J|A|C|S
ARTICLES
Published on Web 04/05/2002

The Effect of Backbone Cyclization on the Thermodynamics of β-Sheet Unfolding: Stability Optimization of the PIN WW Domain

Songpon Deechongkit and Jeffery W. Kelly*

Contribution from the Department of Chemistry and The Skaggs Institute of Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, BCC 506, La Jolla, California 92037

Received October 15, 2001

Abstract: Backbone cyclization is often used in attempts to enhance protein stability, but is not always successful as it is possible to remove stabilizing or introduce destabilizing interactions in the process. Cyclization of the PIN1 WW domain, a 34-residue three-stranded β-sheet structure, removes a favorable electrostatic interaction between its termini. Nevertheless, optimization of the linker connecting the N- and C-termini using information based on the previously determined ensemble of NMR structures leads to β-sheets that are more stable than those derived from the linear sequence. Linkers that are too short or too long introduce strain, likely disrupting native interactions, leading to cyclic folds that are less stable than that of the linear sequence.

Introduction

One strategy for stabilizing the folded conformation of proteins is to covalently link parts of the molecule that are distant in primary structure. This can improve thermal stability, rescue loss of function caused by a mutation, and change the mechanism of folding.[1−5] Such links can be made by disulfide bond formation or by cyclization of the protein through the backbone or the side chains. Nature almost always uses the first strategy in proteins derived from ribosomal synthesis; the biosynthetic machinery needed to achieve the latter only exists for small peptides (the largest, having approximately 30 residues, being derived from plants[6−9]). However, side-chain and backbone cyclization can be achieved chemically, and there has been significant interest in its use in protein engineering over the last two decades.[1,3,10,11] The utility of backbone cyclization in protein engineering hinges on our ability to optimize the thermodynamic stability of cyclized proteins while maintaining the integrity of their folds. It is clear that the chain entropy of the unfolded state will always be reduced in cyclized proteins relative to their linear counterparts. However, cyclic proteins sometimes are less stable than their acyclic counterparts, likely owing to strain introduced by linking the N- and C-termini.[2]

To cyclize the backbone of a folded protein, the N- and C-termini have to be close enough for amide bond formation. Several proteins meet this requirement, allowing the preparation of their cyclic analogues.[12,13] The first successful semisynthesis of a cyclic protein was accomplished by Creighton and Goldenberg through exposure of native BPTI to water-soluble carbodiimide (EDCI).[11] Since then, there have been significant advances in peptide and protein synthesis methodology that have made the preparation of backbone-cyclized proteins more facile, most notably native chemical ligation and the intein approach.[3,14a,15] In this study, we have cyclized the PIN WW domain, a well-studied, cooperatively folded, three-stranded β-sheet miniprotein,[16,17] by inserting a flexible glycine spacer of variable length between the N- and C-termini to address the problem of strain versus stability.

Results

Design. The PIN WW domain is one of over 200 members of a family named after the two conserved tryptophan residues

* To whom correspondence should be addressed. Voice: (858) 784-9605. Fax: (858) 784-9610. E-mail: jkelly@scripps.edu.

(1) Gilon, C.; Halle, D.; Chorev, M.; Selinger, Z.; Byk, G. *Biopolymers* **1991**, *31*, 745−750.
(2) Vogl, T.; Brengelmann, R.; Hinz, H.-J.; Scharf, M.; Loetzbeyer, M.; Engels, J. W. *J. Mol. Biol.* **1995**, *254*, 481−496.
(3) Iwai, H.; Pluckthun, A. *FEBS Lett.* **1999**, *459*, 166−172.
(4) Grantcharova, V. P.; Baker, D. *J. Mol. Biol.* **2001**, *306*, 555−563.
(5) Camarero, J. A.; Fushman, D.; Sato, S.; Giriat, I.; Cowburn, D.; Raleigh, D. P.; Muir, T. W. *J. Mol. Biol.* **2001**, *308*, 1045−1062.
(6) Craik, D. J.; Daly, N. L.; Bond, T.; Waine, C. *J. Mol. Biol.* **1999**, *294*, 1327−1336.
(7) Gustafson, K. R.; Walton, L. K.; Sowder, R. C., Jr.; Johnson, D. G.; Pannell, L. K.; Cardellina, J. H., Jr.; Boyd, M. R. *J. Nat. Prod.* **2000**, *63*, 176−178.
(8) Saether, O.; Craik, D. J.; Campbell, I. D.; Sletten, K.; Juul, J.; Norman, D. G. *Biochemistry* **1995**, *34*, 4147−4158.
(9) Felizmenio-Quimio, M. E.; Daly, N. L.; Craik, D. J. *J. Biol. Chem.* **2001**, *276*, 22875−22882.
(10) Camarero, J. A.; Pavel, J.; Muir, T. W. *Angew. Chem., Int. Ed.* **1998**, *37*, 347−349.
(11) Goldenberg, D. P.; Creighton, T. E. *J. Mol. Biol.* **1984**, *179*, 527−545.
(12) Friedler, A.; Friedler, D.; Luedtke, N. W.; Tor, Y.; Loyter, A.; Gilon, C. *J. Biol. Chem.* **2000**, *275*, 23783−23789.
(13) Friedler, A.; Zakai, N.; Karni, O.; Broder, Y. C.; Baraz, L.; Kotler, M.; Loyter, A.; Gilon, C. *Biochemistry* **1998**, *37*, 5616−5622.
(14) (a) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. *Science (Washington, D.C.)* **1994**, *266*, 776−779. (b) Hackeng, T. M.; Griffin, J. H.; Dawson, P. E. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 10068−10073. (c) Lu, W.; Qasim, M. A.; Kent, S. B. H. *J. Am. Chem. Soc.* **1996**, *118*, 8518−8523.
(15) Camarero, J. A.; Muir, T. W. *J. Am. Chem. Soc.* **1999**, *121*, 5597−5598.
(16) Kaul, R.; Angeles, A. R.; Jaeger, M.; Powers, E. T.; Kelly, J. W. *J. Am. Chem. Soc.* **2001**, *123*, 5206−5212.
(17) Jaeger, M. N. H.; Crane, J. C.; Kelly, J. W.; Gruebele, M. *J. Mol. Biol.* **2001**, *311*, 373−393.

