Supporting Information

An all-atom model for stabilization of α-helical structure in peptides by hydrocarbon staples.

Peter S. Kutchukian, Jae Shick Yang, Gregory L. Verdine, and Eugene I. Shakhnovich[*]

Hamiltonian

The all-atom energy function H in our previous study[1] has been further developed to model cross-linked peptides and consists of the following energy terms:

$$H = E_{\text{con}} + w_{\text{hb}} \times E_{\text{hb}} + w_{\text{bbtor}} \times E_{\text{bbtor}} + w_{\text{sct}} \times E_{\text{sct}} + w_{\text{linktor}} \times E_{\text{linktor}} \quad (1)$$

where E_{con} is the pairwise atom-atom contact potential, E_{hb} is the hydrogen bonding potential, E_{bbtor} is the sequence-dependent backbone torsional potential based on the statistics of sequential amino acid triplets, E_{sct} is the side-chain torsional angle potential, and E_{linktor} is the torsional potential for linkers. The first three energy terms are described in detail in our previous publication.[1] It should be noted that secondary structure information from PSIPRED is not used in this study, as we were interested in studying the thermodynamics of peptide folding rather than attempting to solely predict the folded structure.

The side-chain torsional angle energy was obtained from the same database used in our previous study[1] by

$$E_{\text{sct}} = \sum_i E_{\text{A}_{i-1}\text{A}_i\text{A}_{i+1}}, \quad E_{\text{A}_{i-1}\text{A}_i\text{A}_{i+1}} = \frac{-\mu N_j + (1-\mu)\widetilde{N}_j}{\mu N_j + (1-\mu)\widetilde{N}_j}, \quad (2)$$

where N_j and $\widetilde{N}_j$ are the number of observations in the j-th bin of a side-chain torsional angle χ's of residue A_i and total number of observances subtracted by N_j for a triplet consisting of A_{i-1}, A_i, and A_{i+1}, respectively. The bin width was 30° and the value of $\mu = 0.991$ was chosen to make the net interaction zero.

The torsional potential for hydrocarbon linkers E_{linktor} was derived by fitting quantum mechanical calculations with a fourier series. Energy as a function of dihedral angle for butane (C_4H_{10}) was calculated using the Guassian package (HF/3-21 level of theory). The energy was then normalized such that the maximum value was set to 1.0, and the minimum value was set to -1.0. This is so that the energy scale is similar to that

of the side chain torsions and backbone torsions. The following fourier series was obtained:

$$E_{linktor} = \alpha_0 + \sum_{n=1}^{3} \alpha_n \cos(n\theta) \quad (3)$$

where the α's were derived by summing the real part of the fourier coefficients that are complex conjugates and dividing them by the number of points used during fitting (in this case 12).

Monte Carlo Move Set

There are three kinds of degrees of freedom for stapled peptides: backbone torsional angles, side-chain torsional angles, and linker torsional angles. Inside the macrocycle which is composed of the backbone between stapled residues and the linker, only the local move was applied to the rotation of the backbone and/or linker because a global move would break the macrocycle. A local move consists of rotating seven consecutive backbone/linker torsional angles within a window, with atoms outside this window unchanged. Outside the macrocycle, either global or local moves were used for backbone rotations.[1] A global move rotates the torsional angle of a randomly selected residue. The step sizes of the global and local moves are drawn from a Gaussian distribution with zero mean and standard deviation of 2° and 60°, respectively. In a side-chain move, all side-chain torsional angles in a randomly selected nonproline residue are rotated. The step size of the side-chain move follows a Gaussian distribution of zero mean and standard deviation of 10°. Note that a knowledge-based move[1,2] was not used and that the local move set was modified[3] to keep the detailed balance condition.

Generation of Peptide Structures

Peptides were built in Macromodel using the Maestro GUI.[4] After building peptides in helical conformations, the unnatural cross-links were added, as well as α-methyl carbons. The cross-links were built in the reduced (saturated) form, that is, there were no olefins. This is a reasonable approximation since in the case of the RNAse A stapled peptides, the reduced cross-linked peptides possess similar helical propensities when compared to those containing olefins.[5] The cross-links were locally minimized using the OPLS2005 force field, while all other atoms were held frozen. The entire peptide was then relaxed with 100 steps of minimization.

Monte Carlo Simulation Protocol

A random coil conformation previously obtained from high temperature runs well above the melting temperature of each peptide was used as the initial structure in each simulation. Each run employed a different initial random coil structure, as well as a different seed for the random number generator. For each peptide runs of 100×10^6 steps were carried out. Structures and their corresponding energetic values were written to file every 500,000 steps for analysis. In all calculations, the peptides were allowed to

equilibrate for 50×10^6 steps prior to analysis. Thus, for each peptide 100 conformations per run were obtained for analysis.

Percent helicity calculations

In order to assess if our folding code was capable of recapitulating experimental helical propensities we ran both peptide systems at a temperature (T=0.72) that approximately reproduced the helical propensities of the WT peptide. To test for convergence of the helicities, MC runs were repeated until the computed helicity values converged. Error was obtained by dividing the runs into three groups, and calculating the standard deviation of the group averages. For the RNAse A peptides 120 runs were sufficient for convergence (Fig. S1), while 270 runs were necessary for the BID BH3 peptides (Fig. S2 and S3).

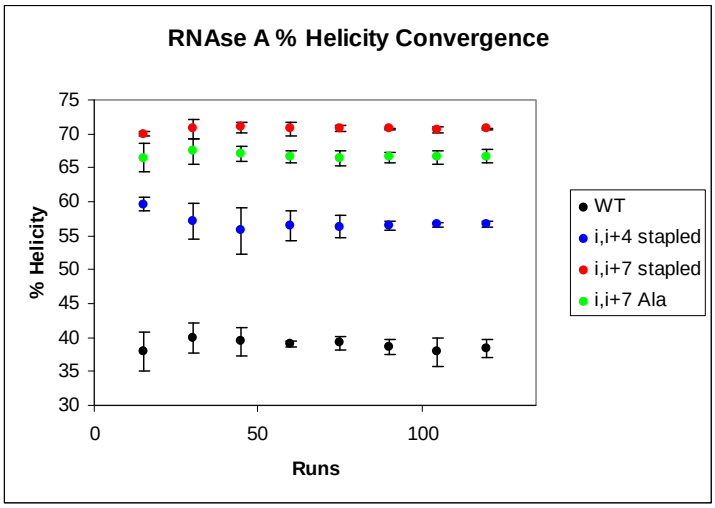

Figure S1. Convergence of percent helicity for WT, stapled, and Ala mutant (where stapling residues are replaced by alanine) RNAse A peptides (T=0.72).

