FOR THE RECORD

Baseline length and automated fitting of denaturation data

DEVON L. ALLEN AND GARY J. PIELAK

Department of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-3290

(RECEIVED December 10, 1997; ACCEPTED February 11, 1998)

Abstract: To understand relationships between protein sequence and stability, we often compare data from proteins that differ by the substitution of one amino acid. Frequently, an amino acid change causes the cooperative denaturation transitions to shift to lower temperatures, diminishing the signal from the native state. Here we show that apparent stability changes, i.e., the free energy of denaturation, ΔG_D , can also be caused by a deficiency of points in the low temperature end of the transition. In addition, we suggest a method for overcoming this problem.

Keywords: baselines; computer fitting; thermodynamic parameters

There are many techniques, including circular dichroism spectropolarimetry (CD), that can probe protein stability as a function of temperature. Investigators often use computer algorithms to fit these signal-versus-temperature plots to a two-state model (Lumry et al., 1966),

$$N \rightleftharpoons D. \quad (1)$$

The application of this model usually involves defining linear baselines for the native and denatured states:

$$y_{N,T} = m_N T + b_N \quad (2)$$

$$y_{D,T} = m_D T + b_D \quad (3)$$

where m_N and b_N are the slope and y-intercept of the native baseline, m_D and b_D are these values for the denatured baseline, and T is the absolute temperature. By comparing the signal, A_T , at any temperature to the native and denatured baselines, the fraction of protein denatured, $\alpha_{D,T}$, can be determined:

$$\alpha_{D,T} = \left(\frac{A_T - y_{N,T}}{y_{D,T} - y_{N,T}} \right). \quad (4)$$

The equilibrium constant for denaturation, $K_{D,T}$, is obtained via Equation 5,

$$K_{D,T} = \frac{[D]}{[N]} = \frac{\alpha_{D,T}}{1 - \alpha_{D,T}} \quad (5)$$

and $\Delta G_{D,T}$ is obtained via Equation 6,

$$\Delta G_{D,T} = -RT \ln K_{D,T} \quad (6)$$

where R is the gas constant. The slope of a $-R \ln K_{D,T}$ -versus- T^{-1} plot is the van't Hoff enthalpy, ΔH_m . The subscript m means that the slope is evaluated at T_m , the temperature at which $K_{D,T} = 1$. More often, however, Equations 2 through 6 are inserted into Equation 7, and the data are fit directly.

$$\Delta G_{D,T} = \Delta H_m \left(1 - \frac{T}{T_m} \right) - \Delta C_p \left[(T_m - T) + T \ln \left(\frac{T}{T_m} \right) \right]. \quad (7)$$

In Equation 7 (Elwell & Schellman, 1977), ΔC_p is the change in heat capacity upon denaturation.

Another simple technique for fitting denaturation curves was developed by Breslauer (1995) through manipulation of the van't Hoff equation to yield:

$$\Delta H_m = -4RT_m^2 \left(\frac{\delta \alpha}{\delta T} \right)_{T_m}. \quad (8)$$

For this method, the baselines are determined manually. A line through the steepest part of the transition $(\delta \alpha / \delta T)_{T_m}$ is drawn and its slope is used in conjunction with these baselines and T_m to obtain ΔH_m .

We have observed that baselines from computer fitting often disagree with those selected by hand. To probe this discrepancy, we chose a large CD thermal denaturation data set (76 points, Fig. 1), truncated it at various intervals, and determined the effect of data removal on automated baseline selection, ΔH_m , T_m , and $\Delta G_{D,T}$.

Nonlinear least-squares regression (SigmaPlot version 3.06, Jandel Scientific) was used to fit the data. Our results are summarized in Table 1 and Figure 1. The data set was truncated by removing 0, 10, 15, and 20 consecutive points from the low temperature end.

Reprint requests to: Gary J. Pielak, Department of Chemistry, University of North Carolina, Chapel Hill, NC 27599-3290; e-mail: gary_pielak@unc.edu.

