Polypeptide Synthesis and Characterization

Materials

N-Fmoc and side chain-protected amino acids, Fmoc-(DMB)Gly-OH, Fmoc-(S,S)-ACPC-OH, Fmoc-(R,R)-ACPC-OH, 2-(6-chloro-1H-benzotriazole-1-yl)-1,1,3,3- tetramethyluronium hexafluorophosphate (HCTU), and 6-chloro-1-hydroxybenzotriazole (Cl-HOBt) were purchased from ChemImpex. β^3-amino acids were purchased from PepTech. NovaPEG Wang resin and 1-(mesitylene-2-sulfonyl)-3-nitro-1,2,4-triazole (MSNT) were purchased from Novabiochem. N,N-dimethylformamide (DMF), methylene chloride (DCM), N,N-diisopropylethylamine (DIEA), trifluoroacetic acid (TFA), N-methylimidazole and acetonitrile (MeCN) were purchased from Sigma-Aldrich. Fritted syringes were purchased from Torviq.

Solid-Phase Synthesis

NovaPEG Wang resin was preloaded with the C-terminal residue (Gly) of the WW domain sequence, to afford a carboxylic acid upon cleavage. The resin loading procedure used 5 equivalents of (MSNT), 5 equivalents of the Fmoc-protected amino acid and 3.75 equivalents of N-methylimidazole in DCM. This reaction was carried out for at least 4 hours (generally overnight), the resin was washed extensively with DMF and then DCM, and the resin was subjected to the same reaction conditions again. Following MSNT coupling of the C-terminal amino acid, the resin was washed with DCM and dried on a vacuum manifold.

Peptides were synthesized using microwave irradiation (CEM MARS), or on a Symphony automated synthesizer (Protein Technologies, Inc.) at the UW-Madison Biotech Center peptide synthesis facility. Coupling reactions carried out using microwave irradiation involved a two-minute ramp to 70 °C and a four-minute hold at that temperature, using

four equivalents of Fmoc-protected amino acid, four equivalents of HCTU, four equivalents of Cl-HOBt and eight equivalents of DIEA, with DMF as the solvent. Coupling reaction volumes were approximately 2 mL per 50 μmol resin (solid-phase syntheses were generally carried out on 50 or 100 μmol scale). Fmoc deprotections carried out using microwave irradiation involved a two-minute ramp to 80 °C followed by a two-minute hold at that temperature, using an excess of 20% piperidine in DMF (generally ~2 mL per 25 μmol resin). After each coupling/deprotection cycle the resin was washed 3X with DMF. Upon completion of the synthesis the resin was washed with DCM and dried prior to cleavage and global deprotection.

Couplings carried out by automated synthesizer (30-60 minute reaction times) used five equivalents of Fmoc-protected amino acid, five equivalents of HBTU and 20 equivalents of NMM, with DMF as the solvent. Deprotection steps used 20% piperidine in DMF for 20 min.

For all peptide sequences reported here, the N-terminal three amino acids (Lys-Leu-Pro) were "double-coupled" – i.e., the resin was charged with activated monomer twice in succession, without intermediate washing, before removal of the Fmoc protecting group in piperidine/DMF solution.

In a preliminary attempt to synthesize variant 2, we observed an impurity *via* HPLC that corresponded to an *m*/*z* value 18 mass units less than the expected mass. This impurity was attributed to aspartimide formation from the Asp-Gly dipeptide unit. To circumvent potential issues arising from aspartimide formation, we incorporated the glycine residue of some Pin1 WW domain sequences based on 2 using a commercially available Fmoc-(DMB)Gly-OH monomer, which is not capable of aspartimide formation. To increase

product yield of peptides synthesized using Fmoc-(DMB)Gly-OH, the next amino acid in the sequence was typically double-coupled at 70 °C, and for 8 minutes per reaction rather than 4 minutes. Following the coupling reactions, the resin was treated with a solution of DMF/DIEA/Ac_2O (8:2:1; approximately 1.1 mL per 25 μmol resin loading) for at least five minutes to cap unreacted N-termini. This protocol was intended to suppress the formation of single-deletion products resulting from unreacted DMB-functionalized N-termini, as these deletion products can complicate HPLC purification of the desired product from the crude mixture. This general procedure was used for the synthesis of sequences 2, 2(A17[S,S]ACPC) and 2(A17β^3A). Sequence 2(G19β^3G) was generated by completing all synthetic steps, including and following Fmoc-Asp deprotection, at room temperature on an automated synthesizer.

Peptide Cleavage

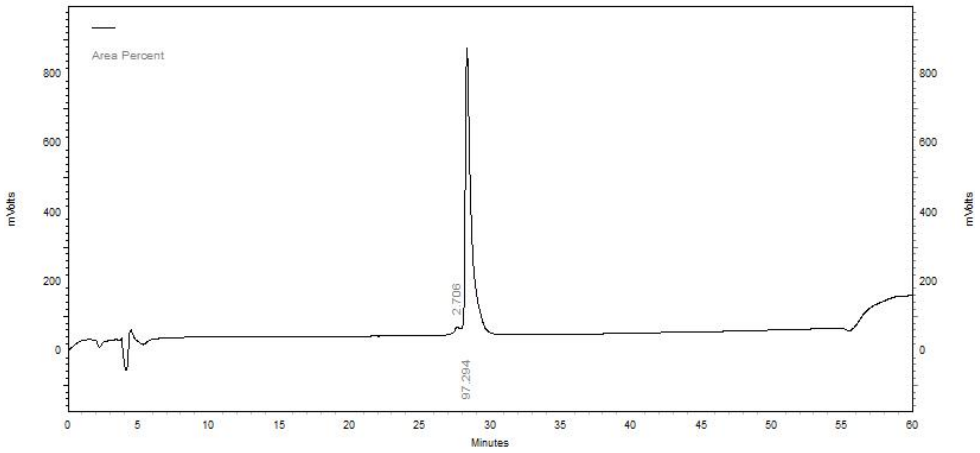
Polypeptides were cleaved from resin and globally side chain deprotected usingRe-agent K conditions [82.5% TFA, 5% thioanisole, 5% water, 5% phenol, 2.5% ethanedithiol (EDT) v/v] to minimize oxidation of methionine side chains to sulfoxides. After drying of resin by passing a stream of air through the fritted syringe, cleavage reactions were carried out for 2-3 h by suspending the resin in the cleavage cocktail and incubating the reaction on a mechanical rocker. A volume of 2 mL cleavage cocktail was used per 25 μmol resin loading. Following the cleavage reaction, TFA solutions were drained into 50 mL tubes, and the resin was washed 2X with several milliliters of TFA. All TFA solutions were com-bined and reduced to ~50 % volume under a stream of nitrogen gas or air. Peptides were precipitated from the TFA solution by addition of cold diethyl ether, then collected by

centrifugation and decanting of the ether. Pellets were then washed with an additional volume of cold ether and collected again via centrifugation. Peptide pellets were dried under nitrogen to remove residual diethyl ether prior to high performance liquid chromatographic (HPLC) purification.

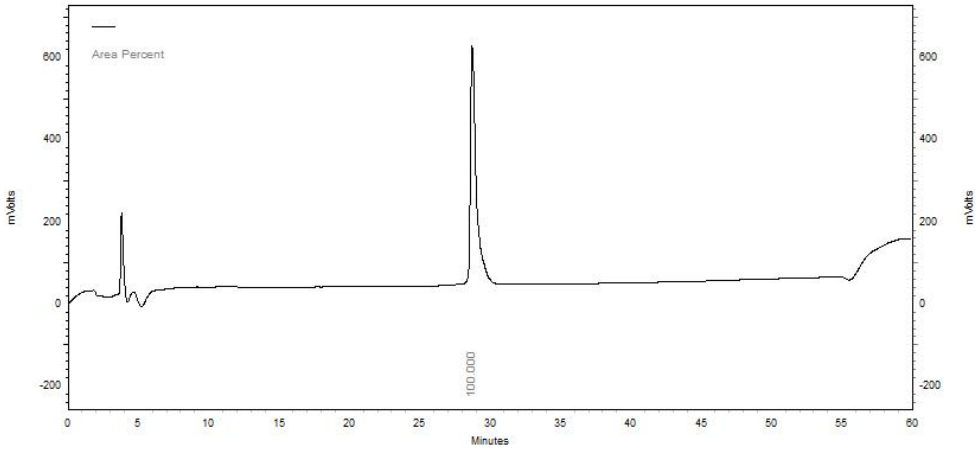
WW Domain Variant Characterization

All peptides were purified on a Shimadzu SCL-10A liquid chromatograph fitted with either a C_5- or C_{18}-functionalized reverse-phase column. The binary solvent system used in purifications used H_2O:TFA (100:0.1 v/v) as A solvent and MeCN:TFA (100:0.1 v/v) as B solvent. Following purification, fractions were pooled, diluted into an equal volume of H_2O, and lyophilized to dryness. Polypeptide identity was established using matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry. Calculated $[M+H]^+$ ions are given below, assuming average mass values unless otherwise noted. Note that peak integration values in the chromatograms shown below should be disregarded.

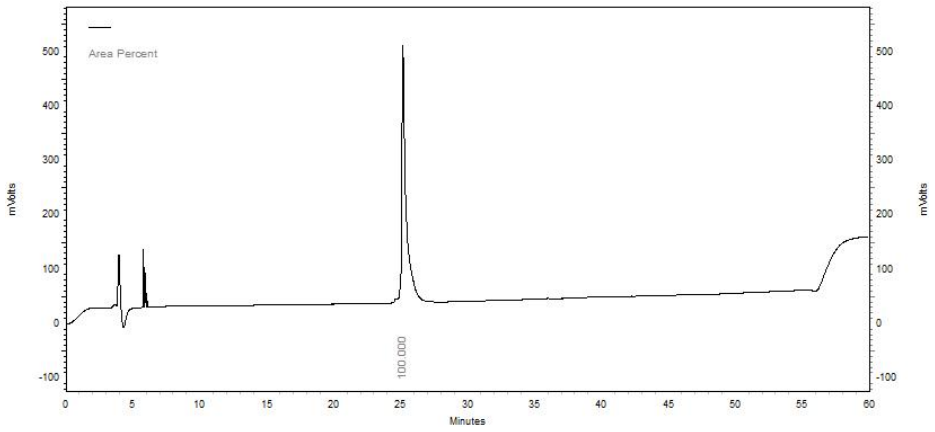
WW(4:6) **1**: C_{18} analytical column (250 X 4.6 mm, Supelco), flow rate 1 mL/min, gradient 10-60% B solvent (MeCN:TFA, 100:0.1 v/v) in A (H_2O:TFA, 100:0.1 v/v) over 50 minutes, retention time of 28 minutes. MALDI-TOF [M+H]+ calculated 4024.5, observed 4024.2.