10.1021/ja0123608 CCC: $22.00

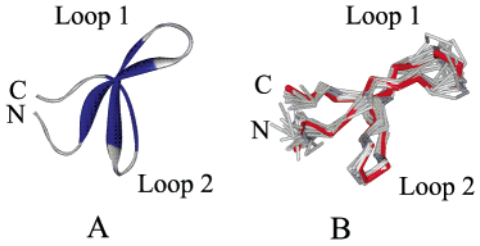


Figure 1. (A) Ribbon diagram of the polypeptide backbone of the isolated WW domain from the PIN1 X-ray crystal structure. (B) Line diagram representation of 20 low energy structures compatible with the NMR constraints for the isolated PIN WW domain (6−39). The red structure represents the average of the structural ensemble.

found in their sequences. WW domains bind peptide ligands that are rich in Pro residues and/or contain a phosphorylated residue, and thus they often mediate protein−protein interactions. The PIN WW domain recognizes a phosphoserine residue, allowing the other domain (the PPIase domain) to effect cis−trans peptidyl prolyl isomerization in the substrate.[18−24] The crystal structure of the entire PIN1 protein has been solved to 1.34 Å resolution.[25] The WW domain portion of PIN1 is shown in Figure 1A. The solution structure of the isolated PIN WW domain has been solved by NMR (Figure 1B).[26] As evident from the figure, the PIN WW domain retains the same fold in isolation from the PPIase domain. This domain is monomeric and functional in that it binds its phosphorylated peptide ligand with the expected affinity.

The primary consideration in determining whether the PIN WW domain can be backbone cyclized is to evaluate the distance between the N- (residue 6) and C-termini (residue 39) of the isolated domain. The termini are separated by 9.1 Å according to the crystal structure (Figure 1A),[25] and by 10.9 Å in the average NMR structure (Figure 1B).[26] This separation is too large for direct linkage of the termini; however, linkage could be facilitated by the inclusion of an additional 2−4 residues (linker sequence) in the primary structure. Cyclic PIN WW domains were prepared with linkers varying from 1 to 7 residues. In the process of reducing the chain entropy for folding, cyclization also removes the charge at the N- and C-termini. To evaluate the influence of charge removal on the stability of cyclic PIN WW domains, an N-terminally acetylated and C-terminally amidated PIN WW domain (capped PIN) was prepared for comparison to the cyclized and wt variants.

Synthesis. The native and capped PIN WW domains were prepared by conventional Fmoc-based solid-phase peptide synthesis using Wang resin and Rink amide resin, respectively. We chose to use the native chemical ligation method introduced by Dawson and Kent for the preparation of cyclic PIN WW domains.[14] This has been used previously for the preparation of a cyclic YAP WW domain by Camarero et al.,[10] who prepared a peptide having an N-terminal cysteine and C-terminal thioester by Boc-based solid-phase peptide synthesis. The use

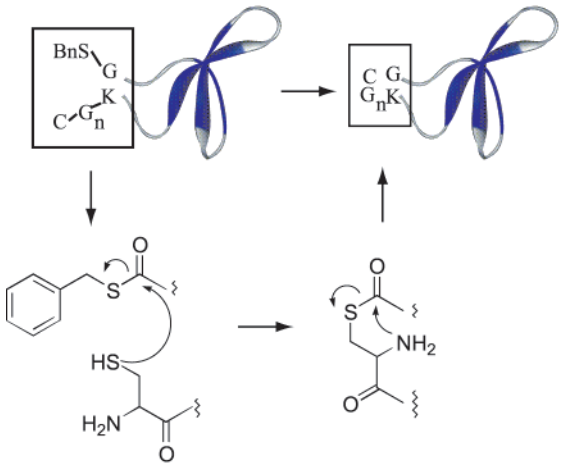
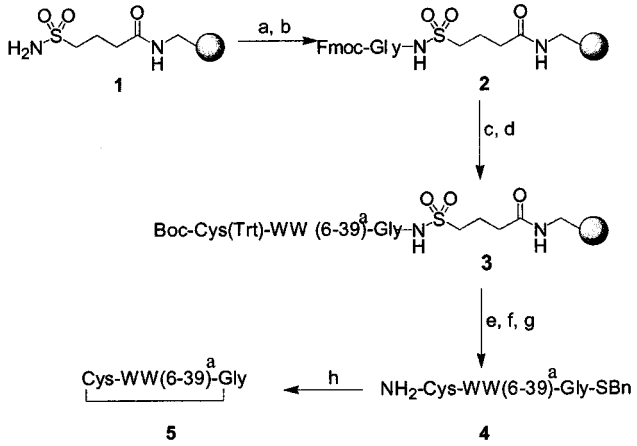


Figure 2. Strategy for backbone cyclization of the PIN WW domain (6−39). A Cys−Gly$_n$ linker (n = 0−6) was added to the N-terminus to react with the C-terminal benzyl thioester. Cyclization under native chemical ligation conditions yields the generic cyclic peptide shown on the right.

Scheme 1. Synthesis of the Cyclic PIN WW Domain Analogues



[a] WW (6−39) denotes the PIN WW domain 6−39 sequence. [b] (a) Fmoc−Gly−OH, PyBOP, DIEA, DMF; −20 °C, 8 h; room temperature, 16 h; (b) Ac_2O, HOBt, DIEA, 24 h; (c) repetitive cycles of deprotection (piperidine) and coupling utilizing Fmoc amino acids activated by HOBt, HBTU, DIEA; (d) Boc−Cys(Trt)−OH, HBTU, DIEA, 30 min; (e) ICH_2CN, DIEA, NMP, 24 h; (f) BnSH, DMF, 24 h; (g) reagent K (0.5 mL of *m*-cresol, 0.5 mL of thioanisole, 0.5 mL of H_2O, 0.25 mL of ethanedithiol, 10 mL of TFA), 4 h; (h) 0.2% PhSH, 0.2% BnSH, 1 mM DTT, 1 mM EDTA, 100 mM Na phosphate (pH 7.7), 24 h.

of native chemical ligation demands that the sequence have a cysteine residue at the point of ligation.[14] Therefore, one cysteine residue was incorporated along with glycine residues to optimize the linker's ability to accommodate whatever conformational demands arise from cyclization of the WW domain (Figure 2). The sequences for the native, capped, and cyclic PIN WW domains are shown below.

