Supporting Information for "**Probing the Lower Size Limit for Protein-like Fold Stability: Ten-residue Microproteins Greater than 97% Folded in Water at 280K",**

Brandon L. Kier & Niels H. Andersen**.**

CD Methods

The Trp/Trp exciton couplet is clear, and dominates the CD spectrum of all species studied in the present work, but it is difficult to derive precise T_m values or χ_F measures from the data. Although we have observed that a number of short peptides (including KWK, GHKW, and Ac-AIIW) display a positive CD band at circa 225 nm with *molar* ellipticities of circa +20,000°, [1] more specific controls for prior W-loop-W constructs suggested that the coil value is circa -11,000°/Trp [2,3] and relatively insensitive to temperature. [3] For the present studies we prepared Ac-WIPGKWTG-NH_2 as an unfolded control. The CD molar ellipticity at ~228 nm was -40,000° and hardly varied with temperature. At 213 nm, the exciton couplet minimum, the control value changed from -60,000 to -100,000° as the temperature was increased from 5 to 86 °C. As a result, we add 40,000° to the 228 nm ellipticity value for CD melts and apply the temperature dependence at 213 nm for similar plots for the exciton couplet minimum. The extraction of precise T_m and χ_F values is still not possible since we have no way to evaluate the slope of the 100%-folded baseline. It is apparent, however, that this value at 228 nm is larger than the previous estimate, a 0.36% loss/°C [3], a value as large as 0.52%/°C may apply.

NMR Methods and Results

Fraction folded estimates from chemical shifts

Fraction-folded measures based on backbone Hα and H_N shifts, employ CSD comparisons. An updated version (http://andersenlab.chem.washington.edu/CSDb) [4] of our previously published [5] method for determining random-coil values and CSDs was used. For H_N's, alternative temperature gradients are used depending on the degree of solvent sequestration present in the folded state and the degree of unfolding observed over the temperature range examined.[6] A new set of near-neighbor corrections for Trp and Tyr have been added.[7] The probes employed are given in the text or in Table 5S.

H/D exchange studies

The complete H_N and aromatic region spectra for two points in the exchange time course for a mixture of Pr-WI**p**GLWTGPS and Pr-WIPGIWTGPS (as the unfolded control) at pD = 3.84 appear in Figure 2S. The resulting protection factors from this experiment and a similar study at pD = 4.74 appear in Table 4S. The study of Pr-WI**p**GIWTGPS appears in the "**Additional Mutational Studies of the –WI(N/p)G-X5-W-X7-G- System**" section.

NMR structure ensemble

A more complete segment of the NOESY spectrum that was employed for the generation of a structure ensemble for Pr-WI**p**GIWTGPS appears in Figure 5S. In Figure 5S, unlike Figure 5 in the article text, the water suppression artifacts and small peaks due to the *cis*-GP isomer (~5%) have not been deleted. The most upfield line corresponds to the methyl group of the propanoyl cap. In this segment, the NOE peaks in the 8.0 – 6.7 ppm span are, in order of decreasing chemical shift: T7H_N,

W6Hζ2, W6Hη2 and W6Hζ3. In the upfield portion of the NOESY spectrum even larger NOEs are observed between the Me signals and T7Hγ1, G8Hα3 and G8H_N.

The NOE intensities were obtained directly from the NOESY spectra and these were converted to distances by an automated program, DIS, which also calculates the allowed ranges.[8] For a typical 3.0 Å constraint the limits from this procedure are circa ±.0.5 Å, with extended upper bounds in the case of sharper peaks and unspecified pro-diastereotopic or unresolved CH_2 groups. When the individual protons of CαH_2, CβH_2 or CδH_2 groups display very different NOE intensities to the same site, either 1) only the distance from the larger peak is used in the constraint table, or 2) the longer distance from the less intense NOE is given a relaxed upper bound to account for possible secondary contributions. The NOE constraints employed appear in Table 3S (using the diastereotopic designation of CNS[9] rather than IUPAC), long range connectivities are in bold. Somewhat surprisingly, the application of the constraints in our CNS-based annealing and minimization protocols[3,7] typically produced structures of which only 50 - 55% passed the acceptance script (no NOE violations > 0.15 Å, negative E_{TOT}, and standard bond length and angles). All of the rejected structures had significantly higher E_{NOE} terms, and a significant number were well-converged about another structure (a popular, very high energy, local minimum). This alternative structure could also be dismissed based on experimental observations since it predicts a number of very strong NOEs for sequence remote protons pairs for which there is no detectable NOE in the spectrum. The structure and violation statistics appear in Table 2S with a comparison to the most similar previous peptide.[7]

Table 2S. NMR Structure Ensemble Statistics Comparison

A. Structure Generation Summary

		Pr-WI**p**GIWTGPS	Ac-WINGKWTG-NH_2
Distance Constraints			
	total	148	106
	intra-residue	47	52
	sequential	34	21
	i/i+n (*n* = 2-4)	30	16
	i/i+n (*n* ≥ 5)	37	17
Ensemble Statistics (with standard deviation)			
	accepted/random starts	26 / 50	47 / 50
	NOE Distance viol RMSD	0.016±0.002	0.043±0.010
	E_{NOE} (kcal/mole)	3.00±1.00	18.71±2.00
	E_{LJ} (kcal/mole)	-37.97±1.01	-28.01±1.84
	Bond RMSD Å	0.003±0.000	0.005±0.000
	Angles RMSD (deg)	0.337±0.009	0.476±0.078
	Improper RMSD (deg)	0.261±0.014	0.337±0.080

B. Intra-ensemble RMSD Values (Å)

	Pr-WI**p**GIWTGPS	Ac-WINGKWTG-NH_2
Backbone	0.30±0.13	0.37±0.20
All heavy atoms	0.59±0.20	0.84±0.30
Backbone, without C-terminal residue	0.08±0.06	0.17±0.07
All heavy atoms, without C-terminal residue	0.17±0.11	n. d.
Indole rings	0.02±0.02	0.06±0.03

Although the Pr-WI**p**GIWTGPS NMR structure ensemble displays a very well-converged turn region, a more quantitative analysis of the NOE intensities reveals a minor inconsistency. A stronger than expected **p**3Hα to G4H_N NOE suggests that the turn has some type II' character (type II' turns are expected to have intense $H\alpha_i/H_{Ni+1}$ peaks on par with those of β strands). This indicates that the **p**3-G4 amide plane may be more flexible than the structure shows. Related studies replacing G4 with chiral D and L residues (which force type I' and II' turns, respectively) suggest that both turn types are tolerated and have essentially no effect on non-turn CSDs, though type I' is dominant in the G4 species.

Table 3S. Distance Constraints for Elucidation of Pr-WI**p**GIWTGPS Structure.

