

A Hyperstable Miniprotein: Additive Effects of D- and L-Ala Substitutions,

D. Victoria Williams, Bipasha Barua, Niels H. Andersen

Electronic Supplementary Information -

Materials and Methods

Materials, Peptide synthesis and purification

Peptides were synthesized on an Applied Biosystem 433A peptide synthesizer using standard Fmoc solid-phase peptide synthesis methods employing Wang resins preloaded with the C-terminal amino acid. Peptides were cleaved from the resin using a 95:2.5:2.5 trifluoroacetic acid (TFA): triisopropylsilane: water mixture. The cleaved peptides were purified by reverse phase HPLC on a Varian C18 or C8 prep-scale column using gradients of water/acetonitrile (having 0.1% and 0.085% TFA respectively). Collected fractions were lyophilized and their identity and molecular weight confirmed using a Bruker Esquire Ion Trap mass spectrometer.

NMR Spectroscopic Methods

All NMR experiments were performed on a Bruker DRX 500MHz spectrometer on peptide samples of 1-1.5 mM concentration in 50 mM phosphate buffer at pH 7, with 10% D₂O. DSS was used as the internal proton reference standard and set to 0 ppm for all conditions. A combination of TOCSY and NOESY 2D NMR spectra recorded at temperatures ranging from 280–335 K were used to assign all resonances. An MLEV-17 spinlock¹ (60 ms) was employed for the TOCSY and a mixing time of 150 ms was used for the NOESY.

CSD calculations and melts

The chemical shift deviations (CSD = $\delta_{\text{obs}} - \delta_{\text{rc}}$) for all the peptides were calculated using the reference random coil chemical shifts (δ_{rc}) for the specific residue with nearest neighbor corrections. With the exception of Gly¹¹, all the H α random coil values were obtained from our automated CSD calculation program², which is available at <http://andersenlab.chem.washington.edu/CSDb>. The reference random coil values used for Gly¹¹-H α 2, Pro¹⁸-H α , Pro¹⁸-H β 3 and Pro-H δ 3, δ 2 were 4.02, 4.69, 2.29, 3.74 and 3.59 ppm respectively; the listed proline β 3, δ 3, and δ 2 values are used for prolines at other positions as well.

NMR melting temperatures reported herein rely on sums of CSDs. For each construct, two T_m-measures, “cage” and “helix”, were calculated. T_m, corresponds to $\chi_{\text{F}} = 0.50$. The helix melt measure employs H α sites within the helix -- Y3, Q5, W6, A8. The “cage” T_m's are reported in Table 1; these are derived from the sum of the chemical shift deviations of six proton sites (L7 α /G11H α 2/P18 α , β 3/P19 δ 2, δ 3)³ experiencing large upfield ring-current shifts associated with the indole ring of Trp⁶. In other cases, the three protons (G11H α 2, P18 α , P18 β 3) with the largest shifts are employed. Stability changes due to mutations are reported as ΔT_{m} values in Table 1 with TC10b as the reference point. The 100% folded values at L7 α and the ring-current shifted sites in R16, P18 and P19 for the G11→D-Ala mutant were assumed to be identical to those observed for TC10b.

NH exchange studies and protection factor (PF) calculations

The NH exchange experiments were performed on a Bruker DRX 500 MHz spectrometer at 280 K by adding pre-cooled D₂O phosphate buffer (pD = 6.9) to the lyophilized peptide sample in a pre-cooled NMR tube and then recording 1D proton spectra at various time intervals. The slow exchanging NH's were used for calculating the exchange rates and protection factors ($PF = k_{rc} / k_{obs}$). Exchange rates (k_{obs}) were obtained from 1D spectra as the slopes of plots of \ln (NH signal intensity) *versus* time. The reference random coil rate constant for exchange (k_{rc}) was calculated using the appropriate Molday factors⁴. The relationship, $\chi_F = 1 - (1/PF)^5$, was employed to convert protection factors to an extent of folding measure and individual ΔG_U^{280} values for each site.