	6	39
wt PIN:	KLPPGWEKRMSRSSGRVYYFNHITNASQWERPSG	
cPINWW	CG$_n$KLPPGWEKRMSRSSGRVYYFNHITNASQWERPSG	
Capped PIN:	Ac-KLPPGWEKRMSRSSGRVYYFNHITNASQWERPSG-NH$_2$	

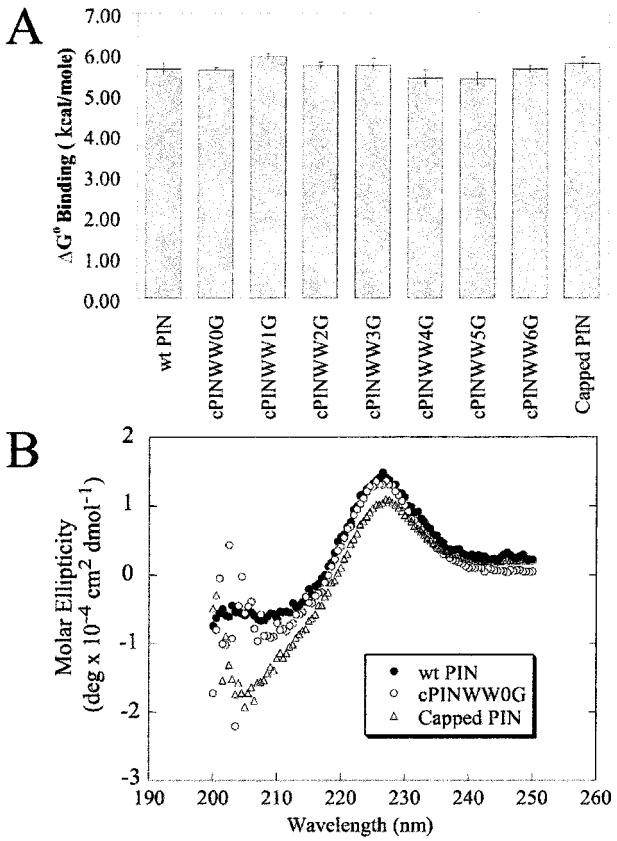
Because resin cleavage and side-chain deprotection are much easier with peptides prepared by Fmoc chemistry (HF is not required), we adapted the method of Pessi et al. (Scheme 1) to prepare linear peptide thioesters utilizing Fmoc-based solid-

(18) Macias, M. J.; Gervais, V.; Civera, C.; Oschkinat, H. *Nat. Struct. Biol.* **2000**, *7*, 375−379.
(19) Verdecia, M. A.; Bowman, M. E.; Lu, K. P.; Hunter, T.; Noel, J. P. *Nat. Struct. Biol.* **2000**, *7*, 639−643.
(20) Sudol, M.; Hunter, T. *Cell* **2000**, *103*, 1001−1004.
(21) Sudol, M. *Prog. Biophys. Mol. Biol.* **1996**, *65*, 113−132.
(22) Staub, O.; Rotin, D. *Structure*. **1996**, *4*, 495−499.
(23) Lu, P. J.; Zhou, X. Z.; Shen, M.; Lu, K. P. *Science (Washington, D.C.)* **1999**, *283*, 1325−1328.
(24) Lu, K. P.; Hanes, S. D.; Hunter, T. *Nature* **1996**, *380*, 544−547.
(25) Ranganathan, R.; Lu, K. P.; Hunter, T.; Noel, J. P. *Cell* **1997**, *89*, 875−86.
(26) Kowalski, J. A.; Liu, K.; Kelly, J. W. *Biopolymers* **2002**, *63*, 111−121.

phase peptide synthesis.[27] The commercially available safety-catch linker (**1**) was acylated using Fmoc−Gly−OH, PyBOP, DIEA in DMF at −20 °C and allowed to warm to room temperature to yield Fmoc−Gly-substituted resin (**2**).[28,29] After capping unreacted resin sites with acetic anhydride, the sequence was assembled by automated Fmoc-based solid-phase peptide synthesis.[30] The final residue incorporated was Boc−Cys(Trt)−OH.[31] The resin-bound peptide (**3**) was treated with iodoacetonitrile (ICH_2CN) which cyanomethylated the sulfonamide nitrogen, activating it toward nucleophilic displacement with benzylmercaptan (BnSH). The fully protected peptide thioester was deprotected using reagent K (82.5% TFA, 5% *m*-cresol, 5% thioanisole, 5% H_2O, and 2.5% ethanedithiol).[32] The resulting deprotected peptide (**4**) was cyclized by slow addition of the peptide thioester to a solution of 100 mM sodium phosphate (pH 7.7), 1 mM EDTA, 1 mM DTT, 0.05% benzylmercaptan, and 0.05% thiophenol via syringe pump (thiophenol facilitates the reaction by trans-thioesterification). The slow addition ensures that the peptide concentration is always low, favoring intramolecular cyclization over intermolecular oligomerization. Typical yields of cyclized products (**5**) were around 60%. While it is conceivable that the activated amino acid thioester could undergo epimerization during native chemical ligation, previous studies have shown that this does not occur.[14b,c] In this study, the activated amino acid is glycine for which epimerization is impossible. The capped and cyclized products were purified by reverse phase HPLC followed by gel filtration chromatography. The identities of the products were confirmed by mass spectrometry.

Biophysical Studies. Sedimentation equilibrium analytical ultracentrifugation (AUC) studies demonstrate that all of the variants prepared for this study were monomeric under the conditions used to evaluate the thermodynamic parameters (see Experimental Section). All these constructs bind to the phosphorylated peptide YSPTpSPS ligand (CTD-S5) with comparable free energies (5−6 kcal/mol; Figure 3A), implying that the cyclized WW domains adopt a native structure. Binding was demonstrated by the fluorescence anisotropy method of Vinson et al. outlined in the Experimental Section.[33] That the cyclic and capped WW variants are folded was further demonstrated by their unusual and characteristic circular dichroism spectra exhibiting the same 227 nm maximum as the wt PIN WW domain.[17] Representative spectra and molar ellipticities at 227 nm for each of the WW domain variants are shown in Figure 3B.

A single thermal denaturation curve can only be used to determine the temperature and enthalpy at the midpoint of unfolding (T_m and ΔH_m, respectively).[34] The free energy of unfolding ($\Delta G°_u$) cannot be determined from a single thermal denaturation curve without knowing the heat capacity of unfolding ($\Delta C_{p,u}$). Because $\Delta C_{p,u}$ is equal to $d\Delta H_m/dT_m$, $\Delta C_{p,u}$ can be measured if a set of conditions can be found under which ΔH_m and T_m vary.[35−38] We found that the stability of the PIN WW domain is pH-dependent. The method of Privalov was therefore used to determine $\Delta C_{p,u}$ from thermal denaturation curves recorded at several pH's, Figure 4.[34] Because the hydrophobic surface area exposed upon unfolding should be very similar for all the WW domains studied herein, we assumed that they should all have the same value of $\Delta C_{p,u}$. A global fit



Polypeptides	Molar Ellipticity (227 nm) (deg cm^2 $dmol^{-1}$)
cPINWW0G	13,600
cPINWW1G	12,200
cPINWW2G	13,300
cPINWW3G	12,200
cPINWW4G	12,800
cPINWW5G	11,100
cPINWW6G	10,400
wt PIN	14,000
Capped PIN	10,900

Figure 3. (A) A plot of summarizing the free energy of binding of the YSPTpSPS ligand to the various PIN WW domain variants at pH 7.0 (25 °C). (B) Representative far-UV CD spectra of WW domains (25 μM) at pH 7 (20 mM sodium phosphate buffer, 1 mM DTT (2 °C)) along with the values of molar ellipticity at 227 nm for each WW variant.