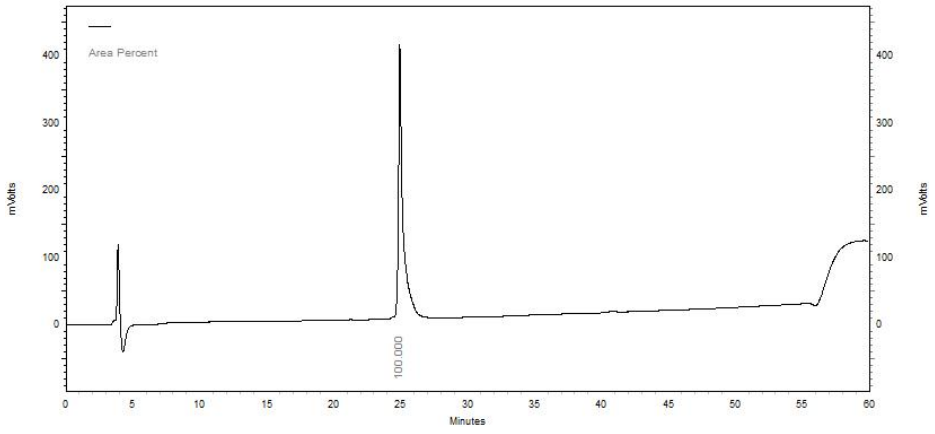
1(R17[S,S]ACPC): C_{18} analytical column (250 X 4.6 mm, Supelco), flow rate 1 mL/min, gradient 10-60% B solvent (MeCN:TFA, 100:0.1 v/v) in A (H_2O:TFA, 100:0.1 v/v) over 50 minutes, retention time of 29 minutes. MALDI-TOF [M+H]+ calculated 3979.3, observed 3977.7.



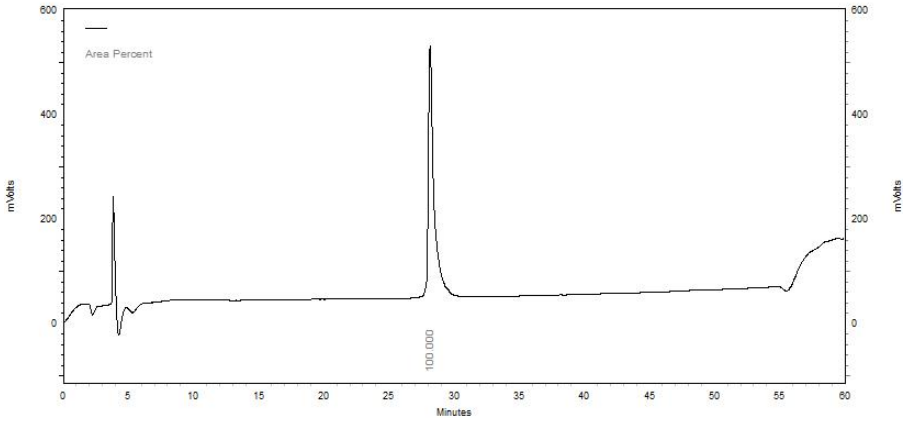
1(S18[S,S]ACPC): C_{18} analytical column (250 X 4.6 mm, Supelco), flow rate 1 mL/min, gradient 10-60% B solvent (MeCN:TFA, 100:0.1 v/v) in A (H_2O:TFA, 100:0.1 v/v) over 50 minutes, retention time of 25 minutes. MALDI-TOF [M+H]+ calculated 4048.4, observed 4047.2.



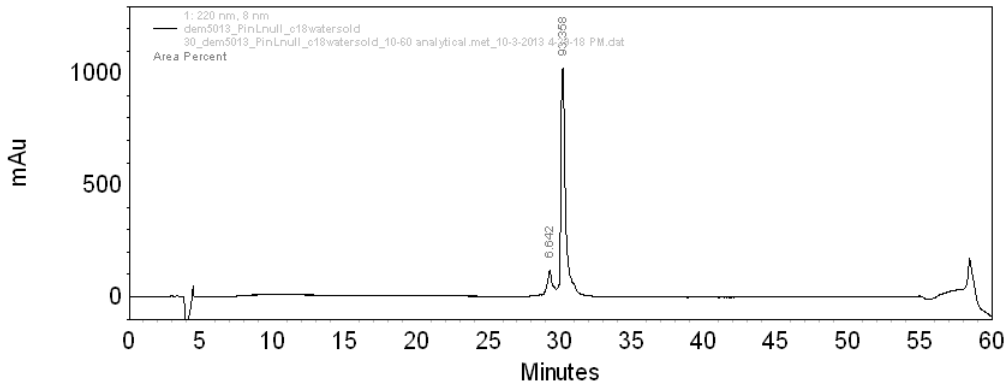
1(S19[S,S]ACPC): C_{18} analytical column (250 X 4.6 mm, Supelco), flow rate 1 mL/min, gradient 10-60% B solvent (MeCN:TFA, 100:0.1 v/v) in A (H_2O:TFA, 100:0.1 v/v) over 50 minutes, retention time of 25 minutes. MALDI-TOF [M+H]+ calculated 4048.4, observed 4050.3.



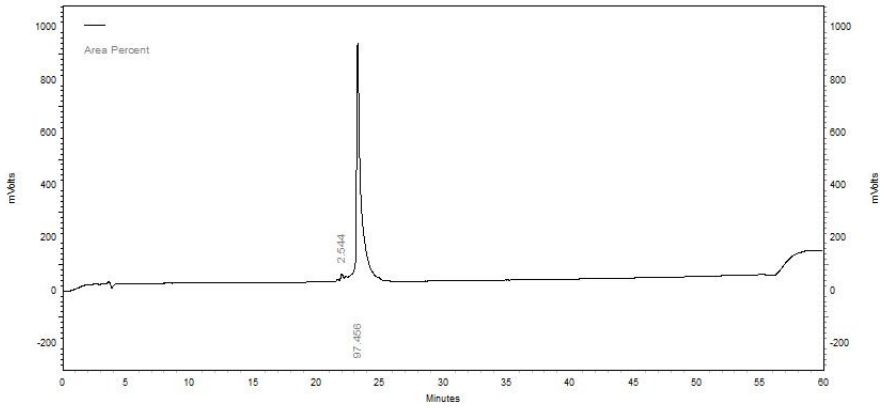
1(G20[R,R]ACPC): C_{18} analytical column (250 X 4.6 mm, Supelco), flow rate 1 mL/min, gradient 10-60% B solvent (MeCN:TFA, 100:0.1 v/v) in A (H_2O:TFA, 100:0.1 v/v) over 50 minutes, retention time of 28 minutes. MALDI-TOF $[M+H]^+_{monoisotopic}$ calculated 4076.1, observed 4076.9.



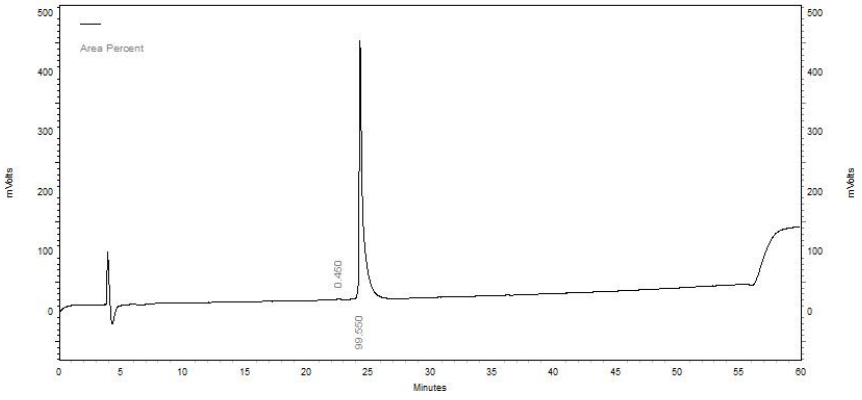
1(G20[S,S]ACPC): C_{18} analytical column (250 X 4.6 mm, Supelco), flow rate 1 mL/min, gradient 10-60% B solvent (MeCN:TFA, 100:0.1 v/v) in A (H_2O:TFA, 100:0.1 v/v) over 50 minutes, retention time of 30 minutes. MALDI-TOF $[M+H]^+_{monoisotopic}$ calculated 4076.1, observed 4076.3.



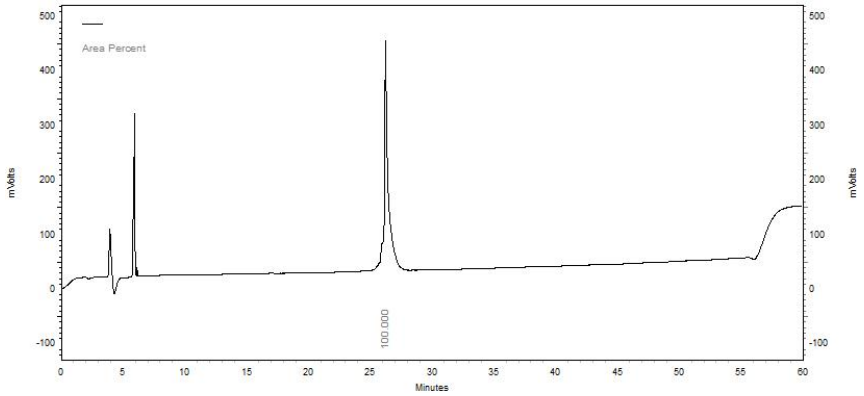
WW(3:5) **2**: C_{18} analytical column (250 X 4.6 mm, Supelco), flow rate 1 mL/min, gradient 10-60% B solvent (MeCN:TFA, 100:0.1 v/v) in A (H_2O:TFA, 100:0.1 v/v) over 50 minutes, retention time of 23 minutes. MALDI-TOF [M+H]+ calculated 3880.3, observed 3879.8.



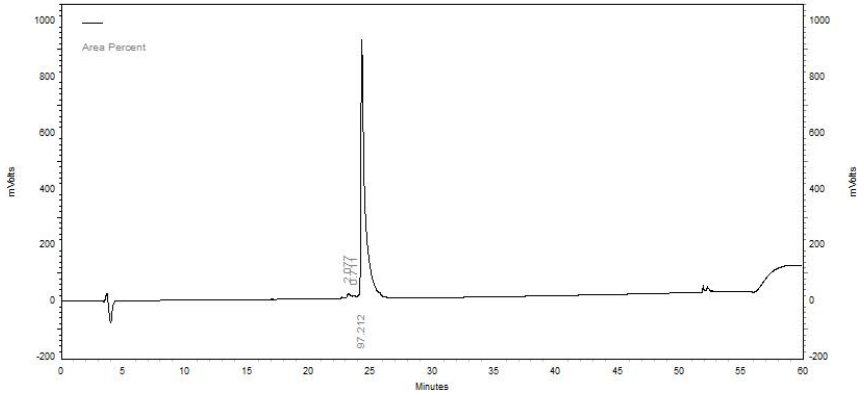
2(A17[S,S]ACPC): C_{18} analytical column (250 X 4.6 mm, Supelco), flow rate 1 mL/min, gradient 10-60% B solvent (MeCN:TFA, 100:0.1 v/v) in A (H_2O:TFA, 100:0.1 v/v) over 50 minutes, retention time of 24 minutes. MALDI-TOF [M+H]+ calculated 3920.2, observed 3919.0.



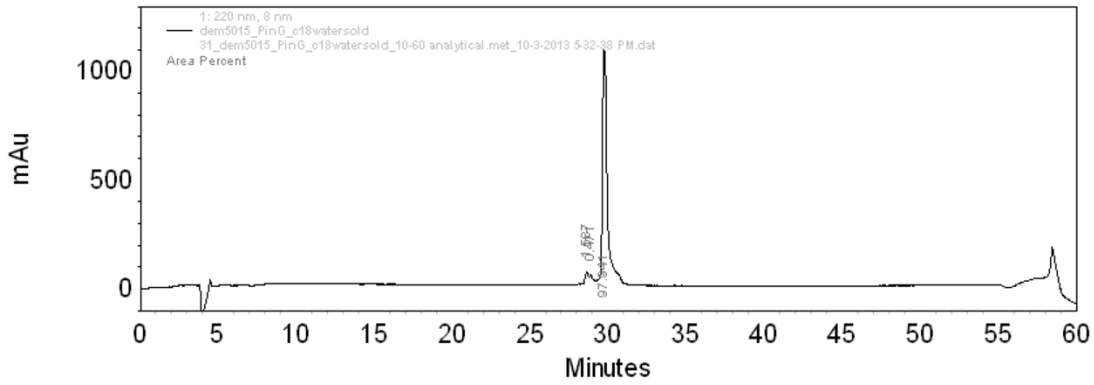
2(D18[S,S]ACPC): C_{18} analytical column (250 X 4.6 mm, Supelco), flow rate 1 mL/min, gradient 10-60% B solvent (MeCN:TFA, 100:0.1 v/v) in A (H_2O:TFA, 100:0.1 v/v) over 50 minutes, retention time of 26 minutes. MALDI-TOF [M+H]+ calculated 3876.2, observed 3876.4.



