interested here in the I_{NC} to N transition. The rate constants k_{UI} and k_{IN} can be measured directly from these results. Other states, both those seen in Fig. 2 and others that are invisible to these measurements (5, 9, 18, 24, 25) do not affect the present considerations.

Experiments like those in Fig. 2 were performed between 1.0 M and 3.5 M GdmCl, where the equilibrium unfolding free

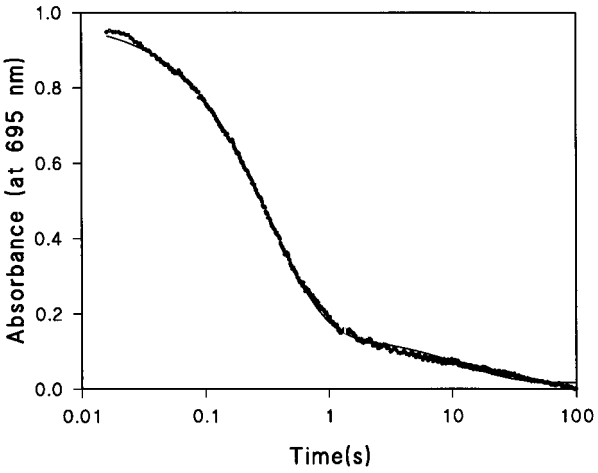

FIG. 3. Kinetic trace of refolding of Cyt *c* monitored at 695-nm wavelength by absorbance in arbitrary unit. The final concentrations of Cyt *c* and GdmCl are 26 μM and 0.7 M, respectively. The major phase of the refolding yields a rate constant k_{IN} of 2.9 sec^{-1}, which is in excellent agreement with the value of 2.7 sec^{-1} from the stopped-flow fluorescence experiment (see Table 1).

energy is known to be a linear function of denaturant. At 1.0 M GdmCl, the fast kinetic phase in Fig. 2*A* gives a rate constant $\lambda_1 = 50$ sec^{-1}, and the second kinetic phase yields an apparent rate constant $\lambda_2 = 3.5$ sec^{-1}. Because λ_1 is 15 times larger than λ_2, the two events are decoupled well kinetically and can be measured separately (26). Because I_{NC} is generated from U, and N is formed subsequently as the final stable product, $\lambda_1 = k_{UI} + k_{IU}$ and $\lambda_2 = k_{IN} + k_{NI}$. The apparent folding rate constant from I_{NC} to N is then obtained directly as $k_{IN} = \lambda_2 = 3.5$ sec^{-1}, because the unfolding rate constant, k_{NI}, is negligibly small ($\approx 10^{-3}$ sec^{-1}) under this native condition (see Fig. 4*B*). Similarly, for most purposes, k_{UI} is generally equal to λ_1.

Indirect Measurement of k_{IU}. To apply the kinetic criterion considered here, we want to compare k_{IN} to k_{IU}. The rate constant k_{IU} refers to melting of the stable I to the U under native conditions. It can only be determined indirectly. The value for k_{IU} can be determined in three independent ways. In all three approaches, $\lambda_1 = k_{UI} + k_{IU}$ is obtained directly from refolding experiments under three-state conditions, as above. ΔG_{IU}, the stability of the I at native conditions, can be obtained from either of two different kinds of hydrogen-exchange experiments and independently from extrapolation of kinetic folding and kinetic unfolding data. The value of k_{IU} can then be obtained by algebra from $\lambda_1 = k_{UI} + k_{IU}$ and $\Delta G_{IU} = -RT \ln K_{IU} = -RT \ln (k_{IU}/k_{UI})$.

Elöve and Roder (27) measured ΔG_{IU} at 3 kcal/mol (pH 6.2, 10°C, 0.7 M GdmCl) by using the hydrogen-exchange pulse-labeling method with variable pulse pH. Independently, the N hydrogen-exchange results of Bai *et al.* (9) yield 2.3 kcal/mol for ΔG_{IU} at pD 7.0, 30°C, 0.7 M GdmCl. Kinetic experiments like those in Fig. 2*A*, done under these same conditions, provide k_{UI} and k_{IN} as described above. These values and the values for k_{IU} derived from them are listed in Table 1.