CD spectroscopy and Melting Analysis

Stock peptide solutions (circa 400 μ M) for CD experiments were prepared in 20mM phosphate buffer at the pH 7.0. The concentration of the stock solution was determined by the UV absorption of tyrosine and tryptophan ($\epsilon = 1280 \text{ M}^{-1} \text{ cm}^{-1}$ and $5580 \text{ M}^{-1} \text{ cm}^{-1}$, respectively at the maximum, $279 \pm 1 \text{ nm}$). CD samples with concentrations of 20-30 μ M were prepared by dilution of the stock solution. Spectra were recorded on a Jasco J715 instrument in cells with a path-length of 0.1 cm. Typical spectral accumulation parameters were a scan rate of 100 nm/min with a 2 nm band-width, and a 0.1 nm step resolution over the wavelength range 185-270 nm with 12 scans averaged for each spectrum at temperatures ranging from 5 to 86 °C. The spectrophotometer was equilibrated at each temperature for 5 minutes before acquisition. The raw ellipticity data was converted into mean residue molar ellipticity units, ($\text{deg} \cdot \text{cm}^2$)/(residue.dmol) using standard Jasco software. Figure 2 in the communication text illustrates a complete CD melt, that of TC16b.

CD melts are presented as plots of mean residue molar ellipticity at 221.8 – 222.8 nm ($[\theta]_{222}$) *versus* temperature. As in previous studies³, the 100% folded baseline was obtained by assuming a temperature gradient ($\delta[\theta]_{222}/\delta T$) that was 0.3-0.32% of the 100%-folded ellipticity. The latter was obtained from the lowest temperature $[\theta]_{222}$ observation adjusting it to 100% folded based on the NMR measures (CSDs) of χ_F . The temperature dependence of the unfolded state CD of Trp-cage from previous studies³ is $[\theta]_{223, U} = -900 - 29 \cdot T$. We assumed that this applies to all constructs with the same L-AA / Gly / aryl-group composition. For the D-Ala mutants, the unfolded baseline was adjusted to -810 and -720 (for one and two D-Ala substitutions, respectively) assuming that each D-alanine completely cancels an L-alanine.

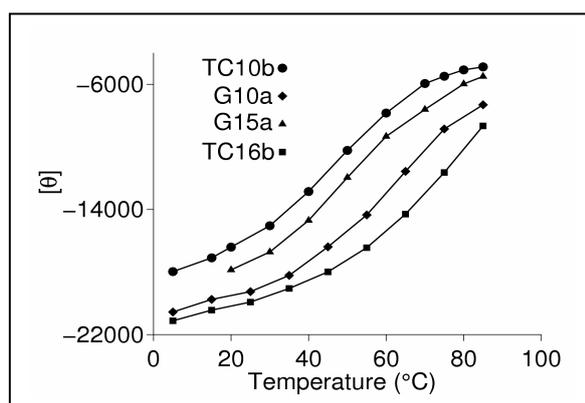


Figure 3S. Raw data, $[\theta]_{222}$ versus T, CD melt comparisons for, in order of increasing T_m , TC10b, (G15a)-TC10b, (G10a)-TC10b, and TC16b. The fraction folded versus T plots in Fig. 3 of the paper text employed the following 100% folded baselines: TC10b, $-18920 + 60 \cdot T$; (G15a)-TC10b, $-19820 + 64 \cdot T$; (G10a)-TC10b, $-21460 + 68.6 \cdot T$; TC16b, $-21930 + 66 \cdot T$.

Converting ΔT_m values to $\Delta\Delta G_U$ estimates.