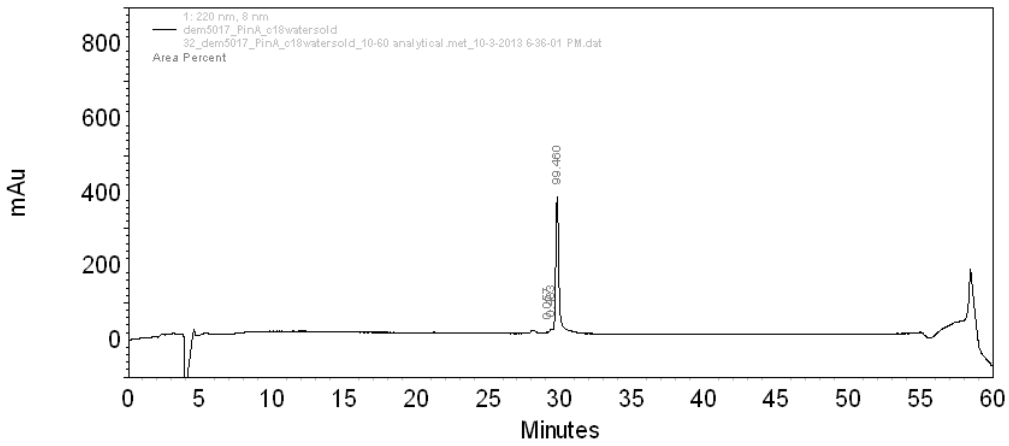
2(G19[R,R]ACPC): C_{18} analytical column (250 X 4.6 mm, Supelco), flow rate 1 mL/min, gradient 10-60% B solvent (MeCN:TFA, 100:0.1 v/v) in A (H_2O:TFA, 100:0.1 v/v) over 50 minutes, retention time of 24 minutes. MALDI-TOF [M+H]+ calculated 3934.3, observed 3935.8.



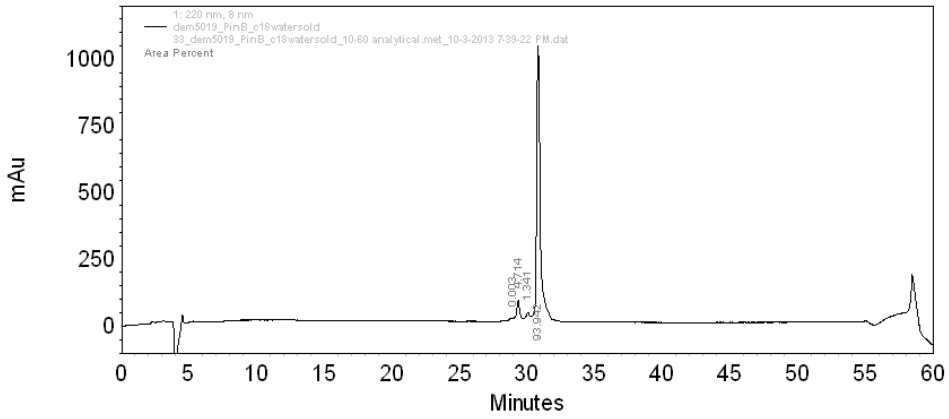
WW(2:2) **3**: C_{18} analytical column (250 X 4.6 mm, Waters), flow rate 1 mL/min, gradient 10-60% B solvent (MeCN:TFA, 100:0.1 v/v) in A (H_2O:TFA, 100:0.1 v/v) over 50 minutes, retention time of 30 minutes. MALDI-TOF [M+H]+ calculated 3808.2, observed 3808.1.



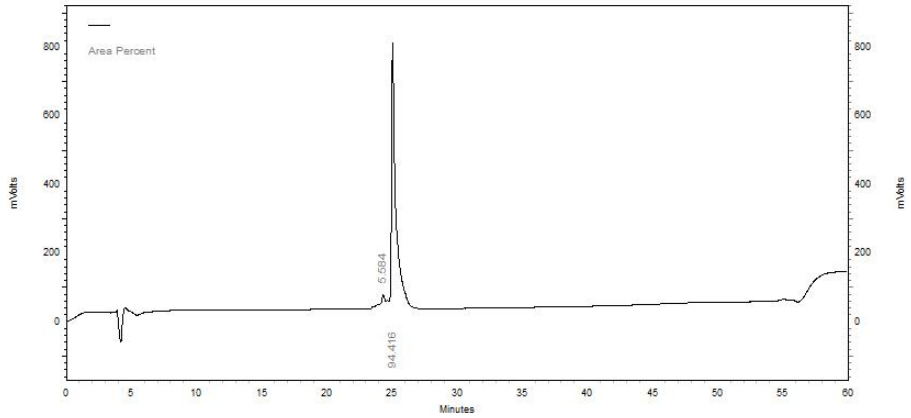
3(N17[R,R]ACPC): C_{18} analytical column (250 X 4.6 mm, Waters), flow rate 1 mL/min, gradient 10-60% B solvent (MeCN:TFA, 100:0.1 v/v) in A (H_2O:TFA, 100:0.1 v/v) over 50 minutes, retention time of 30 minutes. MALDI-TOF [M+H]+ calculated 3805.1, observed 3805.7.



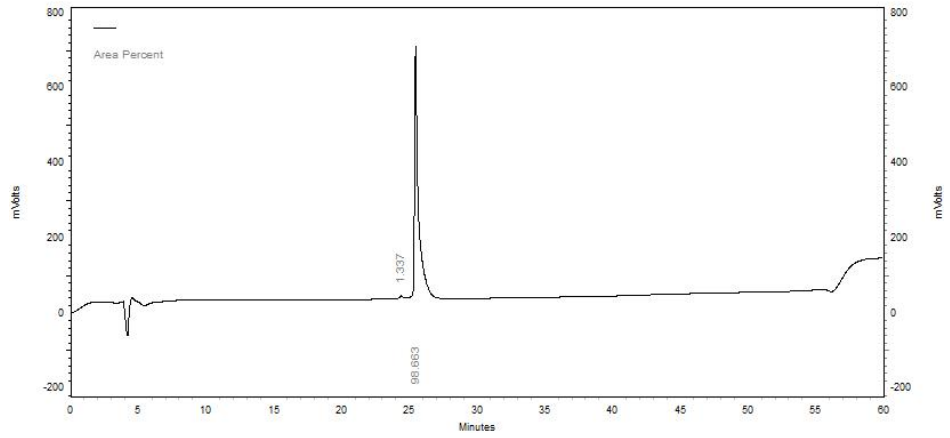
3(G18[R,R]ACPC): C_{18} analytical column (250 X 4.6 mm, Waters), flow rate 1 mL/min, gradient 10-60% B solvent (MeCN:TFA, 100:0.1 v/v) in A (H_2O:TFA, 100:0.1 v/v) over 50 minutes, retention time of 31 minutes. MALDI-TOF [M+H]+ calculated 3862.2, observed 3860.9. *Note that on page 239 of the Ph.D. Thesis of David E. Mortenson (UW-Madison, 2014), due to a technical error introduced during formatting, the chromatogram shown below was inadvertently obstructed, leading to the impurity peak at a retention time of ~29 minutes being partially hidden. The correct chromatogram is shown below.



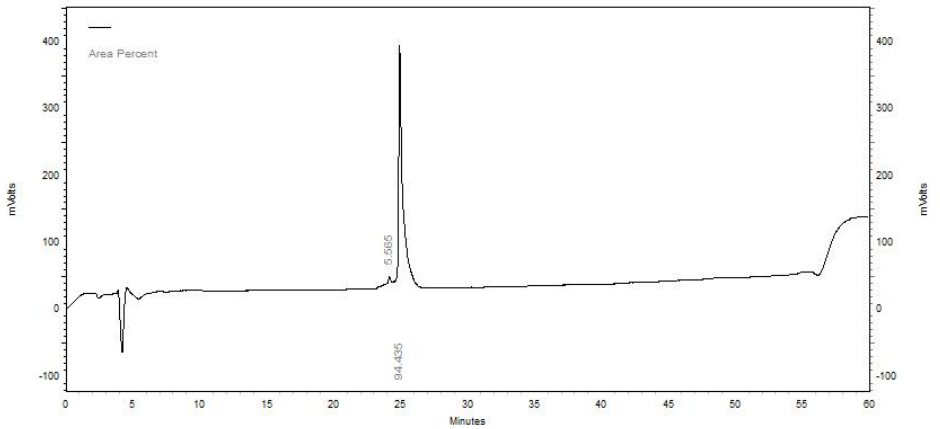
1(R17β^3R): C_{18} analytical column (250 X 4.6 mm, Supelco), flow rate 1 mL/min, gradient 10-60% B solvent (MeCN:TFA, 100:0.1 v/v) in A (H_2O:TFA, 100:0.1 v/v) over 50 minutes, retention time of 25 minutes. MALDI-TOF [M+H]+ calculated 4038.5, observed 4038.8.



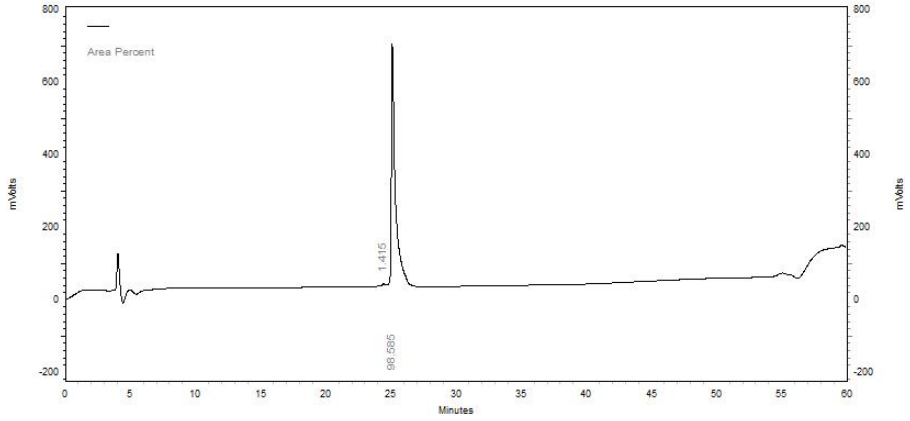
1(S18β^3S): C_{18} analytical column (250 X 4.6 mm, Supelco), flow rate 1 mL/min, gradient 10-60% B solvent (MeCN:TFA, 100:0.1 v/v) in A (H_2O:TFA, 100:0.1 v/v) over 50 minutes, retention time of 26 minutes. MALDI-TOF [M+H]+ calculated 4038.5, observed 4038.6.