In a third approach, ΔG_{IU} was obtained from a linear extrapolation of kinetic data and the relationship $\Delta G_{IU} = \Delta G_{NU} - \Delta G_{NI}$. ΔG_{NU} was obtained from equilibrium GdmCl unfolding and linear extrapolation to 1.0 M GdmCl (Fig. 4*A*; refs. 19, 20). ΔG_{NI} was measured by performing kinetic refolding and unfolding experiments as a function of GdmCl in the transition zone by using the method of Matoushek *et al.* (6). The unfolding rate constant k_{NI} at 1.0 M was obtained based on linear extrapolation of the right limb of the chevron curve in the transition zone where I are not populated (Fig. 4*B*; ref. 6). ΔG_{NI} was then calculated by using $\Delta G_{NI} = -RT \ln(k_{NI}/k_{IN})$. It leads to a value of ΔG_{NI} (1.0 M) 4.2 ± 0.1 kcal/mol. ΔG_{IU} (1.0 M) was then calculated to be 2.5 ± 0.2 kcal/mol. Finally, by using $\Delta G_{IU} = -RT \ln(k_{IU}/k_{UI})$ and $\lambda_1 = k_{UI} + k_{IU}$, the rate constant k_{IU} was determined to be 0.6 sec^{-1} (Table 1). The consistent results from three independent methods show that the value of k_{NI} at 1.0 M GdmCl, based on the above linear extrapolation, seems correct.

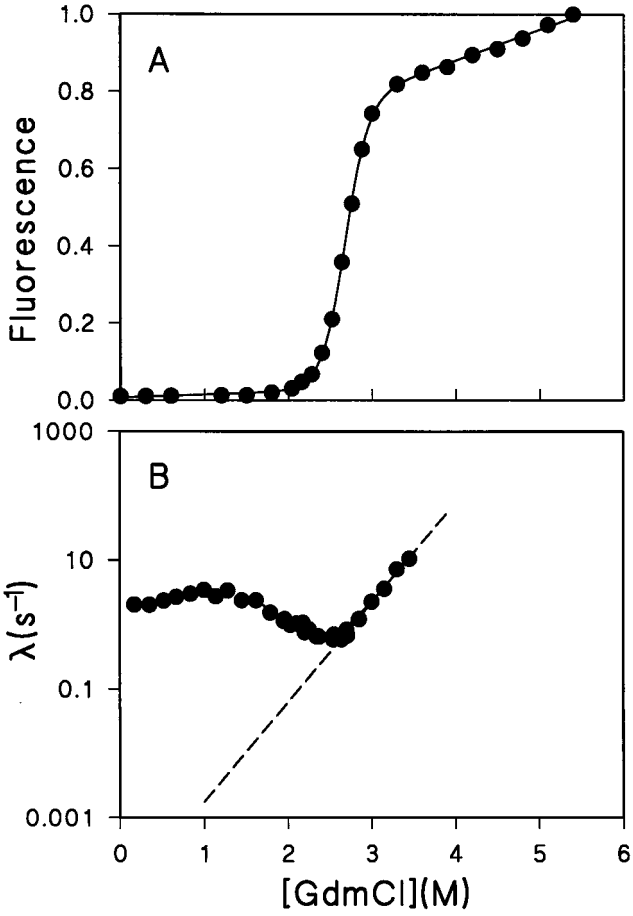

FIG. 4. Equilibrium and kinetic unfolding and refolding of Cyt *c*. (*A*) Equilibrium unfolding by fluorescence. The unit is defined arbitrarily with the last point at 5.4 M GdmCl taken as 1.0. The global unfolding data, fitted with a linear extrapolation (solid line; refs. 19 and 20), yield a ΔG_{NU} (1.0 M) = 6.7 ± 0.2 kcal/mol and a linear equation of ΔG_{NU} ([GdmCl]) = 10.7 ± 0.3 kcal/mol − (4.0 ± 0.1) (kcal/mol·M) [GdmCl](M). (*B*) The rate constants of the kinetic unfolding and the middle phase of the kinetic refolding as a function of GdmCl concentration. The dashed line represents the linear extrapolation of k_{NI} in logarithmic scale to 1.0 M GdmCl, which yields a value of k_{NI} 1.7 ± 0.6 × 10^{-3} sec^{-1}.

The above calculations used linear extrapolation in the short range between 1.0 M and 2.3 M. In the case of Cyt *c*, it is known to be valid in this region (21). In fact, nonlinear behavior of unfolding free energy and the logarithm of kinetic unfolding rate constant as a function of the denaturant concentration do not affect the above analysis. It is because linear extrapolation underestimates the true unfolding free energy (21, 28–30). When a nonlinear effect is considered, it will favor the conclusion even more.

