Original article:

A method for determining states in the course of protein unfolding

Mostafa Rezaei-Tavirani [1,2,*], Sayed-Amir Marashi [3], Seyed Mohammad Mahdavi [1]

[1]Asre-Novin Institute of Research and Industrial Services, Unit 15, Nariman Building, Kamranyeh Junction, Shaheed-Bahonar Ave., Tehran, Iran; [2]Faculty of Medicine, Medical University of Ilam, Ilam, Iran; [3]Department of Biotechnology, University College of Science, University of Tehran, Tehran, Iran. *Corresponding author. E-mail address: rezaei.tavirani@ibb.ut.ac.ir, Fax: +98-21-22815601

ABSTRACT
A simple model is presented for analyzing a set of spectra obtained from spectrophotometric study of protein titration. With this model one can determine the states of (probable) intermediates in the course of protein unfolding. The model is developed based on abundance of native state, intermediate(s) and denatured state, and their contributions to differential absorbance at selected wavelengths. The model is tested for the two-state unfolding of ribonuclease A by urea in formate buffer, and also for the three-state unfolding of α-lactalbumin by guanidine hydrochloride (GdnHCl) in phosphate buffer. It was demonstrated that unfolding of ribonuclease A is matched acceptably with the two-state model, while α-lactalbumin unfolding starts with a two-state mechanism (when [GdnHCl]≥1.6M) followed by a three-state pathway.

Keywords: Protein denaturation, spectrophotometry, two-state denaturation, three-state denaturation, ribonuclease A, α-lactalbumin

INTRODUCTION

Unraveling the cooperating interaction within domains and among protein scaffold is an important problem. The cooperation mechanism can explain the basis for principles of protein architecture, evolution and functioning. Traditionally, the keyway to understand this mechanism has been unfolding the protein from native state to denatured state (Privalov, 1989). The study of the mechanisms through which a protein becomes denatured involves the identification of (probable) intermediates, and their contributions to the denaturation process. Identification of protein intermediates in the course of protein unfolding and also refolding is of central importance in biochemistry, biology, medicine and biotechnology (Yon, 2001).

Different techniques has been traditionally applied to identify protein folding/unfolding states: circular dichroism spectroscopy (Chedad and Van Dael, 2004; Kuwajima, et al., 1976), size exclusion chromatography (Martins and Santoro, 1999; Samuel, et al., 2000) and NMR spectroscopy (Ropson and Frieden, 1992; Samuel, et al., 2000). In the present work, a model is presented to determine states in the course of protein denaturation based on

protein differential spectroscopy at different wavelengths. Besides the simplicity of the model, we demonstrated that it is able to successfully explain the behavior of spectra during protein unfolding.

MATERIALS AND METHODS

Materials
Ribonuclease A, α-lactalbumin, urea and guanidine hydrochloride (GdnHCl) were obtained from Sigma. All other materials and reagents were of analytical grade and purchased from Merck. The solutions were in distilled water. KCl (0.1 M) and phosphate buffer (20 mM, pH=6.8) for α-lactalbumin solution and formate buffer (30 mM, pH=3.55) for ribonuclease A was used. Protein concentration was 0.2 mg/ml and temperature was kept at 300 K during the experiments.

Methods
All experiments were performed using a UV 3100 recording spectrophotometer (Shimadzu, Japan). The spectra were obtained in the range of 250 to 300 nm using denaturant solution as the reference and protein in denaturant solution as the sample. Absorbance change, ΔA_λ, was measured as the difference between the absorbance of the native protein and the absorbance of the denatured state at wavelength λ.

Statistical analyses
All the fittings and statistical analyses were done by Minitab[©] 14 software. Linear regressions were obtained with setting intercepts to zero. Unless clearly stated in the text, the default settings were used.