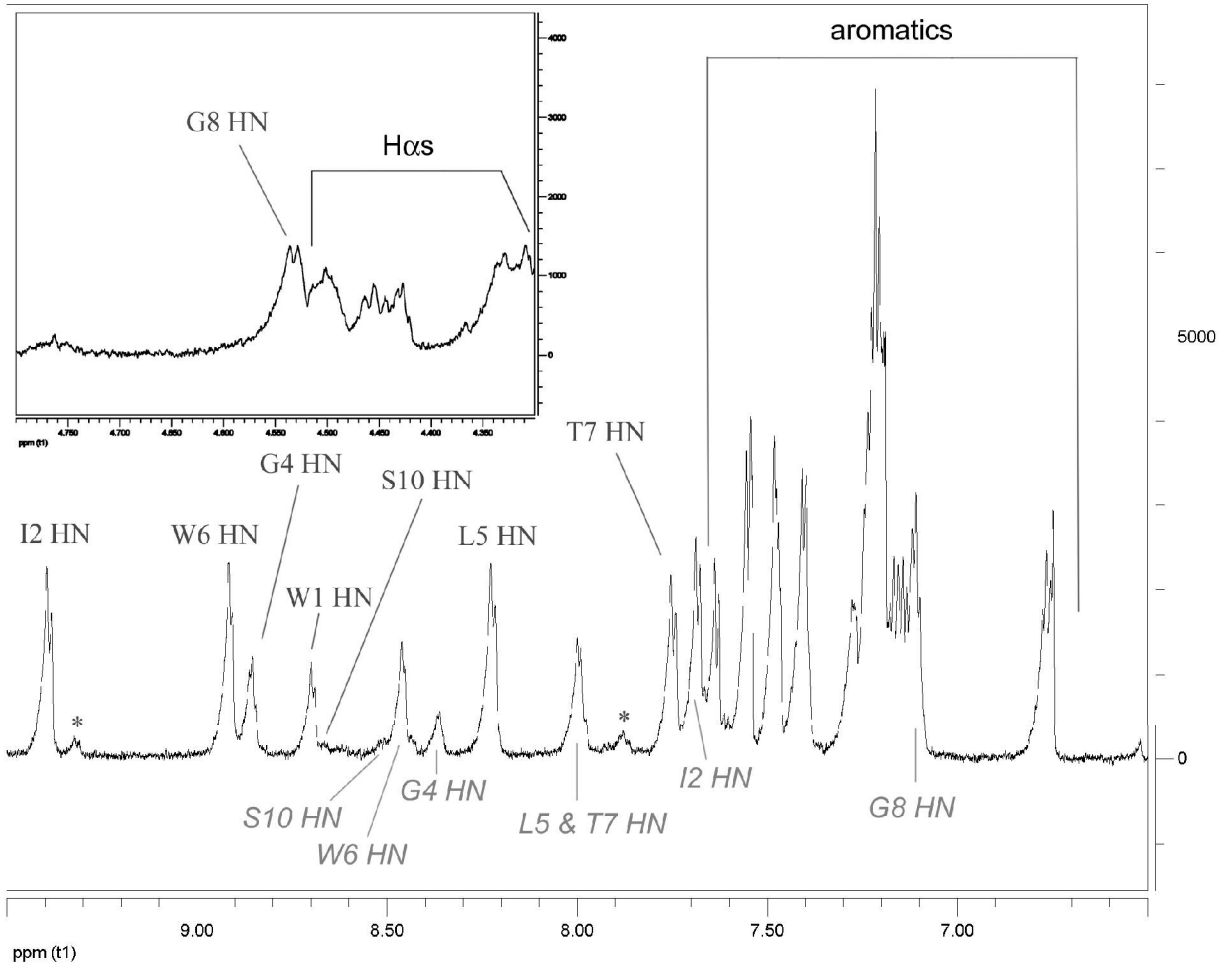
Res1	Atom1	Res2	Atom2	D (Å)	D-	D+
9	ha	10	hn	2.08	0.14	0.18
9	hb1	10	hn	3.46	0.58	0.48
9	hd2	6	hz2	2.87	0.39	0.51
9	hd2	6	he1	3.65	0.64	0.75
10	hb*	10	hn	3.27	0.52	0.42
10	ha	10	hn	3.28	0.52	0.42
10	**hn**	**6**	**he1**	**4.00**	**0.75**	**0.92**
0	ha2	1	hn	2.51	0.28	0.24
0	ha1	1	hn	2.57	0.30	0.25
0	ht#	1	hn	3.54	0.60	0.51
0	**ha1**	**6**	**hz3**	**3.19**	**0.49**	**0.60**
0	**ht#**	**6**	**hh2**	**3.63**	**0.63**	**0.75**
0	**ht#**	**6**	**hz2**	**3.78**	**0.68**	**0.81**
0	**ha2**	**6**	**hz3**	**3.79**	**0.68**	**0.81**
0	**ht#**	**6**	**hz3**	**3.99**	**0.75**	**0.91**
0	**ht#**	**6**	**he1**	**4.57**	**0.93**	**1.29**
0	**ht#**	**7**	**hn**	**3.63**	**0.63**	**0.55**
1	hb2	1	hd1	2.37	0.23	0.42
1	hb2	1	hn	2.44	0.26	0.23
1	hn	1	hd1	2.50	0.28	0.44
1	hb1	1	he3	2.69	0.33	0.47
1	ha	1	hb1	2.71	0.34	0.28
1	ha	1	he3	2.75	0.36	0.49
1	ha	1	hn	2.98	0.43	0.34
1	ha	1	hb2	3.12	0.47	0.38
1	hb2	1	he1	4.41	0.88	1.16
1	hn	1	he1	4.42	0.88	1.17
1	ha	2	hn	2.10	0.15	0.18
1	hb1	2	hn	3.42	0.57	0.47
1	**hz3**	**5**	**hn**	**4.26**	**0.83**	**1.07**
1	**ha**	**6**	**ha**	**2.88**	**0.39**	**0.32**
1	**hn**	**6**	**hz3**	**3.28**	**0.52**	**0.62**
1	**he1**	**6**	**hz3**	**3.36**	**0.55**	**0.85**
1	**ha**	**7**	**hn**	**3.20**	**0.50**	**0.40**
1	**ha**	**7**	**hg2#**	**4.64**	**0.85**	**0.78**
2	hn	1	he3	3.56	0.61	0.72
2	ha	1	hn	4.21	0.82	1.34
2	hn	1	hn	4.35	0.86	0.93
2	hb	2	hn	2.79	0.37	0.30
2	ha	2	hb	2.95	0.42	0.33
2	ha	2	hn	2.99	0.43	0.34
2	ha	2	hg11	3.31	0.53	0.43
2	hg11	2	hn	3.52	0.60	0.50

2	hg2#	2	hn	4.21	0.71	0.60
2	hd1#	2	hn	4.87	0.92	0.90
2	ha	3	hg2	3.91	0.72	0.67
2	ha	4	hn	3.65	0.64	0.55
2	hn	4	hn	4.41	0.88	0.96
2	hn	5	hn	3.17	0.49	0.39
2	ha	5	hn	4.00	0.75	0.71
2	hb	5	hn	4.47	0.90	1.01
2	**hb**	**7**	**hg2#**	**2.79**	**0.26**	**0.75**
2	**hn**	**7**	**hn**	**4.11**	**0.78**	**0.78**
3	ha	2	hg2#	3.83	0.59	0.48
3	hb1	4	hn	3.17	0.49	0.39
3	hg2	4	hn	3.50	0.59	0.50
3	ha	5	hn	3.36	0.55	0.45
4	ha2	1	he3	3.08	0.46	0.56
4	ha1	1	he3	3.46	0.58	0.68
4	ha1	4	hn	2.60	0.47	0.44
4	ha2	4	hn	2.52	0.28	0.24
4	hn	5	hn	2.71	0.34	0.28
4	ha2	5	hn	3.50	0.59	0.50
4	ha1	5	hn	3.57	0.61	0.52
5	**ha**	**1**	**hz3**	**3.01**	**0.44**	**0.55**
5	**hn**	**1**	**he3**	**3.39**	**0.56**	**0.66**
5	**ha**	**1**	**he3**	**3.46**	**0.58**	**0.68**
5	**hg2#**	**1**	**hz3**	**4.84**	**0.91**	**1.12**
5	hb	2	hb	3.23	0.51	0.41
5	hb	2	hn	3.87	0.71	0.65
5	hb	5	hn	2.53	0.28	0.24
5	ha	5	hn	2.75	0.35	0.29
5	ha	5	hg2#	2.78	0.26	0.25
5	hb	5	hg12	2.89	0.60	0.52
5	hb	5	hg11	2.89	0.60	0.52
5	ha	5	hb	3.10	0.47	0.37
5	hg12	5	hn	3.15	0.48	0.38
5	ha	5	hg12	3.20	0.50	0.40
5	hg11	5	hn	3.21	0.50	0.60
5	ha	5	hg11	3.57	0.61	0.52
5	hg2#	5	hn	3.78	0.58	0.46
5	ha	6	hn	2.00	0.12	0.17
5	hg2#	6	hn	3.19	0.39	0.32
5	ha	6	hb1	4.00	0.75	0.72
5	hb	7	hg2#	4.49	0.80	0.71
5	hg2#	7	hn	5.49	1.12	1.31
6	**he3**	**0**	**ha1**	**4.20**	**0.81**	**1.03**
6	**hz3**	**1**	**hd1**	**2.35**	**0.23**	**0.61**
6	**hb2**	**1**	**hz3**	**3.12**	**0.47**	**0.58**
6	**hn**	**1**	**hz3**	**3.14**	**0.48**	**0.58**
6	**he3**	**1**	**hd1**	**3.19**	**0.49**	**0.80**
6	**hb2**	**1**	**hh2**	**3.21**	**0.50**	**0.60**
6	**he3**	**1**	**he1**	**3.26**	**0.52**	**0.82**
6	**hn**	**1**	**he3**	**3.53**	**0.60**	**0.71**
6	**hb2**	**1**	**hz2**	**3.58**	**0.62**	**0.73**
6	**hb2**	**1**	**he3**	**3.70**	**0.66**	**0.78**
6	**he3**	**1**	**hn**	**3.99**	**0.75**	**0.91**
6	**he3**	**1**	**hb1**	**4.25**	**0.83**	**1.06**
6	**he3**	**1**	**hb2**	**4.44**	**0.89**	**1.19**

