Supporting Information

Site-Specific Unfolding Thermodynamics of a Helix-Turn-Helix Protein

Krista E. Amunson, Loren Ackels and Jan Kubelka.

Department of Chemistry, University of Wyoming, 1000 E. University Ave., Laramie, WY 82071

jkubelka@uwyo.edu

Experimental procedures:

The subdomain of the P22 protein from *Salmonella typhimurium bacteriophage* with sequence NH_2-ITGDVSAANK^{10} DAIRKQMDAA^{20} ASKGDVETYR^{30} KLKAKLKGIR^{40}-COOH, and four ^{13}C isotopically labeled variants were synthesized using standard FMOC solid-phase synthesis on a PS-3 automated peptide synthesizer (Proteins Technologies, Inc.). The synthetic proteins were purified by reverse-phase HPLC and sample purity was verified by MALDI-TOF mass spectrometry.

Two peptide fragments corresponding to the individual α-helices: helix 1 (N-terminal) NH_2-ITGDVSAANK^{10} DAIRKQMDAA^{20} A-COOH; and helix 2 (C-terminal) NH_2-VETYRKLKAK^{10} LKGIR-COOH were synthesized and purified using the same procedures.

The circular dichroism (CD) and infrared (IR) spectra were measured from 0 to 85°C in 5°C steps. The CD spectra were collected on a JASCO 810 spectropolarimeter, equipped with a thermoelectrically temperature controlled sample holder, at a concentration of 52 μM in a 5 mM phosphate buffer, pH 6.9. The concentration was determined by the UV absorption of the Tyr residue[1] ε(280nm) = 1209 $cm^{-1}M^{-1}$.

The IR spectra were measured on a Bruker Tensor 27 FTIR spectrometer, equipped with a RT-DLaTGS detector. All protein samples were pre-exchanged with deuterium and the residual trifluoroacetic acid (TFA) from the HPLC purification was removed by dissolving the samples in D_2O with 0.05M HCl, followed by lyophilization. The lyophilized P22 samples were re-dissolved in a 50mM phosphate/D_2O buffer, pH 6.9 at a concentration ~20mg/mL. A custom-made IR cell with CaF_2 windows and a Teflon spacer with nominal pathlength of 50μm, was used for all IR measurements. The temperature was controlled by an external water bath interfaced with the data collection software and calibrated using a thermocouple placed in the sample cell. At each temperature 256 scans were collected at a resolution of $4cm^{-1}$.

Data analysis:

The data processing and analysis was carried out using Matlab (Mathworks, Inc., Mattick MA). The buffer spectra were collected under identical conditions and subtracted from those of the protein samples. The residual baseline was corrected by the offset of the second derivative spectra,[2] which effectively corresponds to the subtraction of the second order polynomial. The anchor points for the baseline correction and truncation of the amide I' were determined from the zero crossings of the first derivative spectra. Finally, the FTIR spectra of all isotopically labeled variants of the P22 subdomain were normalized to the integral amide I' intensity at 85^{o}C.

The sets of temperature-dependent CD and FTIR spectra were analyzed by singular value decomposition (SVD).[3] Three significant components from both CD (98% of the data variance) and FTIR (99% of the data) were retained as significant, based on both the singular values (S) and visual inspection of the temperature-dependent SVD coefficients (V), as illustrated in Fig. S1. The plot of $\ln(S^2)$ in the SVD of the CD data (Fig. S1a top) changes slope at the third singular value, which suggests that the first three SVD components are significant. The first three V values from the SVD of the CD show a systematic, correlated temperature dependence, while the fourth one appears essentially random (Fig. S1a, bottom). For the FTIR data set, four components can be considered significant, from both the S and V values (unlabeled protein FTIR SVD shown in Fig. S1b). However, the additional component in the FTIR decomposition may be a consequence of the intrinsic temperature dependence of the amide I' IR spectra, or other effects, such as buffer subtraction, baseline correction, etc. A similar fourth component is absent in the SVD of the CD data, while the first three FTIR and CD V-vectors are very similar; strong evidence from two independent experimental probes that these three components are indeed associated with the protein unfolding transitions. We also note that an additional justification for the three-component/three-state description can be drawn from the fact that if the third component is discarded, the first two components, either CD or FTIR, cannot be satisfactorily fitted by a two-state thermodynamic model.

The temperature-dependent coefficients (V values, weighted by the corresponding singular values) were simultaneously fit to a three-state thermodynamic model (see text, Fig. 1). The uncertainties in the data points were estimated from the singular values of the "noise" SVD components. The errors in the fitted parameters were estimated from the second derivative of χ^2 near its minimum.[4] The decomposition of the amide I' bands into ^{12}C and ^{13}C components is illustrated in Fig. S3 and S4 below.