MODELS

Consider a protein solution which obeys Beer-Lambert law, with N_0, A_i and ε_i as its concentration, its absorbance and its extinction coefficient at wavelength i, respectively. If the solution is studied within a cell with a light path length of 1 cm, one can write:

$$A_i = \varepsilon_i N_0 \qquad (1)$$

Consider that a titrant with the concentration of j is added to the protein. For a two-state unfolding process (i.e. negligible population of partially unfolded proteins, which normally happens during protein unfolding (Pace, 1990)):

$$N \leftrightarrow D \qquad (2)$$

and one can say that:

$$N_0 = N_j + D_j \qquad (3)$$

where N_j and D_j are native-state and denatured-state concentrations, respectively. Since only native and denatured states are considered to be present in the solution, we can assume that $A_{i,j}$ includes two independent components regarding the native and denatured states:

$$A_{i,j} = \varepsilon_i N_j + \varepsilon_{i,j} D_j \qquad (4)$$

in which $\varepsilon_{i,j}$ is the extinction coefficient of the denatured state. Hence, absorbance change due to protein unfolding is equal to:

$$\Delta A_{i,j} = A_{i,j} - A_i = \varepsilon_i N_j + \varepsilon_{i,j} D_j - \varepsilon_i (N_j + D_j) \qquad (5)$$

$$\Delta A_{i,j} = (\varepsilon_{i,j} - \varepsilon_i) D_j \qquad (6)$$

For another wavelength, *ii*, with similar notations,

$$\Delta A_{ii,j} = (\varepsilon_{ii,j} - \varepsilon_{ii})D_j \qquad (7)$$

Therefore, there must be a constant value like X_i that:

$$\Delta A_{ii,j} = X_i \Delta A_{i,j} \qquad (8)$$

For a three-state unfolding like:

$$N \leftrightarrow I \leftrightarrow D \qquad (9)$$

consider that during unfolding with denaturant the concentration of the intermediate is I_j and its extinction coefficient is $\varepsilon'_{i,j}$. With similar notations as before, one can write:

$$N_0 = N_j + I_j + D_j \qquad (10)$$

$$A_{i,j} = \varepsilon_i N_j + \varepsilon'_{i,j} I_j + \varepsilon_{i,j} D_j \qquad (11)$$

$$\Delta A_{i,j} = A_{i,j} - A_i =$$
$$\varepsilon_i N_j + \varepsilon'_{i,j} I_j + \varepsilon_{i,j} D - \varepsilon_i (N_j + I_j + D_j) \qquad (12)$$

$$\Delta A_{i,j} = (\varepsilon'_{i,j} - \varepsilon_i)I_j + (\varepsilon_{i,j} - \varepsilon_i)D_j \qquad (13)$$

For two other wavelengths, *ii* and *iii*,

$$\Delta A_{ii,j} = (\varepsilon'_{ii,j} - \varepsilon_{ii})I_j + (\varepsilon_{ii,j} - \varepsilon_{ii})D_j \quad (14)$$

$$\Delta A_{iii,j} = (\varepsilon'_{iii,j} - \varepsilon_{iii})I_j + (\varepsilon_{iii,j} - \varepsilon_{iii})D_j \quad (15)$$

Note that in general no constant value like k exists that $\Delta A_{i,j} = k\Delta A_{ii,j}$.

$\Delta A_{i,j}$ and $\Delta A_{ii,j}$ can be considered as two independent vectors, in which I_j and D_j are their horizontal and vertical elements, respectively. Any desired vector like $\Delta A_{iii,j}$ can be written as the linear combination of $\Delta A_{i,j}$ and $\Delta A_{ii,j}$. Hence, two constant values, like X_i and X_{ii}, exist that satisfy the following equation:

$$\Delta A_{iii,j} = X_i \Delta A_{i,j} + X_{ii} \Delta A_{ii,j} \qquad (16)$$

Equations (8) and (16) can be extended for an *n*-state unfolding process:

$$\Delta A_{n,j} = \sum_{k=1}^{n-1} X_k \Delta A_{k,j} \qquad (17)$$

Using a fitting program, one can estimate X_k values in Equations (8), (16) or (17) based on a set of spectra obtained as a function of denaturant concentration.

RESULTS AND DISCUSSION

From Equation (8) it is clear that for a two-state unfolding process, absorbance change at a certain wavelength must be proportional to the absorbance change at any desired wavelength. The validity of this equation was tested for Ribonuclease A unfolding by urea as a putative two-state process (Pace, 1990). Briefly, A solution of ribonuclease A in formate buffer was prepared as described by Pace (1990). Subsequently, increasing concentrations of urea were added and the increase in absorbance was measured at 270 nm. Figure 1 shows the experimental absorbance changes of this process at 270 nm. In addition, the predicted values are calculated based on the relation between ΔA_{270} and ΔA_{250}:

