Chemical Denaturation: Potential Impact of Undetected Intermediates in the Free Energy of Unfolding and *m*-Values Obtained from a Two-State Assumption

Jose L. Soulages
Department of Biochemistry, University of Arizona, Tucson, Arizona 85721 USA

ABSTRACT The chemical unfolding transition of a protein was simulated, including the presence of an intermediate (I) in equilibrium with the native (N) and unfolded (U) states. The calculations included free energies of unfolding, ΔG_u^w, in the range of 1.4 kcal/mol to 10 kcal/mol and three different global *m*-values. The simulations included a broad range of equilibrium constants for the N ↔ I process. The dependence of the N ↔ I equilibrium on the concentration of denaturant was also included in the simulations. Apparent ΔG_u^w and *m*-values were obtained from the simulated unfolding transitions by fitting the data to a two-state unfolding process. The potential errors were calculated for two typical experimental situations: 1) the unfolding is monitored by a physical property that does not distinguish between native and intermediate states (case I), and 2) the physical property does not distinguish between intermediate and unfolded states (case II). The results obtained indicated that in the presence of an intermediate, and in both experimental situations, the free energy of unfolding and the *m*-values could be largely underestimated. The errors in ΔG_u^w and *m*-values do not depend on the *m*-values that characterize the global N ↔ U transition. They are dependent on the equilibrium constant for the N ↔ I transition and its characteristic m_1-value. The extent of the underestimation increases for higher energies of unfolding. Including no random error in the simulations, it was estimated that the underestimation in ΔG_u^w could range between 25% and 35% for unfolding transitions of 3–10 kcal/mol (case I). In case II, the underestimation in ΔG_u^w could be even larger than in case I. In the same energy range, a 50% error in the *m*-value could also take place. The fact that most of the mutant proteins are characterized by both a lower *m*-value and a lower stability than the wild-type protein suggests that in some cases the results could have been underestimated due to the application of the two-state assumption.

INTRODUCTION

Protein unfolding studies are essential to obtain thermodynamic properties of proteins and gain understanding in the energetics of protein folding and structure. The reversible unfolding of a protein is commonly induced by changes in temperature or solvent, and monitored by spectroscopic (Pace, 1986) or calorimetric methods (Privalov, 1979). In chemical unfolding studies, which are the subject of this study, the solvent is usually modified by addition of strong denaturants, such as guanidine hydrochloride or urea. In most of the protein unfolding studies, the unfolding curves are interpreted on the basis of the linear extrapolation method (Greene and Pace, 1974; Pace, 1986; Schellman 1978; Santoro and Bolen, 1992). These studies provide two parameters, which have thermodynamic and structural implications. The main parameter provided is the change in Gibbs free energy upon total unfolding of the native state (N → U), ΔG_u. The second parameter, which is called *m*-value, is given by the derivative of ΔG_u with respect to the denaturant concentration. The *m*-value is related to the difference in solvent-accessible surface area between the unfolded and the native states (Schellman, 1978; Shortle, 1995). The comparison of the values obtained for these two parameters with a wild-type and a mutant protein constitutes a common method to gather information about the role of certain interactions in the structure and stability of the protein (Matthews, 1995; Shortle, 1995). Even though the experiments are relatively straightforward, the validity of the parameters deduced from the analysis of the unfolding curves relies on several criteria, which often are difficult to meet. These have been clearly stated and discussed by several authors (Pace, 1986; Yao and Bolen, 1995; Bolen and Santoro, 1988; Santoro and Bolen, 1988). One of the main conditions resides in the number of states involved in the transition. In most of the reported studies, the transition between the native and the unfolded states is assumed to proceed by a two-state mechanism (N → U). This condition implies the practical absence of unfolding intermediates throughout the transition. To be consistent with a two-state model, the unfolding curves should be undistinguishable when monitored by several independent techniques (Brandts, 1969; Pace, 1986). The validity of this assumption has been confirmed for several small proteins, but not all of them (Privalov, 1979). Increasing evidence indicates that the unfolding of many proteins involves at least three states (Ptitsyn, 1992, 1995). However, most of the reported data on free energies of unfolding and *m*-values have been obtained under a two-state analysis of unfolding studies monitored by one or, at the most, two techniques. This fact makes uncertain the validity of the data reported in many studies. One of the possible explanations for this fact could

Received for publication 27 January 1998 and in final form 31 March 1998.

Address reprint requests to Dr. Jose L. Soulages, Department of Biochemistry, Biosciences West 440, University of Arizona, Tucson, AZ 85721. Tel.: 520-621-1772; Fax: 520-621-9288; E-mail: jlsoulag@u.arizona.edu.

0006-3495/98/07/484/09 $2.00

be the lack of knowledge about the magnitude of the error that could affect the deduced parameters when the two-state assumption is not actually valid. The potential problem of the nonvalidity of the two-state assumption is further increased by the large number of proteins that are being generated by site-directed mutagenesis. These are the proteins that are being used to study the role of certain amino acid residues, or interactions, on the thermodynamic properties of the protein.

The accuracy of the parameters, ΔG_u and *m*-value, obtained through a two-state analysis of a three-state unfolding process, will be dependent on the properties of the intermediate and the concentration reached along the unfolding curve. It will also depend on the technique used to monitor the unfolding transition.

A clear example of the potential misleading results that would result from a two-state assumption was pointed out by Carra and Privalov (1995), who by means of a three-state analysis of calorimetric data were able to solve the apparently contradictory role of certain mutations in the energetics of the staphylococcal nuclease.

The purpose of the current study consisted in the evaluation of the potential errors in the Gibbs free energy of unfolding and the *m*-values that are derived from chemical denaturation studies of three-state unfolding processes, which are analyzed under the two-state unfolding assumption. This study provides the magnitude of the errors that can be expected in a wide range of unfolding energies.