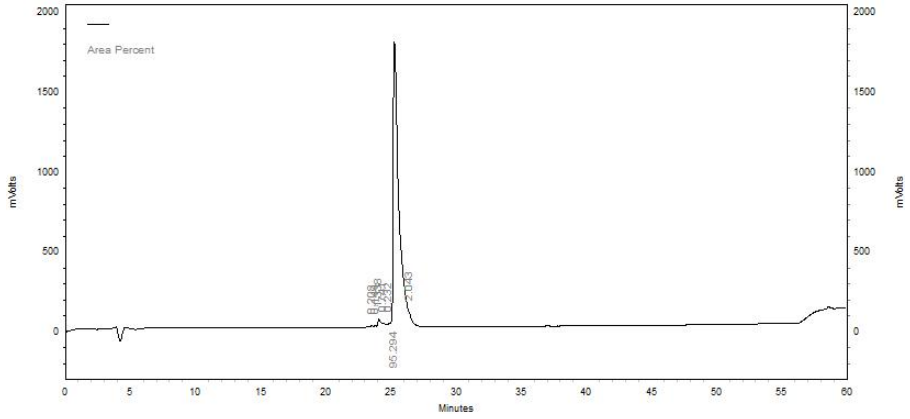
1(S19β^3S): C_{18} analytical column (250 X 4.6 mm, Supelco), flow rate 1 mL/min, gradient 10-60% B solvent (MeCN:TFA, 100:0.1 v/v) in A (H_2O:TFA, 100:0.1 v/v) over 50 minutes, retention time of 25 minutes. MALDI-TOF [M+H]+ calculated 4038.5, observed 4038.8.



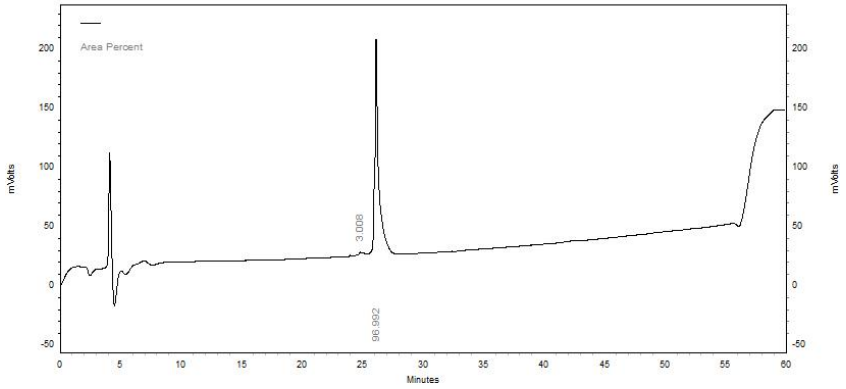
1(G20βG): C_{18} analytical column (250 X 4.6 mm, Supelco), flow rate 1 mL/min, gradient 10-60% B solvent (MeCN:TFA, 100:0.1 v/v) in A (H_2O:TFA, 100:0.1 v/v) over 50 minutes, retention time of 25 minutes. MALDI-TOF [M+H]+ calculated 4038.5, observed 4038.1.



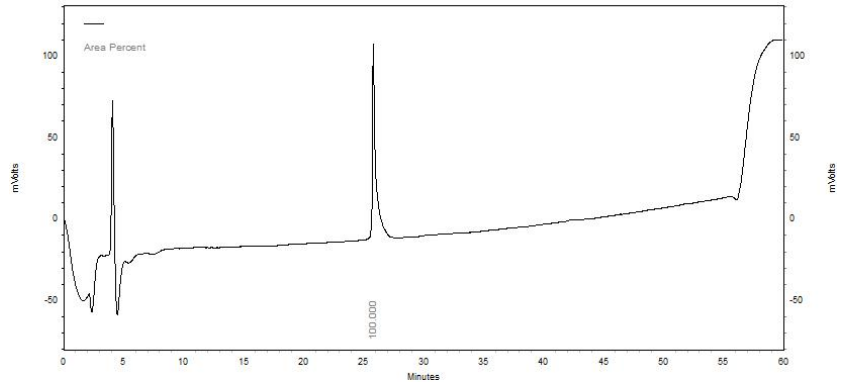
2(A17β^{3}A): C_{18} analytical column (250 X 4.6 mm, Supelco), flow rate 1 mL/min, gradient 10-60% B solvent (MeCN:TFA, 100:0.1 v/v) in A (H_2O:TFA, 100:0.1 v/v) over 50 minutes, retention time of 25 minutes. MALDI-TOF [M+H]+ calculated 3894.3, observed 3893.1.



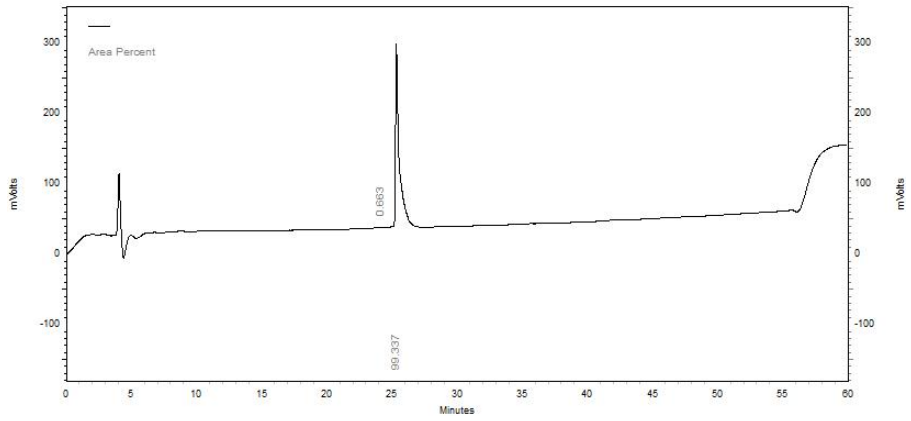
2(D18β[3]D): C_{18} analytical column (250 X 4.6 mm, Supelco), flow rate 1 mL/min, gradient 10-60% B solvent (MeCN:TFA, 100:0.1 v/v) in A (H_2O:TFA, 100:0.1 v/v) over 50 minutes, retention time of 26 minutes. MALDI-TOF [M+H]+ calculated 3894.3, observed 3893.9.



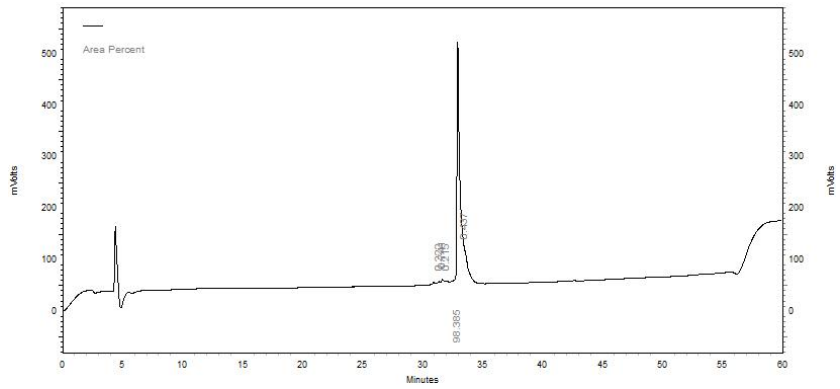
2(G19βG): C_{18} analytical column (250 X 4.6 mm, Supelco), flow rate 1 mL/min, gradient 10-60% B solvent (MeCN:TFA, 100:0.1 v/v) in A (H_2O:TFA, 100:0.1 v/v) over 50 minutes, retention time of 26 minutes. MALDI-TOF [M+H]+ calculated 3894.3, observed 3892.1.



3(N17β^{3}N): C_{18} analytical column (250 X 4.6 mm, Supelco), flow rate 1 mL/min, gradient 10-60% B solvent (MeCN:TFA, 100:0.1 v/v) in A (H_2O:TFA, 100:0.1 v/v) over 50 minutes, retention time of 25 minutes. MALDI-TOF [M+H]+ calculated 3822.2, observed 3820.0.



3(G18βG): C_{18} analytical column (250 X 4.6 mm, Supelco), flow rate 1 mL/min, gradient 10-60% B solvent (MeCN:TFA, 100:0.1 v/v) in A (H_2O:TFA, 100:0.1 v/v) over 50 minutes, retention time of 33 minutes*. MALDI-TOF [M+H]+ calculated 3822.2, observed 3823.5. *The late retention time of this WW domain variant can be attributed to variability in HPLC performance (i.e., to inefficient solvent delivery by the B solvent pump).



Circular Dichroism Experiments

Wavelength scan and thermal denaturation data were collected on an Aviv model 420 spectropolarimeter (Aviv Biomedical) fitted with a 5-cell sample changer. For each Pin1 WW domain variant, CD data were collected from aliquots of a freshly prepared sample divided among three 0.1 mm pathlength quartz cuvettes (Hellma).

Buffer used for circular dichroism studies was prepared by diluting a 100 mM sodium phosphate (pH 7.0) stock by 10-fold. Concentrated peptide stocks were prepared by dissolving lyophilized peptide into 10 mM sodium phosphate buffer, typically at concentrations between 200 and 550 μM. Peptides were dissolved with vortexing, and stocks were centrifuged and decanted to remove insoluble material that was sometimes observed during sample preparation. Peptide concentration was determined by diluting aliquots of peptide stocks 20-fold into 8 M guanidinium chloride (GdmCl) solution and measuring absorbance values at 280 nm ($\varepsilon_{280,Trp}$ = 5690 $M^{-1} cm^{-1}$, $\varepsilon_{280,Tyr}$ = 1280 $M^{-1} cm^{-1}$, $\varepsilon_{280,Pin\ WW}$ = 13940 $M^{-1} cm^{-1}$).[1] UV-Vis data were normalized by subtracting spectra taken for the buffer described above diluted 20-fold into 8 M GdmCl solution. Samples for CD experiments were prepared by diluting peptide stock solutions to ~40 μM concentration, and then filtering through 0.2 μm syringe filters immediately prior to use. Sample filtration was critical to successful sample preparation; in preliminary experiments where samples were not filtered prior to thermal denaturation, we often observed irreversible aggregation/precipitation upon heating.

Thermal denaturations

Maxima in the CD spectra of all Pin1 WW domain variants were observed at approximately 227 nm; consequently, all thermal denaturations were monitored at this wavelength. Thermal denaturation experiments were carried out over a range of 2 to 98 °C, in 4 degree steps. Averaging time for each measurement was 20 seconds. The equilibration time between measurements was 5 minutes.