$$\Delta A_{270} = 1.0805\Delta A_{250}; \quad R^2 = 0.9951 \quad (18)$$
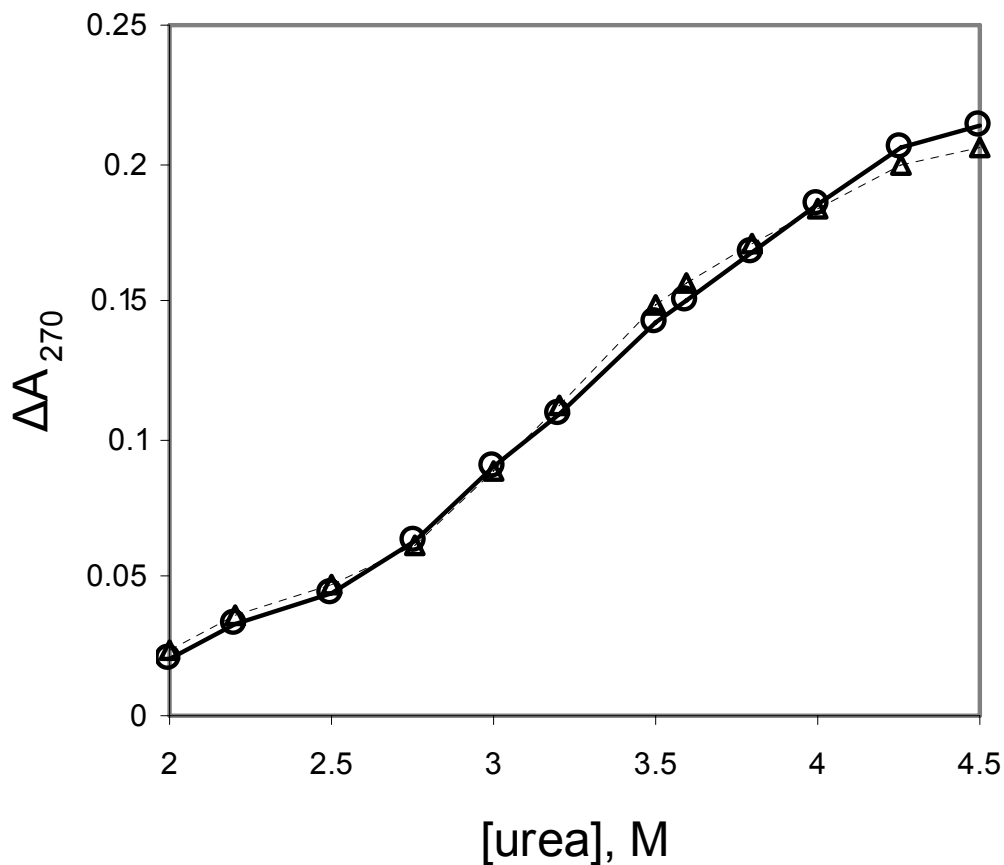
Figure 1: Absorbance change at 270 nm during ribonuclease A unfolding by urea. (○): observed values; (Δ): predicted values based on ΔA_{250} .

The strong correlation between the absorbance changes at the two wavelengths implies that the model successfully works in case of a two-state protein denaturation. We suggest that for any denaturation process, one can test the validity of the two-state denaturation model simply by determining absorbance changes at two different wavelengths and fit them to the linear model of Equation (8).