DISCUSSION

The results obtained here are summarized in Table 1. The three different methods yield the same conclusion, namely that I_{NC} of Cyt *c* moves to N significantly faster (about 7 times faster under the conditions tested) than it moves to U. Thus, I_{NC} cannot go through the U state on its way to N. I_{NC} is not an

Table 1. Thermodynamic and kinetic parameters of I_{NC}

Method	Conditions	ΔG_{IU}, kcal/mol	k_{UI}, sec^{-1}	k_{IU}, sec^{-1}	k_{IN}, sec^{-1}
H/D pulse labeling (27)	pH 6.2, 10°C, 0.7 M GdmCl	3.0	55	0.3	2.7
N hydrogen-exchange (9)	pD 7.0, 30°C, 0.7 M GdmCl	2.3	61	1.3	10.0
Linear extrapolation	pH 6.2, 10°C, 1.0 M GdmCl	2.5	50	0.6	3.5

ΔG_{IU} values, obtained from hydrogen-exchange pulse labeling (27) or by the N hydrogen-exchange method (9) were taken from the literature where indicated or were computed from equilibrium unfolding data and the extrapolation of kinetic folding data as in Fig. 4. Values for k_{UI} were measured directly from the early phase of kinetic stopped-flow experiments done under three-state folding conditions, as in Fig. 2*A*. From these parameters, k_{IU} was evaluated by using $\Delta G_{IU} = -RT \ln (k_{IU}/k_{UI})$ and then compared with k_{IN}, measured from the later phase of the same stopped-flow experiments as for k_{UI}. These multiple comparisons all show that $k_{IN} > k_{IU}$, ruling out the off-pathway option, $I \leftrightarrow U \leftrightarrow N$.

off-pathway I, according to the classical description (2, 31). Although I_{NC} does not unfold fully before folding to N, it may unfold partially, as it seems to do judging from the small but detectable negative dependence on GdmCl seen in Fig. 4*B*. Some unfolding, measured as increased surface exposure, occurs in the transition state leading from I to N. It was interpreted by Sosnick *et al.* (14) as a necessary partial unfolding to reorganize a misfolded and trapped segment before going forward. The present results show that in this partial unfolding, the previously formed native-like N and C helices do not unfold.

The problem of whether an I is on- or off-pathway has been considered since the earliest protein-folding work began (13, 31, 32), but we still do not know how to make this distinction. A number of suggestions as to how we might do so have been put forward: (*i*) compare the ϕ values of the I and transition states (1); (*ii*) look for a lag phase during which I accumulates before the acquisition of the final N (33, 34); and (*iii*) check for a negative dependence of folding rate on denaturant concentration (2, 35). However, in no case has the true situation been shown. Recently, a folding I of hen egg-white lysozyme with the α-domain folded I_α was suggested to be on-pathway, based on the denaturant dependence of rate in the low-concentration range (35). Although an off-pathway I may be consistent with such behavior, so is an on-pathway I with a transition state that exposes increased surface area, as has been suggested for Cyt *c* (14). The lysozyme work also assumed that the logarithms of the folding and unfolding rate constants are linear functions of GdmCl over a broad concentration range, but only a limited data set was used for the fitting. When a complete data set is used, the fitting for the triangle mechanism proposed deviates dramatically from the experimental data (unpublished results).

The approach shown in the present paper characterizes the folding and unfolding events by using multiple experimental methods. If an apparent three-state system can be isolated from other folding/unfolding events because of differences in time scales or populations, then a further test for a possible on-pathway I can be made, as described here. This avoids the uncertainty inherent in brute-force fitting.

Note. Recently, Baldwin and coworkers (36) have shown that the late folding I_N of ribonuclease A is on the folding pathway. This I can also be shown to be on-pathway by using the kinetic criterion presented here, because the rate constant of $I_N \rightarrow N$ is 400 times larger than the rate constant of $I_N \rightarrow U$.

I thank S. Walter Englander for numerous stimulating discussions on the folding pathways of Cyt *c* and for his continuous guidance, Dr. T. R. Sosnick for helpful discussions, Drs. B.-K. Lee and P. Wagner for their careful reading and comments on the paper, and Drs. C. Klee, W. Klee, and H. S. Chen for their comments on the kinetic criterion at an early stage.