Fits to thermal denaturation data were carried out using the following expression to describe the ellipticity signal as a function of temperature ($\Theta_{obs}(T)$):

$$\Theta_{obs}(T) = \frac{\Theta_{denat}(T) + \Theta_{native}(T) \cdot K_f(T)}{1 + K_f(T)}$$

In the expression shown above, $\Theta_{denat}(T)$ and $\Theta_{native}(T)$ are the temperature-dependent ellipticity signals arising from the unfolded and folded states, respectively. $K_f(T)$ is the equilibrium folding constant as a function of temperature. The temperature-dependent ellipticity signals arising from the denatured and native states are described as follows:

$$\Theta_{denat}(T) = \Theta^0_{denat} + m_{denat} \cdot T$$

$$\Theta_{native}(T) = \Theta^0_{native} + m_{native} \cdot T$$

In both expressions shown above, θ^0 is the intercept of the unfolded or folded baseline and *m* is the slope. The equilibrium constant for folding ($K_f(T)$) is described by the following equation, where $\Delta G_f(T)$ is the free energy of folding as a function of temperature and R is the ideal gas constant.

$$K_f(T) = \exp(\frac{-\Delta G_f(T)}{R \cdot T})$$

The temperature-dependent free energy of folding $\Delta G_f(T)$ is described by the following expression, where $\Delta H_f(T_m)$, ΔC_p and T_M are fitted parameters:

$$\Delta G_f(T) = \frac{\Delta H_f(T_m)}{T_m}(T_m - T) + \Delta C_p[T - T_m - T\ln(\frac{T}{T_m})]$$

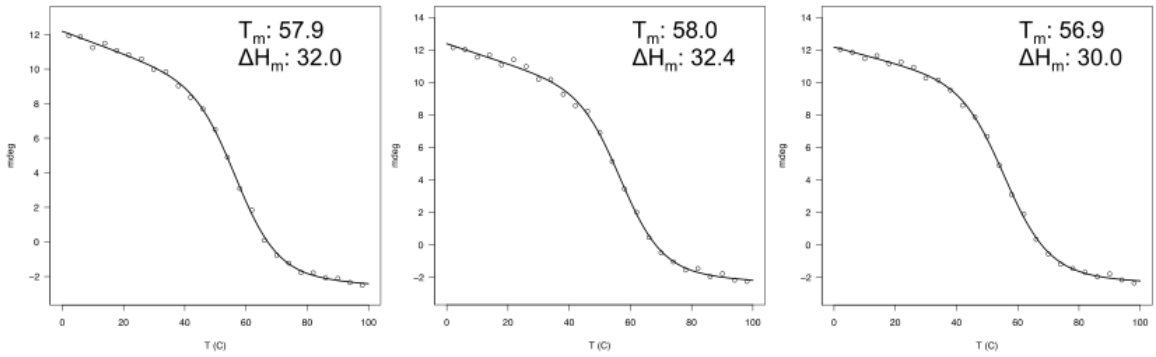
We found that the uncertainty in the values of ΔC_p determined from fitting to observed data was generally of the same order of magnitude as the values themselves. Consequently, we fit thermal denaturation data using the following expression for $\Delta G_f(T)$, where ΔC_p is set to zero:

$$\Delta G_f(T) = \frac{\Delta H_f(T_m)}{T_m}(T_m - T)$$

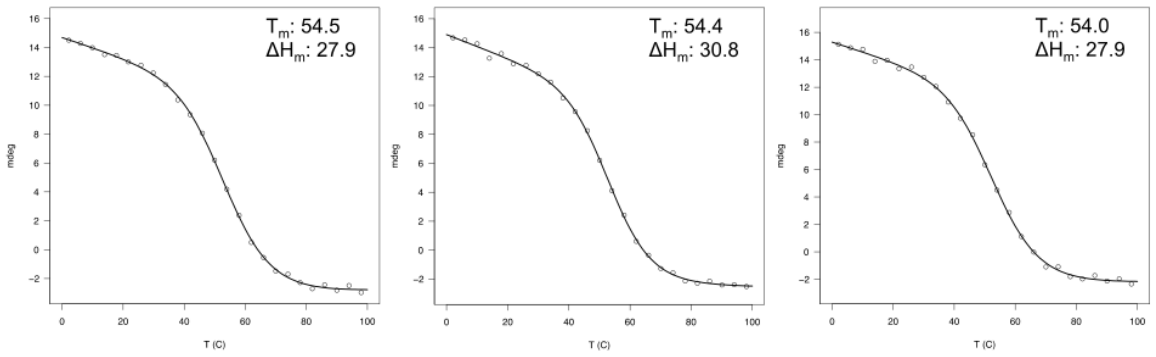
Data from three thermal denaturation measurements of each WW domain variant were fit individually; reported T_M values are the averages of the three measurements. Fits were carried out using the R statistical package. For several WW domain variants without well-resolved unfolded baselines, unfolded baseline slopes were constrained to values derived from closely-related sequences. For derivatives **2**, **2**(A17[S,S]ACPC), **2**(A17β^3A) and **2**(G19[R,R]ACPC), unfolded baseline slopes were set to the average value derived from three fitting sessions for derivative **2**(D18[S,S]ACPC), which is less thermally stable. For derivatives **3** and **3**(N17[R,R]ACPC), unfolded baseline slopes were derived from the slope determined for analog **3**(G18[R,R]ACPC). For derivative **1**(G20[S,S]ACPC), the un-

folded baseline slope was derived from that of derivative **1**. Notably, this procedure produced T_M values for analogs **2** and **3** that are very close to literature values. In contrast, treating the unfolded baseline slope as an adjustable parameter for these derivatives produced T_M values that differed more significantly from previously-determined values. Thermal denaturation data are shown below for all Pin1 WW variants described in this manuscript.

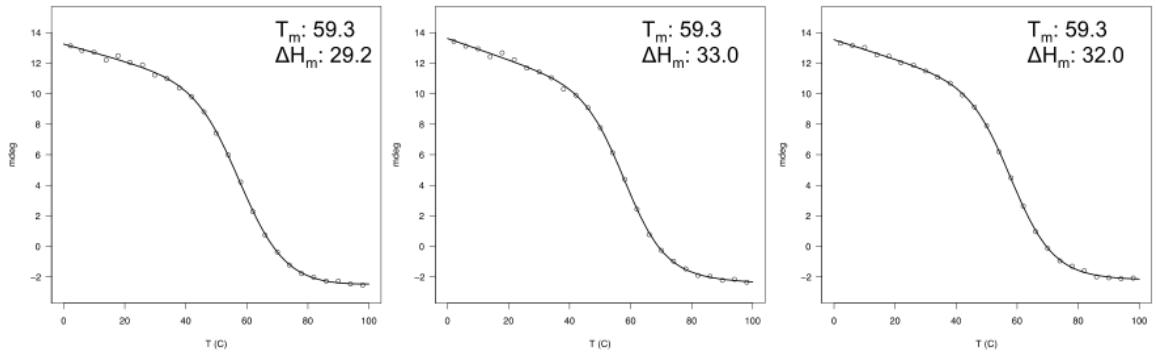
WW(4:6) **1**



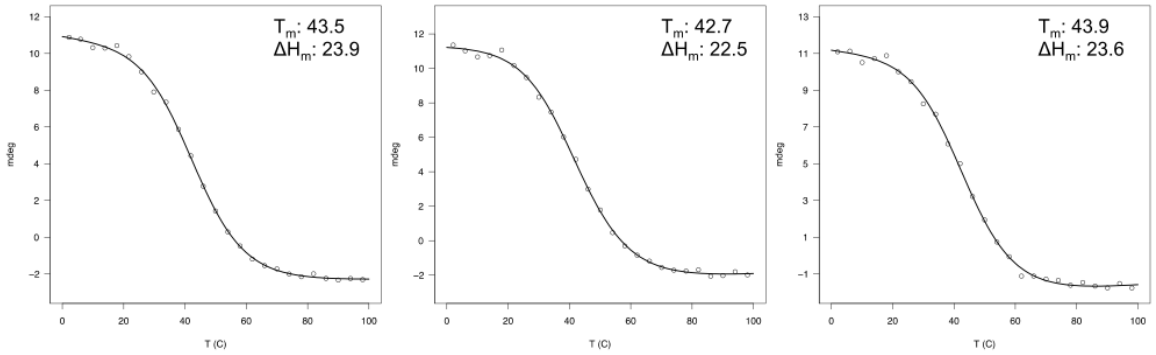
1(R17[S,S]ACPC)



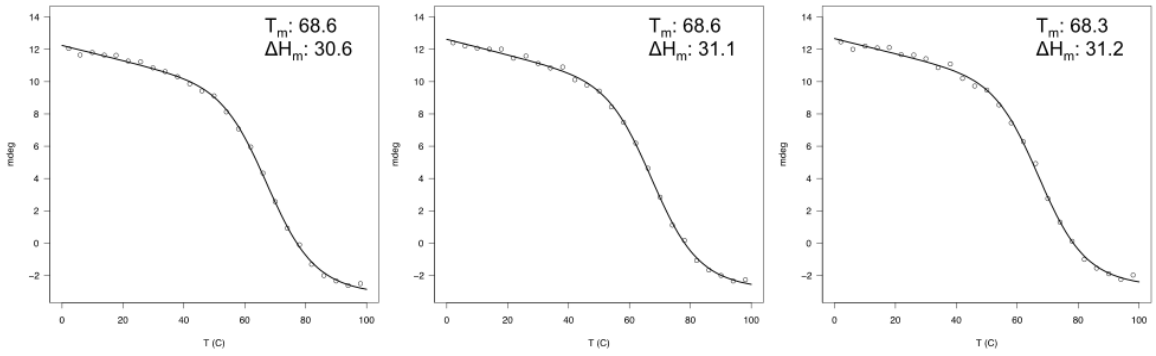
1(S18[S,S]ACPC)



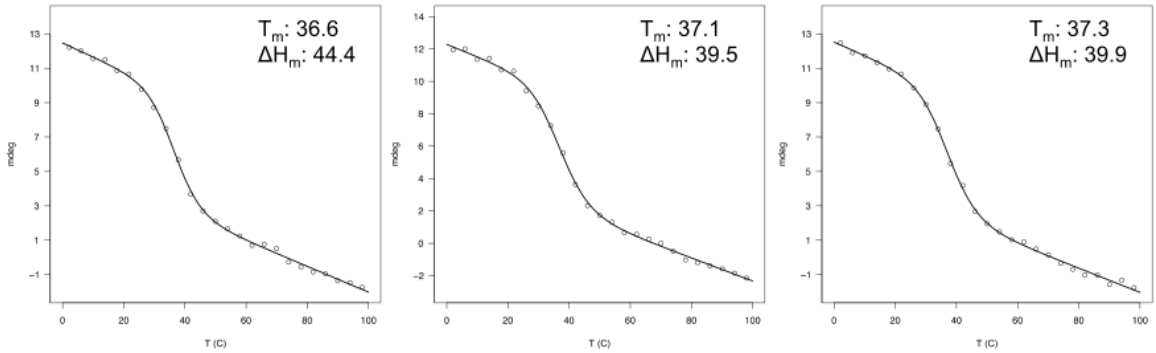
1(S19[S,S]ACPC)



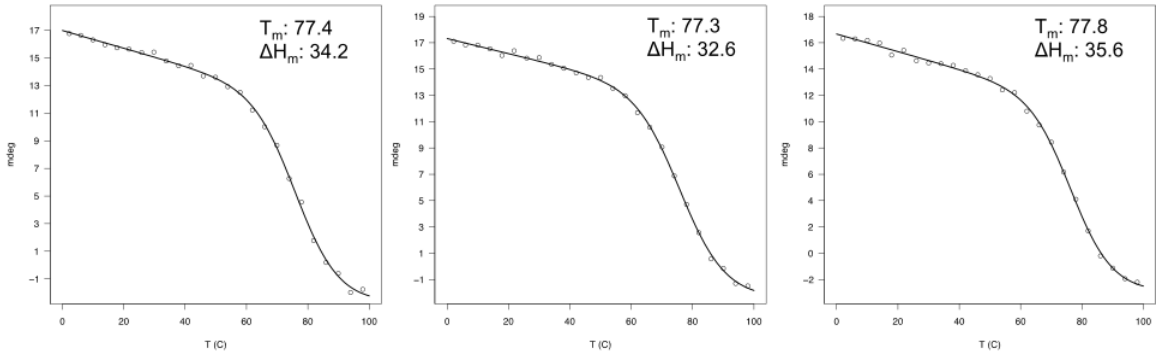
1(G20[R,R]ACPC)