(27) Ingenito, R.; Bianchi, E.; Fattori, D.; Pessi, A. *J. Am. Chem. Soc.* **1999**, *121*, 11369−11374.
(28) Backes, B. J.; Virgilio, A. A.; Ellman, J. A. *J. Am. Chem. Soc.* **1996**, *118*, 3055−6.
(29) Shin, Y.; Winans, K. A.; Backes, B. J.; Kent, S. B. H.; Ellman, J. A.; Bertozzi, C. R. *J. Am. Chem. Soc.* **1999**, *121*, 11684−11689.
(30) Atherton, E.; Sheppard, R. C. *Solid-Phase Peptide Synthesis: A Practical Approach*; IRL Press: Oxford, U.K., 1989.
(31) Schnoelzer, M.; Alewood, P.; Jones, A.; Alewood, D.; Kent, S. B. H. *Int. J. Pept. Protein Res.* **1992**, *40*, 180−193.
(32) Sole, N. A.; Barany, G. *J. Org. Chem.* **1992**, *57*, 5399−5403.
(33) Vinson, V. K.; Cruz, E. M. D. L.; Higgs, H. N.; Pollard, T. D. *Biochemistry* **1998**, *37*, 10871−10880.
(34) Privalov, P. L. *Protein Folding* **1992**, 83−126.
(35) Zaiss, K.; Jaenicke, R. *Biochemistry* **1999**, *38*, 4633−4639.
(36) Viguera, A. R.; Martinez, J. C.; Filimonov, V. V.; Mateo, P. L.; Serrano, L. *Biochemistry* **1994**, *33*, 2142−2150.
(37) Padmanabhan, S.; Laurents, D. V.; Fernandez, A. M.; Elias-Arnanz, M.; Ruiz-Sanz, J.; Mateo, P. L.; Rico, M.; Filimonov, V. V. *Biochemistry* **1999**, *38*, 15536−15547.
(38) Beldarrain, A.; Lopez-Lacomba, J. L.; Furrazola, G.; Barberia, D.; Cortijo, M. *Biochemistry* **1999**, *38*, 7865−7873.

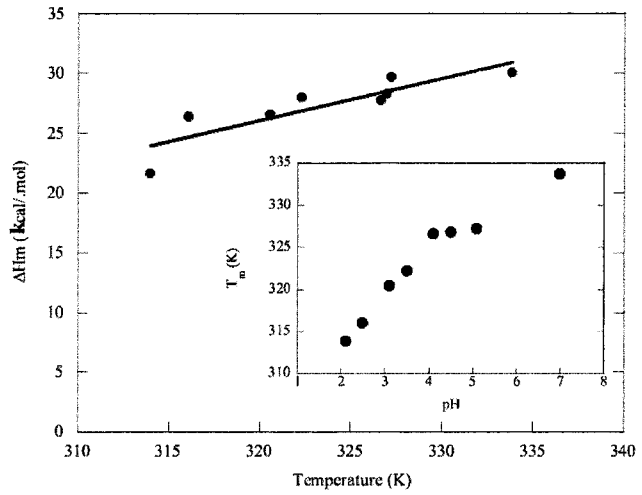


Figure 4. A plot of ΔH_m vs T_m obtained from cPINWW0G thermal denaturation at different pH's. The inset shows the dependence of T_m on pH for cPINWW0G. Analogous plots for all the variants were fitted simultaneously to obtain the optimal $\Delta C_{p,u}$ value.

Table 1. Comprehensive Data Characterizing the Unfolding Thermodynamics of PIN WW Domain Variants

polypeptide	T_m (°C)[a]	ΔH_m[b] (kcal/mol)	$\Delta H°(T)$[c] (kcal/mol)	$\Delta S°(T)T$[c] (kcal/mol)	$\Delta G°(T)$[c] (kcal/mol)
cPINWW0G	60.7	30.1	16.5 ± 1.2	14.1 ± 1.1	2.5 ± 0.1
cPINWW1G	66.5	30.8	15.0 ± 1.4	12.2 ± 1.3	2.8 ± 0.1
cPINWW2G	71.7	37.1	19.3 ± 1.5	15.6 ± 1.4	3.8 ± 0.1
cPINWW3G	70.3	36.6	19.4 ± 1.5	15.7 ± 1.3	3.6 ± 0.1
cPINWW4G	69.5	36.3	19.3 ± 1.4	15.8 ± 1.3	3.6 ± 0.1
cPINWW5G	66.4	32.4	16.7 ± 1.4	13.7 ± 1.3	2.9 ± 0.1
cPINWW6G	63.6	35.0	20.3 ± 1.3	17.2 ± 1.2	3.1 ± 0.1
wt PIN	59.0	34.8	19.5 ± 1.0	17.4 ± 0.9	2.1 ± 0.1
Capped PIN	52.8	30.1	21.9 ± 1.2	19.0 ± 1.1	2.9 ± 0.1

[a] Error is within ±0.5 °C. [b] Error is within ±0.5 kcal/mol. [c] These parameters were calculated using globally fitted ΔC_p of 0.38 ± 0.03 kcal/mol K. The reference condition for all these parameters is at 298 K, 1 atm, and pH 7.0.

of the ΔH_m and T_m data for all the WW domain variants gives $\Delta C_{p,u}$ equal to 0.38 ± 0.03 kcal mol^{-1} K^{-1}. Using this value, $\Delta G°_u$, $\Delta H°_u$, and $T\Delta S°_u$ at 298 K (pH 7.0) were determined as outlined in Table 1. The values of $\Delta G°_u$ for the variants are plotted versus the number of residues intervening between the N- and C-termini in Figure 5.[39]

Discussion

Three characteristics of the unfolding thermodynamics of the WW domain are enlightening. First, removing the charges at the N- and C-termini by acetylation and amidation significantly reduces the stability of the PIN WW domain. Second, cyclizing the PIN WW domain greatly enhances its stability for all linker lengths relative to capped PIN. The degree of stabilization (Table 2) is 0.4 and 1.0 kcal/mol for the shortest and the longest linkers and 1.7 kcal/mol for the optimal three-residue linker (cPINWW2G). Third, the effects of charge removal and cyclization act against each other, such that only the cyclic PIN analogues with linker lengths between 3 and 5 residues (cPINWW2G to cPINWW4G) are more stable than wild-type PIN. These linkers are expected

(39) In addition to the aforementioned method, significant effort was put into determining ΔC_p by differential scanning calorimetry (DSC). However, the method failed, probably because PIN WW domain variants aggregated at the high concentration and temperature required for carrying out the experiment.