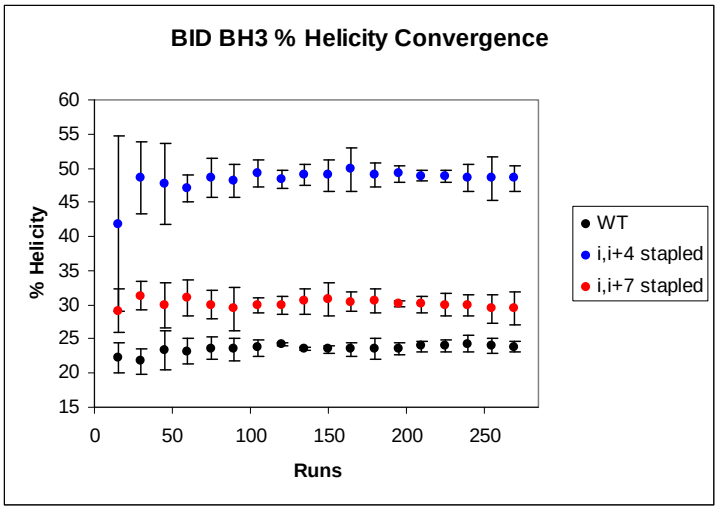

Figure S2. Convergence of percent helicity for WT and stapled peptides BID BH3 peptides (T=0.72).

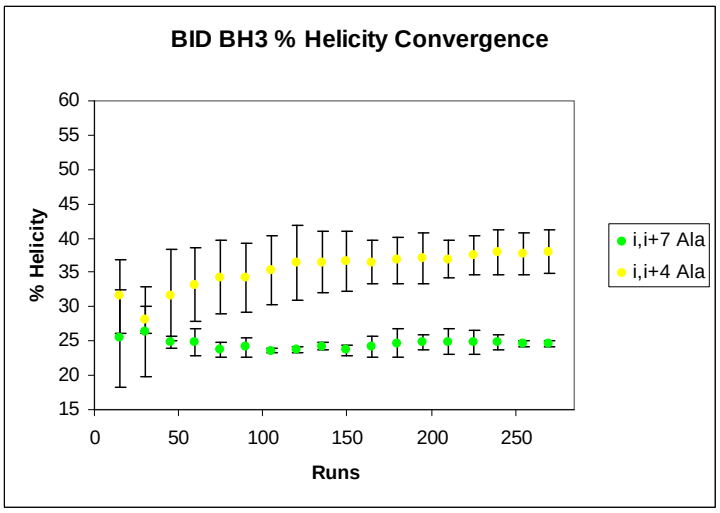

Figure S3. Convergence of percent helicity for Ala mutant (where stapling residues are replaced by alanine) BID BH3 peptides (T=0.72).

Helical content analysis using DSSP

The DSSP program was used to convert pdb files obtained during thermodynamic MC simulations into dssp files.[6] The percent helicity of a peptide conformation was calculated as the number of residues marked as helical "H" by DSSP, divided by the total number of residues. These were averaged over each conformation to obtain the average helicity of a peptide at specific temperatures (Fig. 1A). In order to ascertain the helical propensity of each residue we divided the total times that each residue was marked as

"H" by the total number of structures of that peptide that were analyzed (Fig. 2A and 2B). Alternatively the probability of observing specific helical conformations was obtained by binning helical conformations from 0-100% with 10% increments (Fig. 2C and 2D).

Average Distance of Cross-linked Residues

High temperature (T=1.0) simulations were carried out to determine to what extent the staples constrain peptides relative to the WT peptides. 30 runs were carried out, and the C_α distance of the stapling residues was monitored (Fig. S4).

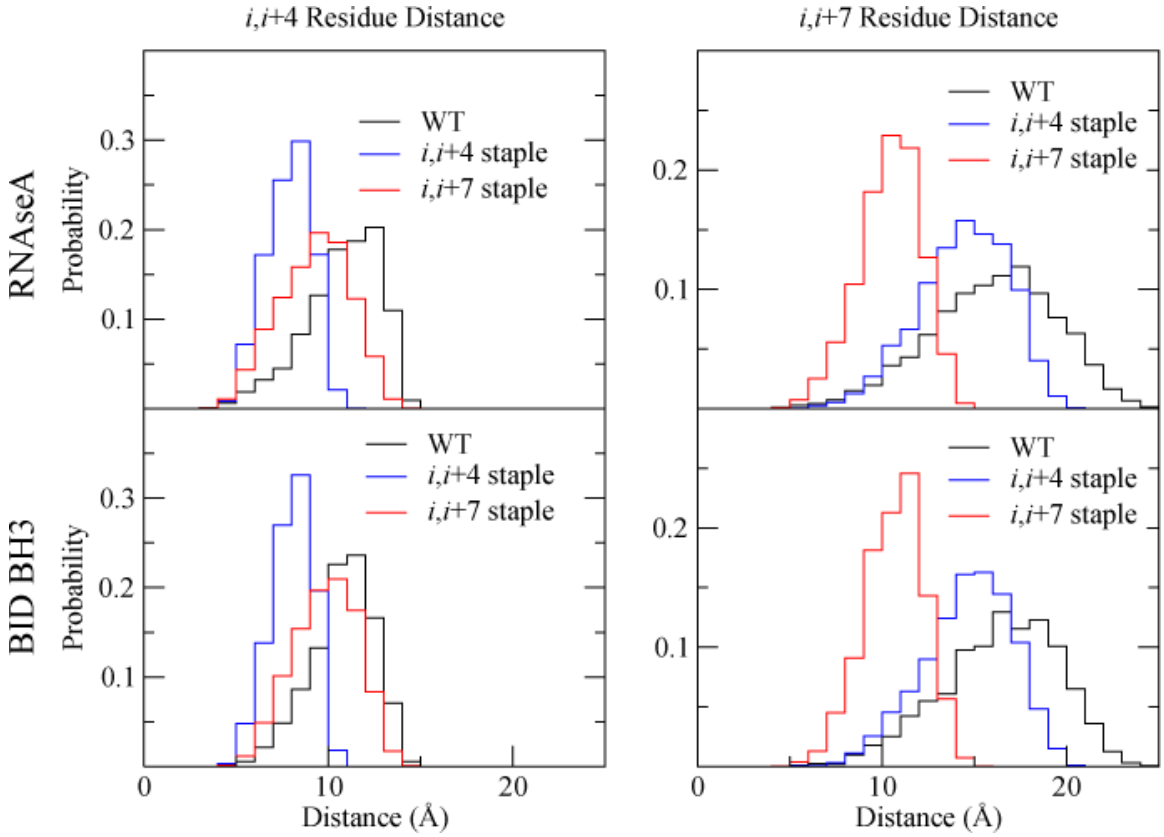

Figure S4. C_α distance distribution of *i*,*i*+4 (left) or *i*,*i*+7 (right) residues during high temperature (T=1.0, T >> T_m) trajectories where peptides are largely in the denatured state.

Melting Temperature of Peptides

It is desirable to carry out thermodynamic simulations in the vicinity of $T \approx T_m$ so that both the folded and unfolded peptide conformations are sampled. As such, we obtained rough melting curves for each peptide by performing 30 MC runs at various temperatures (Fig. S5).