Equation (16) suggests that for a three-state unfolding process, absorbance change at a certain wavelength must be equal to the linear combination of absorbance changes at two desired wavelengths. The validity of this equation was tested for α-lactalbumin unfolding by GdnHCl as a three-state process. Briefly, A solution of α-lactalbumin in phosphate buffer was prepared as described by Kuwajima et al. (1976). Then, aliquots of GdnHCl was added to the solution and the increase in absorbance was measured at 257, 267 and 287 nm. Figure 2A shows the correlation between the observed ΔA_{267} values and predicted ΔA_{267} values based on ΔA_{287} and ΔA_{257} (i.e. the three-state model; Equation 16).
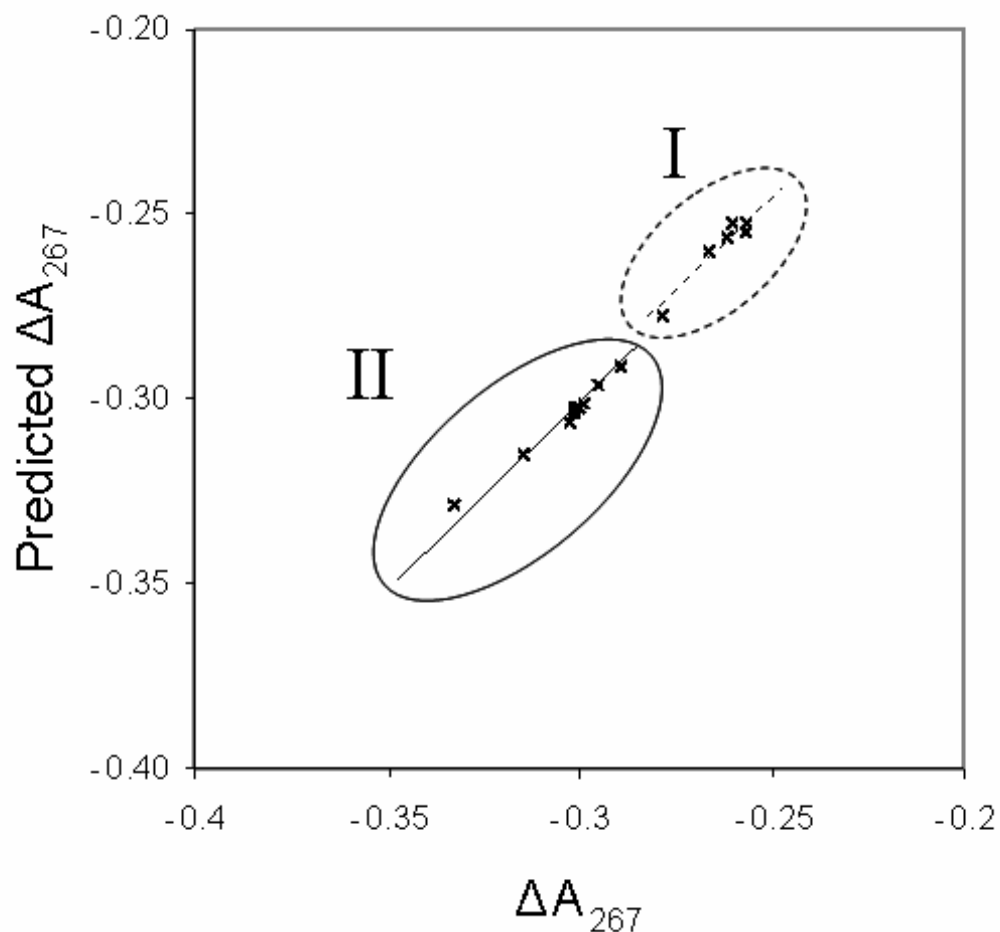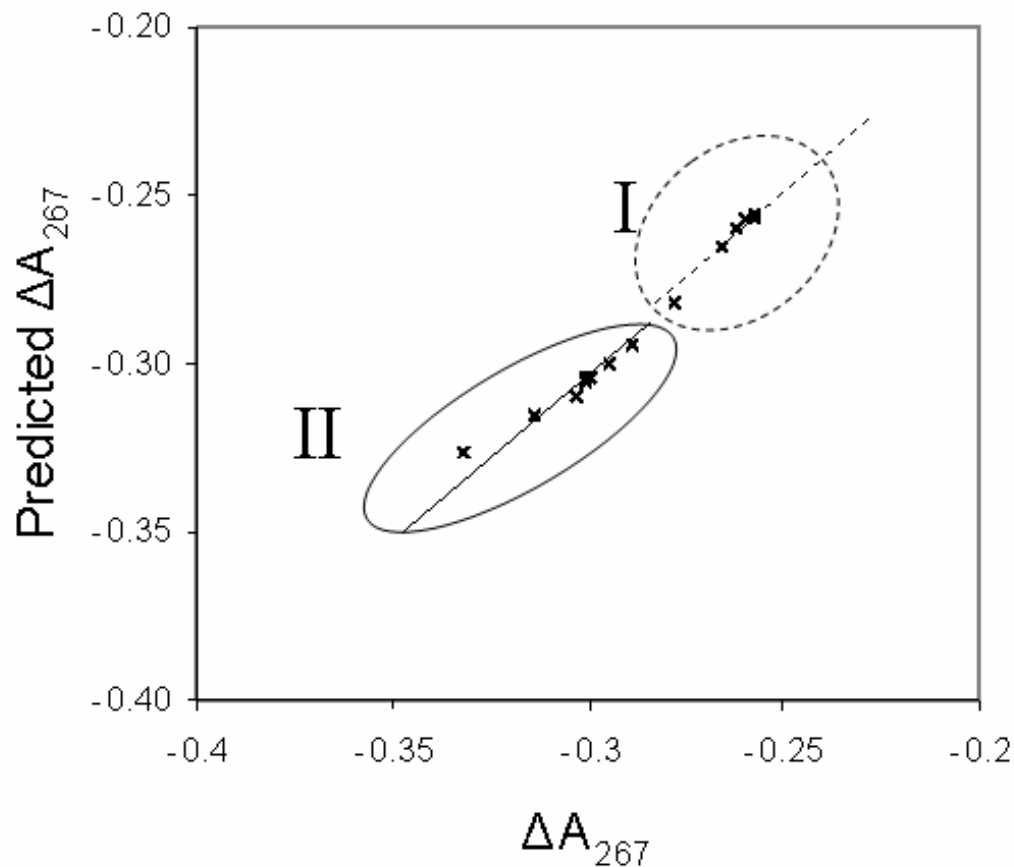
Figure 2A: Predicted vs. observed values for ΔA_{267}. Prediction based on three-state model. Regression lines for cluster I and II are shown by dashed and solid lines, respectively.