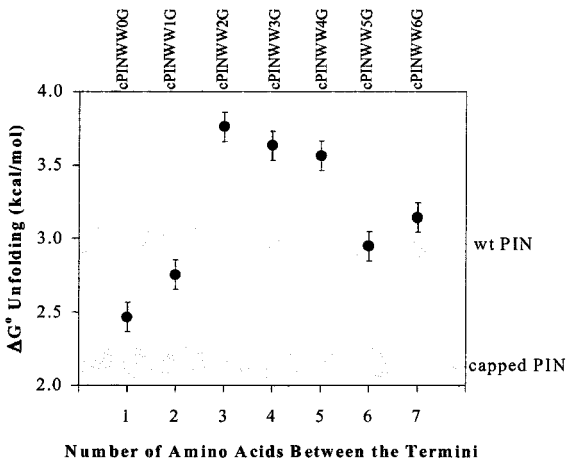
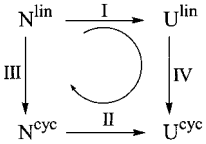


Figure 5. A plot of $\Delta G°$ unfolding for the cyclic PIN WW domain variants vs the number of added residues in the linker. The two gray bands represent the uncertainty in the measured $\Delta G°$ for the wt Pin and capped Pin WW domains.

Table 2. Summary of the Thermodynamic Parameters Characterizing the Unfolding of the PIN WW Domain Variants Relative to Those of Capped PIN

	$\Delta\Delta H°$ kcal/mol	$\Delta(T\Delta S°)$ kcal/mol	$\Delta\Delta G°$ kcal/mol
wt PIN	2.3 ± 1.5	1.5 ± 1.4	0.8 ± 0.1
Capped PIN	0.0	0.0	0.0
cPINWW0G	−3.0 ± 1.6	−3.4 ± 1.5	0.4 ± 0.1
cPINWW1G	−4.6 ± 1.7	−5.2 ± 1.6	0.6 ± 0.1
cPINWW2G	−0.2 ± 1.8	−1.9 ± 1.7	1.7 ± 0.1
cPINWW3G	−0.2 ± 1.8	−1.7 ± 1.6	1.5 ± 0.1
cPINWW4G	−0.2 ± 1.7	−1.7 ± 1.6	1.5 ± 0.1
cPINWW5G	−2.9 ± 1.7	−3.7 ± 1.6	0.8 ± 0.1
cPINWW6G	−0.8 ± 1.6	−0.2 ± 1.5	1.0 ± 0.1

Scheme 2. Thermodynamic Cycle Comprised of Four Processes Associated with the WW Domain[a]



[a] (I) Unfolding of linear peptide, (II) unfolding of cyclic peptide, (III) cyclization of the peptide in the native state, and (IV) cyclization of the peptide in the unfolded state.

to span a distance of 10.5 (three-residue linker) to 17.5 Å (five-residue linker), assuming a translation of 3.5 Å per residue. Linkers comprised of 3−5 residues appear optimal to bridge the shortest and the longest separations between the N- and C-termini in the ensemble of NMR-derived structures (6.4−13.5 Å).[40] This emphasizes the importance of carefully considering the ensemble of structures representing the folded state when choosing the optimal linker length.

The source of the stabilization from cyclization appears to be clear: cyclization decreases the conformational entropy of the unfolded state.[2] However, it is not immediately obvious how changing the linker length further modulates stability. This can be understood with the help of Scheme 2, which depicts a thermodynamic cycle consisting of four processes: I, the unfolding of the linear WW domain; II, the unfolding of the cyclic WW domain; III, the cyclization of the WW domain in

(40) The termini are ill-defined in this NMR structure either because they are flexible or because there were too few constraints. Nevertheless, the range from 6.4−13.5 Å includes all of the likely possibilities for the N- and C-terminal spacing.

the native conformation; and IV, the cyclization of the WW domain from an ensemble of unordered conformations. As with any thermodynamic cycle, the sum of the free energies (or any other thermodynamic state function) around the cycle has to be zero. In the direction indicated by the clockwise arrow in the Scheme 2, this sum is

$$\Delta G^\circ_{\mathrm{I}} + \Delta G^\circ_{\mathrm{IV}} - \Delta G^\circ_{\mathrm{III}} - \Delta G^\circ_{\mathrm{II}} = 0 \qquad (1)$$

This can be rearranged to give

$$\Delta G^\circ_{\mathrm{I}} - \Delta G^\circ_{\mathrm{II}} = \Delta G^\circ_{\mathrm{III}} - \Delta G^\circ_{\mathrm{IV}} \qquad (2)$$

In eq 2, the left-hand side represents the difference between the free energies of unfolding of the linear ($\Delta G^\circ_{\mathrm{I}}$) and cyclic peptides ($\Delta G^\circ_{\mathrm{II}}$). The right-hand side represents the difference between the free energies of cyclization of the native ($\Delta G^\circ_{\mathrm{III}}$) and the unfolded states ($\Delta G^\circ_{\mathrm{IV}}$). The quantity of interest here is $\Delta G^\circ_{\mathrm{I}} - \Delta G^\circ_{\mathrm{II}}$ ($\Delta\Delta G^\circ$), which can be approximated for a given cyclic PIN variant by the difference between that variant's free energy of unfolding and that of capped PIN. Values derived from this approach are listed in Table 2.

To understand how the differences in $\Delta G^\circ_{\mathrm{I}} - \Delta G^\circ_{\mathrm{II}}$ arise, we note that $\Delta G^\circ_{\mathrm{I}} - \Delta G^\circ_{\mathrm{II}}$ can be described in terms of the cyclization thermodynamics of the native state (process III) and the cyclization thermodynamics of the unfolded state (process IV). It is known from studies concerning the cyclization of large rings that the entropy of cyclization is very weakly dependent on ring size for rings consisting of over 40 rotors.[41,42] Given that peptides generally have two rotors per residue, the series of PIN variants studied has 68−82 rotors. Therefore, the entropy of cyclization for the unfolded state does not vary significantly with linker length. The enthalpy of cyclization also should not vary with linker length since the same bond is formed in every case, and the strain energy for such large rings does not change with ring size. Because neither the entropy nor the enthalpy of cyclization for the unfolded state changes with linker length, the free energy of cyclization of the unfolded state ($\Delta G^\circ_{\mathrm{IV}}$) also does not vary with linker length. The differences observed in Table 2 must then be due to the differences in the native state ($\Delta G^\circ_{\mathrm{III}}$).