Prior studies³ of Trp-cage species have revealed that changes in T_m 's (whether determined by NMR or CD melts) in Trp-cage species display a linear correlation with ΔG_U^{280} as determined by amide NH exchange protection studies. That correlation is expressed by Eqn 1,

$$\text{Eqn 1} \quad \Delta\Delta G_U^{280} = 0.26 \cdot (\Delta T_m) \text{ kJ/mol},$$

which was used to obtain $\Delta\Delta G_U$ estimates of single site D-Ala substitution effects from the melting data in Table 1. The NH protection factor derived ΔG_U data in Table 1, suggests that the proportionality constant (0.26 ± 0.02 from ref. 3) may need to be adjusted downward to 0.22.

Sigmoidal fits for CD Melts.

The standard eqn. for a sigmoidal fit, as it would apply to a melt is $\chi_U = 1/(1 + \exp(-k(T-T_m)))$ (Eqn 2). However, this does not provide a perfect fit to χ_U values calculated from the thermodynamic equation for a 2-state unfolding transition (Eqn. 3), even when ΔC_{pU} is set to zero.

$$\text{Eqn. 3} \quad -RT \cdot \ln K_U = \Delta H_U^* - T\Delta S_U^* - \Delta C_{pU} \cdot [T^* - T(1 - \ln(T/T^*))]$$

Setting T^* to T_m , efforts to fit χ_U values from Eqn 3 to Eqn 2 revealed that k is proportional to ΔH^{T_m} and that a perfect fit could not be obtained. Significant deviations appear as T moves further from T_m . The case with $T_m = 338 \text{ K}$ and $\Delta H_U^* = 67350 \text{ J/mol}$ is illustrated in Figure 4S. The difference plot, $\chi_U(\text{Eqn 2}) - \chi_U(\text{Eqn 3})$, is shown in panel **b**.

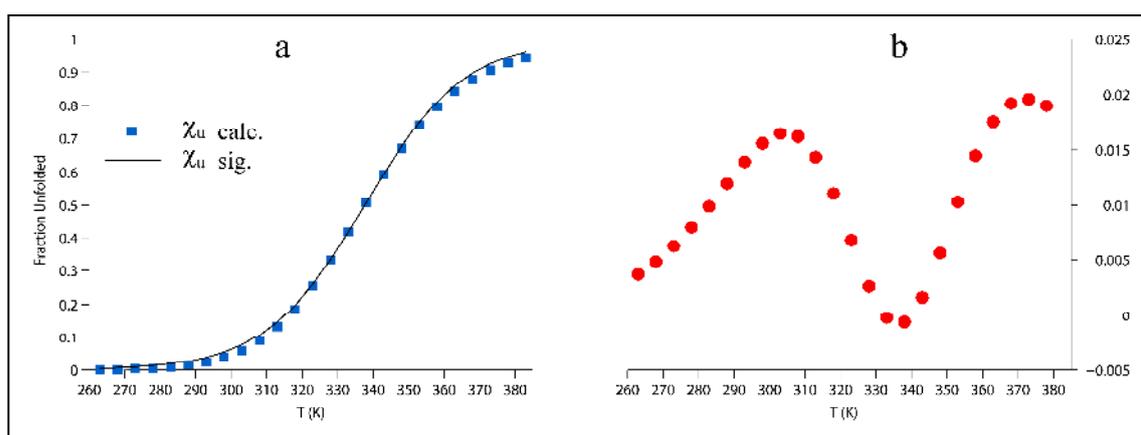


Figure 4S. Idealized fits between strict sigmoidal behavior (Eqn 2) and Eqn 3. In panel **a**, the blue points are calculated using the thermodynamic relationship (Eqn 3). Panel **b** shows the difference, $\chi_U(\text{Eqn 2}) - \chi_U(\text{Eqn 3})$.