In the estimation of the errors, two different experimental situations were considered. First, it was assumed that the physical property determined to monitor the denaturation does not distinguish between the native and the intermediate state. This is the case for denaturation experiments monitored by far-UV circular dichroism (CD) and, in some cases, by UV absorption or fluorescence when the intermediate is a molten globular-like state. Second, it was assumed that the physical property does not distinguish between the intermediate and the unfolded state. This would be the case for some proteins studied by UV absorption or by fluorescence.

An increasing awareness about the potential magnitude of these errors is likely to increase the number of meaningful thermodynamic data and, with them, our chances of understanding the energetics of protein structure.

One-step and two-step unfolding transition models

For a single-step unfolding process, N ↔ U, where N is the native state and U is the unfolded state, the equilibrium constant K_u is

$$K_u = X_u/X_n, \tag{1}$$

with X_u and X_n being the molar fractions of U and N, respectively.

To determine K_u, it is necessary to count with a physical method to determine the concentration of molecules in the native and unfolded states. Because K_u is usually small, $K_u \ll 1$, and the properties of both conformational species must be known, the determination of the equilibrium constant requires the perturbation of the system. This is accomplished by the addition of a denaturant. The denaturant increases K_u, allowing the determination of the physical property for the individual species and, with them, K_u at different concentrations of denaturant. If the physical property used is linearly related to the concentration of the species, then

$$\vartheta_{obs} = \vartheta_n X_n + \vartheta_u X_u \quad \text{and} \quad K_u = (\vartheta_{obs} - \vartheta_n)/(\vartheta_u - \vartheta_{obs}), \tag{2}$$

where ϑ_{obs}, ϑ_n and ϑ_u represent the observed property, the property of the native-state, and the property of the unfolded state, respectively.

The free energy of unfolding in the absence of denaturant, ΔG_u^w, is obtained by the linear extrapolation model (Greene and Pace, 1974; Pace, 1986). The relationship between the concentration of denaturant *D* and ΔG_u is approximated by the following equation:

$$\Delta G_u = \Delta G_u^w - m^*D, \tag{3}$$

where $\Delta G_u = -RT \ln(K_d)$ and $m = \delta\Delta G_u/\delta D$.

For a two-state unfolding process, the ΔG_u^w and the *m* parameter can be obtained by either linear regression using the equilibrium constants of the transition region of the denaturation curve or by nonlinear regression to an equation of the form (Santoro and Bolen, 1988):

$$\vartheta_{obs} = [a + b^*D + (c + d^*D)e^{(-\Delta Gu,w+m^*D)/RT}]/(1 + e^{(-\Delta Gu,w+m^*D)/RT}) \tag{4}$$

In the previous equation, *a* and *c* are ϑ_n and ϑ_u, respectively. The parameters *b* and *d* are given by $(\delta\vartheta_n/\delta D)$ and $(\delta\vartheta_u/\delta D)$, respectively, and represent the dependence of the physical property of the native and unfolded states with the concentration of denaturant.

Linear or nonlinear regression should, in principle, render the same results. Nonlinear regression has the advantage that it allows the curve fitting when the stability of the native state is very low and its physical property cannot be directly determined.

For an unfolding transition involving one intermediate, the observed property ϑ_{obs} is given by

$$\vartheta_{obs} = \vartheta_n X_n + \vartheta_i X_i + \vartheta_u X_u \tag{5}$$

We will analyze two particular cases: 1) the physical property ϑ has similar or identical values for the native and intermediate states, $\vartheta_n = \vartheta_I$ (case I), and 2) the physical property does not distinguish between intermediate and unfolded states, $\vartheta_u = \vartheta_I$ (case II). For these two experimental situations, the observed property will be related to the concentration of the species in equilibrium by Eqs. 6 and

7, respectively.

$$\vartheta_{obs} = \vartheta_n(X_n + X_i) + \vartheta_u X_u \quad (6)$$

$$\vartheta_{obs} = \vartheta_n X_n + \vartheta_u(X_u + X_i) \quad (7)$$

The concentration of the three species are related by the equilibrium constants that define the $N \leftrightarrow I \leftrightarrow U$ process:

$$X_n = 1/(1 + K_1 + K_1K_2); \quad X_I = K_1/(1 + K_1 + K_1K_2); \quad X_u = K_1K_2/(1 + K_1 + K_1K_2), \quad (8)$$

where K_1 and K_2 correspond to the $N \leftrightarrow I$ and $I \leftrightarrow U$ equilibrium, respectively.

$$K_1 = X_I/X_n; \quad K_2 = X_U/X_I$$

Replacing the concentration of the three species in Eqs. 6 and 7 by the expressions given in Eqs. 8 we obtain the relationship between the observed physical property and the equilibrium constants for cases a and b, respectively.

$$\text{Case I:} \quad \vartheta_{obs} = [\vartheta_n + \vartheta_u(K_u/(1 + K_1))]/[1 + (K_u/(1 + K_1))] \quad (9)$$

$$\text{Case II:} \quad \vartheta_{obs} = [\vartheta_n + \vartheta_u(K_u + K_1)]/[1 + K_u + K_1], \quad (9)$$

where K_u is the denaturation constant for the total process $N \leftrightarrow U$, or $K_u = K_1{*}K_2 = X_u/X_n$.

For a three-state denaturation, Eq. 4 must be modified as follows:

$$\vartheta_{obs} = [a + b{*}D + (c + d{*}D){*}A{*}e^{(-\Delta Guw + m{*}D)/RT)}]/(1 + A{*}e^{(-\Delta Guw + m{*}D)/RT)}), \quad (10)$$

where ΔG_u^w is the free energy of unfolding of the native state in buffer, and A will be given by

$$A = 1/(1 + K_1) = [1 + K_{01}^{*}\exp(m_1^{*}D/RT)]^{-1} \quad \text{when } \vartheta_n = \vartheta_I$$

or by

$$A = [(K_{01}/K_u){*}\exp((m_1 - m){*}D/RT)] + 1 \quad \text{when } \vartheta_u = \vartheta_I$$

K_{01} is the equilibrium constant for the $N \leftrightarrow I$ transition in the absence of denaturant, and m_1 is the m-value defined in Eq. 3 but corresponding to the $N \leftrightarrow I$ process.