The native state influence that causes linker length dependent stability could be explained in several ways; here we propose a likely possibility. We note that when the linker length is shorter than the distance separating the N- and C-termini in the NMR-based structural ensemble, bringing the two ends together likely creates strain in the native state. On the other hand, when the linker is significantly longer than the distance between the termini observed in structural ensemble, the linker could force the termini apart, thereby distorting the structure. When the linker is in the range predicted to be optimal by the NMR-derived structural ensemble, the stability optimum can be achieved. In the absence of NMR information, the crystal structure could be used for optimizing linker length provided that a range of linker lengths encompassing the observed separation between the termini is explored.

Conclusions

Backbone cyclization can increase the thermodynamic stability of a protein. However, care must be taken to avoid introducing forces that destabilize the native state or remove forces that stabilize the native state. In the WW domain, cyclization removes favorable electrostatic interactions between its termini. However, this can be more than compensated for through cyclization (reducing the entropy of the unfolded states) by matching the linker length to the predicted distance between the N- and C-termini in the ensemble of structures derived from NMR data. Linkers that are too short or too long are unsuccessful from the perspective of stabilizing the protein relative to the linear folded sequence.

Table 3. Mass Spectrometry Data and Hydrodynamic-Based Molecular Weight Information Obtained from Sedimentation Equilibrium Analytical Ultracentrifugation Studies on the PIN WW Domain Sequences

polypeptide	expected mass	observed mass	hydrodynamic weight
cPINWW0G	4108.6	4109.3	4300 ± 300
cPINWW1G	4165.6	4166.3	4200 ± 200
cPINWW2G	4222.6	4222.8	4400 ± 200
cPINWW3G	4279.6	4280.3	4600 ± 90
cPINWW4G	4336.6	4337.3	4800 ± 200
cPINWW5G	4393.7	4398.3	5200 ± 300
cPINWW6G	4450.7	4451.3	5120 ± 90
wt PIN	4005.5	4006.4	3900 ± 200
capped PIN	4167.6	4168.3	3370 ± 90

Experimental Section

Materials and Methods. All Fmoc-protected amino acids, Boc−Cys(Trt)−OH, 4-sulfamybutyryl AM resin, Rink amide resin, 2-(1H-benzotriazol-1-yl)-1,1,3,3,-tetramethyluronium hexafluorophosphate (HBTU), 1-hydroxybenzotriazole (HOBt), and (1H-benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate (PyBOP) were purchased from Novabiochem. Diisopropylethylamine (DIEA), benzyl mercaptan, thiophenol, iodoacetronitrile, *tert*-butyl methyl ether (MTBE), *m*-cresol, thioanisole, and ethanedithiol (EDT) were purchased from Aldrich Chemical Co. Trifluoroacetic acid was purchased from Advanced Chemtech. Reagent grade dichloromethane (DCM), reagent grade *N*,*N*-dimethylformamide (DMF), and HPLC-grade acetonitrile were purchased from Fisher. 1-Methyl-2-pyrrolidine (NMP) and the rest of peptide synthesis reagents were purchased from Applied Biosystems.

Reverse-phase HPLC (RP-HPLC) was performed on a Waters HPLC system with model 600 pumps and model 486 and 2487 detectors with 214 and 280 nm UV detection using Vydac C_{18} column. The flow rate for analytical HPLC was 1 mL/min, while the flow rate for preparative HPLC was 10 mL/min. Acid buffers employed were buffer A (5% acetonitrile, 95% water, 0.1% TFA) and buffer B (95% acetonitrile, 5% water, 0.1% TFA). Basic buffers employed were buffer A (5% acetonitrile, 95% water, 0.1% 0.5 M ammonium bicarbonate) and buffer B (95% acetonitrile, 5% water, 0.1% 0.5 M ammonium bicarbonate). The purified peptides were identified by ESI-MS performed on a Hewlett-Packard LC-MS (MSD1100). The mass spectrometry data are summarized in Table 3.

All PIN WW domain variants were subjected to size exclusion chromatography, which was accomplished using a Pharmacia Model UPC900/P-920 FPLC. The stationary phase was Superdex 30. The eluent was 20 mM sodium phosphate (pH 7.0), 100 mM NaCl, and 1 mM DTT. The flow rate was 0.5 mL/min. The eluted peptide was collected and dialyzed against 20 mM sodium phosphate (pH 7.0) and 1 mM DTT. The concentrations of all PIN WW domain variants were determined by UV ($\epsilon_{280} = 13\,940\ \mathrm{M^{-1}\ cm^{-1}}$).

General Method for Solid-Phase Peptide Synthesis (SPPS). Automated solid-phase peptide synthesis was performed on an Applied Biosystems 433A peptide synthesizer. All syntheses were performed on a 0.1 mmol scale using the standard Fmoc-based FastMoc coupling chemistry provided by the system's software. Briefly, the coupling

(41) Mandolini, L. *Adv. Phys. Org. Chem.* **1986**, *22*, 1−111.
(42) Galli, C.; Mandolini, L. *Eur. J. Org. Chem.* **2000**, 3117−3125.

reactions were carried out in *N*-methylpyrrolidone (NMP) using 10 equiv of amino acid and the activating agents 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (10 equiv) and 1-hydroxybenzotriazole (10 equiv) in the presence of diisopropylethylamine (10 equiv). The N-terminal Fmoc deprotection was achieved using 20% piperidine in DMF for 20 min. For the syntheses of cyclic PIN variants, the last amino acid, Boc−Cys(Trt)−OH, was manually coupled for 30 min to the rest of the peptide using 10 equiv of amino acid, 10 equiv of HBTU, and 15 equiv of DIEA in 5 mL of DMF.