The deviation shown in panel **b** of Figure 3S are nearly identical to a plot of the difference between Eqn 3 results for ΔC_{pU} set to zero versus $-300 \text{ JK}^{-1}/\text{mol}$. Since we wished to derive ΔC_{pU} values, we examined modifications of Eqn 2 that would provide a better fit to the calculated χ_U at $\Delta C_{pU} = 0$ based on Eqn 3. The best fits were obtained with Eqn 5.

$$\text{Eqn 5} \quad \chi_U = 0.975 / [1.01 + 1.025 \cdot \exp(-k \cdot (1.3 + T - T_m))]$$

With Eqn 5, the deviations relative to Eqn 3 ($\Delta C_{pU} = 0$) dropped to levels that are below the typical errors in CD melts. The difference plot is shown as panel **a** of Figure 5S. In panel **b**, the difference plots observed for non-zero ΔC_{pU} values are shown; such a difference plot for experimental data can serve to identify the sign and magnitude of ΔC_{pU} .

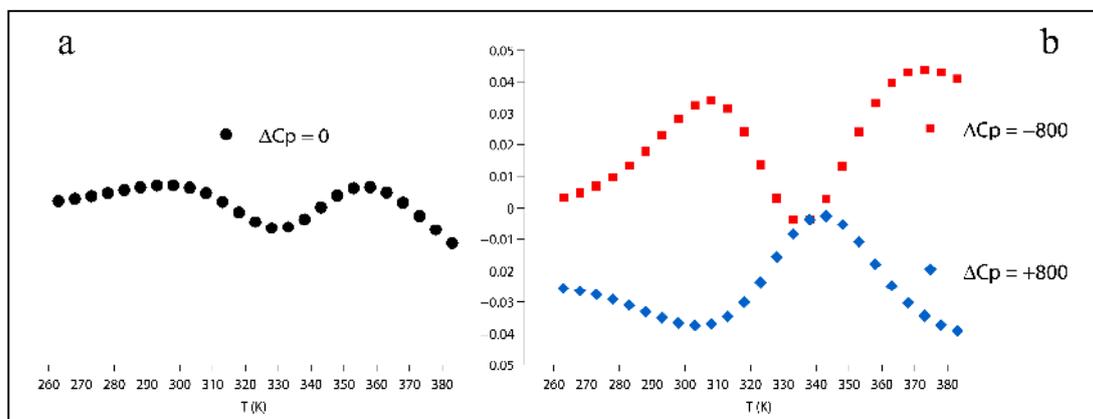


Figure 5S. The effect of ΔC_{pU} on a ‘sigmoidal’ fit (Eqn 5): the difference, $\chi_U(\text{Eqn 5}) - \chi_U(\text{Eqn 3})$, is plotted.

Thermodynamic Parameters from CD Melts.

Determining the sign and upper limit of ΔC_{pU} is the first stage of the analysis. The deviations observed when fitting the CD melts with Eqn 5 implied small negative values of ΔC_{pU} . To confirm this, we applied a method previously applied to two-state hairpin unfolding transitions⁶. Rearrangement of Eqn 3 provides Eqn 6. With experimental χ_F values, a linear plot for $(\ln K_U)$ versus T implies that ΔC_{pU} is indistinguishable from zero. Curvature implies a measurable ΔC_{pU} . We determine the range of ΔC_{pU} values that produced a linear relationship for Eqn. 6 (with T^* set to the apparent T_m value).

$$\text{Eqn 6} \quad \left[RT \ln \left(\frac{\chi_F}{1 - \chi_F} \right) \right] + \Delta C_p \left[T^* - T \left(1 - \ln \frac{T}{T^*} \right) \right] = \Delta H_U^* - T \Delta S_U^*$$

This analysis confirmed that ΔC_{pU} must be negative for all of Trp-cage species examined. The maximum allowed values, the point at which curvature in the wrong direction appears, were: -450 (TC10b), -1000 (G10a-TC10b), -850 (G15a-TC10b), and $-650 \text{ JK}^{-1}/\text{mol}$ (TC16b).