Note that if the model is completely successful, we will observe a linear y=x relation between the observed and the predicted ΔA_{267} values. Predicted values in this plot are calculated based on the following equation:

$$\Delta A_{267} = 2.05\Delta A_{287} - 0.743\Delta A_{257} ;$$
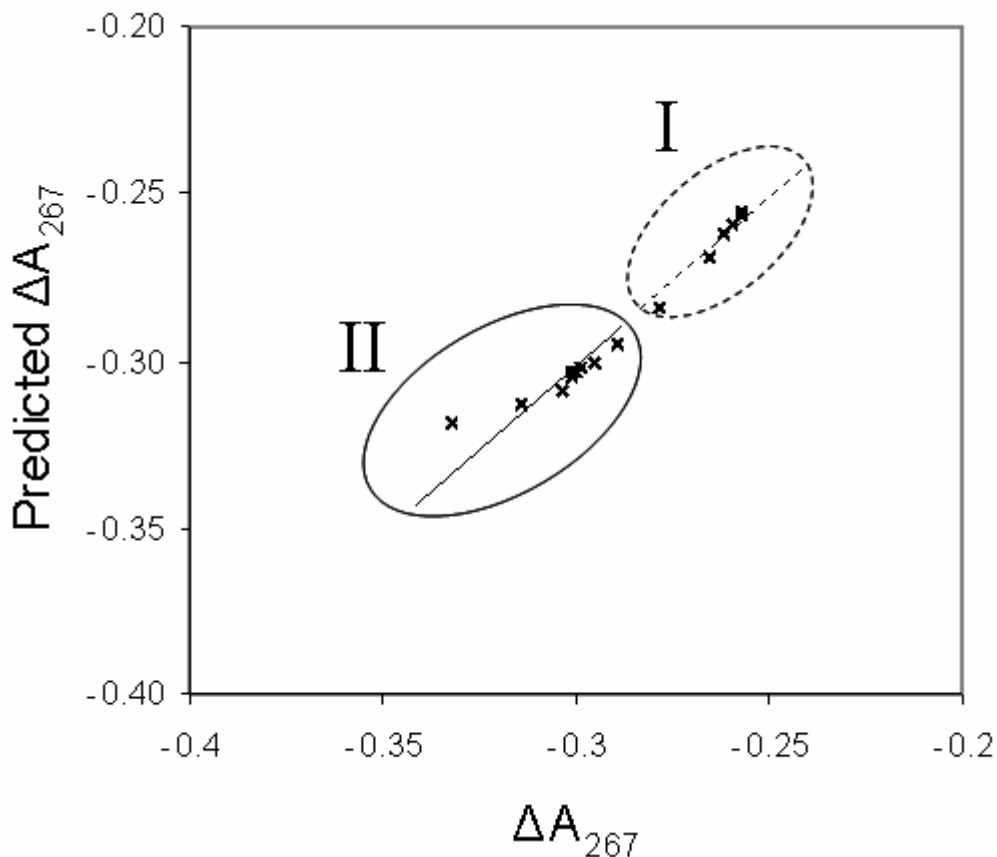$$R^2 = 0.9780 \qquad\qquad (19)$$

Figure 2B: Predicted vs. observed values for ΔA_{267}. Prediction based on two-state model using ΔA_{287} values. Regression lines for cluster I and II are shown by dashed and solid lines, respectively.