Supplementary figures:

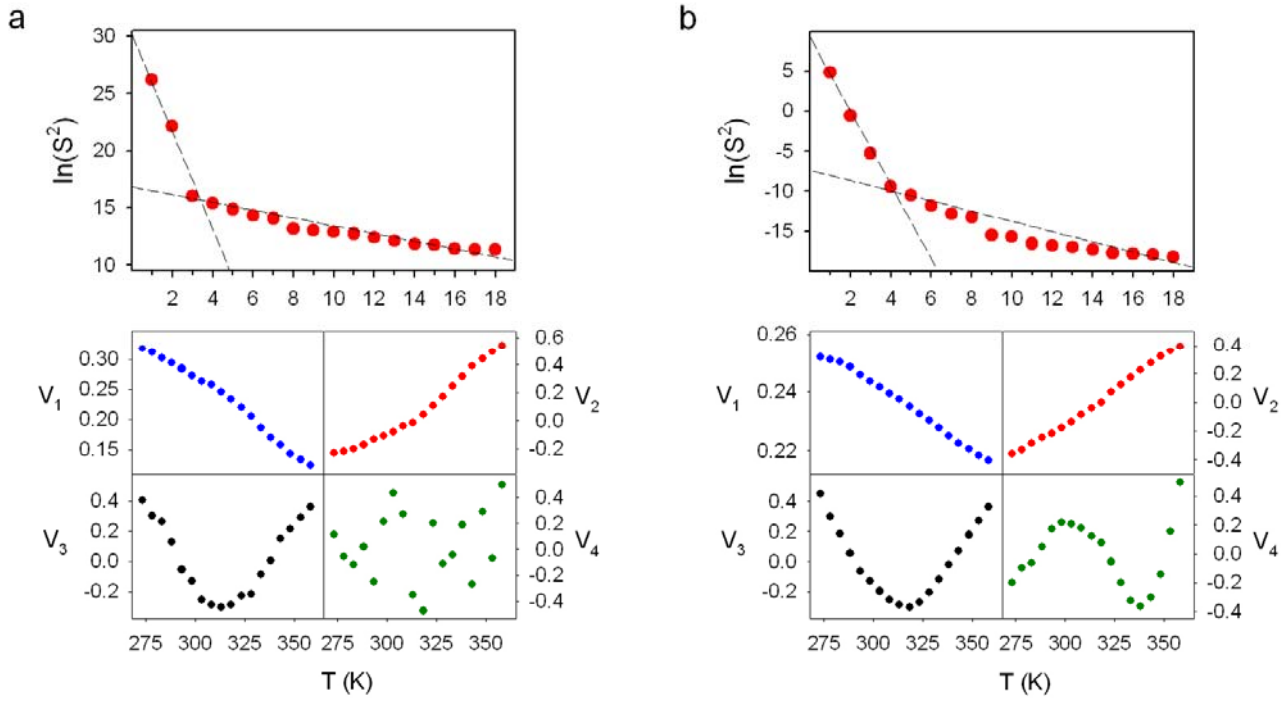

Figure S1. Singular value decomposition of the experimental data (a) CD and (b) FTIR (unlabeled P22 subdomain). The singular values S (top panels) show a characteristic change in the slope of the $\ln(S^2)$ plot, which was used as one of the criteria for the determination of the number of significant components. The examination of the V values as a function of temperature (bottom panels) is another criterion to confirm that at least three SVD components in both the data sets are significant, since they show systematic, correlated temperature dependence. The CD V_4 (a, bottom right) is an example of a "noise" component, since it is essentially random. The three significant SVD components of CD and FTIR account, respectively, for 98% and 99% of the experimental data.

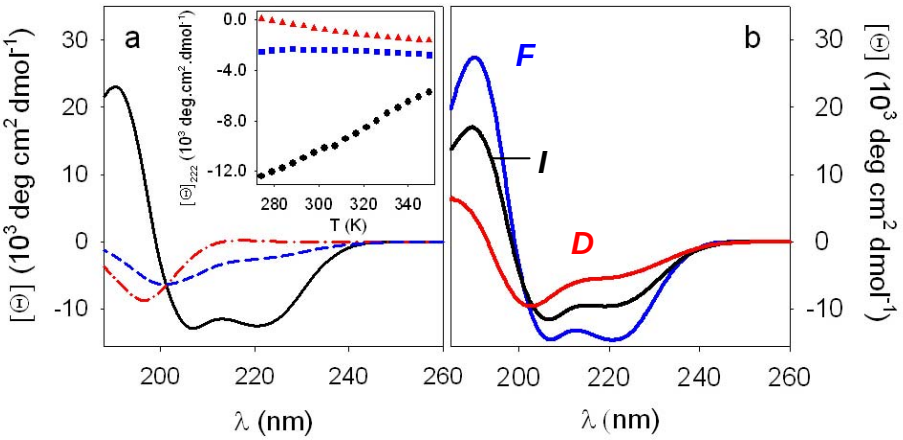

Figure S2. (a) Comparison of the CD spectra of the full P22 subdomain and its constituent α-helical fragments at 0°C: full P22 subdomain (solid black line), helix 1 (dashed blue line) and helix 2 (dash-dot red line). Inset: temperature dependence of the mean residue molar ellipticity at 222 nm. Full P22 subdomain (black circles), helix 1 (blue squares) and helix 2 (red triangles). (b) CD basis spectra of three thermodynamic substates of P22 subdomain: folded (***F*** blue line), intermediate (***I*** black line) and denatured (***D*** red line) showing the partially-helical nature of the intermediate state.