6	**hb2**	**1**	**he1**	**4.49**	**0.91**	**1.23**
6	**ha**	**2**	**hn**	**3.50**	**0.59**	**0.50**
6	**he3**	**2**	**hn**	**4.51**	**0.91**	**1.24**
6	hb1	6	hd1	2.34	0.22	0.41
6	hb1	6	hn	2.42	0.25	0.23
6	hb2	6	hn	2.61	0.31	0.26
6	ha	6	hn	2.89	0.57	0.49
6	ha	6	hb2	2.82	0.38	0.30
6	he3	6	ha	2.91	0.40	0.52
6	he3	6	hb2	3.09	0.76	0.57
6	ha	6	hd1	3.61	0.63	0.74
6	ha	7	hn	2.25	0.20	0.20
6	he3	7	hn	3.39	0.56	0.66
6	hb2	7	hn	3.80	0.69	0.62
6	ha	7	hg2#	3.96	0.63	0.51
7	**hg1**	**0**	**ht#**	**3.09**	**0.46**	**0.37**
7	**hg2#**	**2**	**hn**	**3.64**	**0.53**	**0.42**
7	**hg1**	**2**	**hg11**	**4.48**	**1.40**	**1.02**
7	**hb**	**2**	**hd1#**	**4.67**	**1.16**	**1.10**
7	**hg1**	**2**	**hd1#**	**4.72**	**1.17**	**1.12**
7	**hg1**	**2**	**hb**	**4.79**	**1.00**	**1.28**
7	ha	5	hg2#	4.47	0.80	0.70
7	hg2#	5	hn	5.17	1.02	1.07
7	hg1	6	ha	3.95	0.73	0.69
7	hn	6	hn	4.38	0.87	0.95
7	hg1	7	hn	2.53	0.28	0.24
7	ha	7	hg2#	2.68	0.23	0.24
7	hg1	7	hg2#	3.19	0.39	0.32
7	hg2#	7	hn	3.27	0.42	0.34
7	hg1	7	hb	3.50	0.79	0.50
7	hb	7	hn	3.92	0.72	0.67
8	ha2	9	hd1	2.62	0.34	0.28
8	ha1	9	hd1	2.90	0.40	0.32
8	**ha1**	**0**	**ht#**	**3.19**	**0.49**	**0.40**
8	**hn**	**0**	**ht#**	**3.34**	**0.54**	**0.44**
8	**ha2**	**0**	**ht#**	**3.40**	**0.56**	**0.46**
8	hn	6	he1	2.98	0.43	0.54
8	hn	6	hd1	3.02	0.44	0.55
8	ha1	6	hz2	3.05	0.45	0.56
8	ha1	6	he1	3.26	0.51	0.62
8	hn	6	hz2	3.88	0.71	0.86
8	ha1	6	hd1	3.98	0.74	0.90
8	hn	7	hn	2.57	0.30	0.25
8	ha1	7	hn	4.27	0.83	0.87
8	hn	8	ha1	2.62	0.31	0.26
3	hd1	2	hg2#	2.81	0.51	0.46
2	ha	3	hd1	2.33	0.22	0.21
2	ha	3	hd2	2.40	0.24	0.22
3	hd*	4	hn	3.13	0.61	0.53

Figure 2S. More complete NMR traces from the single-tube Pr-WI**p**GLWTGPS & Pr-WIPGLWTGPS exchange experiment, pH 3.85: peaks due to the unfolded control (Pr-WIPGLWTGPS) are labeled below the trace in panel A; those of the folded pG turn species are labeled above the trace in both panels - "*" denotes the location of resonances due to the cis-GP (Gly8-Pro9) amide species which is also folded. The W6 H_N peak for *cis*-Pr-WI**p**GLWTGPS is coincident with the G4 H_N peak of the folded *trans* isomer.

Panel A, at t=1.16 h



Note – The 1.56 h point represents partial exchange for the unfolded PG control. At the onset of the experiment the peak labeled L5 & T7 H_N integrated to twice the area of the folded species I2H_N peak. The integral of I2 H_N peaks decreased very slowly, even after 15.7 h the loss in signal intensity was negligible (within error.) The definition of the I2H_N protection factor required data at the higher pD value where the intrinsic exchange rates were greater.

Figure 2S, panel B. "After significant exchange" (t=15.571h.)