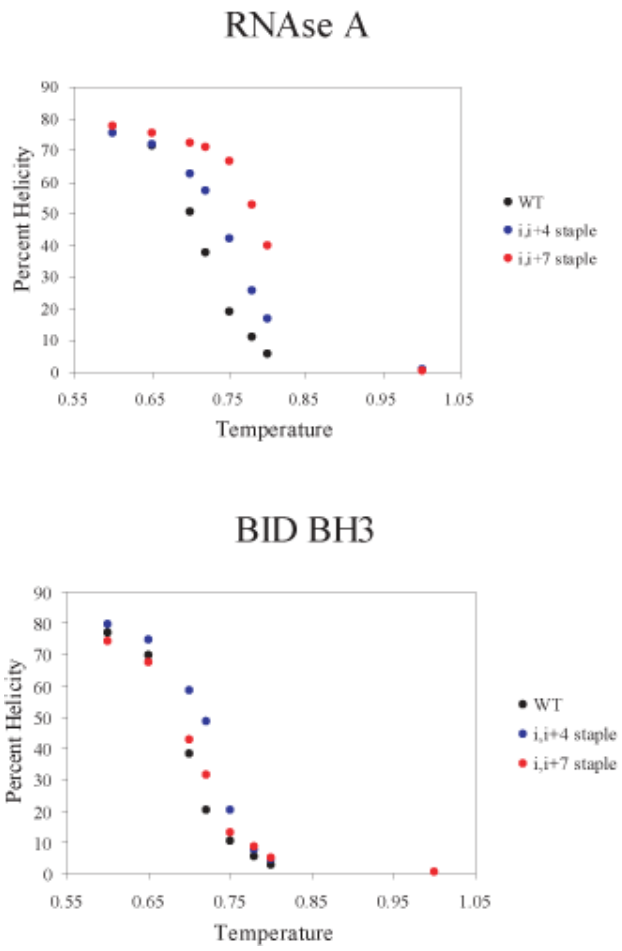

Figure S5. Melting curves for WT (black), *i*,*i*+4 stapled(blue), and *i*,*i*+7 (red) stapled peptides.

Free Energy Folding Curve

We modified a method previously used in our group to determine the free energy folding curve of helix 1 from Crambin.[7] In the current method, the number of helical residues h is used as the order parameter, as opposed to the backbone RMSD from the native structure which was previously used. The average energy as a function of the number of helical residues $H(h)$ was measured, and the probability of observing each helical state $p(h)$ was determined. By using the following statistical mechanical relation:

$$p(h) = \frac{W(h)e^{-\beta H(h)}}{Z} \tag{4}$$

we can find the density of states $W(h)$ for each helical state and subsequently the entropy of each state $S(h)$ if the partition function Z is known since:

$$S(h) = \ln\left(W(h)\right) \tag{5}$$

The partition function Z is determined in the helical state $h = N - 2$, where N is the number of residues, and $N - 2$ is the maximum number of helical residues per peptide as defined by DSSP. As such, when $h = N - 2$, $W(h) = 1$ and the partition function can be

explicitly determined by Eq. (4). We can then determine the free energy at temperature T from $G(h) = H(h) - TS(h)$.

In order to calculate thermodynamic values of interest, we needed to carefully select a temperature for each peptide. Both the fully helical bin ($h = N - 2$) and the unfolded bin ($h = 0$) must be occupied a statistically significant amount of time for each peptide. We did this by choosing a temperature near the T_m of the peptides of interest. Thus we chose T=0.78 for the RNAse A peptides and T=0.70 for the BID BH3 peptides.

The MC simulation protocol was similar to that used to determine percent helicity. For each peptide runs of 100×10^6 steps were repeatedly carried out, until sufficient convergence was achieved. In order to test for convergence, the free energy of folding to the helical state ΔG was plotted versus the number of runs performed. From a preliminary assessment of 30 runs, it was apparent that the folded state of RNAse A was h=11 and BID BH3 was h=18. While testing for convergence, runs were divided into three groups of runs, and the standard deviation of the group averages were used as a measure of error. For RNAse A, 150 runs were sufficient (Fig. S6), while for BID BH3 600 runs were necessary (Fig. S7 and S8). After convergence was reached, the free energy of each helical state was determined (Fig. S9).

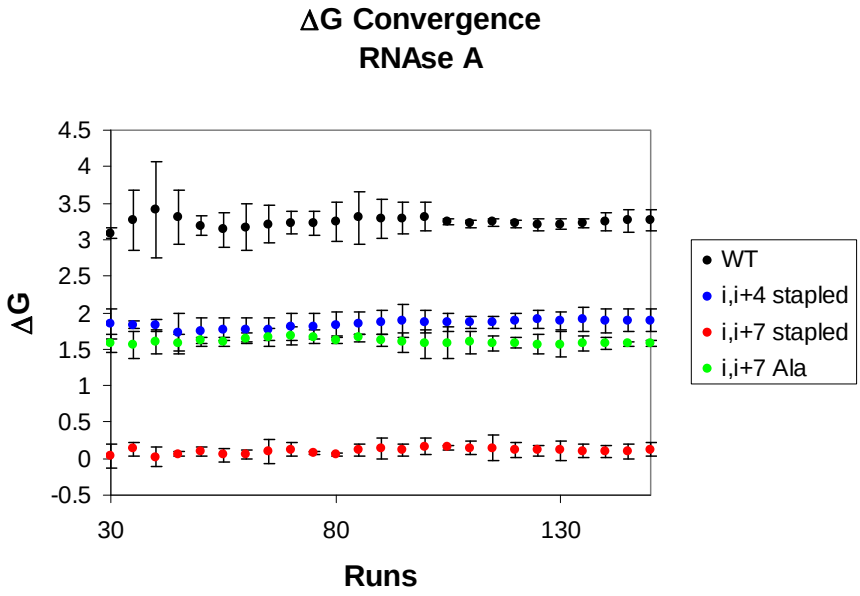

Figure S6. Free energy of folding to the helical state (h=11) versus number of runs performed for RNAse A

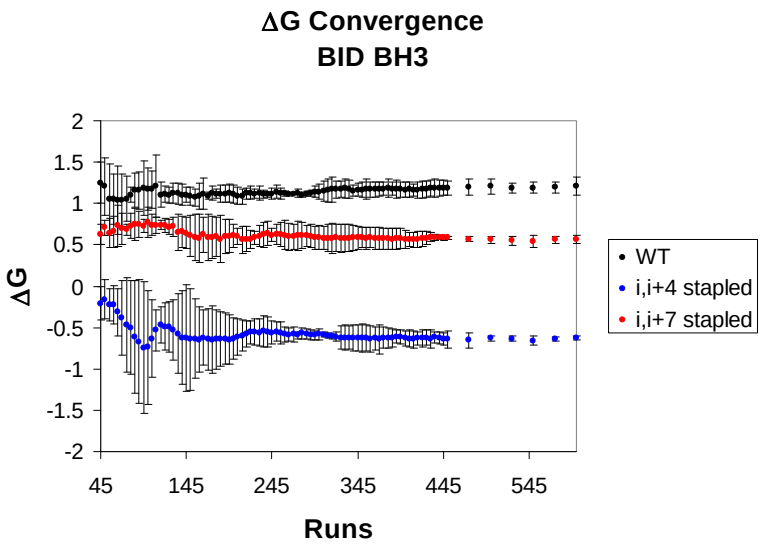

Figure S7. Free energy of folding to the helical state (h=18) versus number of runs performed for BID BH3