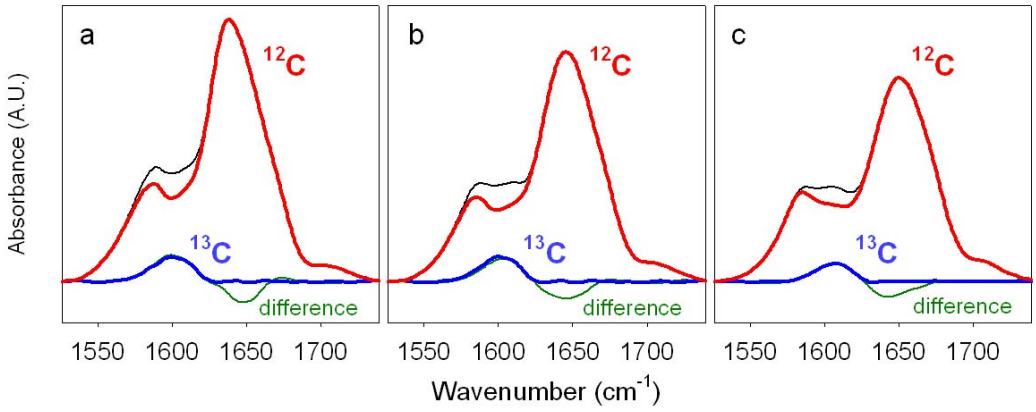

Figure S3. Decomposition of the amide I' FTIR for the isotopically labeled P22 subdomain into the ^{13}C and ^{12}C contributions by a target transformation.[5] The helix 2 ^{13}C labeled P22 subdomain (**h2**) is shown as an example. The positive side of the difference spectrum (green lines) – the best guess for the ^{13}C signal – is the "target" spectrum. Basis functions from the simultaneous SVD of both the helix 2 labeled and unlabeled P22 amide I' spectra are used to best approximate the "target" spectrum. The complete IR spectrum (black) is decomposed into the resulting best fit to the ^{13}C signal (blue lines) and complementary ^{12}C signal (red lines). The decomposition is performed for the three basis spectra describing the P22 unfolding thermodynamics: (a) folded, (b) intermediate and (c) denatured.

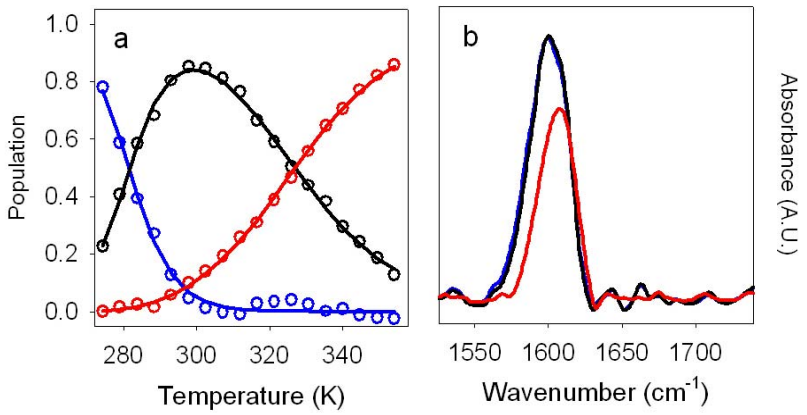

Figure S4. Analysis of the ^{13}C amide I' components. As in Fig. S2, the helix 2 ^{13}C labeled P22 subdomain (**h2**) is shown as an example. The decomposition of the amide I' band (Fig. S2) yields three ^{13}C spectral components, whose corresponding temperature dependent weights were fitted to a three state model (a). The fitting procedure and estimation of the errors in the parameters were identical to those described above (data analysis) for the global fitting. The resulting basis spectra (b) for the first two states (blue and black lines) are virtually identical. As a consequence, the first transition does not correspond to any significant structural change and the local unfolding can be described as a two-state process with the red trace in (a) corresponding to the local unfolding transition.

REFERENCES

(1) Fasman, G. D., *Handbook of chemistry and molecular biology. Section A: Proteins.* 3rd ed.; CRC Press: Boca Raton, FL, 1976.

(2) Coates, J. P., *Appl. Spectrsoc. Rev.* **1999,** 34, 121-138.

(3) Henry, E. R.; Hofricheter, J., *Methods Enzymol.* **1992,** 210, 129-192.

(4) Bevington, P. R.; Robinson, D. K., *Data reduction and error analysis for the physical sciences*. 2nd ed.; WCB: McGraw-Hill USA, 1992.

(5) Pelikán, P.; Čeppan, M.; Liška, M., *Applications of numerical methods in molecular spectroscopy*. CRC Press: Boca Raton, FL, 1994.