Resin Loading Fmoc-Gly-4-sulfamylbutyryl AM Resin. To a 50 mL round-bottom flask were added 4-sulfamylbutyryl AM resin (500 mg, 0.45 mmol), DMF (20 mL), DIEA (1.8 mL, 10.1 mmol), and Fmoc−Gly−OH (1.34 g, 4.5 mmol). The reaction mixture was stirred at room temperature for 10 min before cooling to −20 °C. After 20 min, PyBOP (1.76 g, 3.4 mmol) was added to the mixture. After stirring at −20 °C for 8 h, the mixture was allowed to react at ambient temperature for 16 h. Afterward, the resin was filtered and washed with DMF and DCM. The coupling was repeated. The percent loading was determined using a standard Fmoc quantification method.[30]

Activation and Cleavage. The resin-bound peptide was prepared for activation and cleavage by the addition of NMP (4 mL) and DIEA (0.2 mL, 1.1 mmol). Iodoacetonitrile (0.18 mL, 2.5 mmol), prefiltered through a plug of basic alumina, was added to the mixture with the exclusion of light. The resin was agitated for 24 h, filtered, and washed with NMP (5 × 10 min), DCM (3 × 1 min), and DMF (3 × 1 min). The peptide was cleaved from the resin using 10% benzylmercaptan in DMF overnight. The resin was washed with DCM (3 × 1 min, 5 mL portion). All the filtrate washes were combined.

Side-Chain Deprotection. The side-chain protecting groups on the crude peptide were removed with 1 mL of reagent K per 0.025 mmol of crude peptide (reagent K: 82.5% TFA, 5% *m*-cresol, 5% thioanisole, 5% water, and 2.5% EDT) for 4 h at room temperature.[32] Crude peptide was precipitated in *tert*-butyl methyl ether, centrifuged, redissolved in DMF, and purified by RP-HPLC in acidic buffer using a 0−100% linear gradient in solvent B over 45 min).

Native Chemical Ligation. A solution of linear peptide thioester (0.25 mM) in 10 mM sodium phosphate (pH 6.0) and 1 mM DTT (5 mL) was slowly added to a 20 mL solution of 125 mM sodium phosphate (pH 7.7), 1 mM EDTA, 1 mM DTT, 0.05% (v/v) of thiophenol, and 0.05% (v/v) of benzylmercaptan. Slow addition was accomplished using syringe pump over a period of 24 h (3.5 μL/min). The resulting mixture was concentrated to 10 mL using a Millipore Centriprep membrane (MW 3000 cutoff). The cyclic peptide was purified using RP-HPLC in acidic buffer with a linear gradient of 0−80% solvent B over 45 min.

Capped PIN WW Domain. An N- and C-capped PIN WW domain (linear) was synthesized on Rink Amide resin. The first residue, Fmoc−Gly−OH, was manually loaded onto the resin (0.1 mmol) using 10 equiv of amino acid, 10 equiv of HOBt, 10 equiv of HBTU, and 20 equiv of DIEA in NMP for 24 h. After Fmoc quantification, the loaded resin was capped with acetic anhydride (80 mg of HOBt, 0.9 mL of DIEA, 1.9 mL of acetic anhydride in 10 mL of NMP for 24 h). The rest of the sequence was assembled by automated chain elongation. The acetyl group was introduced afterward by treating the immobilized peptide with acetic anhydride (80 mg of HOBt, 0.9 mL of DIEA, 1.9 mL of acetic anhydride in 10 mL of NMP for 24 h). The peptide was cleaved, deprotected, and purified by the aforementioned methods.

YSPTpSPS (CTD-S5). Fmoc−Ser(*t*Bu)−OH was preloaded onto Rink amide resin and capped with acetic anhydride by the aforementioned methods. The rest of the amino acids, including phosphoserine (pS), were coupled by automated chain elongation. The peptide was cleaved from the resin and deprotected by agitating the resin in 3 mL of reagent K for 2 h. After precipitation with MTBE, the peptide was purified by RP-HPLC in basic buffer employing a linear gradient of 0−20% solvent B over 30 min. *m*/*z* calcd was 815.3; found 816.3 (ESI-MS). The peptide was stored in 15 mM HEPES buffer (pH 7.7) at 4 °C.

Tetramethylrhodamine-Labeled YSPTpSPS (CTD-S5). The peptide was labeled with the amine-reactive reagent tetremethylrhodamine-5-(and-6)-isothiocyanate using a 2:1 molar ratio of rhodamine to peptide in 0.1 M sodium bicarbonate (pH 9.0) at ambient temperature over 10 h. Peptide was purified by RP-HPLC in basic buffer employing a linear gradient of 0−100% solvent B over 30 min. *m*/*z* calcd was 1261.2; found 1262.3 (ESI-MS). The peptide was stored in 15 mM Na-HEPES buffer (pH 7.7) at 4 °C. The concentration of labeled peptide was determined by UV absorption (ϵ_{544} = 84 000 M^{-1} cm^{-1}).

CD Studies. CD spectra were recorded using an AVIV model 202SF stopped flow circular dichroism spectrometer equipped with Peltier temperature-controlled cell holder using a 0.2 cm path length Suprasil quartz cell (Hellma, Forest Hills, NY). Far-UV CD spectra were recorded from 200 to 250 nm at 2 and 25 °C. The wavelength step size was 0.5 nm, and the averaging time was 2 s per scan at each wavelength step. The peptide sample was dissolved in 20 mM sodium phosphate (pH 7.0), 1 mM DTT (25 μM). Peptides in 20 mM buffers at various pH's (glycine-HCl for pH 2−3.6, sodium acetate for pH 3.6−5.0, and sodium phosphate for pH 5−7) were subjected to thermal denaturation studies. Thermal denaturation was monitored at 227 nm. The temperature range was from 2 to 98 °C with a 2 °C step and a 90 s equilibration time. Each data point was averaged for 30 s. After the highest temperature was reached, the sample was cooled to 25 °C, and another wavelength scan was taken. Fraction unfolding (f_u) was determined using the baseline extrapolation method.[43,44] T_m was determined from the fraction unfolding curve (T_m = temperature at which f_u = 0.5).

From fitting the baselines of the unfolding curves (ellipticity [Θ, mdeg] vs temperature [K]) with linear functions (m = slope of the curve), the following baseline eqs 3 and 4 could be obtained for native and unfolded PIN WW domain variants, respectively.

$$\Theta_N(T) = \Theta_N^\circ + m_N T \tag{3}$$

$$\Theta_U(T) = \Theta_U^\circ + m_U T \tag{4}$$

Thus, f_U could be obtained from normalization of the unfolding curve.

$$f_U(T) = [\Theta_N(T) - \Theta(T)]/[\Theta_N(T) - \Theta_U(T)] \tag{5}$$

The fraction unfolding curves for all the variants as well as the relative T_m of all variants with respect to wt PIN are shown in Figure 6.

Calculation of Thermodynamic Parameters. Assuming that equilibrium between the native (N) and unfolded (U) states was established at each temperature step during thermal denaturation,

$$N \leftrightarrow U \tag{6}$$

the equilibrium constant for unfolding can be expressed as

$$K = f_U/f_N = f_U/(1 - f_U) \tag{7}$$

where f_U is fraction unfolded, and f_N is fraction native.