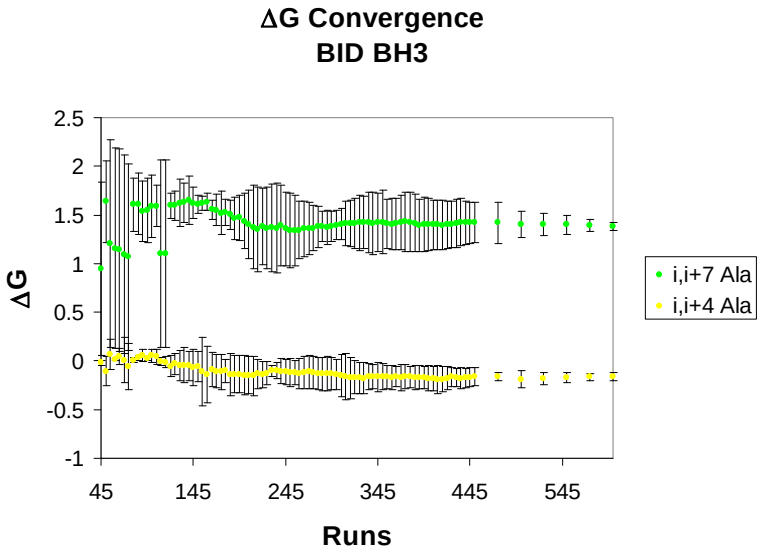

Figure S8. Free energy of folding to the helical state (h=18) versus number of runs performed for BID BH3

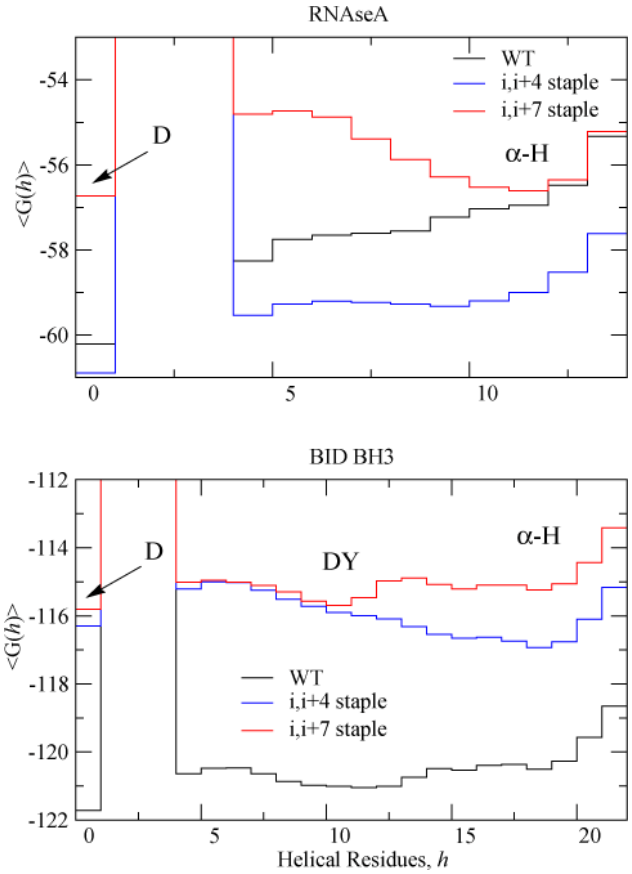

Figure S9. Average free energy <G> of RNAse A peptides at T=0.78 and BID BH3 peptides at T=0.70 where T~T_m.

We defined the denatured state as $h = 0$. The α-helical and decoy states were defined by visually inspecting the free energy folding curve, and their values are below (Table S1).

Table S1. Helical bins (h) used to define the random coil (RC), decoy (D), and α-helix (α-H) states.

	h		
	D	DY	α-H
RNAse A	0	N/A	11
BID BH3	0	10	18

The change in free energy upon folding to the α-helical state is then defined as:

$$\Delta G_{\text{folding}} = G_{h=h_{\alpha-\text{H}}} - G_{h=0} \quad (6)$$

The change in enthalpy and entropy of peptides are calculated in a similar manner (Table 1). The change in a thermodynamic quantity compared to the wild type peptide was also calculated (Table 1). For example, the change in free energy of folding for the BID BH3 *i*,*i*+4 stapled peptide relative to the WT peptide is defined as:

$$\Delta\Delta G_{\text{folding}}^{\text{BIDBH3},i,i+4} = \Delta G_{\text{folding}}^{\text{BIDBH3},i,i+4} - \Delta G_{\text{folding}}^{\text{WT}} \quad (7)$$

Table S2. Contributions of each potential term to overall enthalpic change ΔH for folding to the helical state, as well as the enthalpic change relative to the WT peptide $\Delta\Delta H$.

		WT	*i*,*i*+4 staple	*i*,*i*+4 Ala	*i*,*i*+7 staple	*i*,*i*+7 Ala
RNAse A	ΔH[a]	-29.2_ ± 0.6_	-30.0_ ± 0.5_		-33.25 ± 0.06	-32.59 ± 0.09
	ΔH(con)	-1.05 ± 0.05	-0.75 ± 0.08		-0.37 ± 0.02	-0.57 ± 0.07
	ΔH(hb)	-11.4_ ± 0.5_	-11.9_ ± 0.1_		-14.03 ± 0.07	-13.9_ ± 0.1_
	ΔH(bbtor)	-16.76 ± 0.09	-17.3_ ± 0.3_		-18.50 ± 0.09	-17.83 ± 0.09
	ΔH(sct)	0.0_ ± 0.3_	-0.14 ± 0.09		-0.28 ± 0.04	-0.2_ ± 0.1_
	ΔH(linktor)		0.1_ ± 0.1_		-0.07 ± 0.09	
	$\Delta\Delta H$[a]		-0.8 _± 0.5_		-4.0_ ± 0.6_	-3.4_ ± 0.5_
	$\Delta\Delta H$(con)		0.3_ ± 0.1_		0.69 ± 0.05	0.49 ± 0.06
	$\Delta\Delta H$(hb)		-0.5_ ± 0.4_		-2.7_ ± 0.5_	-2.6_ ± 0.6_
	$\Delta\Delta H$(bbtor)		-0.6_ ± 0.4_		-1.7_ ± 0.2_	-1.1_ ± 0.2_
	$\Delta\Delta H$(sct)		-0.1_ ± 0.4_		-0.3_ ± 0.3_	-0.2_ ± 0.2_
	$\Delta\Delta H$(linktor)		0.1_ ± 0.1_		-0.07 ± 0.09	
BID BH3	ΔH[a]	-47.0_ ± 0.4_	-48.3_ ± 0.3_	-49.2_ ± 0.4_	-47.5_ ± 0.2_	-46.5_ ± 0.4_
	ΔH(con)	-0.3_ ± 0.1_	-0.43 ± 0.07	-0.5_ ± 0.1_	-0.4_ ± 0.1_	-0.6_ ± 0.1_
	ΔH(hb)	-24.3_ ± 0.2_	-24.85 ± 0.09	-24.9_ ± 0.1_	-23.20 ± 0.1_	-22.7_ ± 0.1_
	ΔH(bbtor)	-22.6_ ± 0.2_	-23.3_ ± 0.3_	-23.6_ ± 0.1_	-24.0_ ± 0.2_	-23.9_ ± 0.2_
	ΔH(sct)	0.2_ ± 0.2_	0.2_ ± 0.2_	-0.2_ ± 0.3_	0.41 ± 0.07	0.7_ ± 0.4_
	ΔH(linktor)		-0.01 ± 0.02		-0.25 ± 0.05	
	$\Delta\Delta H$[a]		-1.4_ ± 0.5_	-2.2_ ± 0.3_	-0.5_ ± 0.5_	0.4_ ± 0.5_
	$\Delta\Delta H$(con)		-0.2_ ± 0.1_	-0.19 ± 0.03	-0.16 ± 0.09	-0.36 ± 0.04
	$\Delta\Delta H$(hb)		-0.6_ ± 0.2_	-0.6_ ± 0.4_	1.1_ ± 0.1_	1.6_ ± 0.1_
	$\Delta\Delta H$(bbtor)		-0.6_ ± 0.4_	-1.0_ ± 0.3_	-1.4_ ± 0.3_	-1.3_ ± 0.3_
	$\Delta\Delta H$(sct)		0.0_ ± 0.2_	-0.4_ ± 0.3_	0.2_ ± 0.2_	0.5_ ± 0.5_
	$\Delta\Delta H$(linktor)		-0.01 ± 0.02		-0.25 ± 0.05	

