Supporting Material for "β-Sheet ^{13}C Structuring Shifts Appear only at the H-bonded Sites of Hairpins" I. Shu, J. M. Stewart, M. Scian, B. L. Kier and N. H. Andersen

Methods and materials

Peptide Models: Selection and Synthesis

Most of the peptides examined in the present study have been prepared previously, the present study includes ^{13}C=O isotopomers of the previous constructs. In all cases, the peptide series have been thoroughly characterized by NMR and CD studies (Fesinmeyer, 2005a; Andersen, 2006; Kier, 2008; Eidenschink, 2009a; Kier, 2010).[1-5] A wide variety of hairpin fold stabilities are included as a result of both strand and turn mutations in the MrH peptide series (Maynard, 1998; Fesinmeyer, 2005a; Eidenschink, 2009a).[1,4,6] Unstructured coil reference peptides, including species with a higher proportion of β-branched residues that might be expected to prefer extended-strand configurations, were constructed with and without ^{13}C' labels at valine and alanine; these appear in Table S1.

All peptide hairpins were synthesized (at a 100 or 250 μmol scale) on an Applied Biosystem 433A peptide synthesizer using standard Fmoc solid-phase peptide synthesis methods. Wang resins preloaded with the C-terminal amino acid were employed.

Table 1. Control Peptides Examined.	
Random coil control	
RCA	Ac-GKA**A**AK-NH$_2$
βRCA	Ac-KI**A**VSAK-NH$_2$
βRCV	Ac-KIT**V**SAK-NH$_2$
RCV	Ac-GKA**V**AAK-NH$_2$
MrH2	Ac-GKKITVSA
RCV2	Ac-KAAVAA

C-terminal amides were prepared similarly but using Rink resins. ^{13}C'-labeled valine and alanine were converted to their Fmoc derivative using Fmoc-OSu (N-fluorenylmethyl succinimidyl carbonate) in acetone-water mixtures containing NaHCO$_3$ (16 h with stirring). N-terminal acetylation was performed by adding the peptide bound resin to the 3mL DMF (N,N-dimethylformamide)/95 μL acetic anhydride/140 μL triethylamine mixture and shaking for 1 hr. Peptides are cleaved from the resin using a 95:2.5:2.5 trifluoroacetic acid (TFA): tri-isopropylsilane: water mixture. The cleaved peptides were purified by reverse phase HPLC on a Varian (or Agilent) C18 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. Sequence and purity were verified by ^{1}H NMR.

NMR Data Collection

All NMR samples are prepared at 0.5 – 2 mM peptide concentration in 50 mM, pH 6.0 potassium phosphate buffer, with 10% D$_2$O, DSS (2,2-dimethyl-2-silapentane-5-sulfonate) and/or ^{13}C urea as internal reference standards. Deuterated hexafluoroisopropanol (HFIP) or β-trifluoroethanol (TFE) was added to the vol-% as noted; the aqueous portion was added by pipette with the volatile fluoroalcohol

delivered by glass microliter syringes. For natural abundant ^{13}C experiments, the samples were prepared using D_2O buffers.

The [13]C' chemical shifts of isotopically labeled valine and alanine residues were obtained from 1D ^{13}C experiment on a Bruker AV500 instrument at 125.72 MHz with a 30 ppm spectral width (32K points), centered at 169 ppm, and 64 – 512 scans depending on sample concentration and signal to noise. ^{13}C-urea served as the internal shift reference as previously described (Fesinmeyer, 2005b).[7] For aqueous medium without cosolvent addition, the ^{13}C-urea shift (in ppm) is given by $\delta = 165.609 - 0.0056 \times T$ (T = °C).

[13]C_α and [13]C_β chemical shifts of all residues were measured by a natural abundance 2D ^{1}H-^{13}C HSQC experiment (Vuister, 1992)[8] on Bruker DRX instrument (^{1}H/[13]C 499.85/125.69 MHz), spectral width 10/85 ppm, center frequency at 3.80/37.00 ppm, with 1024/256 time domain points and 80 scans. [13]C nuclei were assigned by the cross peaks associated with the covalently bonded ^{1}H frequencies based on previously reported data or proton assignments through a combination of 2D [1]H-^{1}H TOCSY and NOESY experiments with WATERGATE (Piotto, 1992)[9] solvent suppression. TOCSY employed a 80ms MLEV-17 spinlock (Bax, 1985)[10] and NOESY a 150ms mixing time for 8 and 16 scans, respectively.