To obtain the best fit thermodynamic parameters we returned to fitting the experimental CD χ_U values to Eqn 3. To facilitate fits in a simple spreadsheet format, Eqn 3 was restated as

$$\ln K_U = 0.12027 \cdot \{ T \Delta S_U^* + \Delta C_{pU} \cdot [T^* - T((1 - \ln(T/T^*)))] - \Delta H_U^* \} / T$$

$$\chi_U = \exp(\ln K_U) / (1 + \exp(\ln K_U))$$

Initial trial values of $\Delta H_U^{T_m}$ were obtained from the k values in Eqn 5 that provided the best ‘sigmoidal’ fits to the CD data, with ΔS_U^* constrained to equal $\Delta H_U^{T_m}/T_m$. Trial ΔC_{pU} values went from the upper limit, defined above, to smaller negative values and $\Delta H_U^{T_m}$ was varied within a 4% range. The best fit thermodynamic parameters for $T^* = T_m$ appear below.

	$T^*(\text{K})$	ΔH_U^* (kJ/mol)	ΔS_U^* ($\text{JK}^{-1}/\text{mol}$)	ΔC_{pU} ($\text{JK}^{-1}/\text{mol}$)
TC10b	329.8	63.1	191.3	-350 ± 100
G15a	335.7	66.4	198	-300 ± 150
G10a	345.2	65.6	190	-740 ± 180
TC16b	356	66.8	187.6	-500 ± 150

Given potential errors in the folded and unfolded CD baselines, all of these should be viewed as estimates until supplemented with calorimetric data.

Supplementary Tables. Chemical shift data for Trp-cage mutants: bold data indicates large upfield shifts associated with cage (or half-cage) formation, G11H α 2 (or its equivalent) is in italics. Other shifts that are unusual are shown in a highlighted **outline** font. The full set of proline residues is included in each case. The proline β and δ resonances are listed in the order β 3, β 2 and δ 3, δ 2; an asterisk indicates that the diastereotopic assignment is not firm.

(G15A)-TC10b 280K pH 7.0						
DAYAQWLKDGPPSSARPPPS						
#	Res	HN	H α	H β	H γ	H δ
7	Leu	8.268	3.562	1.737, 1.400	1.522	0.910, 0.832
10	Gly	7.743	4.043, 3.598			
11	Gly	8.177	3.239, 2.161			
12	Pro		4.526	2.411, 2.062,	2.042,2.015	3.401,3.559 *
15	Ala	8.229	4.325	1.461		
16	Arg	8.162	4.596	1.796, 1.688	1.702, 1.556	3.216
17	Pro		4.677	2.311, 1.803	1.985	3.810, 3.573
18	Pro		4.438	2.298, 2.009	1.927	3.629, 3.562
19	Pro		4.352	2.232, 1.958	1.926, 1.860	3.394, 3.219

For the G15A mutant, the shifts observed at P18 H α and H β 3 are nearly at statistical coil norms, even though large upfield shift at G11H α 2 (nearly 2 ppm) and modest shifts at P19- δ 2, δ 3 implies structuring. G15A shows some half-cage diagnostics.

(D1Ac, G10A)-TC10b 280K pH 7.0						
Ac-AYAQLKDGPPSSGRPPPS						
#	Res	HN	H α	H β	H γ	H δ
7	Leu	8.154	3.817	1.596, 1.250	1.447	0.885, 0.845
10	Ala	8.066	4.198	1.244		
11	Gly	7.817	3.676, 3.434			
12	Pro		4.455	2.348	2.010	3.597, 3.333 *
15	Gly	8.393	3.947, 3.828			
16	Arg	8.191	4.647	1.847, 1.739	1.655, 1.649	3.199
17	Pro		4.636	2.279, 1.84	2.007, 1.836	3.837, 3.595
18	Pro		4.083	1.929, 1.77	1.768	3.677, 3.550
19	Pro		4.404	2.278, 1.94	1.964	3.586, 3.424

The G10A analog shows only small ring current shifts indicating a circa 0.2 fold population.