[a]Energetic terms are divided into the pair-wise contact (con), hydrogen bonding (hb), sequence dependent backbone torsion (bbtor), side chain torsion (sct), and linker torsion (linktor) components. Errors are reported as standard deviations of averages of 3 groups of 50 runs (RNAse A) or 200 runs (BID BH3). Simulations were carried out at T=0.78 for RNAse A, and T=0.70 for BID BH3. The folded state is defined as h=11 for RNAse A and h=18 for BID BH3.

Enthalpy of Specific States Relative to WT

After obtaining the thermodynamic quantities of enthalpy, entropy, and free energy at each helical state, we were interested to see how the enthalpic values of each state changed upon stapling relative to the WT peptide. We defined the change in enthalpy of a specific helical state as:

$$\delta H_{\text{state}}^{\text{peptide}} = H_{\text{state}}^{\text{peptide}} - H_{\text{state}}^{\text{WT}} \quad (8)$$

For example, the change in enthalpy of the random coil state (h=0) for a specific peptide would be:

$$\delta H_{\text{h=0}}^{\text{peptide}} = H_{\text{h=0}}^{\text{peptide}} - H_{\text{h=0}}^{\text{WT}} \quad (9)$$

The change in enthalpy for each state was obtained, as well as the change in enthalpy of each energetic term employed in the Hamiltonian (Fig. S10).

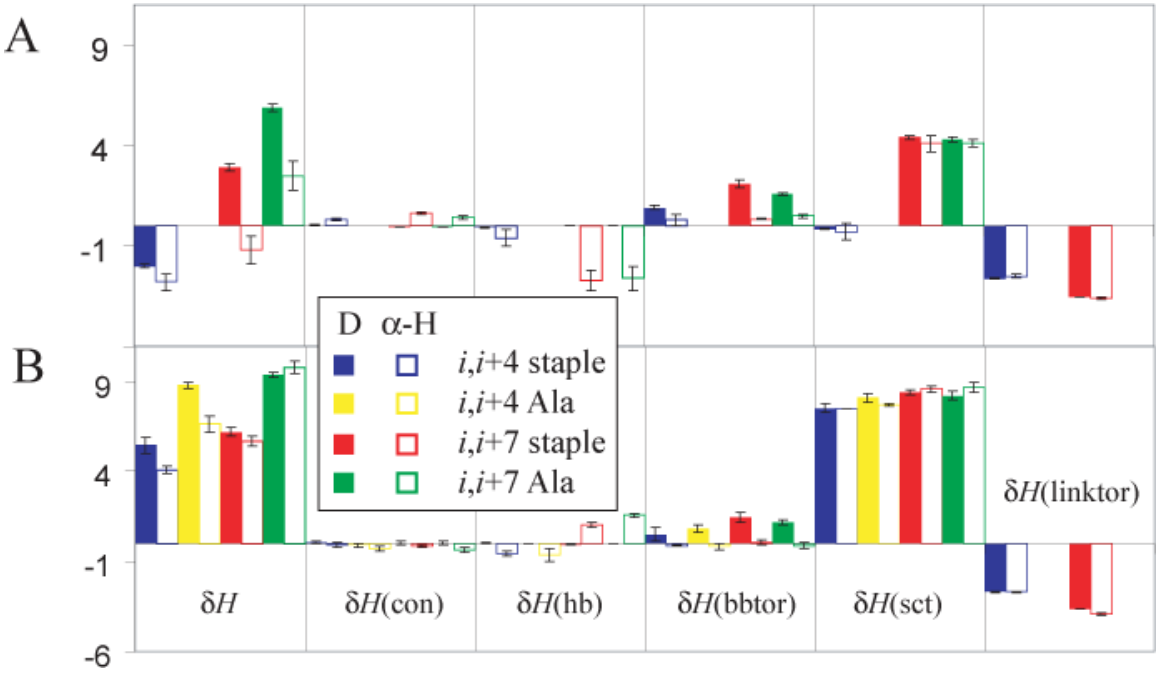

Figure S10. Changes in energy δH as well as changes in specific energetic contributions of the denatured state (D) and folded state (α-H) of stapled peptides relative to the corresponding WT peptide. Energetic contributions are pair-wise atom contacts δH(con), hydrogen bonds δH(hb), sequence-dependent backbone torsions δH(bbtor), side chain torsions δH(sct), and crosslink torsions δH(linktor). Stapled and Ala control peptides (where stapling residues are replaced by alanine) are depicted for RNAse A at T=0.78 (A), and for BID BH3 at T=0.70 (B).

Statistical Analysis of Denatured State

Structures of the denatured state (h=0) were were examined for secondary and tertiary structure content. All conformations from the RNAse A (T=0.78, 150 runs) and BID BH3 (T=0.70, 600 runs) simulations with 0 helical residues were selected for analysis. Files were converted to DSSP files, and the DSSP assignment of each residue was assessed for secondary structure. The most significant changes observed were in bend propensity, with only minor changes observed in hydrogen bonded turns, and 3_{10} helices (Fig. S11).

To assess tertiary structure in the denatured state, a contact map of each peptide in the folded or denatured state was determined. Residues were defined as being in contact

with one another if the C_β-C_β distance was within 7.0 Å. For glycine, the C_α atom was used. The probability that residues were in contact with one another assessed, and for visual analysis the folded and unfolded maps were fused at the main diagonal (Fig. S12), since each individual map is symmetrical along the main diagonal. The most significant changes we see in the *i*,*i*+4 system is an increase in the contact between residues *i* and *i*+4 in the denatured state. We also see a slightly higher propensity for residues lying in between the stapling residues to be in contact with each other in the denatured state. In the *i*,*i*+7 system, we see a slight increase in contacts between residues that lie in between the stapling residues.