Fig. 1. CD signal-versus-temperature plot showing the thermal denaturation data (o) and the results of automated fitting to Equation 7 (solid curves). The baselines were calculated from the full data set (···) and after removing 10 points (---), 15 points (— · —), and 20 points (—) from the low temperature end.

The curves from nonlinear least-squares analysis and the computer-selected baselines are shown in Figure 1. The plots of the fit are unchanged by removing data points and are superimposable on the raw data. In contrast, the computer-selected baselines depend on the number of data points in the pre-transition region. Specifically, as points are eliminated from the low temperature end, the slopes of the native baselines increase while the denatured baselines remain unchanged.

Table 1. Thermodynamic parameters from nonlinear least-squares fitting as a function of the number of data points removed from the pre-transition region^a

Points removed	ΔH_m (kcal/mol)	T_m (K)	ΔG_{300K} (kcal/mol)
0	80.0 ± 3.8	326.0 ± 0.2	4.9 ± 0.3
10	85.5 ± 4.4	326.1 ± 0.2	5.4 ± 0.4
15	91.1 ± 4.7	326.3 ± 0.2	5.8 ± 0.4
20	94.9 ± 5.4	326.5 ± 0.2	6.2 ± 0.4

^aUncertainties from data fitting are shown and are similar to values from reproducibility, ±4 kcal mol⁻¹ for ΔH_m and ±1 K for T_m (Cohen & Pielak, 1994).

We also investigated the effect of data truncation on the values of ΔH_m , T_m , and ΔG_{300K} . As shown in Table 1, ΔH_m increases as points are removed, yet T_m remains the same. We observed the opposite effect (i.e., decreasing ΔH_m , unchanging T_m) when points were truncated from the high temperature end (data not shown). Using Equations 2 through 7, the derived values for ΔH_m and T_m , and a ΔC_p of 1.37 ± 0.06 kcal mol⁻¹ K⁻¹ (Cohen & Pielak, 1994), we calculated ΔG_{300K} as a function of data set length. ΔG_{300K} increases as the number of data points in the pre-transition region decreases (Table 1).

We performed a similar analysis using Breslauer's method. Values of ΔH_m and T_m are within the measured uncertainties of the first entry in Table 1, even when as many as 20 points are removed. Furthermore, different investigators obtained the same values to within the measured uncertainties. These observations suggest that application of Equation 8 is a quick and useful method for confirming values derived from computer fitting, and a means of acquiring thermodynamic parameters from truncated data.

In summary, the number of pre- or post-transition data points influences the slope of the native or denatured baselines from automated fitting. For experiments with poorly defined native or denatured baselines, hand fitting is more useful for the determination of thermodynamic parameters. Alternatively, baselines obtained by inspection can be used as fixed parameters in the automated method.

Materials and methods: The protein used was the C102T variant of yeast iso-1-ferricytochrome *c* (Cutler et al., 1987). Data were acquired at a protein concentration of ≈30 μM in 50 mM sodium acetate buffer, pH 4.6. Data were collected between 5 and 85 °C on an Aviv 62 DS circular dichroism spectropolarimeter at 222 nm.

Acknowledgments: We thank the Pielak group and Dorothy Erie for helpful discussions. This work was supported by NIH grant GM42501. D.L.A. was partially supported by a GAANN Fellowship from the Department of Education.

References

- Breslauer KJ. 1995. Extracting thermodynamic data from equilibrium melting curves for oligonucleotide order-disorder transitions. *Methods Enzymol* 259:221–242.
- Cohen DS, Pielak GJ. 1994. Stability of yeast iso-1-ferricytochrome *c* as a function of pH and temperature. *Protein Sci* 3:1253–1260.
- Cutler RL, Pielak GJ, Mauk AG, Smith M. 1987. Replacement of cysteine-107 of *Saccharomyces cerevisiae* iso-1-cytochrome *c* with threonine: Improved stability of the mutant protein. *Protein Eng* 1:95–99.
- Eiwell ML, Schellman JA. 1977. *Biochim Biophys Acta* 494:367–383.
- Lumry R, Biltonen R, Brandts JF. 1966. Validity of the "two-state" hypothesis for conformational transitions of proteins. *Biopolymers* 4:917–944.