METHODS

The unfolding curves were generated with the Prism program by GraphPad Software (San Diego, CA). No random error was included in the data. In all cases, the unfolding temperature was 25°C. Nonlinear regression was carried out with the same program, which uses the algorithm of Levenburg and Marquardt.

RESULTS

This study considered two limiting, but common, situations. First, we will present the results obtained under the assumption that the native and the intermediate states cannot be distinguished by the physical property used to determine the denaturation curve (case I). Second, we will show the results obtained under the assumption that unfolded and intermediate states are undistinguishable (case II). In both cases, we also assumed that the chemical equilibrium between native and intermediate states could be described by an equation similar to that describing the unfolding of the native state (Eq. 3). Under this assumption, the equilibrium between the native and the intermediate state is defined by two parameters, K_{01} and m_1, which were described in Eq. 10. To study the effect of the presence of an intermediate in the parameters deduced under the assumption of a two-state unfolding process, we studied the impact of K_{01} and m_1 on the unfolding curves. The impact of the intermediate is expected to be dependent on the energy of unfolding of the global process $N \leftrightarrow U$ and its characteristic m-value. Therefore, 12 sets of unfolding curves were generated, according to Eq. 9, for K_u values of 10^{-1}, 10^{-3}, 10^{-5}, and 10^{-7} and m-values of 1500, 3000, and 6000 cal mol^{-1} M^{-1}. Each data set included calculations for 6 m_1-values and at least 10 K_{11} values in the 0–1 range. The m_1-value cannot be larger than the m-value for the global unfolding process. Previous three-state studies on the chemical unfolding of proteins reported m-values of the order of $0.5m$ for the unfolding of the intermediate (Barrick and Baldwin, 1993; Kuwajima et al., 1976). The calculations of this study included m_1-values in the range of $0m$ to $0.5m$.

The unfolding curves generated for each set of parameters, K_u, m, m_1, and K_{01}, were fitted to a two-state unfolding process (Eq. 4). This fitting provided the apparent free energies and m-values of unfolding. These apparent values were compared with the values used in the simulation and the differences expressed as percent error in ΔG_u ($\Delta\Delta G_u$%) and m-values (Δm%).

Case I: $\vartheta_n = \vartheta_I$

This is the case that would apply when the intermediate has the same secondary structure as the native state and the unfolding is monitored by far-UV CD spectroscopy. Independently of the secondary structure, it would also take place when the fluorescence quantum yield or the UV absorption properties of the intermediate and native state are indistinguishable and the unfolding is monitored by the corresponding techniques.

Fig. 1 shows the effects of the magnitudes of K_{01} and m_1 on the error in ΔG_u for three different values of K_u, which cover the free energies of unfolding of most proteins, 3–10 kcal/mol. Fig. 2 shows the error in the m-value obtained for the same set of parameters used in Fig. 1. The inspection of Fig. 1 shows that when m_1 is small, $m_1 \leq 0.2m$, only minor errors in ΔG_u, $\Delta\Delta G_u\% \leq 10\%$, will be introduced by the presence of an undetected intermediate. However, for larger m_1-values, $0.2m < m_1 < 0.5m$, the curves reach a minimum at a certain value of K_{01}, where the potential error in ΔG_u becomes maximal and significant. Conversely, the error curves for the m-value (Fig. 2) show a continuous increase

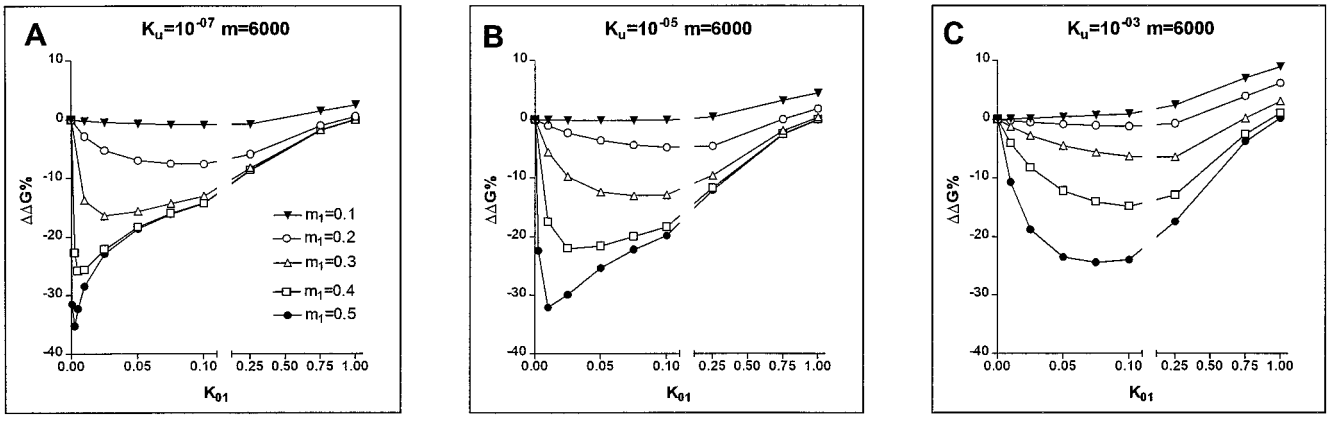

FIGURE 1 Impact of the presence of an undetected intermediate on the estimation of the free energy change for the total unfolding of the native state, $\Delta G_u(N \rightarrow U)$, calculated under the two-state assumption (case I). Effect of the equilibrium constant for the $N \leftrightarrow I$ process, K_1, and the m_1 value on the error in $\Delta G_u(N \rightarrow U)$, $\Delta\Delta G_u\%$. The error in ΔG_u is defined as $\Delta\Delta G_u\% = 100 \times [\Delta G_u(\text{apparent}) - \Delta G_u(\text{actual})]/\Delta G_u(\text{actual})$. (*A–C*) Results obtained for unfolding transitions ($N \rightarrow U$) of 9.6 kcal/mol, 6.8 kcal/mol, and 4.1 kcal/mol, respectively. The calculations shown were obtained for case I, (ϑ), assuming an *m*-value for the global process $N \rightarrow U$ of 6000 cal mol^{-1} M^{-1}.

in error for increasing K_{01} values. The error in the *m*-value also shows a faster increase with K_{01} than the error in ΔG_u. The error in *m*-values rapidly reaches a plateau, where the error is maximal and equal to the m_1/m ratio.