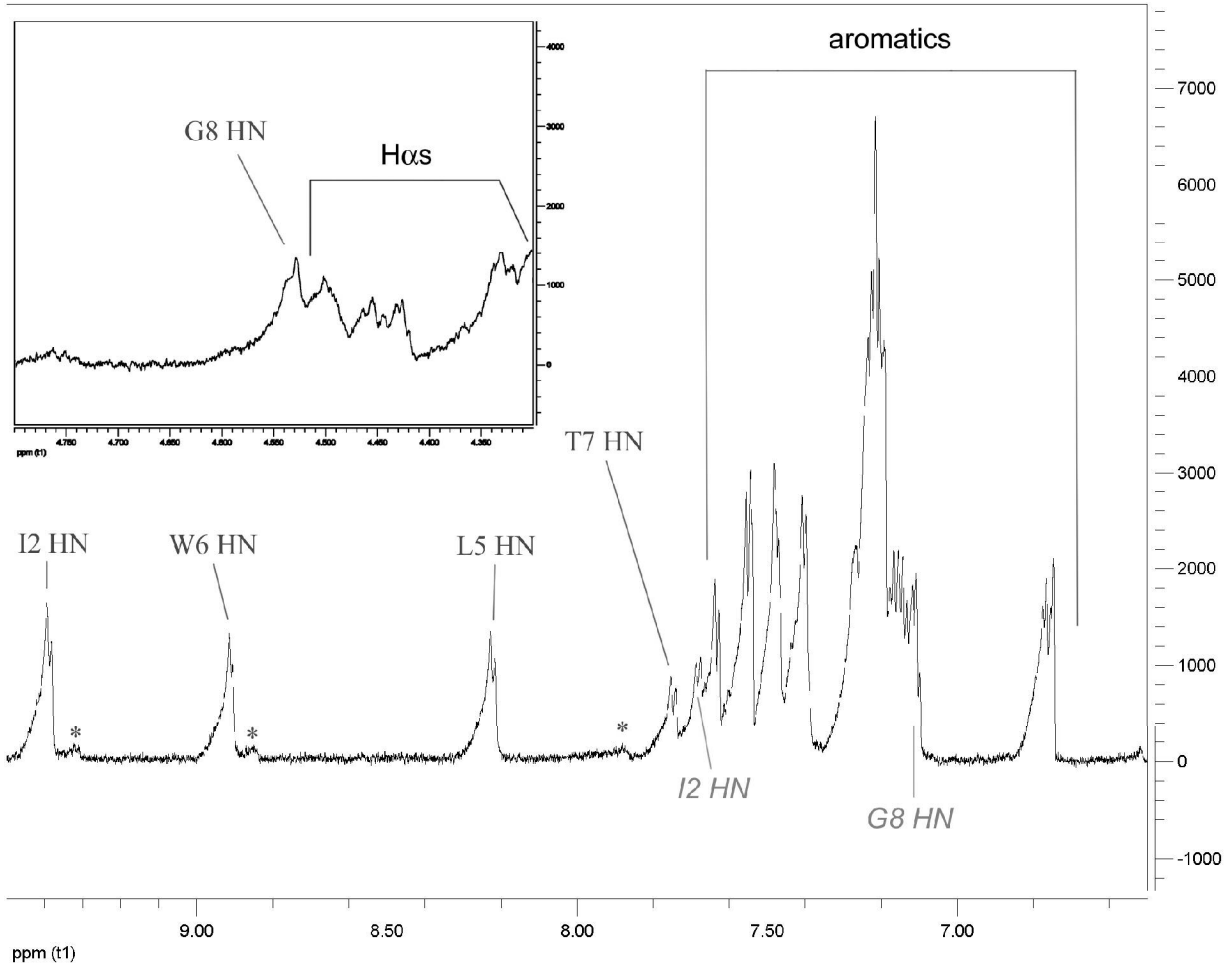


Table 4S. Protection factors (PF's) for the amide protons of Pr-WIpGLWTGPS

Amide site	W1	I2	G4	L5	W6	T7	G8	S10
PF at pD 3.85	~1 (3)	>> 35	2.6 (4)	≥45 (38)	26.2[a] **(21)**	16.7 (22)	~45[b]	0.96 (1.0)
PF at pD 4.65		~45 (52)		45 (38)	~11 (4)	(24)	(713)	

Notes

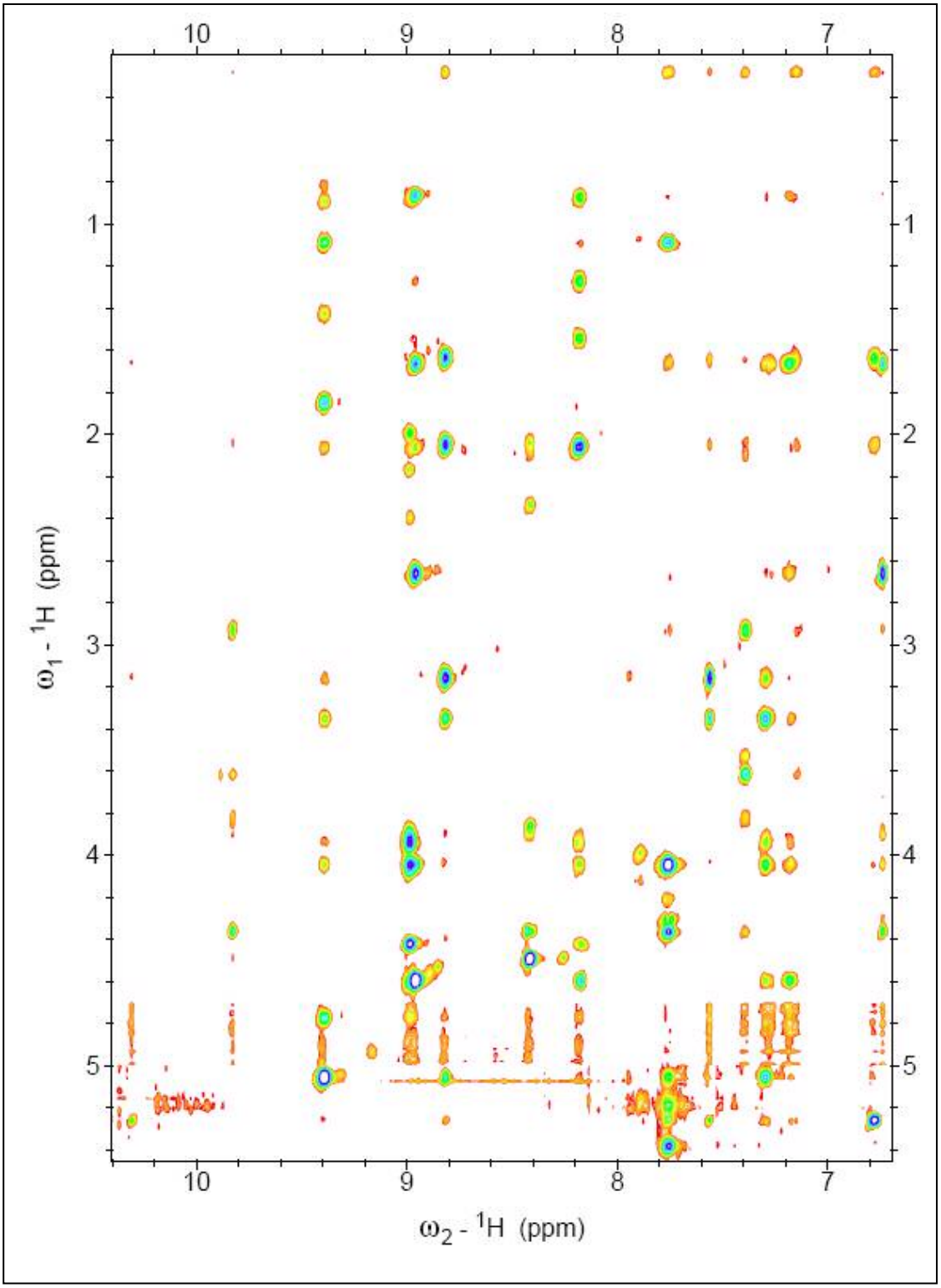
Approximate values exist due to overlap concerns in the 1-D spectrum, the very long half-lives of I2 and G8 NH in the folded peptide, and the very short half life of W1 NH in the unfolded control. The parenthetic values were derived from Molday factors rather than as ratios of the observed rates for the folded system versus the unfolded control.

[a] *Protection factor significantly decreased when pD changed from 3.85 to 4.65; likely offered more protection from acid catalyzed exchange due to H-bonded carbonyl of W6.*

[b] *Compared to Molday factors, some protection is apparent for the G8 amide proton of the otherwise unfolded control, presumably due to the local structuring interaction with the i-2 Trp. Overlap was likewise an issue; ~45x is only a rough estimate.*

Some protection factors remain partly ambiguous, due to peak overlap in the 1-D spectrum. The greatest protection factor that could be determined with precision was that of L5: 45. I2 and G8 were at least as protected, but partial peak overlap (for the unfolded control) made precise quantitation impossible.

Figure 5S. The raw NOESY spectrum (Pr-WI**p**GIWTGPS; 270K, 20% glycol, pH 6.4) from which Figure 5 was generated. Here the connectivities for amide and aromatic H's extends upfield to include sites in the aliphatic sidechains; the topmost horizontal row is signals due to the methyl group of the propanoyl cap.



Addition of 20 vol-% glycol does not alter the conformational preference or greatly increase the folded state population *(added at the request of a reviewer)*

TOCSY melts for **Ac-WIpGKWTGPS** were performed in 20% glycol (270 – 310K) and aqueous buffer lacking glycol (280 – 310K). The Ha CSDs at 270, 280 and 300K that resulted from these experiments are shown in Figure 8S.