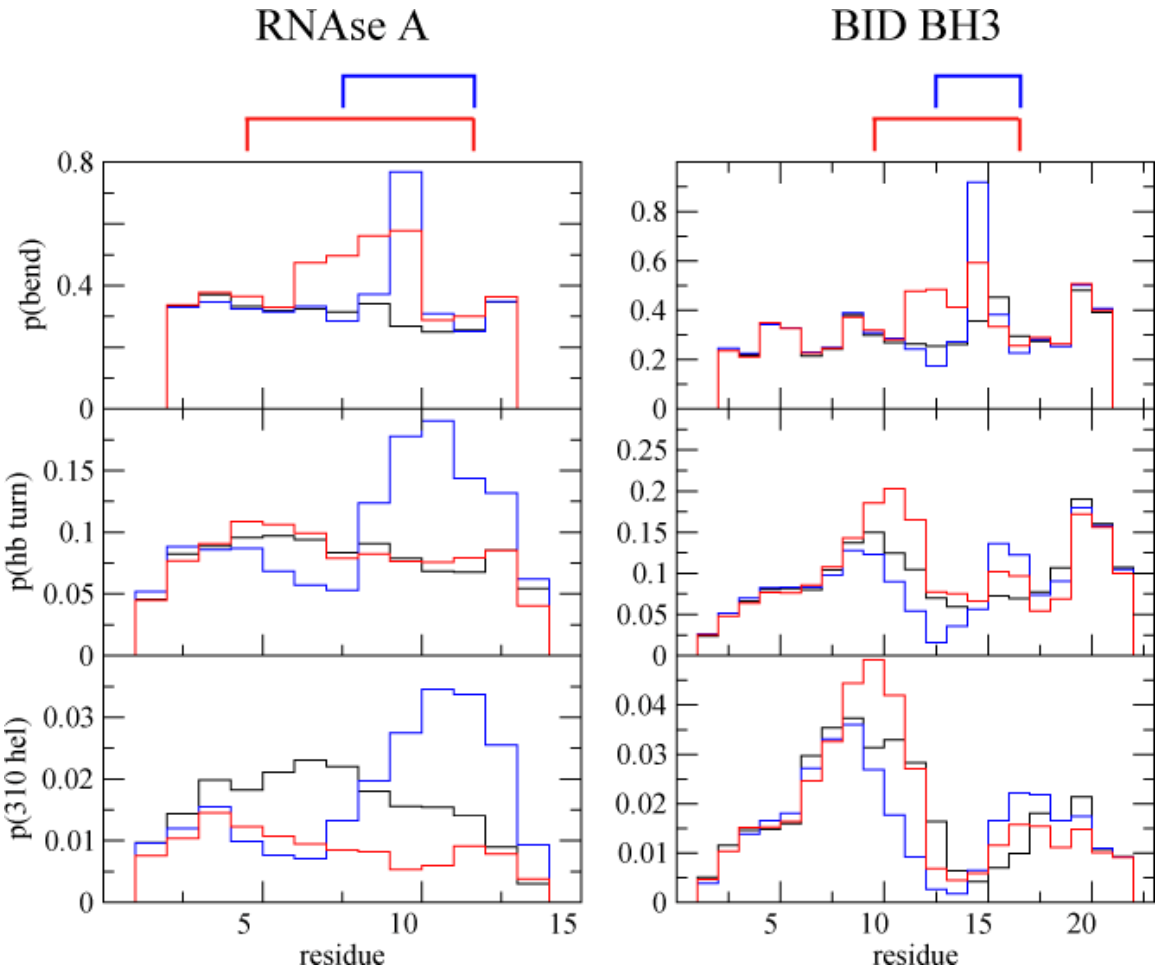

Figure S11. Secondary structure analysis for denatured state (h=0) of WT(black), *i*,*i*+4 stapled (blue), and *i*,*i*+7 stapled (red) peptides. The probability that a residue is classified as in a bend, hydrogen bonded turn, or 3_{10} helix is evaluated.

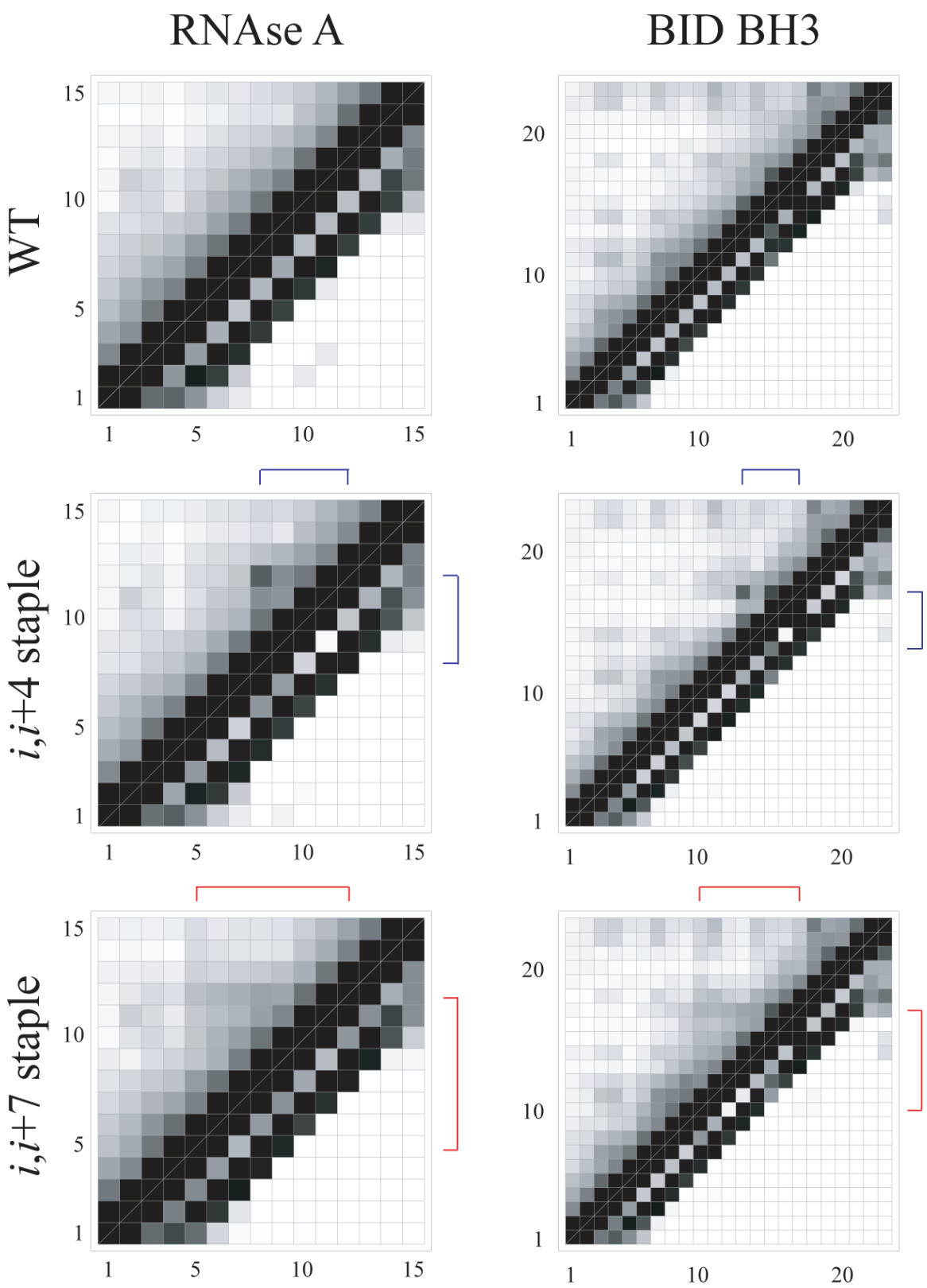

Figure S12. Contact maps of denatured (upper left triangle) and α-helical (lower right triangle) states.