It is interesting to note that, even though the apparent concentration of the native state, $[N]_{app} = [N] + [I]$, is overestimated and the unfolding curves are shifted to higher concentrations of denaturant, the free energy of unfolding will be, in general, underestimated. This is so because the rate of decrease of the apparent *m*-values is faster than the rate of increase in the midpoint of the transition curve.

To exemplify the effect of K_{01} in the shape and position of the unfolding curves, Fig. 3 shows the effect of a very low, an intermediate, and a high K_{01} in the concentration of N, I, and U for a given unfolding constant and m_1-value. For very low K_{01} values, highly unstable intermediates, we will not observe the accumulation of the intermediate along the denaturation, and the unfolding curve will resemble that of a two-state unfolding process (see Fig. 3 *C*). In this case, the two-state unfolding assumption will be valid and both ΔG_u and *m* will be accurately estimated. On the other extreme, for intermediates of similar stability to that of the native state, large K_{01} values, we have a rapid accumulation of intermediate driven by m_1. This accumulation of intermediate produces a major shift in the denaturation curve toward higher concentrations of denaturant (Fig. 3 *A*). The observed transition will take place in the absence of the native state and will correspond to the $I \leftrightarrow U$ equilibrium. However, given that the unfolding energy of the intermediate is similar to that of the native state, $K_{01} \rightarrow 1$, $\Delta G_2 = -RT\ln(K_u/K_{01}) = -RT\ln(K_2) \approx \Delta G_u$, the free energy of unfolding will be only marginally underestimated. It is important to remark that, even though the error in ΔG_u will be small, the two-state assumption will not be valid and we will still have a major error in the estimation of the *m*-value. A simple inspection of the slopes of the denaturation curves

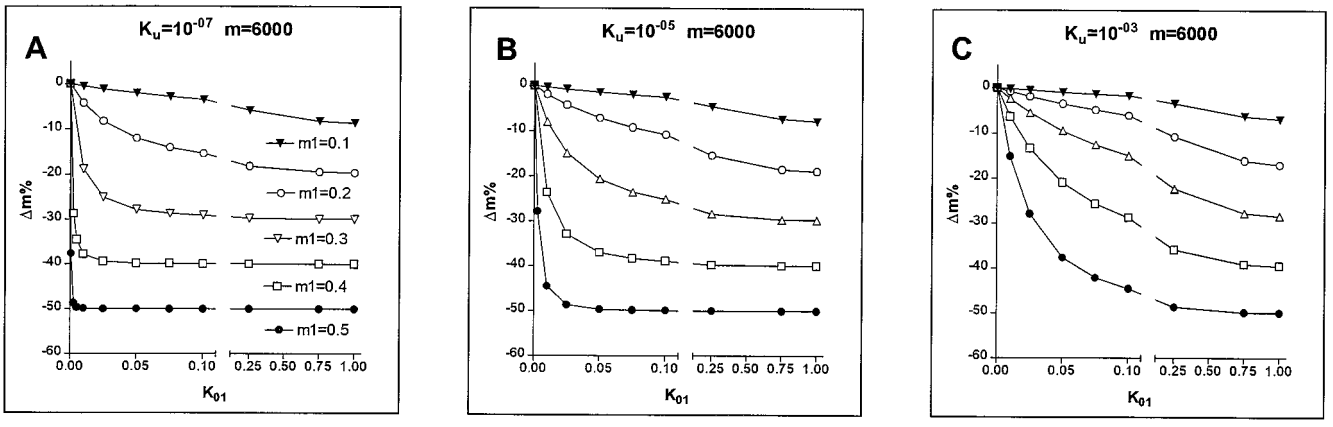

FIGURE 2 Case I: impact of the presence of an undetected intermediate on the estimation of the *m*-value for the total unfolding of the native state, $\Delta G_u(N \rightarrow U)$, calculated under the two-state assumption. The error in the *m*-value is defined as $\Delta m\% = 100 \times [m\text{-value(apparent)} - m\text{-value(actual)}]/m\text{-value(actual)}$. (*A–C*) Results obtained for unfolding transitions ($N \rightarrow U$) of 9.6 kcal/mol, 6.8 kcal/mol, and 4.1 kcal/mol, respectively.