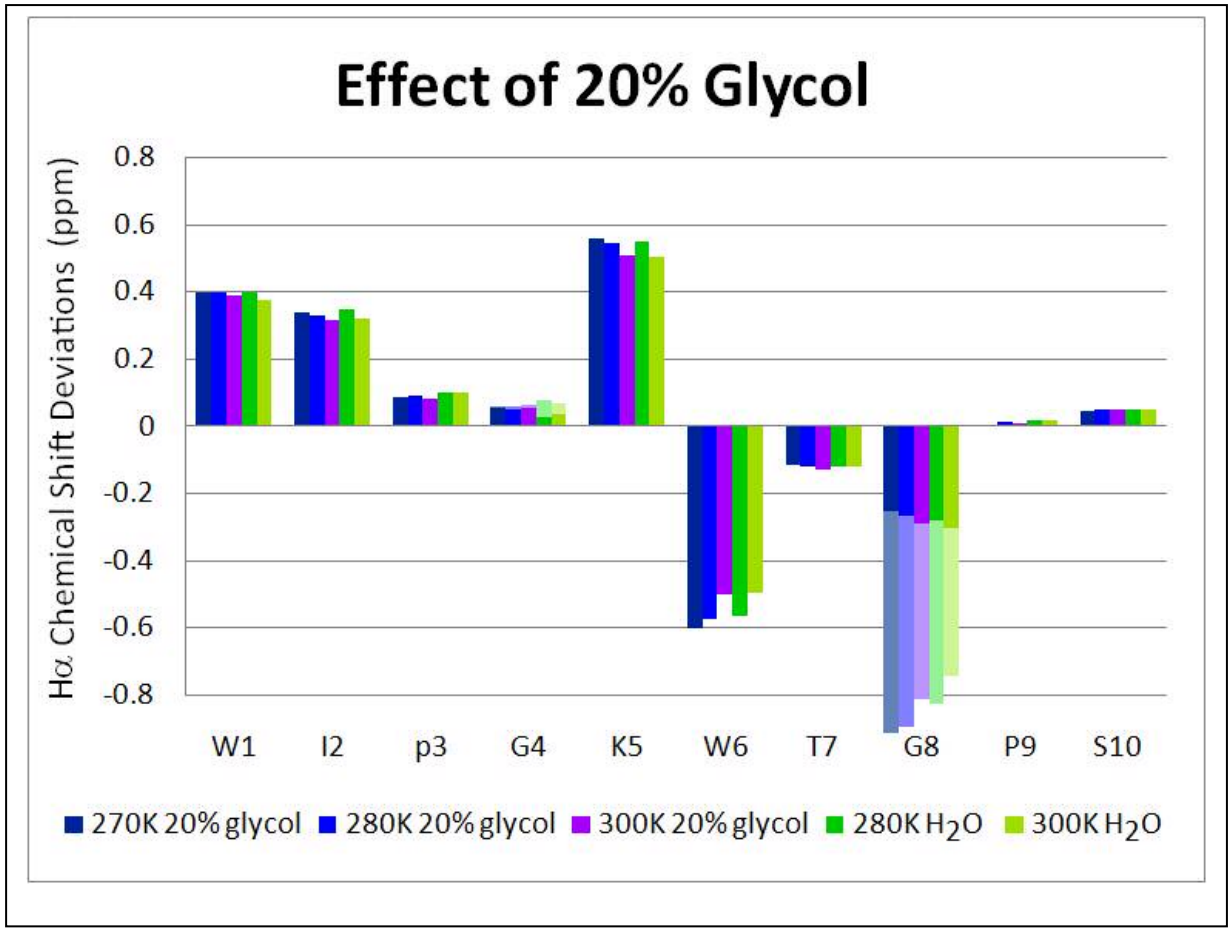


Figure 8S. The addition of glycol does not change the diagnostic shifts for **Ac-WIpGKWTGPS**. For glycines, both Hα protons are shown. At G4, the diastereotopic shift difference was very small.

We conclude that the co-solvent addition, and cooling to 270K as used for the NOESY structure elucidation reported in the paper, does not change the CSDs other than as would result from a slight increase in fold population associated with the lower temperature. The only significant changes in the same temperature comparisons are in the upfield shifts at Gly8: the small shift change at Hα is mirrored by a similar increase for the amide NH (not shown). The strand Hα shifts (and the large downfield shift at $I2H_N$) are not affected by glycol addition. As a result, we conclude that the hairpin structure and fold population are unaltered by the co-solvent. The small effect on the indole ring shielding of Gly8 sites likely reflects some small change in the dynamics of the hairpin capping interaction. There were no significant changes (< 4%) in the large CSDs for W6 Hε3 (2.05 ppm) and Hβ3 (1.4 ppm) at 280K upon glycol addition (data not shown).

Additional Evidence of the Monomeric State of CH_3CH_2CO-WIpGIWTGPS in aqueous solution *(added at the request of a reviewer)*

NMR spectra of were recorded in both D_2O and H_2O for this microprotein construct at 40 μM and 4 mM peptide concentrations to ascertain whether the solution state was monomeric or a mixture of monomer with dimeric and/or oligomeric forms. The spectral comparison appears in Figure 9S.

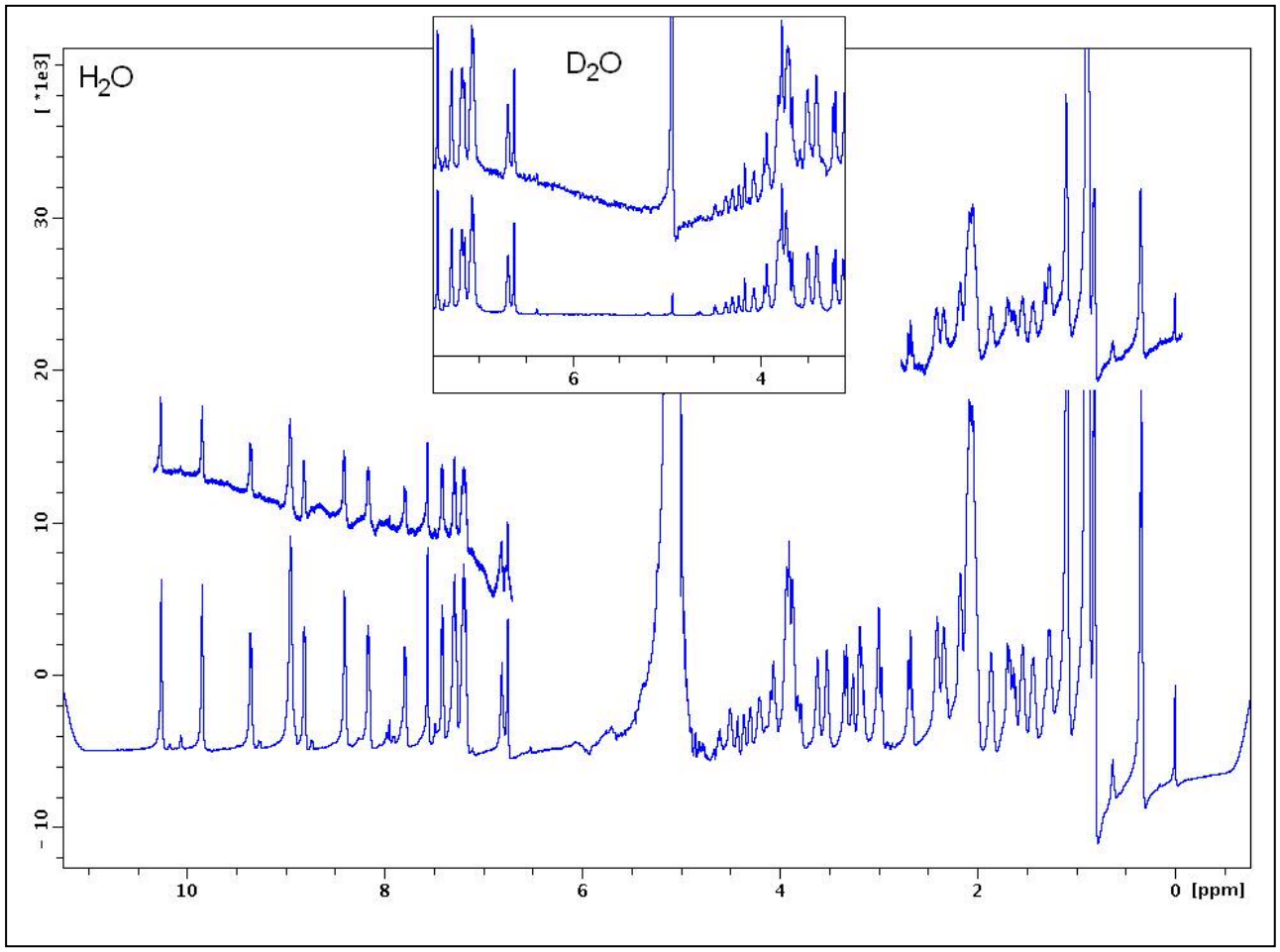


Figure 9S. 600 MHz NMR trace comparisons for **CH_3CH_2CO-WIpGIWTGPS** at 40 μM and 4 mM concentrations. The region near the water signals is shown as observed in nominal 99.8% D_2O medium; the remainder of the spectral regions are from the H_2O experiment. In each panel, the 40 μM NMR appears above the 4 mM spectrum. The W6Hε3 and W1Hα resonances, δ = 5.35 and 5.06 ppm, respectively, were too close to the suppressed water frequency for observation. The G8H_N signal (4.54 ppm) can be observed in the lowest trace, but could not be observed at 40 μM.