With the assumption of a two-state unfolding model, one can estimate ΔA_{267} based on ΔA at a preferred wavelength (Equation 8); the results of such estimations are illustrated in Figure 2B and C, based on ΔA at 287nm and 257nm, respectively:

$$\Delta A_{267} = 1.120 \Delta A_{287}; \quad R^2 = 0.9767 \quad (20)$$

$$\Delta A_{267} = 0.884 \Delta A_{257}; \quad R^2 = 0.9541 \quad (21)$$

Figure 2C: Predicted vs. observed values for ΔA_{267}. Prediction based on two-state model using ΔA_{257} values. Regression lines for cluster I and II are shown by dashed and solid lines, respectively.

The correlation coefficient of the three-state model is greater than that of two-state models. But one may hesitate about the significance of difference between R^2 values in Figure 2. In order to elucidate whether there exists more than one trend for different titrant concentrations, a K-means cluster analysis was performed for each model. Briefly, this method aims to classify observations into groups when the groups are initially unknown. This procedure uses non-hierarchical clustering of observations according to MacQueen's algorithm (see (Johnson and Wichern, 1998)). The results demonstrated that in all cases (i.e. Fig. 2A–C), with the assumption of the existence of two clusters, the first six observations were grouped together (cluster I), and the remaining observations were also clustered together (cluster II). This implies that there are two distinct trends in all of these plots: cluster I, for which [GdnHCl] ≤ 1.8 M, and cluster II, for which [GdnHCl] > 1.8 M.

Table 1 summarizes the information of linear regressions to cluster I and II in case of Figure 2A–C. For Fig. 2A, the fitted line to the first cluster (y = 0.983x) considerably deviates from y=x line, and thus, the three-state model is not a good

predictor of the observed behavior. However, the second cluster fits acceptably to this model.

In case of Fig. 2B–C (i.e. two-state models), although the regression lines are close to y=x, cluster II is not of acceptable R^2, while fittings are satisfactory in case of cluster I. Overall, two distinct patterns (i.e. two- vs. three-state) fit to the unfolding of α-lactalbumin by GdnHCl.
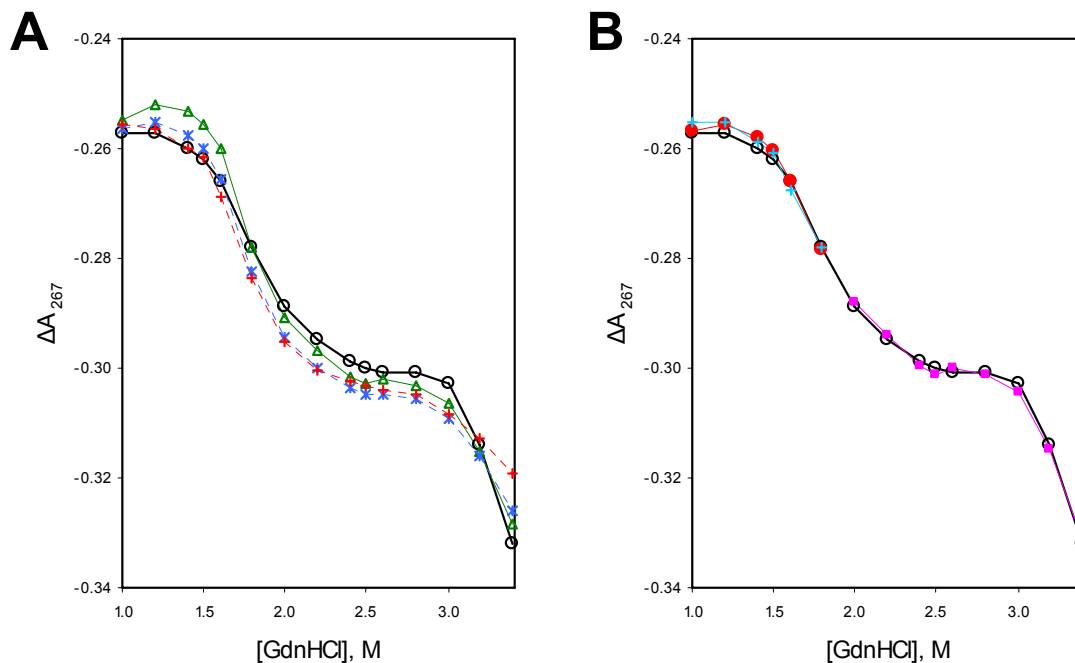
Figure 3:
(A) Absorbance change at 270 nm during α-lactalbumin unfolding by GdnHCl. (○): observed values; (Δ): predicted values based on three-state model (Eqn. 19); (*): predicted values based on two-state model (Eqn. 20); (+): predicted values based on two-state model (Eqn. 21).
(B) Absorbance change at 270 nm during α-lactalbumin by GdnHCl, compared to the appropriate models when $[GdnHCl] \leq 1.6\,M$ or $[GdnHCl] \geq 1.6\,M$. (○): observed values; (■): predicted values based on three-state model (Eqn. 19). (●): predicted values based on two-state model (Eqn. 20); (+): predicted values based on two-state model (Eqn. 21).