According to the van't Hoff equation,

$$d(\ln K)/d(1/T) = -\Delta H_m/R \tag{8}$$

thus

$$\ln K = -(\Delta H_m/R) \times (1/T) + c \tag{9}$$

or

(43) Gursky, O.; Atkinson, D. *Biochemistry* **1998**, *37*, 1283−1291.
(44) Sinha, A.; Yadav, S.; Ahmad, R.; Ahmad, F. *Biochem. J.* **2000**, *345*, 711−717.

$$\ln [f_u/(1 - f_u)] = -(\Delta H_m/R) \times (1/T) + c \quad (10)$$

For each variant, stability was modulated by changing pH. At each pH, T_m and ΔH_m were calculated. By plotting T_m versus ΔH_m, $\Delta C_{p,u}$, the heat capacity of unfolding, could be obtained from the slope of the line of best fit:

$$d\Delta H/dT = \Delta C_p \quad (11)$$

thus

$$\Delta H_m = \Delta C_{p,u} T_m + c \quad (12)$$

Because the hydrophobic surface area exposed upon unfolding should be very similar for all the WW domains studied herein, they should all have the same value of $\Delta C_{p,u}$. Instead of fitting individual plots of T_m versus ΔH_m for each WW sequence to eq 12, thereby obtaining values of $\Delta C_{p,u}$ and c for each variant, the plots of T_m versus ΔH_m for all the WW variants were fitted simultaneously to eq 12 to obtain a common value of $\Delta C_{p,u}$. Only the values of c differed among WW sequences. With $\Delta C_{p,u}$, $\Delta H°$, $\Delta S°$, and $\Delta G°$ for the unfolding process could be obtained as follows:

$$\Delta H°(T) = \Delta H_m + \Delta C_{p,u} (T - T_m) \quad (13)$$

$$\Delta S°(T) = \Delta H_m/T_m + \Delta C_{p,u} \ln(T/T_m) \quad (14)$$

$$\Delta G°(T) = \Delta H°(T) - T*\Delta S°(T) \quad (15)$$

Analytical Ultracentrifugation. The monomeric nature of the samples was confirmed by sedimentation equilibrium measurements carried out employing a temperature-controlled Beckman XL-I Analytical Ultracentrifuge equipped with an An-60 Ti rotor and a photoelectric scanner (Beckman Instrument Inc., Palo Alto, CA). Protein samples were loaded in a double sector cell equipped with a 12 mm Epon centerpiece and a sapphire optical window. The reference compartment was loaded with the matching 20 mM sodium phosphate buffer with 1 mM DTT (140 μL). The samples were monitored at 280 nm employing a rotor speed of 3000−40 000 rpm at 20 °C and analyzed by a nonlinear squares approach using Origin software (Microcal Software Inc., Northampton, MA). The data were fitted to the Lamm equation for a single species model

$$A_r = A_0 \exp[\omega^2/2RT*M(1 - \bar{v}\rho)](r^2 - r_0^2) \quad (16)$$

where A_r is radial absorbance, A_0 is the baseline absorbance, ω is the rotor speed (s^{-1}), R is the gas constant (J mol^{-1} K^{-1}), T is the temperature (K), υ is the partial specific volume (mL g^{-1}), ρ is the density of solvent (g mL^{-1}), r is the variable radius, and r_0 is the meniscus radius. The hydrodynamic masses of all the peptides synthesized in this study are summarized in Table 3.

Ligand Binding Assay. Fluorescence anisotropies of the samples were measured on an Aviv model ATF105 automated differential/ratio spectrofluorometer. Samples composed of 0, 10, 25, 50, and 100 μM PIN WW domain variants having a constant ligand concentration (10 μM) in 20 mM sodium phosphate (pH 7.0) 1 mM DTT were evaluated. An excitation wavelength of 544 nm (1 nm bandwidth) was employed, whereas emission was detected at 571 nm (4 nm bandwidth). Anisotropy versus peptide concentration was fitted to the following equation to obtain K_d:

$$r = r_f + (r_b - r_f)((K_d + [\text{peptide}] + [\text{ligand}]) - (K_d + [\text{peptide}] + [\text{ligand}])^2 - 4[\text{peptide}][\text{ligand}])^{1/2}/2[\text{peptide}] \quad (17)$$

where r is anisotropy, r_f is anisotropy of free ligand, and r_b is anisotropy of bound ligand.

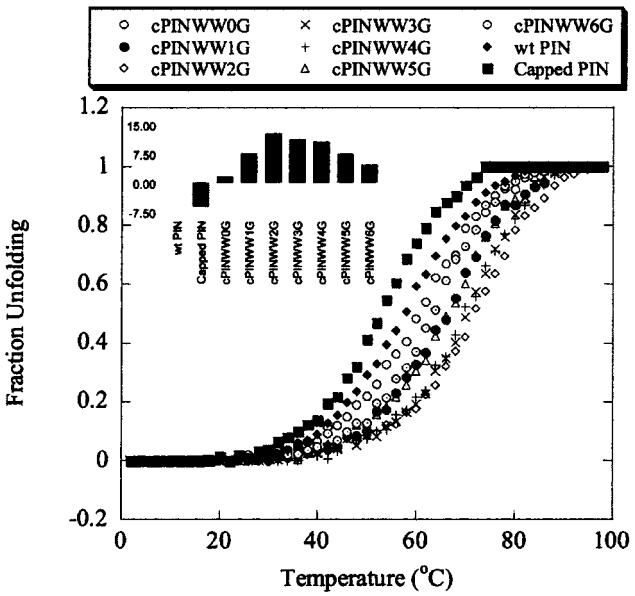


Figure 6. Thermal denaturation curves showing the fraction PIN WW domain variants unfolded as a function of temperature monitored by far-UV CD. The protein sample (25 μM) in 20 mM sodium phosphate buffer (pH 7.0; 1 mM DTT) was heated from 2 to 98 °C with a 90 s equilibration time at each new temperature. A profile of the difference in midpoint temperature (ΔT_m) for each of the variants with respect to T_m of wt PIN is shown in the inset.

Acknowledgment. We gratefully acknowledge financial support from NIH (GM 51105), The Skaggs Institute of Chemical Biology, The Lita Annenberg Hazen Foundation, and scientific advice from Professor Ernesto Freire (John Hopkins University), Professor Evan T. Powers, Professor Philip E. Dawson, and Dr. Marcus Jäger.

JA0123608