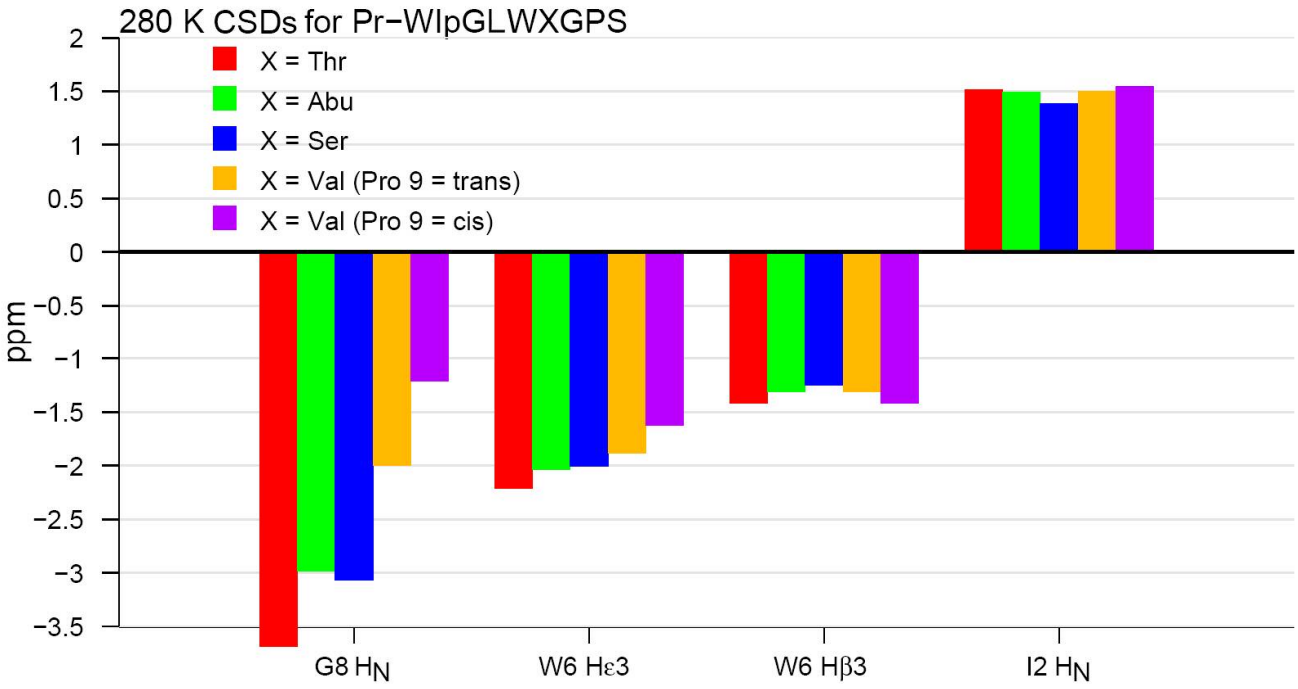
The absence of concentration effects on chemical shifts and line widths over a 100-fold concentration change is viewed as evidence for a strictly monomeric state.

Additional Mutational Studies of the –WI(N/p)G-X5-W-X7-G- System.

The effects of substitutions at Xaa7 (replacement for Thr7)

Though the existence of a sidechain hydroxyl H-bond is demonstrated and likely the source of stabilization, the added stability conferred by threonine at position 7 could also derive from, or be enhanced by, hydrophobic packing with the isoleucine at position 2. To investigate this possibility, T7→S, T7→V, and T7→Abu (α-aminobutyric acid) mutants were prepared.

Figure 10S. Tracking the effects of mutations at T7 by the changes in the four largest CSDs



Though the relative magnitudes of the CSDs observed for the Ser mutant were the same as those for the Thr species (indicative of a near-identical folded state) the serine mutant is less stable. Changes in the relative magnitudes of the CSDs were observed for the other mutants. Based on the $I2H_N$ CSD, which reflects close alignment of the hairpin strands, the Abu and Val mutants have comparable hairpin fold populations to the Thr species. The Thr species has the largest upfield shift at $G8H_N$. The changes in relative CSD magnitude were most pronounced in the Val mutants, in which the Thr hydroxyl is replaced by a Me. The Val mutant was strangely behaved in another way as well; the Gly8-Pro9 amide bond became ~50% cis, (compared to ≤10% for the other constructs) and there were significant differences in the folded structures of both *cis* and *trans* isomers.

The effects of substitutions at Xaa5 (Lys vs Leu vs Ile)

Only the key, single step, mutations are given here. The effect of a K5L mutation was first examined for the Ac-WI**p**GKWTGPS. The CSDs of G8 H_N, G8 Hα3, W6 Hε3, W6 Hβ3 and I2 H_N were employed as the measures of folding (and fold rigidity) with the following results (data for 280K; the differences were more dramatic at 320K with the Lys species approaching its T_m value):

Sequence	G8 H_N	G8 Hα3	W6 Hε3	W6 Hβ3	I2 H_N
Ac-WI**p**GKWTGPS	-3.177	-0.830	-2.010	-1.422	1.368
Ac-WI**p**GLWTGPS	-3.400	-0.886	-2.111	-1.410	1.497

A similar effect, though somewhat muted due to the fold stabilization resulting when the acetyl cap is replaced by propanoyl, was observed for the same Xaa5 mutations. In order to accentuate any subtle changes in folding propensity and allow them to be seen as CSD changes, the L5I mutation was first examined in the Ac-WINGXWTGPS series. CSDs at 5 backbone sites ($I2H_N$, $N3H_N$, $G8H_N$, $G4\Delta\delta\alpha$, and $G8H\alpha3$) at 280K increased by **12.7 ± 3.8%** upon the L→I mutation. The Ile species had a higher melting point and was the only NG turn species to display a Thr sidechain hydroxyl signal. These and other mutational results appear in **Table 5S**.

Table 5S. CSDs for the 7 most shifted proton sites for a complete set of peptide mutants. *The highest temperature at which the sidechain hydroxyl proton could be observed in 2D TOCSY spectra is also recorded.*