The G11A analog data, top of the next page, indicates no formation of the full cage structure since the P18 and P19 sites that are usually upfield are near statistical coil norms. Ring current shifts are observed at A11H α (nearly 1 ppm) and H α of R16 is 0.5 ppm upfield from its location in TC10b. The NOEs observed for the G11A mutant confirmed the close association of Ala¹¹ and Arg¹⁶ with the Trp indole ring. All of the sidechain protons of Arg¹⁶ are upfield by 0.2 – 0.3 ppm while the C-terminal PPPS unit displays statistical coil shifts and no NOEs to sites elsewhere in the sequence. This indicates a very different orientation between R16 and W6 in the folded state of (G11A)-TC10b. An R/W cation- π interaction capping a ‘half-cage structure’³ could rationalize these observations.

(G11A)-TC10b 280K pH 7						
DAYAQWLKDGAPSSGRPPPS						
#	Res	HN	H α	H β	H γ	H δ
7	Leu		3.815	1.639, 1.380	1.434	0.834, 0.735
10	Gly	8.008	3.890, 3.688			
11	Ala	8.090	3.429	0.972		
12	Pro		4.435	2.253, 2.008	2.007	3.462
15	Gly	8.376	4.044, 3.901			
16	Arg	8.007	4.130	1.713, 1.560	1.562, 1.356	3.114
17	Pro		4.613	2.261, 1.825	1.915	3.589, 3.326
18	Pro		4.633	2.269, 1.902	2.000	3.747, 3.565
19	Pro		4.401	2.262, 1.97	1.973	3.745, 3.592

Comparisons for species forming a full Trp-cage, appear below

TC10b 280K pH 7.0						
DAYAQWLKDGGPSSGRPPPS						
#	Res	HN	H α	H β	H γ	H δ
7	Leu	8.369	3.387	1.901, 1.363	1.635	0.995, 0.866
10	Gly	7.513	4.172, 3.454			
11	Gly	8.507	3.128, 0.647			
12	Pro		4.649	2.541, 2.080	2.176	3.480, 3.842
15	Gly	7.984	4.296, 3.795			
16	Arg	8.177	5.086	1.918, 1.810	1.810, 1.652	3.312, 3.222
17	Pro		4.771	2.365, 1.785	1.997	3.880, 3.678
18	Pro		2.401	0.213, 1.312,	1.718, 1.657	3.518 (both)
19	Pro		4.343	2.213, 1.997	1.853, 1.788	3.151, 2.93

(G11a)-TC10b forms the full cage but there are significant structural differences. These are reflected CSDs observed for D-Ala11, the shielding of Arg¹⁶ sites, and the greater shielding at P19 δ 3.

(G11a)-TC10b 280K pH 7.0						
DAYAQWLKDGaPSSGRPPPS						
#	Res	HN	H α	H β	H γ	H δ
7	Leu	8.268	3.602	1.776, 1.431	1.560	0.899, 0.835
10	Gly	7.808	3.858, 3.644			
11	D-Ala	8.212	2.109	0.892		
12	Pro		4.308	2.365	2.239	3.316
15	Gly	8.030	4.165, 3.912			
16	Arg	8.058	4.716	1.939, 1.803	1.696	3.266
17	Pro		4.667	2.296, 1.810	2.008	3.883, 3.638
18	Pro		3.145	1.075, 1.492	1.776, 1.724	3.537, 3.484
19	Pro		4.317	2.198, 1.890	1.829	2.934 (both)

Comparisons for species forming Trp-cage folds with nearly identical backbone dihedrals appear on the final page.