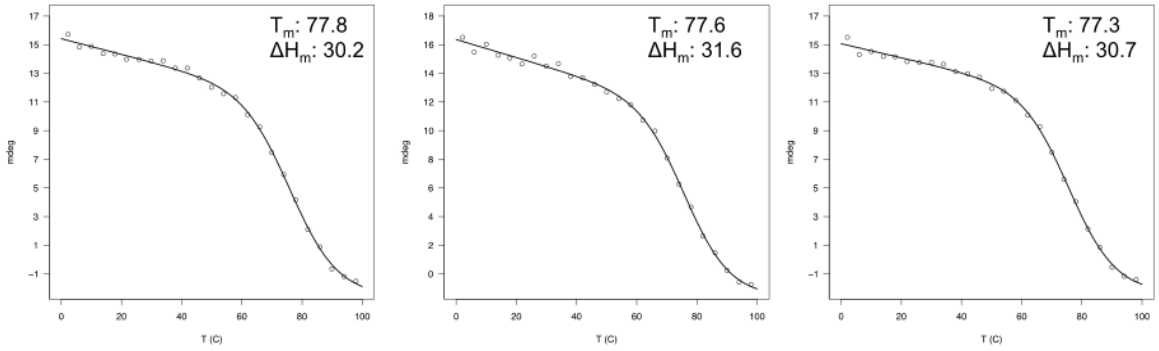
1(G20[S,S]ACPC)



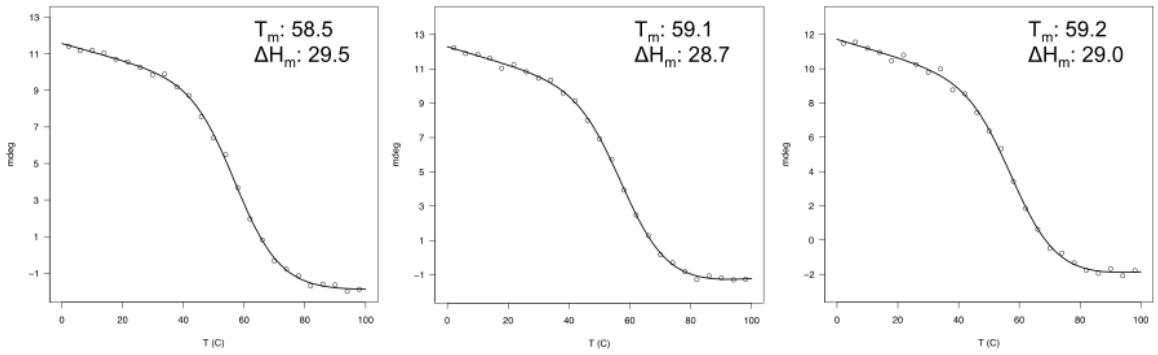
WW(3:5) **2**



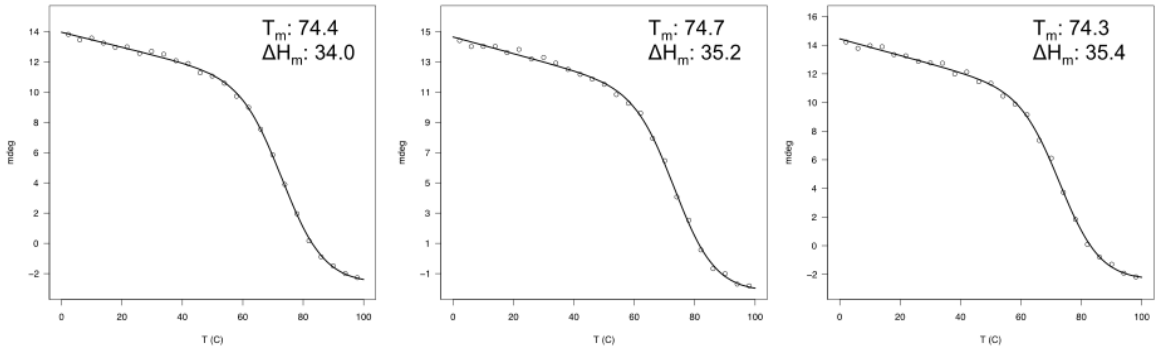
2(A17[S,S]ACPC)



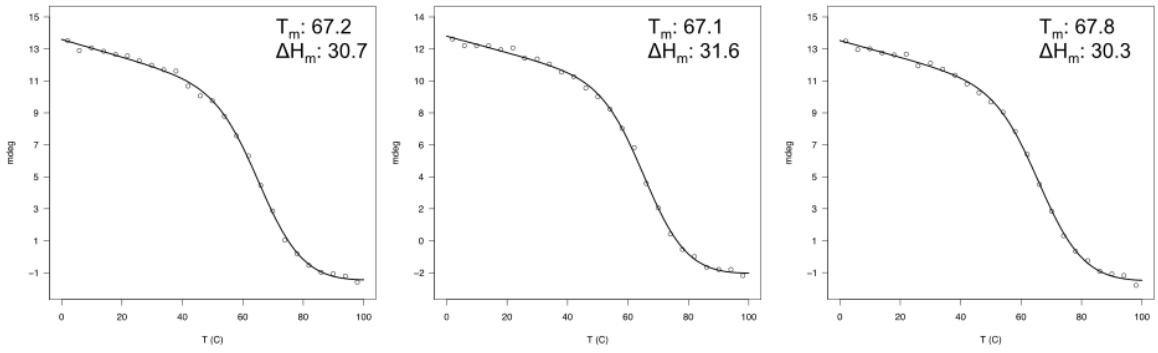
2(D18[S,S]ACPC)



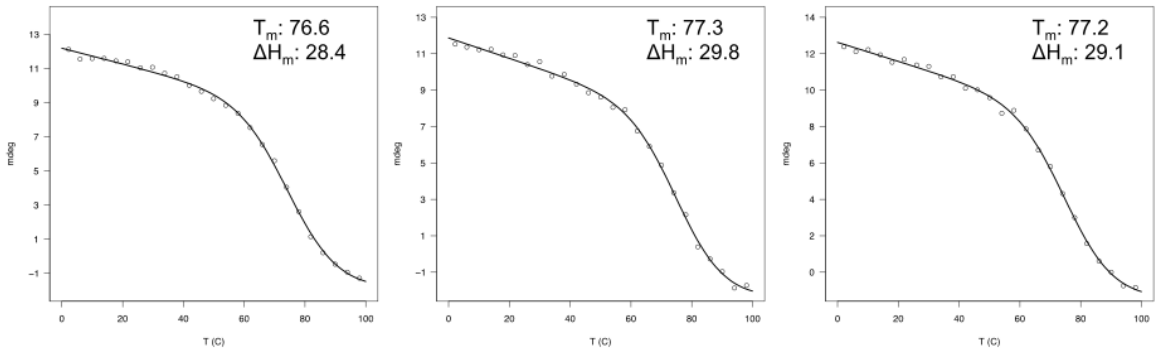
2(G19[R,R]ACPC)



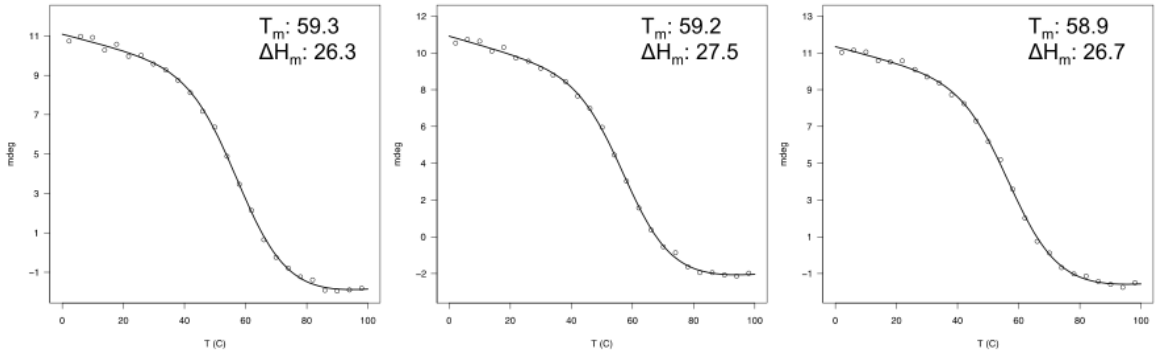
WW(2:2) **3**



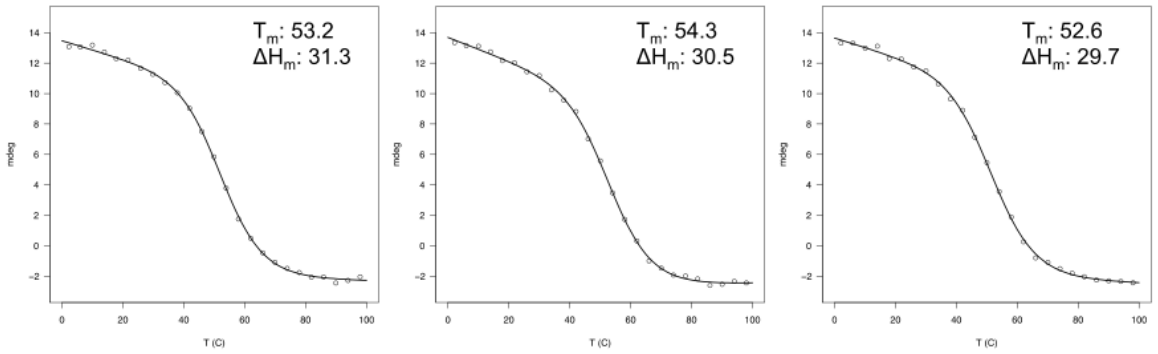
3(N17[R,R]ACPC)



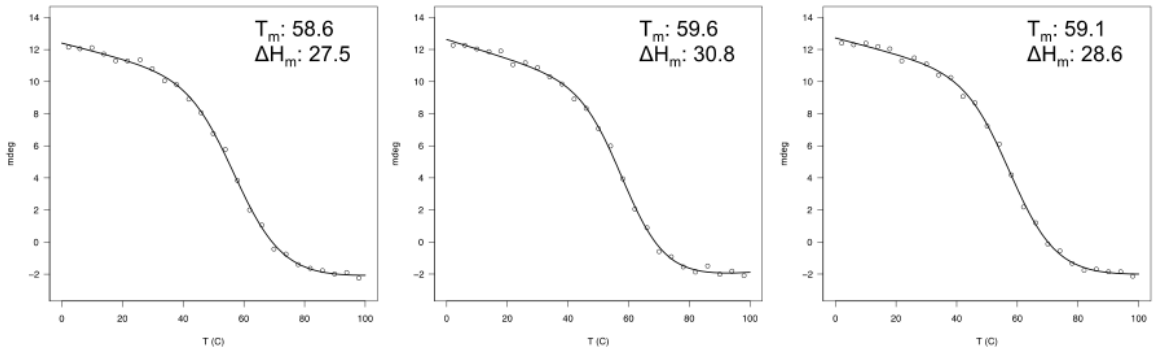
3(G18[R,R]ACPC)



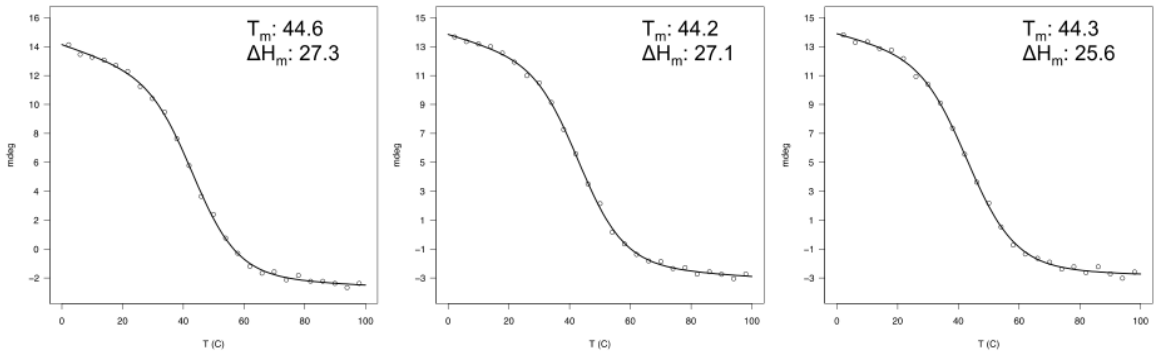
1(R17β^3R)