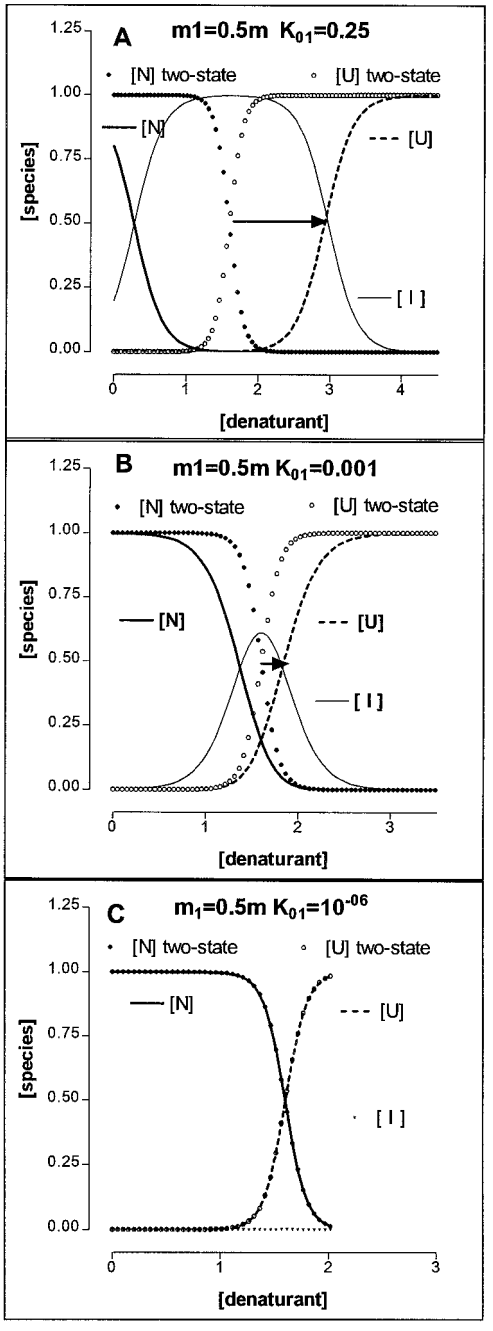

FIGURE 3 Case I: effect of the magnitude of the equilibrium constant for the N ↔ I process, K_{01}, on the shape and position of the unfolding curves and comparison with a two-state unfolding. The dependence of concentrations of unfolded and native states with the concentration of denaturant was calculated for a true two-state unfolding process ([N] two-state and [U] two-state) of 9.6 kcal/mol and an *m*-value of 6000 cal mol^{-1} M^{-1}. These curves are compared with the curves that would be obtained for three-state unfolding processes of the same unfolding free energy but differing in the K_{01} value. (*A*) $K_{01} = 0.25$; (*B*) $K_{01} = 10^{-3}$; (*C*) $K_{01} = 10^{-6}$. In all cases, the m_1-value was assumed to be 3000 cal mol^{-1} M^{-1}. The arrows indicate the expected shift in the midpoint of the denaturation curves. Symbols are indicated in the figure.

for the unfolded state, in the presence and in the absence of intermediate, allows us to see the effect of large K_{01} values in the apparent *m*-value (Fig. 3 *A*).

The major errors in the estimation of ΔG_u will take place for intermediates with a stability that is lower than that of the native state (Fig. 3 *B*) and considerably higher than that of the unfolded state. In this broad region of K_{01} values, the potential error in ΔG_u becomes significant and, for an unfolding transition of 9.6 kcal/mol and $m_1 = 0.5m$, could reach up to 35% of the actual ΔG_u (Fig. 1 *A*). The larger error in ΔG_u observed in this region of K_{01} values is given by the fact that, at moderate K_{01} values, the apparent *m*-values change much faster than the shifts in midpoints of the denaturation curves.

Interestingly, for a given m_1-value, the maximal potential error in both ΔG_u and *m*-values increases with the energy of the unfolding process (Fig. 4). The error in the *m*-value, $|\Delta m\%|$, is much larger than the percent error in ΔG_u. Also, conversely to the error in ΔG_u, the maximal percent error in *m*-values is not affected by the magnitude of K_u. It depends only on the ratio of m_1/m.

Another feature of the potential error in *m*-values and ΔG_u is their independence on the *m*-value of the global process. They are dependent only on the relative magnitude of the m_1-value compared with the *m*-value of the global process. That is why the error curves in Figs. 1 and 2 were identified by the m_1/m ratio.

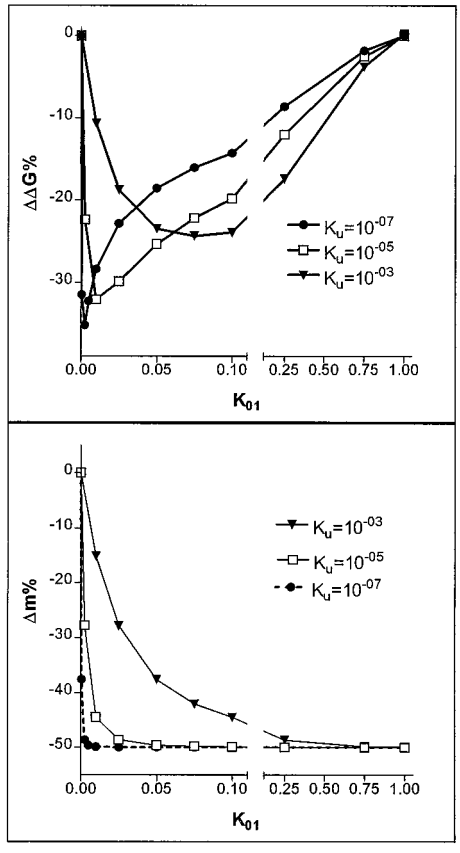

FIGURE 4 Case I: dependence of the potential errors in ΔG_u and *m*-values on the free energy of unfolding of the process. The top panel shows the error in ΔG_u for three different energies of unfolding. The bottom panel shows the error in *m*-value for the same unfolding energies.

A particular case not commented on yet is observed for highly unstable native states ($K_u > 0.1$). In these cases, depending on the magnitudes of K_{01} and the m_1-value, the ΔG_u could be significantly overestimated. Fig. 5 shows the errors expected for $K_u = 0.1$. In this example, we have 10% of unfolded protein in the absence of denaturant. Because K_{01} should be larger than K_u, a significant concentration of intermediate will also be present in the absence of denaturant. For instance, if $K_u = 0.1$, $K_{01} = 1$, and m_1-value = $0.5m$, the overestimation of ΔG_u will be approximately 16%. The m-value will be, as always, underestimated. In this example, an underestimation of 45% will take place. Note that, conversely to the results observed for higher energies of unfolding, in this case, the absolute value of the error in ΔG_u increases for lower m_1-values.

Case II. ($\vartheta_u = \vartheta_I$)

This situation could be easily found when fluorescence, UV absorption, or near-UV CD is used to monitor the unfolding transition. We will show the results obtained assuming that the m_1-value represents 50% of the m-value ($m_1 = 0.5m$). This m_1-value is consistent with the three-state unfolding studies of α-lactalbumin (Kuwajima et al., 1976) and apomyoglobin (Barrick and Baldwin, 1993). As shown for case I, lower m_1-values would produce lower errors and larger m_1-values would produce larger errors.