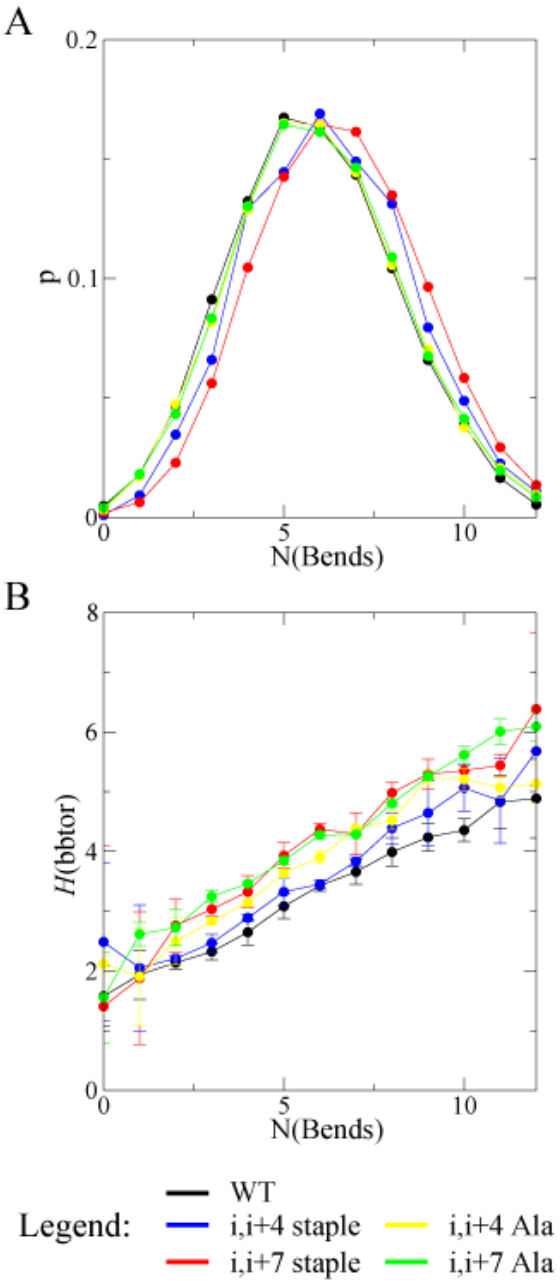

Figure S13. A: Probability of finding conformations with a certain number of "bend" residues in the denatured state for BID BH3 peptides. **B**: The average backbone torsional energy *H*(bbtor) of BID BH3 conformations in the denatured state with versus number of "bend" residues. Simulations were carried out at T=0.70. Colors of peptides are as follows: WT (black), *i*,*i*+4 stapled (blue), *i*,*i*+7 stapled (red), *i*,*i*+4 Ala (yellow),and *i*,*i*+7 Ala (green).

Identification of Representative Decoy Structures

A minimum was observed in the BID BH3 free energy plot for states with 9-11 helical residues (Fig S9). As such, BID BH3 peptide structures with 9-11 helical residues were selected from MC runs carried out at T=0.70 for clustering analysis. Since clustering all structures with 9-11 helical residues would be extremely computationally costly, we randomly selected 500 conformations from the total number of conformations prior to clustering. Single linkage hierarchical clustering was employed using alpha carbon RMSD between each conformation as the clustering metric. The size of the largest cluster, the giant component (GC), was monitored versus RMSD cutoff. Cutoffs were selected at the GC transition, when the GC contains half of the total structures. Representative structures of the GC were selected by choosing the structure with the greatest connectivity *k*.

Table S3. BID BH3 peptides with 9-11 helical residues were clustered. The cutoffs, number of clusters, and connectivity *k* of the GC representative structure are reported.

	Total Structures[a]	Cutoff (Å)	Clusters	*k*
WT	13492	2.37	206	38
SAHB A *i*,*i*+4 staple	5573	2.42	182	15
SAHB D *i*,*i*+7 staple	16016	2.05	211	26

[a]500 structures were selected randomly from the total number of structures prior to clustering.

AGADIR helical propensity prediction

The hypothesis that helical propensities of stapled peptides could be predicted using the simplified assumption that the unnatural stapling residues can be modeled as alanine was tested using the web-based AGADIR program,[8-11] a standard in helical propensity prediction. In the case of RNAse A, the *i*,*i*+4 Ala is identical to the WT, so there is no change in helicity. AGADIR predicts that placing alanine in the *i*,*i*+7 positions should increase helicity moderately. In the case of BID BH3, AGADIR predicts that and *i*,*i*+7 Ala mutant (15.05 %) should be more helical than the WT BID BH3 (7.52 %), while an *i*,*i*+4 does not significantly change the helical propensity (7.39 %). These predictions suggest that predicting the helical propensities of stapled peptides is not a trivial task.

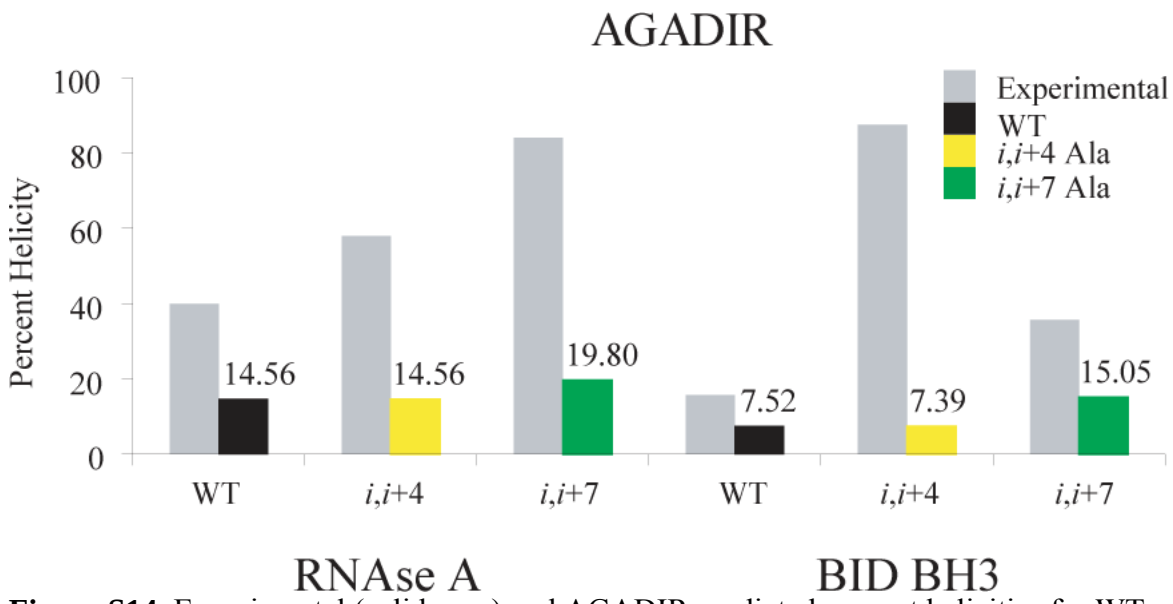

Figure S14. Experimental (solid grey) and AGADIR predicted percent helicities for WT (solid black) and alanine mutants (open yellow or green) at T=278 K.