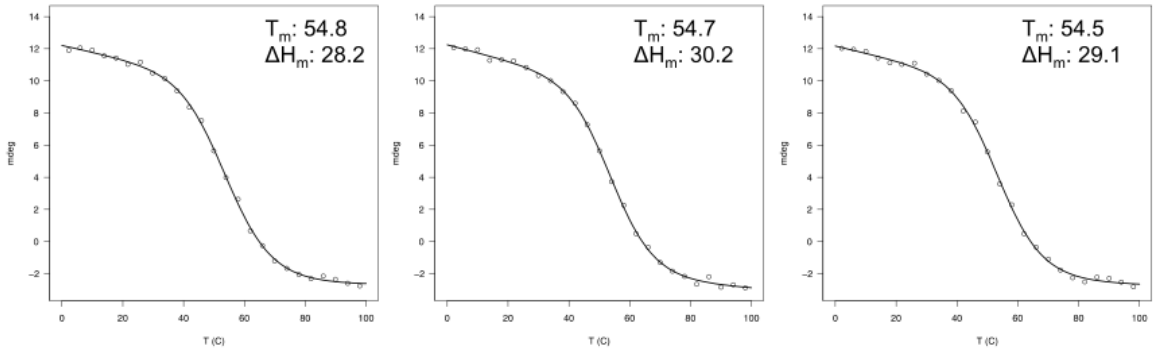
1(S18β^3S)



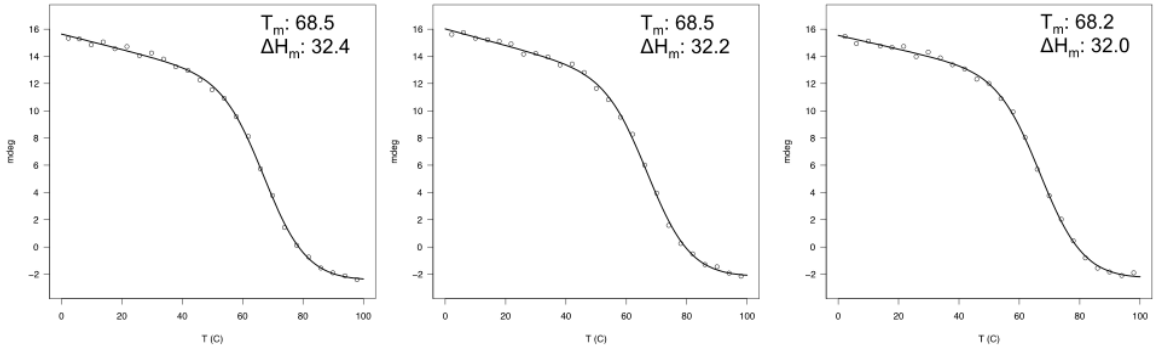
1(S19β^3S)



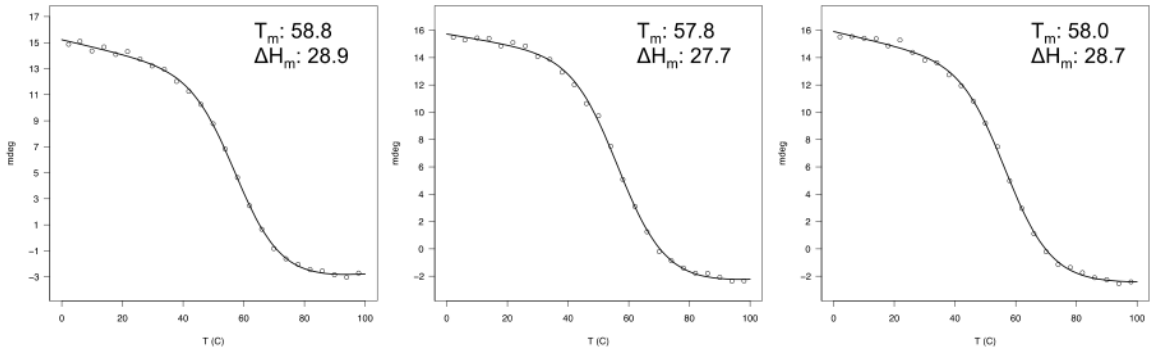
1(G20β^3G)



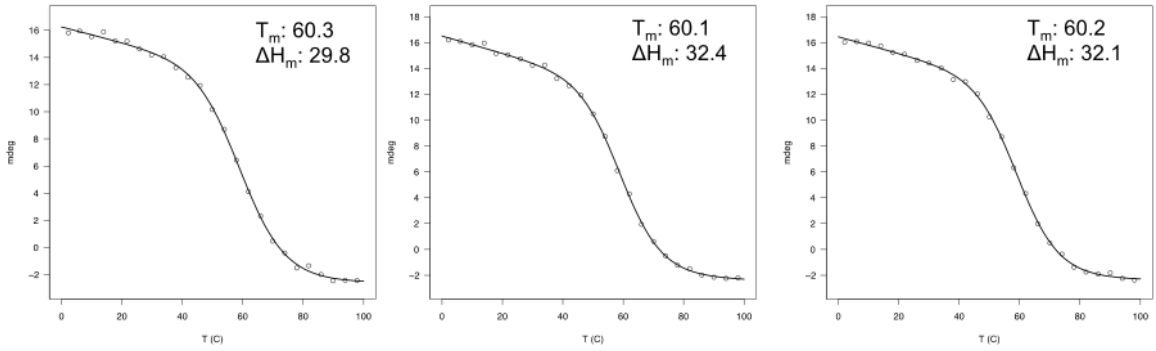
2(A17β^3A)



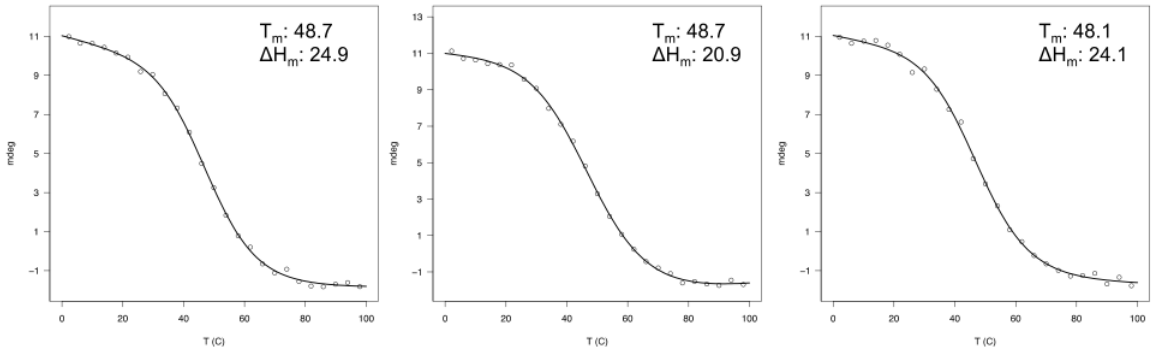
2(D18β^3D)



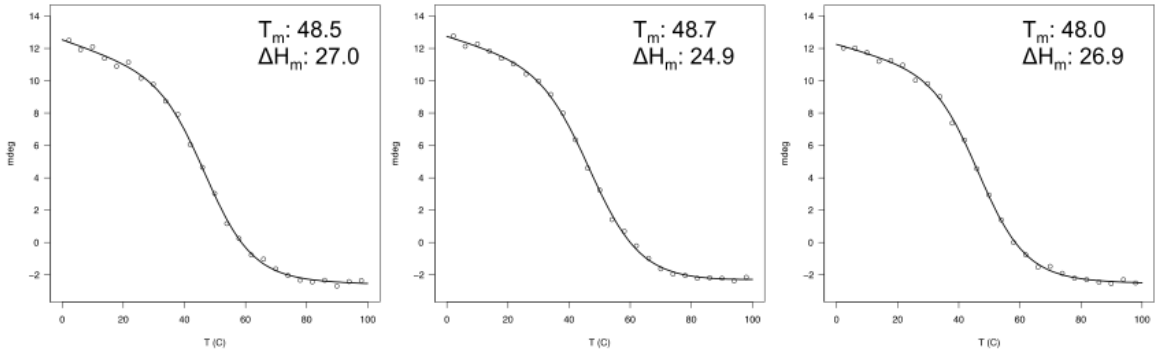
2(G19β^3G)



3(N17β^3N)



3(G18β^3G)



Crystallization Experiments and Diffraction Data Collection

*WW Domain Variant **1**(S18[S,S]ACPC)*

Initially, crystals of **1**(S18[S,S]ACPC) grew from a 1:1 quasiracemic solution with the D-polypeptide form of peptide **1** (3.5 mg/mL total peptide concentration), using condition #2 from the Hampton Index screen (0.1 M sodium acetate pH 4.5, 2.0 M ammonium sulfate). Following preliminary data collection, structural solution and model refinement, it became clear that **1**(S18[S,S]ACPC) had crystallized without its quasi-enantiomer, leading us to set up further crystallization experiments with **1**(S18[S,S]ACPC) alone. Crystals of Pin1 WW domain variant **1**(S18[S,S]ACPC) grew overnight from hanging drops containing 1:1 ~4 mg/mL peptide stock solution and condition #2 in the Index sparse-matrix screen (Hampton Research), as described above. Crystals of **1**(S18[S,S]ACPC) were transferred to a 4:1 solution of precipitant:glycerol before freezing in liquid N_2 for storage.

Initially, a low-resolution diffraction data set was collected from a crystal of **1**(S18[S,S]ACPC) using a rotating anode X-ray source (Bruker). Diffraction data were collected on beam 21-ID-F (LS-CAT sector 21, λ: 0.97872 Å) at the Advanced Photon Source, Argonne National Lab. A data set of 360 frames was collected at a crystal-to-detector distance of 125 mm, with 0.3 second exposures and increments of 1° of rotation per frame. Frames were indexed and integrated using XDS[2], and reflection intensities were scaled in scala[3]. The space group was assigned as $P4_32_12$ using pointless.[4] Intensities were converted to amplitudes using the French and Wilson method implemented in truncate.[5]

*WW Domain Variant **1**(S18β^3S)*

Crystals of Pin1 WW domain variant **1**(S18β^3S) grew from hanging drop vapor diffusion experiments with two conditions identified in sparse-matrix screening: Crystal Screen II #36 (0.1 M HEPES pH 7.5, 4.3 M NaCl) and Index #11 (0.1 M HEPES pH 7.5,

3.0 M NaCl). The concentration of peptide used for crystallization studies was 4 mg/mL. Crystals were observed in both conditions after approximately 6 months, suggesting that an increase in salt concentration was necessary for crystal growth to occur. A crystal grown from Crystal Screen II condition #36 was briefly transferred to a 4:1 solution of precipitant:glycerol before freezing in a cryostream.

Diffraction data were collected in the Molecular Structure Laboratory in the UW Madison Department of Chemistry using an APEXII detector and Mo IμS source (Bruker), to a high-resolution limit of 1.99 Å. Frames from three different oscillations were integrated and scaled in the APEXII software package (Bruker). Data were merged in XPREP. Systematic absences consistent with a $6_1/6_5$ screw axis were observed in the data, leading to the assignment of space group $P6_522$.