Conversely to the case I, when $\vartheta_u = \vartheta_I$ the unfolding curves will be shifted toward lower concentrations of denaturant. Fig. 6 shows the effect of K_{01} on the position and shape of the unfolding curves that would be obtained for a transition of 9.6 kcal. Fitting of the unfolding curves to a two-state model will always translate into underestimated values of ΔG_u and m-values. As seen in case I, the potential error in both parameters ΔG_u and m-value increases with the energy of the unfolding process (Fig. 7). Also, as in case I, the error in m-value rapidly reaches a plateau at an apparent m-value $m_{app} = m - m_1$. In the 4–10 kcal/mol energy range and for a given K_{01} and m_1-value, the error in m-values are approximately the same for cases I and II (Fig. 7 *B*). However, the dependence of the error in ΔG_u with K_{01} shows two major differences with the results obtained for case I ($\vartheta_n = \vartheta_I$). In case II ($\vartheta_u = \vartheta_I$), the error in ΔG_u shows a continuous increase with K_{01}, whereas a minimum is observed in case I (Fig. 7 *A*). Also, at any given K_{01} value, the potential underestimation of ΔG_u is larger for case II. The potential error in ΔG_u is larger in case II because there is no possibility for a compensation, as happens in case I, between the changes in position and slope of the transition curves. If the physical property used to monitor the unfolding does not distinguish between intermediate and unfolded states, the presence of an intermediate translates into a shift of the unfolding curve toward a lower concentration of denaturant and a simultaneous decrease in the slope of the transition. Consequently, both changes contribute to the underestimation of ΔG_u.

In Fig. 8, we compared the errors obtained in cases I and II for unfolding transitions of low energy. Two differences

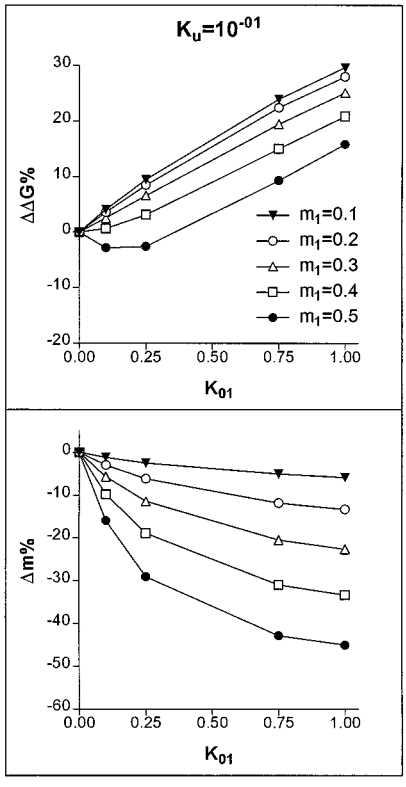

FIGURE 5 Case I: potential errors for highly unstable native states. The top panel shows the effect of the equilibrium constant for the N ↔ I process, K_{01}, and the m_1-value on the error in ΔG_u(N → U), $\Delta\Delta G_u$%, for an unfolding process of 1.4 kcal/mol ($K_u = 0.1$) and an m-value = 6000 kcal/mol. The m_1-values are expressed as a fraction of the global m-value. The bottom panel shows the errors in m-value corresponding to the conditions indicated above.

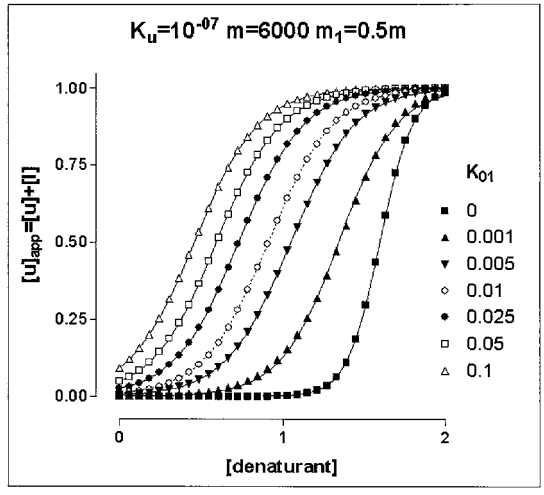

FIGURE 6 Case II: effect of the magnitude of the equilibrium constant for the N ↔ I process, K_{01}, on the shape and position of the unfolding curves when ϑ. In the example, $K_u = 10^{-7}$, m-value = 6000 cal mol^{-1} M^{-1}, and the m_1-value = 3000 cal mol^{-1} M^{-1}. The values of K_{01} are given in the figure. The curve obtained for $K_{01} = 0$ represents the two-state unfolding curve.

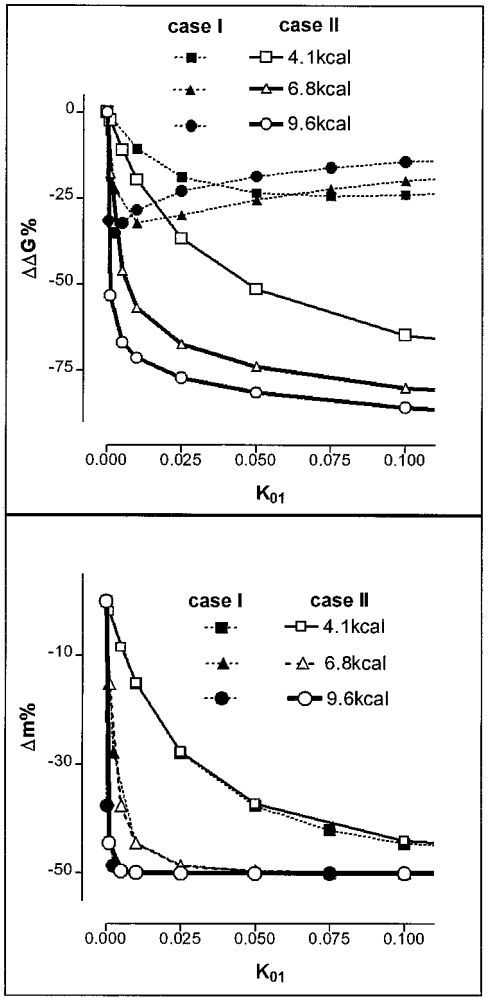

FIGURE 7 Case II: impact of the magnitude of the equilibrium constant for the N ↔ I process on the potential errors in ΔG_u (*top panel*) and *m*-values (*bottom panel*) on the free energy of unfolding of the process. The top panel shows the error in ΔG_u for three different energies of unfolding. The bottom panel shows the error in *m*-value for the same unfolding energies. The errors that would be obtained under case I were also included in the figure to allow comparison between cases I and II.