Hairpin Fold Population Determination

Previously published proton random coil values and near-neighbor sequence corrections (Fesinmeyer, 2005a; Eidenschink, 2009a)[1,4] are used throughout to determine [1]H_α and [1]H_N CSDs ($\delta_{obs} - \delta_{random\ coil}$). Diagnostic 100%-folded ^{1}H CSD reference values have been established (Eidenschink, 2009a, Kier, 2008, 2010)[4,5] for the MrH, and βcap scaffolds employed herein. The diagnostic sites employed are cross-strand directed H_α's, H_N's (Fesinmeyer, 2005a)[1], and protons with larger shifts (> 1 ppm) due to ring current effects. Fold population (fraction folded, χ_F) is thus determined as $CSD_{obs}/CSD_{100\%}$ from each of the diagnostic protons and averaged. In the case of the βcap hairpins, the 100% folded CSDs have been verified by backbone amide exchange protection factors; there are representatives of these folds with $\chi_F >$ 0.97 based on the exchange protection factors (Kier, 2008, 2010)[3,5.]

[13]C CSD Calculations

A number of compilations of experimental ^{13}C random coil shift values have appeared and each was examined for applicability to the present study. As proved the case of ^{1}H shifts, we found the values reported by Wishart et al. (1995)[11] using Ac-GGXAGG-NH$_2$ and Ac-GGXPGG-NH$_2$ peptide models provided the near zero ^{13}C CSDs for our control peptides. The differences for [13]$C_{\alpha/\beta}$ shifts were all < 0.2 ppm. In the case of our [13]C' reference peptides (Table 1S), the apparent CSD calculated using the literature coil shifts was −0.32 ± 0.13 ppm. This difference may represent the absence of the urea denaturant in our medium. Solvent induced changes in [13]C' shifts were context dependent and quite large (Table 2S); as a result, we used direct differencing between observed shifts for hairpin models using the most similar controls at the matching solvent conditions and temperature to derive our CSDs. HFIP addition effects on [13]C_α and [13]C_β shifts were, to the extent examined, similar to those observed for [13]C';

these will be detailed in the subsequent full account of this work. Based on this data, all $^{13}C_{\alpha/\beta}$ CSDs were only calculated for aqueous medium lacking co-solvent addition. For the present account, the observed C' shifts of βRCA and βRCV in each medium were used to calculate all of the ^{13}C' CSDs reported.

Table 2S. Solvent effects on ^{13}C' shifts referenced to DSS at 280K

	Δδ (cosolvent), ppm		
	8% HFIP	20% HFIP	30% TFE
^{13}C=O shifts			
KA**A**AK	+0.26	+1.41	−2.34
KI**A**VS	+0.21	+1.58	−2.62
GK**A**V**A**A	+0.26		
KA**V**AA	+0.16	+1.03	−2.64
KT**V**SK	+0.22		
KI**V**TS	+0.26	+1.16	
^{13}C-Urea	+0.45	+1.05	−2.33

Results

The observed 13C' CSDs and fraction folded measures for the isotopically labeled species examined appear in Table 3S.

Table 3S : ^{13}C=O labeled residue chemical shifts in β constructs under aqueous and co-solvent conditions (as specified). Sites that would be cross-strand H-bonded in the hairpin form are underlined. Values in red are CSDs for NHB sites.

Name	Sequence	χ_F (280K)	CSDs (ppm) N-Val	CSDs (ppm) C-Val	Co-solvent	χ_F (280K)	CSDs (ppm) N-Val	CSDs (ppm) C-Val
MrH3b	KKYT**V**SI-pG-KKIT**V**SA	0.604	-1.74	-1.14	8% HFIP	0.817	-1.99	-1.27
MrH5b-2	KKYT**V**SI-pG-KKVT**V**SA	0.554	-1.63	-1.04	8% HFIP	0.831	-2.09	-1.27
MrH5b	KKYT**V**SI-pG-KK**V**TVSA	0.554	-1.60	-0.91	8% HFIP	0.831	-2.09	-1.32
MrH4e	KKLT**V**SI-UG-KKIT**V**SA	0.519	-1.48	-0.50	20% HFIP	≥ 0.96	-2.08	-1.25
MrH4b	KKLT**V**SI-pG-KKIT**V**SA	0.481	-1.34	-0.49	8% HFIP	0.830	-1.92	-1.17
MrH6e	KKLT**V**SI-UG-KKI**V**TSA	0.470	-1.06	*0.43*	20% HFIP	0.525	-1.14	*0.40*
MrH3b-VTS	KKY**V**TSI-pG-KKI**V**TSA	0.318	*-0.17*	*0.32*	8% HFIP	0.464	*-0.15*	*0.28*
MrH4a	KKLT**V**SI-NG-KKIT**V**SA	0.232	-0.52	-0.18	n.a.			.
MrH4a-T13A	KKLT**V**SI-NG-KKI**A**VSA	0.13	-0.33	-0.24	n.a.			
MrH3d	KKYT**V**SI-**P**G-KKIT**V**SA	≈0.10	n.d	n.d	30% TFE	0.52	-0.96	-0.94
Ac-MrH3d [(a)]	Ac-KKYT**V**SI-**P**G-KKIT**V**SA	≤ 0.096	-0.05	-0.05	8% HFIP / 30% TFE	≤ 0.108 / 0.33	-0.09 / -0.73	-0.09 / -0.62
			A	**V**				
βcap6-NG(A)	Ac-WI**A**VTI-NG-KKIRVWTG-NH$_2$	0.71	*-0.47*	-1.55	n.a.			
control [(b)]	Ac--I**A**VTI-NG-KKIRVWTG-NH$_2$	<0.11	+0.22	+0.16	n.a.			
βcap6-HG(A)	Ac-WIT**A**TI-HG-KKIR**V**WTG-NH$_2$	0.90	-2.23	-2.14	n.a.			