*WW Domain Variant **3**(N17[R,R]ACPC)*

Two crystal forms of Pin1 WW domain variant **3**(N17[R,R]ACPC) were grown *via* hanging drop vapor diffusion experiments from two different conditions identified in sparse-matrix screening: Index #3 (0.1 M BIS-TRIS pH 5.5, 2.0 M ammonium sulfate) and Index #12 (0.1 M Tris pH 8.5, 3.0 M NaCl). Crystals grown from Index condition #12 had a square rod morphology, whereas crystals grown from Index condition #3 had a boat- or football-shaped morphology. The concentration of peptide used for crystallization studies was 2.8 mg/mL. Crystals of both forms described above were treated with Paratone-N prior to freezing, after preliminary attempts at cryoprotection of crystals grown from Index condition #3 (using a 3:1 solution of precipitant:glycerol) were unsuccessful. Notably, efforts to collection diffraction data at room temperature from these crystals (mounted in capillaries) were unsuccessful; freezing in Paratone-N was required for diffraction.

Initially, crystals of **3**(N17[R,R]ACPC) grown from Index condition #3 were used for diffraction data collection on a rotating anode X-ray generator (Bruker). Crystals of **3**(N17[R,R]ACPC) grown from Index condition #3 were then used for synchrotron data collection on beam 21-ID-F (LS-CAT sector 21, λ: 0.97872 Å) at the Advanced Photon Source, Argonne National Lab. Diffraction data were collected to a useful resolution limit of approximately 1.7 Å, based on I/σ_I of 2 to 3 and R_{sym} of approximately 0.7 in the highest-resolution shell in the processed data sets. Several sets of frames were collected at crystal-to-detector distances of 140 or 160 mm, with 2 or 4 second exposures and increments of 1° of rotation per frame. Images were indexed, integrated and scaled in the XDS package.[2] Preliminary data processing in XPREP suggested an orthorhombic P-lattice with cell constants of *a*: 42.54 Å, *b*: 85.00 Å, *c*: 100.32 Å, $\alpha = \beta = \gamma = 90°$. Systematic absences were indicative of either one or two 2_1 screw axes, consistent with space groups $P222_1$ and $P2_12_12$, respectively. Matthews analysis of this crystal form suggested that the asymmetric unit contains approximately 10 symmetry-independent copies of **3**(N17[R,R]ACPC) (consistent with ~45% bulk solvent). Preliminary attempts at structural solution *via* molecular replacement were unsuccessful, likely due to the large number of independent molecules expected in the asymmetric unit.

Crystals of **3**(N17[R,R]ACPC) grown from Index condition #12 were used for synchrotron data collection on beam 21-ID-F (LS-CAT sector 21, λ: 0.97872 Å) at the Advanced Photon Source, Argonne National Lab. Diffraction data were collected to a useful resolution limit of 1.9 Å. A data set of 360 frames was collected at a crystal-to-detector distance of 130 mm, with 1.0 second exposures and increments of 1° of rotation per frame. Frames were indexed and integrated using XDS[2], and reflection intensities were scaled in

scala[3]. The space group was assigned as $P4_32_12$ using pointless[4]. Intensities were converted to amplitudes using the French and Wilson method implemented in truncate.[5]

Structure Solution and Model Refinement

*WW Domain Variant **1**(S18[S,S]ACPC)*

The structure of variant **1**(S18[S,S]ACPC) was initially solved by molecular replacement in Phaser[6] with PDB entry 4GWT[7] as an initial search mode, using a data set collected to 2.4 Å resolution on a rotating anode source (Bruker). Following synchrotron data collection, the model was re-refined to higher resolution, with the set of reflections flagged for R_{free} calculation carried over from the low-resolution data set. The structure of **1**(S18[S,S]ACPC) was refined at 1.50 Å resolution *via* maximum likelihood methods in Refmac5[8], using anisotropic thermal parameters and hydrogens generated at calculated positions. This structure has been deposited in the PDB with accession ID 5VTJ.

*WW Domain Variant **1**(S18β^3S)*

The structure of variant **1**(S18β^3S) was solved by molecular replacement in Phaser[6] using a model derived from the structure of WW domain variant **1**(S18[S,S]ACPC) (*vide supra*). Following structural solution, model refinement was carried out *via* maximum likelihood methods in Refmac5[8] using isotropic thermal parameters and hydrogens generated at calculated positions. This structure has been deposited in the PDB with accession ID 5VTK.

*WW Domain Variant **3**(N17[R,R]ACPC)*

The structure of WW domain variant **3**(N17[R,R]ACPC) was solved by molecular replacement in Phaser using a data set collected on a rotating anode source. The model used for structural solution was derived from the structure of racemic Pin1 WW domain

crystallized with DL-malate (PDB: 4GWT[7]). Following structural solution, the model of **3**(N17[R,R]ACPC) was refined to a high-resolution limit 2.2 Å resolution against home source-derived data, and then re-refined against synchrotron-derived data at 1.80 Å resolution. The same set of reflections was used for the calculation of R_{free} in both refinements. This structure structure has been deposited in the PDB with accession ID 5VTI.

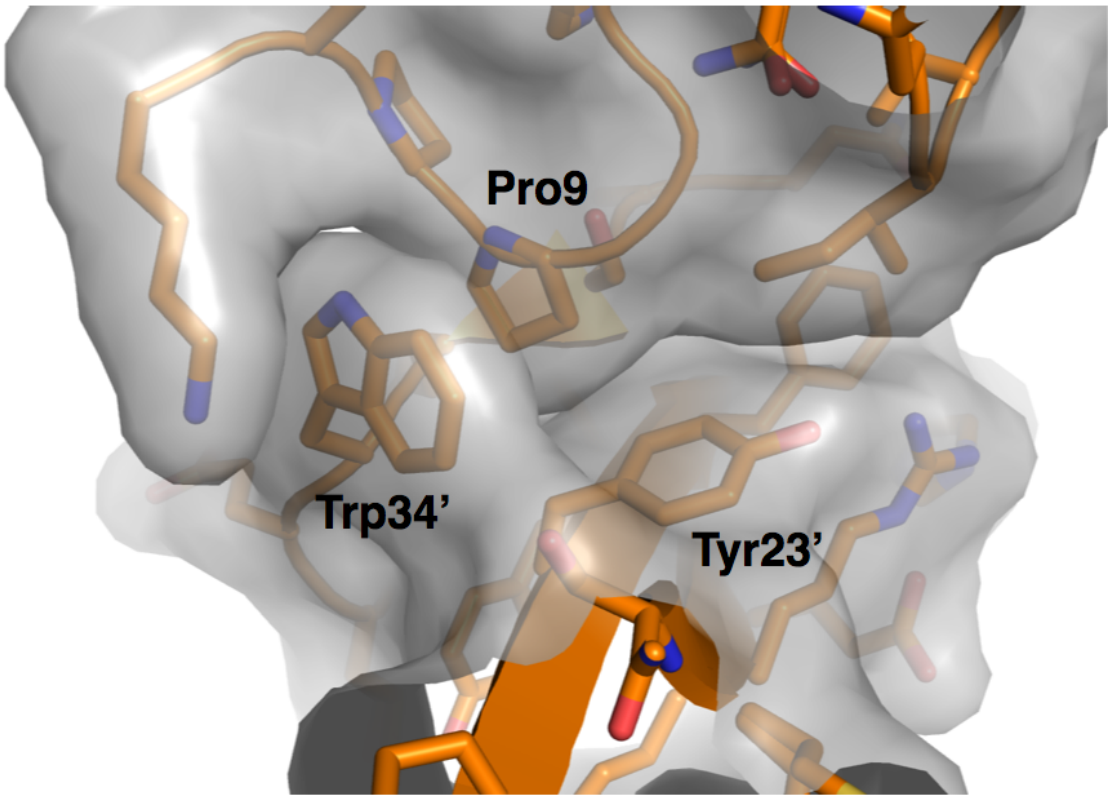
Data Processing and Model Refinement Statistics

Statistics from diffraction data processing and model refinement of the three structures described here shown below. Values in parentheses are for reflections in the highest-resolution shell of a given data set.

Table S1

	1(S18[S,S]ACPC)	1(S18β^{3}S)	3(N17[R,R]ACPC)
PDB Code	5VTJ	5VTK	5VTI
Data Collection Temp. (K)	100	100	100
Beamline	APS 21-ID-F	Bruker Mo	APS 21-ID-F
Wavelength (Å)	0.97872	0.70930	0.97872
Resolution range	47.33 – 1.50 (1.58 – 1.50)	34.12 – 1.99 (2.02 – 1.99)	48.17 – 1.80 (1.90 – 1.80)
Unique reflections	7363 (1038)	3112 (126)	4398 (617)
Completeness (%)	99.1 (98.6)	99.7 (97.7)	99.9 (100.0)
R_{sym}	0.044 (1.934)	0.102 (0.324)	0.076 (1.902)
I/σ(I)	43.3 (2.4)	38.41 (11.27)	25.1 (2.3)
Redundancy	27.9 (28.5)	40.8 (24.6)	26.4 (28.1)
Wilson B-factor (Å^2)	23.1	10.2	31.6
Space group	$P4_32_12$	$P6_522$	$P4_32_12$
Unit cell parameters			
(a/b/c) (Å)	47.33/47.33/38.55	47.63/47.63/60.74	48.17/48.17/37.21
(α/ β/ γ) (°)	90/90/90	90/90/90	90/90/90
Refinement:			
# Non-H Atoms	342	346	308
Resolution range (Å)	33.47 – 1.50 (1.54 – 1.50)	34.12 – 1.99 (2.04 – 1.99)	34.06 – 1.80 (1.85 – 1.80)
R(work)	0.185 (0.301)	0.163 (0.151)	0.231 (0.298)
R(free)	0.253 (0.322)	0.221 (0.293)	0.275 (0.278)
RMS bond length (Å)	0.016	0.019	0.012
RMS bond angle (°)	1.75	1.86	1.71
Coord. ESU (ML; Å)	0.092	0.092	0.119
Mean B Value (Å^2)	31.0	14.4	43.7

Figure S1. Conserved interaction between Pro9 and hydrophobic cleft formed by Trp34 and Tyr23 on symmetry-related monomer (denoted by primes below), from x-ray structure of **3**(N17[R,R]ACPC):



PDB Survey of β-Residue Backbone Dihedrals

PDB accession codes of structures used in survey of β^3-amino acid (185) and ACPC (46) replacements in folded proteins are given below:

5HFY
4OZA
4OZB
2OXJ
3HEW
5I1N
5I1O
5I1P
2OXK
5HG2
3O42
3CGN
3GCO
4WPB
4HJD
4DZV
2YJ1
3O3Z
3O40
3HET
3G7A
3F4Z
4KGS
4KGR
4OZC
5HI1

References

1. Edelhoch, H., *Biochemistry* **1967,** *6* (7), 1948-1954.
2. Kabsch, W., Acta Crystallographica Section D: Biological Crystallography **2010,** 66 (Pt 2), 125-132.
3. Evans, P., Acta Crystallographica Section D **2006,** 62 (1), 72-82.
4. Evans, P. R., Acta Crystallographica Section D **2011,** 67 (4), 282-292.
5. French, S.; Wilson, K., Acta Crystallographica Section A **1978,** 34 (4), 517-525.
6. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., *Journal of Applied Crystallography* **2007,** *40* (4), 658-674.
7. Mortenson, D. E.; Kreitler, D. F.; Yun, H. G.; Gellman, S. H.; Forest, K. T., *Acta Crystallographica Section D: Biological Crystallography* **2013,** *69* (Pt 12), 2506-2512.
8. Murshudov, G. N.; Vagin, A. A.; Dodson, E. J., *Acta Crystallographica Section D* **1997,** *53* (3), 240-255.