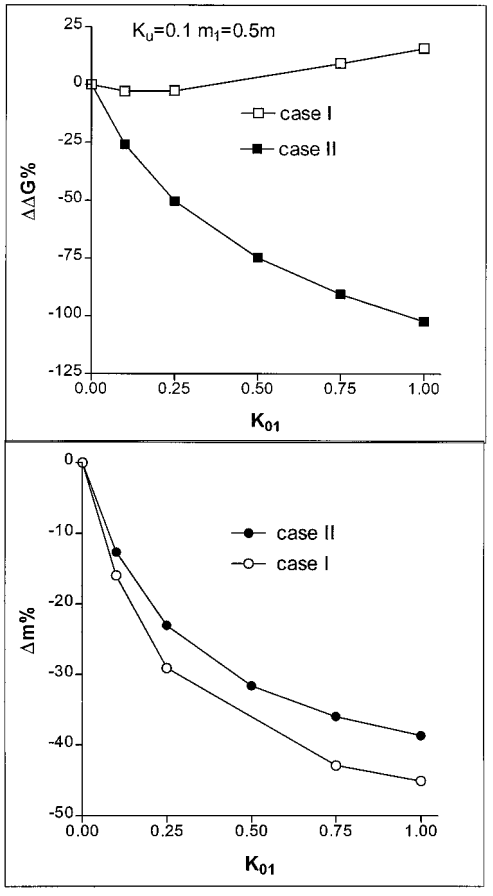

FIGURE 8 Case II: potential errors for highly unstable native states. The top panel shows the effect of the equilibrium constant for the N ↔ I process, K_{01}, on the error in ΔG_u(N → U), for an unfolding process of 1.4 kcal/mol (K_u = 0.1) and an *m*-value of 6000 cal mol^{-1} M^{-1}. The m_1-value was 0.5*m* or 3000 cal mol^{-1} M^{-1}. The bottom panel shows the errors in *m*-value corresponding to the conditions indicated above. The errors that would be obtained under case I were also included in the figure to allow comparison between cases I and II.

were found. In case II, the ΔG_u is likely to be largely underestimated, whereas in case I, we could have a significant overestimation of ΔG_u. The *m*-values will be underestimated in both cases, but the underestimation will be slightly smaller in case II.

DISCUSSION

The presence of three states in equilibrium is not a problem by itself. Systems of three components allow the estimation of thermodynamic parameters if the concentrations of the intermediates can be determined along the denaturation curve. This has been done at least on two occasions in chemical unfolding studies (Kuwajima et al., 1976; Barrick and Baldwin, 1993). Problems arise when it is assumed that the unfolding proceeds through a two-state mechanism and/or the physical properties used to monitor the unfolding, if more than one is used, do not reflect the structural changes that accompany the formation of the intermediate. The easy access to the production of mutants has increased the number of proteins enormously and, with that, the chances to have intermediates at significant concentrations along the unfolding curves. It is important to take into account that a mutation can affect the stability of the native state as much as it can affect the stability of the intermediate state. It is also possible that only the stability of the intermediate state could be affected by a mutation. Even in the case that a wild-type protein had been shown to unfold by a two-state like mechanism, it is not guaranteed that the mutants will unfold through a two-state-like transition. This situation could also be triggered by a change in pH or temperature. This problem has not been clearly seen yet because most of the studies using recombinant proteins limit the characterization of the protein to few physical-chemical studies or functional properties, which are not likely to detect the presence of intermediates. There is one report that

described this problem with a mutant of the staphylococcal nuclease. The contradictory thermodynamic data obtained for one of the mutants of staphylococcal nuclease under the two-state assumption were solved when the involvement of a third component was included in the analysis of the unfolding (Carra and Privalov, 1995).

There is a general consent that protein folding proceeds through an intermediate state that has a remarkably native-like structure (Kim and Baldwin, 1990). Increasing experimental evidence indicates that a large number of wild-type proteins can adopt conformations of intermediate stability, between the native and the unfolded states, under diverse conditions of pH, temperature, and solvent (Ptitsyn, 1995; Fink, 1995). It is worth noting the increasing number of proteins (probably more than 100) that have been reported to adopt molten globular states under diverse conditions of solvent and temperature. There were more than 60 reports in 1995 (Ptitsyn, 1995). Many of these partially folded conformations, such as the molten globule, share common physical properties with both the native state and the unfolded states. Therefore, the chance of detecting the presence of an intermediate will depend on the protein properties and the number and type of techniques used to characterize and monitor the unfolding study. For instance, when the intermediate is a molten globule, it usually has the same, or a very similar, secondary structure as the native state (Ptitsyn, 1995). In this case, a far-UV CD study could not distinguish between the native state and the molten globule. An unfolding experiment monitored by this technique would be overestimating the concentration of the native state. It would be detecting the sum of the concentrations of the native and intermediate state (case I). Moreover, even if affected by a change in secondary structure, an intermediate could have the same quantum yield of fluorescence or UV absorption properties as either the native or the unfolded protein states. Therefore, to have a reasonable confidence about the two-state nature of the unfolding process, it is always convenient to monitor the unfolding curve by means of at least three different techniques.