[(a)] Ac-MrH3d, with an L-Pro replacing the usual D-Pro in the turn, serves as an additional coil control in water: using the βRCV shift values, the CSDs are insignificant (-0.05 ppm). Consistent with prior conclusions based on ^{1}H CSD changes (Fesinmeyer, 2005a), MrH3d (and its acetylated form) populate the hairpin fold in 30% TFE.

[(b)] The N-terminal Trp deletion destroys the β-cap, yielding a peptide sequence that has no detectable hairpin fold population in water.

The CSD data for MrH peptides appearing in Table 3S was employed to generate the correlation appearing as Figure 2 in the communication text. The CSDs for the H-bonded valine sites in the N- and C-terminal strands were averaged and then plotted versus the fraction folded obtained from the validated ^{1}H CSDs. Extrapolation to $\chi_F = 1.0$ leads to a fully-folded reference value of -2.19 ppm for H-bonded ^{13}C' sites. The average, and standard error of the 100% folded values is -2.09 ± 0.61 ppm. When only the more shifted HB-valine sites in the N-terminal strand are used, an equally linear correlation is obtained; and the 100% folded value is -2.54 ± 0.29 ppm. These site-specific effects are discussed below.

The ^{13}C shift effects are conveniently categorized using the β-hairpin nomenclature presented below. T indicates turn positions (which can be from 2 to 4 in number), and S indicates strand positions numbered from the turn locus. S ± even-numbered positions are non-H-bonded and have their H_α's directed inward; S ± odd-numbered strand positions (with the exception of S − 1) are designated as H-bonded sites. An $(S - 1, T1, T2, S + 1)$-β-turn sequence is often described as the *i, i+1, i+2* and *i+3* positions of a four residue turn in a [2:2]- or [2:4]-hairpin (Sibanda and Thornton, 1991)[12]. This nomenclature maintains the S ± odd / S ± even

Figure for β-hairpin nomenclature

designations for H-bonded versus non-bonded sites remote from the turn. The S+1 sites also have an H-bonded carbonyl and can be included in the HB category. The S-1 site is H-bonded in some tight turns as well as in [3:5]- and [4:6]-hairpins.
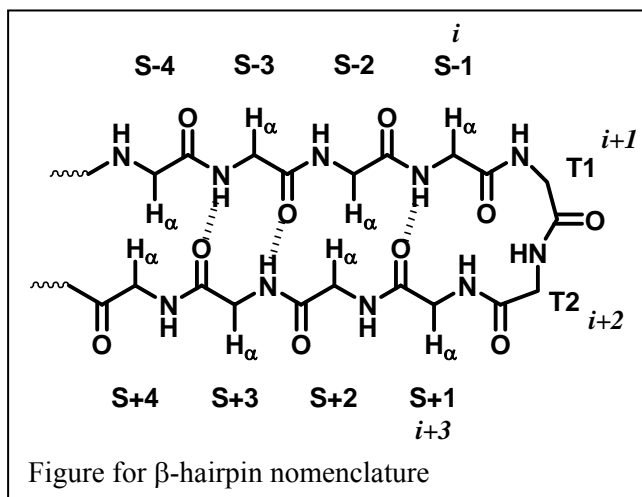
For the MrH peptides, ^{13}C' labels were predominantly at the S − 3 and S + 5 positions, where the two labeled residues are not cross-strand hydrogen bonded with each other. Regardless of the fold population which varied based on mutations and media fluoroalcohol-content, the N-terminal ^{13}C' has a greater CSD magnitude than the C-terminal site. The trend was observed in the MrH5b peptides even shown by the different hydrogen bonded positions in the 2nd strand. MrH5b-2 and MrH5b have an identical sequence, but the C-terminal Val is labeled at S + 5 and S + 3 (hydrogen bonded with the N-terminal Val), respectively. In aqueous solution, the N-terminal Val CSD is -1.6 ppm, and the C-terminal V12, V14 are both smaller in CSD magnitudes, -1.04 and -0.91 ppm. Although the fold population is improved from 55% to 83% when 8% HFIP is added, the C-terminal carbonyls are still less upfield shifted (CSDs were -1.27 and -1.32 ppm) than the N-terminal (-2.09 ppm). We can not generalize this phenomena for all hairpins, since we have examined only a few isotopically labeled hairpins that are not based on MrH model.