1. Fersht, A. R. (1993) *FEBS Lett.* **325,** 5–16.
2. Baldwin, R. L. (1996) *Fold. Des.* **1,** 1–8.
3. Creigton, T. E. (1997) *Biol. Chem.* **378,** 731–744.
4. Pande, P. S., Grosberg, A. Y., Tanaka, T. & Roshsar, D. S. (1998) *Curr. Opin. Struct. Biol.* **8,** 68–79.
5. Roder, H., Elöve, G. & Englander, S. W. (1988) *Nature (London)* **335,** 700–704.
6. Matouschek, A., Kellis, J. T., Jr., Serrano, L., Bycroft, M. & Fersht, A. R. (1990) *Nature (London)* **346,** 440–445.
7. Jennings, P. A. & Wright, P. E. (1993) *Science* **262,** 892–896.
8. Dobson, C. M., Evans, P. A. & Radford, S. E. (1994) *Trends Biochem. Sci.* **19,** 31–37.
9. Bai, Y., Sosnick, T. R., Mayne, L. & Englander, S. W. (1995) *Science* **269,** 192–197.
10. Raschke, T. M. & Marqusee, S. (1997) *Nat. Struct. Biol.* **4,** 298–304.
11. Laurents, D. V. & Baldwin, R. L. (1998) *Biophys. J.* **75,** 428–434.
12. Brems, D. N. & Stellwagen, E. (1983) *J. Biol. Chem.* **258,** 3655–3660.
13. Ridge, J. A., Baldwin, R. L. & Labhardt, A. M. (1981) *Biochemistry* **20,** 1622–1630.
14. Sosnick, T. R., Mayne, L. & Englander, S. W. (1994) *Nat. Struct. Biol.* **1,** 149–156.
15. Elöve, G. A., Bhuyan, A. K. & Roder, H. (1994) *Biochemistry* **33,** 6925–6935.
16. Yeh, S. & Rousseau, D. L. (1998) *Nat. Struct. Biol.* **5,** 222–228.
17. Yeh, S., Takahashi, S., Fan, B. & Rousseau, D. L. (1997) *Nat. Struct. Biol.* **4,** 51–56.
18. Sosnick, T. R., Shtilerman, M. D., Mayne, L. & Englander, S. W. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 8545–8550.
19. Pace, C. N. (1986) *Methods Enzymol.* **131,** 266–280.
20. Santoro, M. M. & Bolen, D. M. (1992) *Biochemistry* **31,** 4901–4907.
21. Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1994) *Proteins* **20,** 4–14.
22. Sosnick, T. R., Mayne, L. & Englander, S. W. (1996) *Proteins* **24,** 413–426.
23. Schechter, E. & Saludjian, P. (1967) *Biopolymers* **5,** 788–790.
24. Shastry, M. C. R. & Roder, H. (1998) *Nat. Struct. Biol.* **5,** 385–392.
25. Bhuyan, A. K. & Udgaonkar, J. B. (1998) *Biochemistry* **37,** 9147–9155.
26. Hagerman, P. J. & Baldwin, R. L. (1976) *Biochemistry* **15,** 1462–1473.
27. Elöve, G. A. & Roder, H. (1991) *ACS Symp. Ser.* **470,** 50–63.
28. Nozaki, Y. & Tanford, C. (1970) *J. Biol. Chem.* **245,** 1648–1652.
29. Makhatadze, G. I. & Privalo, P. L. (1992) *J. Mol. Biol.* **226,** 491–505.
30. Johnson, C. M. & Fersht, A. R. (1995) *Biochemistry* **34,** 6795–6804.
31. Ikai, A. & Tanford, C. (1971) *Nature (London)* **230,** 100–102.
32. Utiyama, H. & Baldwin, R. L. (1986) *Methods Enzymol.* **131,** 51–71.
33. Creighton, T. E. (1994) *Nat. Struct. Biol.* **1,** 135–138.
34. Heidary, D. K., Gross, L. A., Roy, M. & Jennings, P. A. (1997) *Nat. Struct. Biol.* **4,** 725–731.
35. Wildegger, G. & Kiefhaber, T. (1997) *J. Mol. Biol.* **270,** 294–304.
36. Laurents, D. V., Bruix, M., Jamin, M. & Baldwin, R. L. (1998) *J. Mol. Biol.* **283,** 669–678.