Go Model Simulations

To check if the Gō potential can reproduce experimental trends in helicities, Gō model simulations were performed similar to our MC procedure except the Gō potential instead of our potential was used. Native contacts were given interaction energies of -1.0, and all non-native contacts were given interaction energies of zero. Accordingly, temperatures were scanned to find out melting temperature of each peptide by running 30 runs at various temperatures (Fig. S15), and simulated temperatures were set to be 3.0 (RNAase A) and 3.5 (BID BH3) where simulated helicities match experimental values for WT (Fig. S16). Runs were divided into three groups of runs, and the standard deviation of the group averages were used as a measure of error.

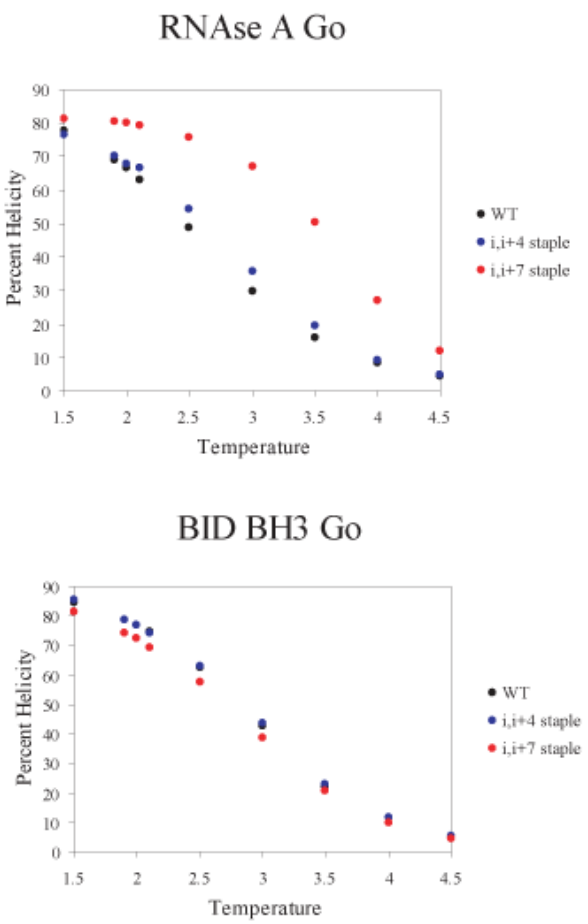

Figure S15. Melting curves for WT (black), *i*,*i*+4 stapled(blue), and *i*,*i*+7 (red) stapled peptides using the Gō model.

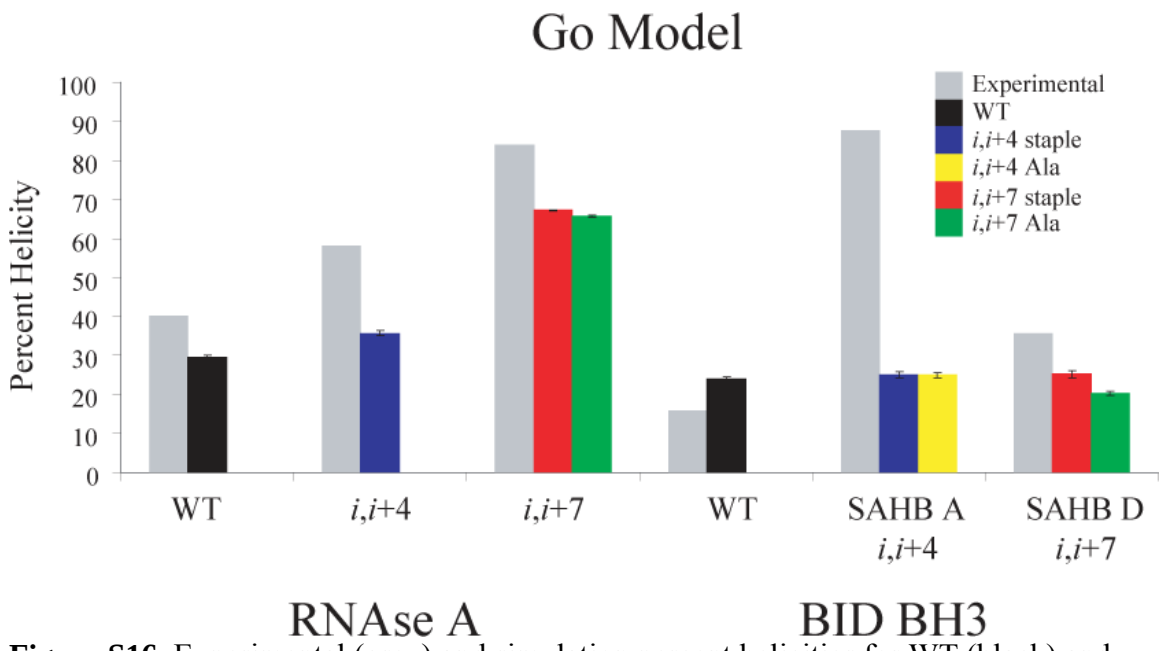

Figure S16. Experimental (grey) and simulation percent helicities for WT (black) and stapled (blue, green) peptides. Simulation results for Alanine mutants (yellow, green) are also depicted. The temperature chosen for the RNAse A (T=3.0) and the BID BH3 (T=3.5) differed in order to match the WT helical propensities.

References

(1) Yang, J. S.; Chen, W. W.; Skolnick, J.; Shakhnovich, E. I. *Structure* **2007**, *15*, 53-63.
(2) Chen, S.-T. C., H.-J; Yu, H.-M.; Wang,K.-T. *J. Chem. Res. (S)* **1993**, 228-229.
(3) Dodd, L. R.; Boone, T. D.; Theodorou, D. N. *Mol. Phys.* **1993**, *78*, 961-996.
(4) Macromodel, S., LLC, New York, NY, 2005.
(5) Schafmeister, C. E.; Po, J.; Verdine, G. L. *J. Am. Chem. Soc.* **2000**, *122*, 5891-5892.
(6) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577-2637.
(7) Shimada, J.; Kussell, E. L.; Shakhnovich, E. I. *J Mol Biol* **2001**, *308*, 79-95.
(8) Munoz, V.; Serrano, L. *Journal of Molecular Biology* **1995**, *245*, 275-296.
(9) Munoz, V.; Serrano, L. *Journal of Molecular Biology* **1995**, *245*, 297-308.
(10) Munoz, V.; Serrano, L. *Biopolymers* **1997**, *41*, 495-509.
(11) Lacroix, E.; Viguera, A. R.; Serrano, L. *Journal of Molecular Biology* **1998**, *284*, 173-191.

Complete ref 30:

(30) Danial, N. N.; Walensky, L. D.; Zhang, C. Y.; Choi, C. S.; Fisher, J. K.; Molina, A. J.; Datta, S. R.; Pitter, K. L.; Bird, G. H.; Wikstrom, J. D.; Deeney, J. T.; Robertson, K.; Morash, J.; Kulkarni, A.; Neschen, S.; Kim, S.; Greenberg, M. E.; Corkey, B. E.; Shirihai, O. S.; Shulman, G. I.; Lowell, B. B.; Korsmeyer, S. J. *Nat Med* **2008**, *14*, 144-153.