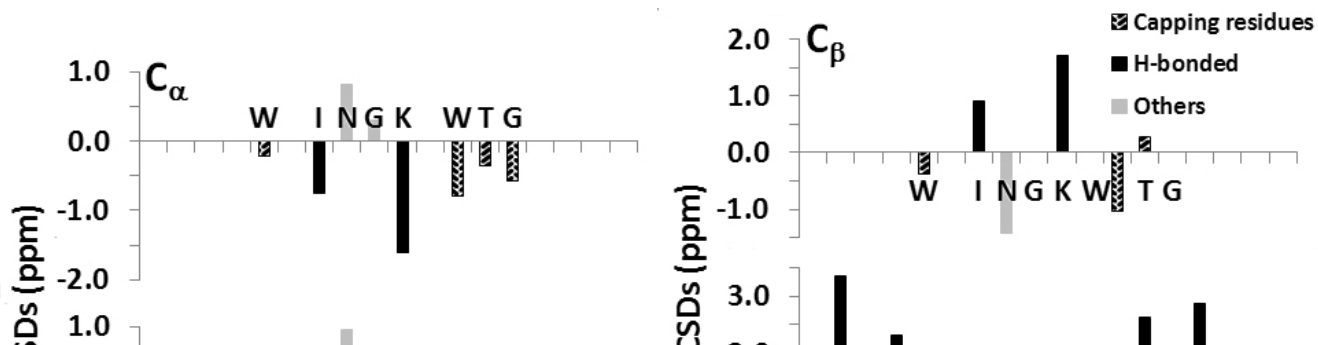*Integrity of the $^{13}C_{\alpha/\beta}$ Pattern in Another β Sheet Model*

The data for the β-capped species and the corresponding capped turn microprotein (Kier, 2008)[3] provide an example (Table 4S and Figure 4S). These hairpins, with the same β-capping motif, have relatively immobilized termini when compared to non-capped isolated β hairpins (Kier, 2010)[5]. As a result, the terminal residues adjacent to the capping motif, which have to be in the register allowing the cross-strand hydrogen bonds to form, do not fray as much as is observed in many hairpins. If the alternating $^{13}C_{\alpha/\beta}$ shifts reflect, in some manner, the H-bonding status of strand sites, the pattern should be even clearer in these peptides. Indeed, upfield C_α and downfield C_β shifts are observed at the H-bonded sites and are more intense near the stabilizing β-cap. The shifts appearing in the capping motif are also retained in the extended hairpin, but are excluded from the present discussion; the TG residues do not reside at φ/ψ torsional angles conforming to β structure.

Table 4S: Testing the generality $^{13}C_{\alpha/\beta}$ CSD patterns. Red underlined sites are expected to display the Cα downfield/Cβ upfield pattern; the black underlined positions are not as clearly of the H-bonded type.

Peptide	Sequence
βcap-INGK	Ac-W--INGK-WTG-NH₂
βcap-IHGK	Ac-WITVT--IHGK-KIRVWTG-NH₂

References –

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

2. 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.

3. Kier, B. L.; Andersen, N. H. *J. Am. Chem. Soc.* **2008,** *130,* 14675-14683.

4. Eidenschink, L. A.; Kier, B. L.; Huggins, K. N. L.; Andersen, N. H. *Proteins Struct. Funct. Bioinf.* **2009,** *75,* 308-322.

5. Kier, B. L.; Shu, I.; Eidenschink, L. A.; Andersen, N. H. *Proc. Natl. Acad. Sci. U.S.A.* **2010,** *107,* 10466-10471.

6. Maynard, A. J.; Sharman, G. J.; Searle, M. S. *J. Am. Chem. Soc.* **1998,** *120*, 1996-2007.

7. Fesinmeyer, R. M.; Peterson, E. S.; Dyer, R. B.; Andersen, N. H. *Protein Sci.* **2005,** *14,* 2324-2332.

8. Vuister, G. W.; Bax, A. *J. Magn. Reson.* **1992,** *98,* 428-435.

9. Piotto, M.; Saudek, V.; Sklenar, V., *J. Biomol. NMR* **1992,** *2,* 661-665.

10. Bax, A.; Davis, D. G., *J. Magn. Reson.* **1985,** *65,* 355-360.

11. Wishart, D. S.; Bigam, C. G.; Holm, A.; Hodges, R. S.; Sykes, B. D. *J. Biomol. NMR* **1995,** *5,* 67-81.

12. Sibanda, B. L.; Thornton, J. M. *Methods Enzymol.* **1991,** *202,* 59-82.