T (K) § All 280K, pH 6.4 H_2O

280 if not specified	Proton Sequence	I2 HN	W6 HN	W6 Hα	W6 Hβ3	W6 Hε3	T/S7 OH to	G8 HN	G8 $H\alpha_{up}$
	Ac-WIPGKWTG-NH_2	**-0.19**	**0.217**	**0.065**	**0.003**	**0.05**		**-1.081**	**-0.319**
	Pr-WIPGLWTGPS	**-0.24**	**0.179**	**0.08**	**0.013**	**-0.031**		**-1.132**	**-0.34**
	Ac-WINGKWTG-NH_2	**1.008**	**0.520**	**-0.522**	**-1.216**	**-1.556**		**-2.705**	**-0.470**
	GWIpGKWTG-NH_2	**1.125**	**0.671**	**-0.449**	**-1.321**	**-1.941**		**-0.66**	**-0.278**
	Ac-WIpGKWTG-NH_2	**1.344**	**0.595**	**-0.569**	**-1.438**	**-1.869**		**-3.056**	**-0.545**
	Pr-WIpGKWTG-NH_2	**1.347**	**0.582**	**-0.625**	**-1.501**	**-1.884**	**280K**	**-3.406**	**-0.752**
	Ac-WIpGKWTGPS	**1.368**	**0.611**	**-0.568**	**-1.422**	**-2.01**	**290K**	**-3.177**	**-0.83**
270 §	**Ac-WIpGKWTGPS**	**1.404**	**0.588**	**-.603**	**-1.479**	**-2.167**		**-3.613**	**-0.925**
	Ac-WIpGIWTGW-NH_2	**1.309**	**0.512**	**-0.648**	**-1.472**	**-1.925**		**-2.866**	**-0.693**
	Ac-WINGLWTGPS	**1.028**	**0.475**	**-0.468**	**-1.146**	**-1.645**		**-2.859**	**-0.716**
	Ac-WINGIWTGPS	**1.134**	**0.567**	**-0.622**	**-1.416**	**-1.92**	**280K**	**-3.132**	**-0.813**
320	**Ac-WINGIWTGPS**	**0.566**	**0.355**	**-0.235**	**-0.720**	**-0.891**		**-1.739**	**-0.443**
	Ac-WIpGLWTGPS	**1.497**	**0.586**	**-0.499**	**-1.41**	**-2.111**	**300K**	**-3.4**	**-0.886**
	PWIpGLWTGPS	**1.308**	**0.634**	**-0.439**	**-1.499**	**-2.243**		**-0.995**	**-0.373**
	AcPWIpGLWTGPS*1	**0.87**	**0.329**	**-0.352**	**-0.727**	**-0.898**		**-2.149**	**-0.662**
	AcPWIpGLWTGPS1**	**1.281**	**0.544**	**-0.473**	**-0.891**	**-1.511**		**-2.616**	**-0.813**
	Pr-WYpGYWTGPS	**1.552**	**0.608**	**-0.54**	**-1.325**	**-1.851**	**280K**	**-3.007**	**-0.837**
	Pr-WIpGLWSGPS	**1.384**	**0.717**	**-0.548**	**-1.247**	**-2.004**	**280K**	**-3.066**	**-0.82**
X=Abu	**Pr-WIpGLWXGPS**	**1.495**	**0.534**	**-0.518**	**-1.315**	**-2.039**		**-2.982**	**-0.955**
	Pr-WIpGLWVGPS*2	**1.503**	**0.494**	**-0.487**	**-1.315**	**-1.885**		**-1.996**	**-0.717**
	Pr-WIpGLWVGPS2**	**1.546**	**0.549**	**-0.762**	**-1.416**	**-1.621**		**-1.213**	**-1.376**
	Pr-WIpGLWVGPS †	**1.637**	**0.115**	**0.052**	**-0.877**	**-2.057**		**-3.035**	**-1.186**
	Pr-WIpGLWTGPS-NH_2	**1.516**	**0.622**	**-0.548**	**-1.436**	**-2.169**	**300K**	**-3.611**	**-0.98**
	Pr-WIpGLWTGPS	**1.514**	**0.609**	**-0.547**	**-1.421**	**-2.212**	**300K**	**-3.684**	**-1.017**
320	**Pr-WIpGLWTGPS**	**1.317**	**0.525**	**-0.385**	**-1.156**	**-1.718**		**-2.776**	**-0.832**
320	**Pr-WIpGIWTGPS**	**1.308**	**0.497**	**-0.478**	**-1.288**	**-1.792**		**-2.865**	**-0.866**
	Pr-WIpGIWTGPS	**1.457**	**0.579**	**-0.642**	**-1.557**	**-2.224**	**310K**	**-3.679**	**-1.026**
270 §	**Pr-WIpGIWTGPS**	**1.472**	**0.528**	**-0.667**	**-1.594**	**-2.320**		**-3.938**	**-1.096**
	Ac-WIpnKWTG-NH_2	**1.33**	**0.559**	**-0.614**	**-1.53**	**-1.928**	**280K**	**-3.20**	**-0.555**

§ Shifts at 270K are for 20 vol-% d6-glycol media, † Spectrum recoded in 20 vol-% hexafluoroisopropanol
*1 trans-Ac-Pro1, **1 cis-Ac-Pro1, *2 trans-Gly8-Pro9, **2 cis-Gly8-Pro9

Since some of the protection factors of Pr- WI**p**GLWTGPS (Table 4S) remained somewhat ambiguous due to peak overlap in the 1-D spectrum, we also examined Pr- WI**p**GIWTGPS. The greatest protection factor that could be determined with precision for Pr- WI**p**GLWTGPS was PF(L5) = 45. The L5 to Ile mutant was examined at pD = 6.15 (280K) without the addition of a specific coil control. The higher pH was employed since we would be relying on Molday factors for the extraction of protection data. At this pH, only the base-catalyzed exchange process contributes and a single set of well-determined Molday factors could be used in the analysis. At this higher pD, the W6 Hε1 peak exchange slowed to the point where rates could be followed for 3 half-lives; the rate was roughly 30 times the expected random coil value (W1 Hε1 was never visible.) This suggests that either the G8H_N H-bond or shielding by aliphatic sidechain and backbone moieties* imparts some protection. When compared to expected exchange rates predicted using Molday factors reported by Bai et. al.[10], the amides protons of I2 and I5 were protected by factors of 51 and 91, respectively. The greater protection factor for I5 can be justified by its position nearest the **D**-Pro Gly turn; it may also reflect partial protection in the 'unfolded' state since turns with a **D**-Pro Gly locus can be populated to some extent (15 – 50%) in the absence of hairpin strands. The I2H_N protection factor of 51 corresponds to 98% folded.

Efforts to quantitate the contribution of the FtE Trp/Trp interaction to fold stability

A fraction folded comparison between Ac-TINGKWTG-NH_2 and Ac-WINGKWTG-NH_2 appeared to present the best opportunity to evaluate the energetic contribution of a FtE indole/indole interaction. In this comparison, the complete WTG unit, together with the acetyl that could provide the carbonyl needed for the bifurcated H-bonding of the Thr7 H_N and Hγ is intact. The f_F value of Ac-WINGKWTG is known: 0.77 at 280K. The H_N CSDs observed for the Ile and Asn in INGK turns were selected as folding measures for Ac-TINGKWTG: 0.22 and 0.14 ppm, respectively. Numerous well-folded peptide hairpins with an INGK turn yield 100% folded values for the Ile and Asn on the order of 0.88 and 0.83 ppm, respectively.[6,8] In a previous study[7] of hairpins with WINGKA and AINGKW units at the turn, the Ile-H_N/Asn-H_N CSDs (extrapolated to 100% folding) were: 0.88/0.95 and 1.57/1.1 ppm, respectively. The AINGKW is viewed as the best analogy for Ac-TINGKWTG; on this basis, the W1T mutation reduces the fold population from 0.77 to 0.15 (N3H_N) or 0.25 (I2H_N): $\Delta\Delta G_U$ = 6.9 and 5.3 kJ/mol. Of this Trp to Thr mutational effect, circa 2 kJ/mol[7] could be associated with the greater β propensity of Trp relative to Thr.

---

* Alternative rationales for the protection observed at W1 Hε1 can also be suggested from the NMR structure ensemble: a weak H-bond with the Pro9 carbonyl, or close spatial proximity to the C-term carboxylate which would exert a Coulombic effect opposing the hydroxide attack required for amide H/D exchange at this pD.

Tables 6S a-f; NMR Chemical Shift Assignments

Pr-WIpGIWTGPS

pH 6.4 20 vol-% d6 ethylene glycol, pH = 6.4, 270K

#	Res	HN	Hα	Hβ (Hβ')	Others
0	*Pr-*		2.051,1.638	0.28	
1	Trp	8.816	5.053	3.352,3.156	Hδ1:7.560,Hε1:10.304,Hε3:7.300,Hζ2:7.267, Hζ3:7.183,Hη2:7.196
2	Ile	9.392	4.771	1.845	Hγ12:1.429, Hγ13:1.085,Hγ2:0.896,Hδ1:0.813
3	dPro		4.418	1.989,2.393	Hγ:2.086,2.172,Hδ:3.891,3.939
4	Gly	8.980	4.050,3.934		
5	Ile	8.183	4.594	2.055	Hγ12:1.541, Hγ13:1.267,Hγ2:0.860,Hδ1:0.860
6	Trp	8.958	4.043	2.661,1.666	Hδ1:6.738,Hε1:9.828,Hε3:5.260,Hζ2:7.392, Hζ3:6.780,Hη2:7.150
7	Thr	7.758	4.303	4.209	Hγ1:5.379,Hγ2:1.087
8	Gly	4.362	3.827,2.930		
9	Pro		4.498	2.089,2.337	Hγ:2.025,2.033,Hδ:3.524,3.611
10	Ser	8.414	4.358	3.901,3.862	