The *m*-values constitute one of the main pieces of information used to study the role of specific mutations on the structural stability of the protein (Shortle, 1995, 1989). The *m*-values and free energies of unfolding of more than 100 mutants of staphylococcal nuclease have been determined by chemical denaturation (Shortle, 1995). Most of the mutants showed a $\Delta m < 0$ ($\Delta m = m_{\text{mutant}} - m_{\text{native}}$) as well as a decreased stability (Shortle, 1995). Several of the mutants with decreased *m*-values were later found to unfold by a three-state mechanism with increasing temperature (Carra et al., 1994; Carra and Privalov, 1995). The results of our study suggest that the large number of mutants displaying decreased *m*-values could be partially due to the lack of validity of the two-state assumption, at least for some of the mutants. Similarly, the apparently excessive number of destabilizing mutations observed with staphylococcal nuclease and T4 lysozyme (Matthews, 1995) could also be affected by the nonvalidity of the two-state assumption. As indicated by the results obtained in this study, the presence of an intermediate will almost always imply a significant underestimation of both the *m*-values and the free energy of unfolding. It is important to note that these results apply whether or not the mutations stabilize or destabilize the native state of the protein. If a significant concentration of intermediate accumulates along the unfolding curve, a moderately stabilizing mutation could appear as a destabilizing mutation and a destabilizing mutation could appear much more destabilizing than it actually is.

One question that we can ask is whether the results of ΔG_u and *m*-value obtained for a given mutant, by a single physical method, could be affected by the presence of an undetected intermediate. If the unfolding of the wild-type protein is shown to be consistent with a two-state process, the comparison of the results obtained with the wild-type and the mutant protein sometimes can answer that question. For instance, if the mutation is apparently stabilizing, and the stability of the mutant is in the 3–12 kcal/mol range, then it is likely that the mutation was truly stabilizing. However, the extent of the stabilization could still be underestimated. To discard the presence of an intermediate, we should observe no change in the *m*-value or a slight increase. On the other hand, the validity of apparently destabilizing mutations can be confirmed only when the *m*-value of the mutant is similar to the *m*-value of the wild type. If the *m*-value of the mutant is smaller than that of the wild type, the apparent destabilization could only be the result of the presence of an intermediate.

Concluding remarks

The presence of a significant concentration of intermediate along the unfolding curve will always translate into an underestimation of the *m*-values for the $N \leftrightarrow U$ process. The magnitude of the maximal potential error is limited by the *m*-value that characterizes the $N \rightarrow I$ transition.

In most non-two-state transitions, the ΔG_u will also be underestimated.

In general, we can expect lower errors in ΔG_u for unfolding transitions that are monitored by far-UV CD (case I) rather than by fluorescence or UV absorption.

A mutant that has a significantly reduced *m*-value may indicate the presence of a non-two-state unfolding process.

The magnitude of the errors in both ΔG_u and the *m*-value is not affected by the procedure chosen for fitting. Linear regression of ΔG versus [denaturant] using the transition region of the unfolding curve or nonlinear regression including the data points that are before and at the onsets of the transitions will result in the same percent error.

I thank Drs. John Rupley and Gordon Tollin for their useful comments and suggestions and Roberto Nussenzveig for his help with the calculations. This work was supported by National Institutes of Health grant GM 55622.

REFERENCES

Barrick, D., and R. L. Baldwin. 1993. Three-state analysis of sperm whale apomyoglobin folding. *Biochemistry*. 32:3790–3796.

Brandts, J. F. 1969. Conformational transitions of proteins in water and aqueous mixtures. *In* Structure and Stability of Biological Macromolecules. S. N. Timasheff and Fasman, G. D., editors. Marcel Dekker, New York. 213–290.

Carra, J. H., E. A. Anderson, and P. L. Privalov. 1994. Three-state thermodynamic analysis of the denaturation of staphylococcal nuclease mutants. *Biochemistry*. 33:10842–10850.

Carra, J. H., and P. L. Privalov. 1995. Energetics of denaturation and m-values of staphylococcal nuclease mutants. *Biochemistry*. 34: 2034–2041.

Fink, A. L. 1995. Molten globules. *Methods Mol. Biol.* 40:343–360.

Greene, R. F. J., and C. N. Pace. 1974. Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, alpha-chymotrysin and beta-lactaglobulin. *J. Biol. Chem.* 249:5388–5393.

Kim, P. S., and R. L. Baldwin. 1990. Intermediates in the folding reactions of small proteins. *Annu. Rev. Biochem.* 59:631–660.

Kuwajima, K., K. Nitta, M. Yoneyama, and S. Sugai. 1976. Three-state denaturation of α-lactalbumin by guanidine hydrochloride. *J. Mol. Biol.* 106:359–373.

Matthews, B. W. 1995. Studies on protein stability with T4 lysozyme. *Adv. Prot. Chem.* 46:249–278

Pace, N. C. 1986. Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* 131:266–280.

Privalov, P. L. 1979. Stability of proteins: small globular proteins. *Adv. Prot. Chem.* 33:167–197.

Ptitsyn, O. B. 1992. The molten globule state. *In* Protein Folding. T. E. Creighton, editor. Freeman, New York. 243–300.

Ptitsyn, O. B. 1995. Molten globule and protein folding. *Adv. Protein Chem.* 47:83–229.

Santoro, M. M., and D. W. Bolen. 1988. Unfolding free energy changes determined by the linear extrapolation method. I. Unfolding of phenylmethanesulfonyl α-chymotrypsin using different denaturants. *Biochemistry*. 27:8063–8068.

Santoro, M. M., and D. W. Bolen. 1992. A test of the linear extrapolation of unfolding free energy changes over an extended denaturant concentration range. *Biochemistry*. 31:4901–4907.

Schellman, J. 1978. Solvent denaturation. *Biopolymers*. 17:1305–1322.

Shortle, D. 1989. Probing the determinants of protein folding and stability with amino acid substitutions. *J. Biol. Chem.* 264:5315–5318.

Shortle, D. 1995. Staphylococcal nuclease: a showcase of m-value effects. *Adv. Prot. Chem.* 46:217–278.

Yao, M., and D. W. Bolen. 1995. How valid are the denaturant-induced unfolding free-energy measurements? level of conformance to common assumptions over an extended range of ribonuclease A stability. *Biochemistry*. 34:3771–3781.