TC10b = DAYAQWLKDGGPSSGRPPPS 280K pH 7.0						
#	Res	HN	H α	H β	H γ	H δ
7	Leu	8.369	3.387	1.901, 1.363	1.635	0.995, 0.866
10	Gly	7.513	4.172, 3.454			
11	Gly	8.507	3.128, 0.647			
12	Pro		4.649	2.541, 2.080	2.176	3.480, 3.842
15	Gly	7.984	4.296, 3.795			
16	Arg	8.177	5.086	1.918, 1.810	1.810, 1.652	3.312, 3.222
17	Pro		4.771	2.365, 1.785	1.997	3.880, 3.678
18	Pro		2.401	0.213 , 1.312	1.718, 1.657	3.518 (<i>both</i>)
19	Pro		4.343	2.213, 1.997	1.853, 1.788	3.151, 2.93

(G15a)-TC10b = DAYAQWLKDGGPSSaRPPPS 280K pH 7.0						
#	Res	HN	H α	H β	H γ	H δ
7	Leu	8.364	3.401	1.905, 1.360	1.64	1.000, 0.866
10	Gly	7.512	4.191, 3.447			
11	Gly	8.523	3.206, 0.568			
12	Pro		4.618	2.542, 2.053	2.223	3.536, 3.853
15	D-Ala	7.355	4.421	1.491		
16	Arg	8.174	5.061	1.915, 1.822	1.642	3.324, 3.222
17	Pro		4.761	2.336, 1.780	2.005	3.859, 3.685
18	Pro		2.379	0.209 , 1.297	1.726, 1.661	3.513 (<i>both</i>)
19	Pro		4.338	2.212, 1.987	1.827, 1.782	3.136, 2.920

(G10a)-TC10b = DAYAQWLKDaGPSSGRPPPS 280K pH 7.0						
#	Res	HN	H α	H β	H γ	H δ
7	Leu	8.372	3.344	1.908, 1.336	1.636	0.997, 0.878
10	D-Ala	7.329	4.278	1.219		
11	Gly	8.597	3.136, 0.523			
12	Pro		4.673	2.548, 2.087	2.191	3.530, 3.848
15	Gly	7.985	4.308, 3.793			
16	Arg	8.155	5.112	1.911, 1.805	1.647	3.272, 3.215
17	Pro		4.781	2.348, 1.772	1.994	3.873, 3.681
18	Pro		2.334	0.118 , 1.281	1.704, 1.656	3.507 (<i>both</i>)
19	Pro		4.351	2.155, 2.098	1.827, 1.778	3.150, 2.927

TC16b = DAYAQWLADaGPASaRPPPS 280K pH 7.0						
#	Res	HN	H α	H β	H γ	H δ
7	Leu	8.453	3.356	1.897, 1.351	1.636	1.001, 0.889
10	D-Ala	7.425	4.312	1.256		
11	Gly	8.633	3.203, 0.488			
12	Pro		4.616	2.547, 2.048	2.224	3.607, 3.804
15	D-Ala	7.254	4.401	1.482		
16	Arg	8.167	5.087	1.961, 1.859	1.824, 1.636	3.343, 3.207
17	Pro		4.771	2.350, 1.772	1.999	3.854, 3.685
18	Pro		2.322	0.112 , 1.284	1.688, 1.629	3.50 (<i>both</i>)
19	Pro		4.340	2.213, 2.005	1.825	3.161, 2.926

References

1. Bax, A., Davis, D. G. *J. Magn. Reson.* **65**, 355-360 (1985).
2. Fesinmeyer, R. M., Hudson F. M., Andersen, N. H. *J. Am. Chem. Soc.* **126**, 7238-7243 (2004).
3. Barua, B., Lin, J. C., Williams, D. V., Kummeler, P., Neidigh, J., Andersen, N. H. *PEDS* **21**, 171-185 (2008).
4. Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S.W. *Proteins* **17**, 75–86 (1993).
5. Fezoui, Y., Braswell, E. H. *Biochemistry* **38**, 2796-2804 (1999).
6. Andersen, N. H., Olsen, K. A., Fesinmeyer, R. M., Tan, X., Hudson, F. M., Eidenschink L. A., Farazi, S. R. *J. Am. Chem. Soc.* **128**, 6101-6110 (2006).