Pr-WIpGIWTGPS pH 6.4 aqueous, 280K

#	Res	HN	Hα	Hβ (Hβ')	Others
0	*Pr-*		2.064,1.645	0.354	
1	Trp	8.769	5.060	3.333,3.172	Hδ1:7.560,Hε1:10.244,Hε3:7.312,Hζ2:7.285, Hζ3:7.191,Hη2:7.226
2	Ile	9.334	4.784	1.859	Hγ12:1.427, Hγ13:1.099,Hγ2:0.903,Hδ1:0.818
3	dPro		4.430	2.008,2.407	Hγ:2.086,2.172,Hδ:3.891,3.939
4	Gly	8.910	4.088,3.928		
5	Ile	8.158	4.608	2.048	Hγ12:1.541, Hγ13:1.272,Hγ2:0.876,Hδ1:0.876
6	Trp	8.927	4.061	2.689,1.703	Hδ1:6.750,Hε1:9.838,Hε3:5.356,Hζ2:7.415, Hζ3:6.811,Hη2:7.180
7	Thr	7.790	4.301	4.207	Hγ1:5.362,Hγ2:1.097
8	Gly	4.539	3.801,2.993		
9	Pro		4.498	2.094,2.341	Hγ:2.037,2.049,Hδ:3.528,3.618
10	Ser	8.367	4.362	3.915,3.867	

Pr-WIpGIWTGPS pH 6.4 aqueous, 320K

#	Res	HN	Hα	Hβ (Hβ')	Others
0	*Pr-*		2.056, 1.77	0.482	
1	Trp	8.233	5.034	3.303,3.118	Hδ1:7.450,Hε1:10.034,Hε3:7.403,Hζ2:7.310, Hζ3:7.161,Hη2:7.200
2	Ile	9.013	4.711	1.856	Hγ2:0.889
3	dPro		4.417	1.980,2.357	Hγ:2.060,2.138,Hδ:3.852,3.844
4	Gly	8.434	3.988,3.989		
5	Ile	8.038	4.533	2.023	Hγ:1.520,1.240,0.861,Hδ:na
6	Trp	8.517	4.197	2.776,1.972	Hδ:6.758,Hε:9.771, Hε3:5.788,Hζ:7.394, Hζ3:6.783,Hη:7.150
7	Thr	7.751	4.254	4.162	Hγ2:1.069
8	Gly	5.025	3.699,3.125		
9	Pro		4.458	2.047,2.300	Hγ:2.045,2.045,Hδ:3.495,3.570
10	Ser	7.962	4.330	3.880,3.850	

Pr-WIpGLWTGPS pH 6.4 aqueous, 280K

#	Res	HN	Hα	Hβ (Hβ')	Others
1	Trp	8.716	5.115	3.383,3.156	Hδ1:7.548,Hε1:10.225,Hε3 etc.: overlap
2	Ile	9.391	4.764	1.886	Hγ1:1.454,1.102, Hγ2:0.912,Hδ:0.837
3	dPro		4.491	2.336	Hδ:3.524,3.616
4	Gly	8.867	4.003		
5	Leu	8.220	4.890	1.817,1.661	Hγ:1.490,Hδ:0.973,0.923
6	Trp	8.927	4.126	2.720,1.839	Hδ1:6.762,Hε1:9.833,Hε3:5.368,Hζ2:7.413, Hζ3:6.774,Hη2:7.169
7	Thr	7.727	4.309	4.202	Hγ1:5.360, Hγ2:1.061
8	Gly	4.534	3.795,3.002		
9	Pro		4.457	1.990,2.379	Hδ:3.849,3.940
10	Ser	8.346	4.362	3.916,3.868	

Ac-WIpGKWTGPS

pH 6.4 20 vol-% d6 ethylene glycol, pH = 6.4, 270K

#	Res	HN	Hα	Hβ (Hβ')	Others
1	Trp	8.953	5.057	3.358,3.138	Hδ1:7.528,Hε1:10.290,Hε3:7.328,Hζ2:7.267,Hζ 3:7.138,Hη2:7.188
2	Ile	9.324	4.762	1.846	Hγ1:1.445,1.103,Hγ2:0.905,Hδ:0.830
3	dPro		4.433	1.990,2.387	Hγ:2.076,2.173,Hδ:3.939,3.861
4	Gly	9.003	3.999,3.999		
5	Lys	8.239	4.861	1.941,1.809	Hγ:1.426,1.335,Hδ:1.692,1.692,Hε:2.968,2.968,Hζ:7.684
6	Trp	8.998	4.107	2.681,1.781	Hδ1:6.728,Hε1:9.866,Hε3:5.413,Hζ2:7.397, Hζ3 6.777,Hη2:7.162
7	Thr	7.723	4.248	4.146	Hγ1:5.545, Hγ2:1.046
8	Gly	4.687	3.771,3.101		
9	Pro		4.475	2.030,2.330	Hδ:3.622,3.529
10	Ser	8.380	4.351	3.891,3.858	

Ac-WIpGKWTGPS pH 6.4 aqueous, 280K

#	Res	HN	Hα	Hβ (Hβ')	Others
1	Trp	8.860	5.049	3.341,3.141	Hδ1:7.513,Hε1:10.216,Hε3:7.339,Hζ2:7.287, Hζ3:7.159,Hη2:7.205
2	Ile	9.245	4.764	1.858	Hγ1:1.446,1.117, Hγ2:0.906,Hδ:0.834
3	dPro		4.438	2.001,2.392	Hγ:2.171,2.082,Hδ:3.864,3.935
4	Gly	8.900	4.034,3.965		
5	Lys	8.207	4.844	1.936,1.811	Hγ:1.434,1.342,Hδ:1.700,1.700,Hε:2.995,2.995
6	Trp	8.939	4.135	2.712,1.838	Hδ1:6.739,Hε1:9.879,Hε3:5.570,Hζ2:7.414,Hζ3:6.813,Hη2:7.182
7	Thr	7.749	4.237	4.136	Hγ:5.519, Hγ 1.056
8	Gly	5.041	3.737,3.189		
9	Pro		4.479	2.055,2.330	Hγ:2.091,2.039,Hδ:3.526,3.617
10	Ser	8.320	4.348	3.891,3.866	

(1) Woody, R. W. *European Biophysics J.* **1994**, *23*, 253-262.

(2) The molar ellipticity reported for Ac-W-NH_2 is $[\theta]_{227}$ = +12,900°(as reported in ref #1). Theoretical calculations indicates positive ellipticities for the indole B_b band (circa 225 nm) for many peptide chain and Trp sidechain conformations, but a negative band is predicted for β strand conformations.[1]

(3) Andersen, N. H.; Olsen, K. A.; Fesinmeyer, R. M.; Tan, X.; Hudson, F. M.; Eidenschink, L. A.; Farazi, S. R. *J. Am. Chem. Soc.* **2006**, *128*, 6101-6110.

(4) Fesinmeyer, R. M.; Hudson, F. M.; Andersen, N. H. *J. Am. Chem. Soc.* **2004**, *126*, 7238-7243.

(5) Andersen, N. H.; Neidigh, J. W.; Harris, S. M.; Lee, G. M.; Liu, Z.; Tong, H. *J Am Chem Soc* **1997**, *119*, 8547-8561.

(6) Fesinmeyer, R. M.; Hudson, F. M.; White, G. W. N.; Olsen, K. A.; Euser, A.; Andersen, N. H. *J. Biomol. NMR* **2005**, *33*, 213-231.

(7) Eidenschinck, L. A.; Kier, B.L.; Huggins, K.; Andersen, N. H. *Proteins: Structure, Function & Genomics*, in revision per reviewer critiques.

(8) Fesinmeyer,R.M. **2005** "Chemical Shifts Define the Structure and Folding Thermodynamics of Polypeptides", Ph. D. thesis, University of Washington.

(9) CNS: http://cns-online.org/v1.2/. (a) Brunger, A. T.; Adams, P. D.; Clore, G. M.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J.-S.; Kuszewski, J.; Nilges, N.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L.; *Acta Cryst.* **1998**, *D54*, 905-921. (b) Brunger, A. T. *Nature Protocols* **2007**, *2*, 2728-2733.

(10) Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S.W. *Proteins* **1993**, *17*, 75–86.