Figure 3A demonstrates together the real and the predicted curves of α-lactalbumin titration with GdnHCl. When $[GdnHCl] > 1.8\,M$, three-state model unambiguously is a better estimation for the real data, while two-state models perform better when $[GdnHCl] \leq 1.8\,M$ (i.e. cluster I). Figure 3-B illustrates that the best possible fitting happens when the data are fitted by a two-state model in the range of $[GdnHCl] \leq 1.8\,M$ and a three-state model in the range of $[GdnHCl] > 1.8\,M$. This observation might be explained with the consideration of the

unfolding process in more details. When the protein unfolds to the intermediate with the addition of a denaturant, at the beginning of the process a considerable fraction of native proteins might become transformed to intermediate state, while little unfolded protein is formed yet. More specifically, at $[GdnHCl] = 1.8\,M$, about 35–40% of the native protein is changed, while less than 10% of the protein is in the unfolded form (see Figure 2 in (Kuwajima, et al., 1976)). In other words, the beginning of the three-state unfolding might behave like a two-state process.

		Regression line	R^2
A	Cluster I	y = 0.983x	0.9163
	Cluster II	y = 1.005x	0.9615
B	Cluster I	y = 0.998x	0.9410
	Cluster II	y = 1.011x	0.8203
C	Cluster I	y = 1.004x	0.9347
	Cluster II	y = 1.006x	0.2719

Table 1. Information related to linear regressions displayed in Fig. 2A–C. See the text for more details.

In conclusion, we presented a novel method by which one can infer the number of states in the course of protein unfolding. This method eliminates the need for rather costly facilities, since it is based on the data obtained by spectrophotometry. In addition, using this method one can infer the condition of states, even when the behavior of denaturation process is best explained by multiple patterns (e.g. combination of two- and then three-state denaturation).

REFERENCES

Chedad, A. and VanDael, H. (2004). Kinetics of folding and unfolding of goat α-lactalbumin. *Proteins* **57,** 345-356.

Johnson, R. A. and Wichern, D. W. (1998). *Applied Multivariate Statistical Methods.*, Fourth Edition. Prentice-Hall.

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

Martins, N. F. and Santoro, M. M. (1999). Partially folded intermediates during trypsinogen denaturation. *Braz. J. Med. Biol. Res.* **32,** 673-682.

Pace, C. N. (1990). Measuring and increasing protein stability. *Trends Biochem. Sci.* **8,** 93-98.

Privalov, P. L. (1989). Thermodynamic problems of protein structure. *Annu. Rev. Biophys. Biophys. Chem.* **18,** 47-69.

Ropson, I. J. and Frieden, C. (1992). Dynamic NMR spectral analysis and protein folding: Identification of a highly populated folding intermediate of rat intestinal fatty acid-binding protein by ^{19}F NMR. *Proc. Nati. Acad. Sci. USA* **89,** 7222-7226.

Samuel, D., Kumar, T. K. S., Srimathi, T., Hsieh, H. C. and Yu, C. (2000). Identification and characterization of an equilibrium intermediate in the unfolding pathway of an all β-barrel protein. *J. Biol. Chem.* **275,** 34968-34975.

Yon, J. M. (2001). Protein folding: a perspective for biology, medicine and biotechnology. *Braz. J. Med. Biol. Res.* **34,** 419-435.